Energy & Buildings 225 (2020) 110305

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Supervised machine learning of thermal comfort under different indoor temperatures using EEG measurements

Xin Shan, En-Hua Yang *

School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 11 April 2020
Revised 10 June 2020
Accepted 10 July 2020
Available online 15 July 2020

Keywords:
Machine learning
electroencephalogram (EEG)
Supervised learning
Thermal comfort
Human sensing

ABSTRACT

In this paper, machine learning techniques in conjunction with passive EEG (electroencephalogram) measurement were explored to classify occupants' real-time thermal comfort states, which have the potential in the future for energy saving through adopting time varying set points when real-time changes in thermal comfort can require less energy input. The performances of different machine learning techniques were compared, and the method to select linear continuous features for class interpolation was also explored. For the full-set features, the performances of different classifiers were satisfactory, with classification rates all above 90%. The LDA classifier had the best performance. The second best was the NB classifier, and the relatively worst was the KNN classifier. The linear continuous EEG features were selected by interpolation and can be found for all human subjects. Higher selection threshold led to less selected features but higher average performance of these features. In general, the EEG based machine learning methods can classify occupants' real-time thermal comfort states, and could potentially lead to more building energy saving through comfort-driven time varying set points.

1. Introduction

Building sector contributes to more than 30% of energy consumption globally [1], and air conditioning and mechanical ventilation (ACMV) system is one major contributor to building energy consumption [2]. The purpose of ACMV system is mainly to provide thermally comfortable indoor environment. Existing thermal comfort models are mainly based on standards such as ASHRAE standard 55 [3], which recommends comfortable thermal environment through the joint effects of temperature, relative humidity, air speed and human factors based on the predicted mean vote. However, existing thermal comfort models do not take into account some important time varying factors [4], such as occupants' acclimatization or varying physical conditions during the day [5–8]. Current ACMV systems often adopt time invariant set points based on the existing models and standards, which could cause energy inefficiency [9]. Therefore, potential energy savings could be achieved through obtaining occupants' real-time thermal comfort states and adopting time varying set points accordingly when real-time changes in thermal comfort can require less energy input.

Some studies have tried to obtain occupants' real-time thermal comfort states through online questionnaire methods. For instance, online reporting tools have been developed to obtain real-time feedbacks from occupants regarding thermal comfort and other indoor environment aspects [10]. Jazizadeh et al. [11] also tried to integrate occupants' real-time thermal comfort votes into the air conditioning system control. Another study also used online data acquisition combined with Bayesian optimal classifier for modeling and quantifying personalized thermal comfort [12]. However, survey based methods require active continuous input from occupants regarding their real-time thermal comfort states, which are difficult to implement in practice. Some other studies have shown that physiological measurements could be used to passively obtain the state of thermal comfort, such as end-tidal partial CO2 (ETCO2), arterial blood oxygen saturation (SPO2), biomarkers in saliva and tear film quality [13–14], but many of these measurements are also difficult to implement in practice on a real-time basis, as they may require bio-samples to be collected and analyzed in lab (such as saliva or tear film), or exhaled gases to be measured in-situ.

Machine learning techniques, when used in conjunction with passive real-time measurements, could be potential methods to better classify occupants' real-time thermal comfort states. Machine learning techniques can find patterns and make decisions based on data, and these techniques include model-based statisti-

* Corresponding author.
E-mail address: ehyang@ntu.edu.sg (E.-H. Yang).

https://doi.org/10.1016/j.enbuild.2020.110305
0378-7788/

cal learning (supervised or unsupervised), or model-free reinforcement learning or more general artificial neural network [15]. Some studies have already explored in this way. Ghahramani et al. [16] used hidden Markov based machine learning technique to identify thermal comfort (with three conditions: uncomfortably warm, comfortable, uncomfortably cool) utilizing infrared thermography of human face, and could well classify the uncomfortable conditions. Other study used Support Vector Machine, K-Nearest Neighbor, Random Forest, and Subspace Linear Discriminant classification algorithm to classify thermal comfort using air temperature sensor, skin temperature device and thermal camera, and found that the combination of sensors gave the best classification result [17].

Another potential real-time passive measurement of occupants is EEG (electroencephalogram), which is an electrophysiological technique that records brain activity. EEG has already been explored in previous thermal comfort studies. Yao et al. [18] and Lan et al. [19] used EEG technique to study the effects of different temperatures, and showed that the EEG frequency powers were different as a result of temperature variation. Shan et al. [20] studied correlations of EEG indices with subjective perception/task performance under different indoor temperatures. On the other hand, in the EEG research domain for emotion [21–23] and mental workload [24,25], machine learning methods were also often used to classify different mental states. Thermal comfort is also defined as the condition of mind that expresses satisfaction with a thermal environment [26]. Therefore, based on these previous studies, there's a potential way to bridge the machine learning-based techniques in EEG research domain with the thermal comfort of building occupants.

For the machine learning techniques in these EEG studies, supervised learning was often used. The basic idea of supervised learning is to train a classifier (model-based or model-free) with labeled data (classes are clear beforehand), and this classifier then predicts the classes of similar but unlabeled data. Classes are often pre-defined in a particular area of study, such as different classes of emotions in the emotion studies, or different classes/levels of mental workloads in the workload studies. For the machine learning-based EEG emotion studies, the two dimensional Arousal-Valence model [27] was commonly used, where arousal denotes the intensity of emotion and valence denotes the positive/negative characteristics of emotion. Different emotions can be elicited through the International Affective Picture System (IAPS) [28] or the International Affective Digitized Sounds [29], and these databases map different sounds or pictures that denote different classes of emotion onto the 2D Arousal-Valence space. The transition among emotions is assumed as continuous, and these databases have large amount of classes that cover the 2D space. Typical classes of different emotions are low-arousal-positive-valence class (*e.g.,* calm), high-arousal-negative-valence class (*e.g.,* angry), or other predefined narrower classes. For the machine learning-based EEG mental workload/vigilance research, the basic idea is similar. Different levels of workload [24] or vigilance [25] were elicited by mental tasks of different difficulties. In these studies the number of levels/classes was often small (*e.g.,* only high, intermediate and low levels), and the mental states were also implicitly assumed as continuous. However, this limited amount of established levels/-classes might pose problems for future application when more precise levels within or beyond these established levels are required. For instance, if more precise identification of mental workload between the established high and intermediate levels is needed, then the EEG data of this mental workload level needs to be elicited and collected through additional experiment. Depending on the precision required, more differentiated mental workload levels can be established through more experiments, but this could be constrained by limited resources.

The goal of the current study is to investigate the supervised machine learning of thermal comfort under different indoor temperatures with EEG measurement by addressing the above mentioned issues and problems, with data from a previous experiment [20]. In the current study, firstly the suitability of machine learning methods was explored and the performances of different machine learning techniques were compared. The classes in this study are thermal comfort states in either rest or task conditions under various indoor temperatures. Secondly, the method to interpolate different classes was explored, which can be a potential way to establish more classes without conducting additional experiments. As for the long term vision, the machine learning methods with EEG measurement could potentially contribute to better thermal comfort of occupants and also to energy saving when real-time changes in thermal comfort can require less energy input.

2. Methodology

2.1. Experiment

The data used in this study were from a previous experiment in an office room [20]. The room is 4.7 m in length, 3.1 m in width, and 2.6 m in height (floor to false ceiling). The air conditioning and mechanical ventilation (ACMV) is fan-coil system, and the air temperature in the room could be adjusted with controls of the ACMV and a stand-alone heater. The impact of radiant temperature was regarded negligible, as the room has no sunlight exposure, windows or radiant asymmetry. As a result, air temperature was taken as operative temperature in this study [3]. Three temperature levels were investigated in the experiment: 23, 26 and 29 °C. Other environmental parameters were maintained relatively the same. In the following sections, these three temperature levels are labeled as cool, neutral, and warm states, respectively.

Twenty-two human subjects were recruited from college students. They were asked to dress typical local clothing (short-sleeve shirt and long trousers) that has a 0.57clo level [3]. Each participant experienced all three thermal conditions (within-subject). There were three experimental timeslots every day: 10:30–12:30, 13:00–15:00, and 15:30–17:30. Each participant attended for three days, and was required to select the same session for each day to minimize possible confounding effects.

Questionnaires were used to investigate subjective perception of indoor environment aspects [3,30–33]. Computerized tasks (*e.g.* short-term memory tasks) were also used to assess work performance [34–35]. EEG data were monitored in rest and task states. The data were collected by the Emotiv EPOC equipment (EPOC+, Emotiv Inc. USA), which is a high resolution, non-intrusive, and portable wireless headset. It has fourteen channels located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 in accordance with the International 10–20 system. Two extra channels at CMS and DRL are references. The headset's internal sampling rate is 2048 Hz, and it then down-samples and outputs at 128 Hz. The noise from building's electrical power system (50/60 Hz) is already filtered in the headset. The data are sent to laptop via Bluetooth through proprietary USB operating at 2.4 GHz. Contact pads are moistened with saline solution to improve conductivity. The Software Development Kit (SDK) has a packet count function to check no data are lost, and also has a real-time sensor contact quality display. Headset setup and channel locations are shown in Fig. 1.

The major experimental steps in each session are shown in Fig. 2. For the Rest EEG condition, the participants were asked to sit relaxed with eyes open, and not to do anything. For the Task EEG condition, the participants were asked to do the same basic

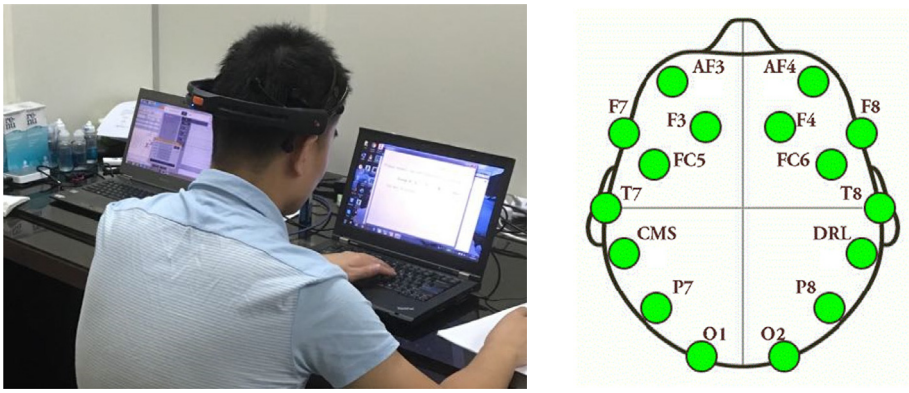

Fig. 1. Emotiv EPOC: (1) headset setup; (2) channel locations.

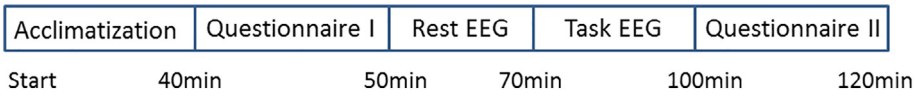

Fig. 2. Major procedures for each timeslot.

mental tasks such as short-term memory task. The room temperature levels were balanced across participants to reduce sequential effects, such as fatigue.

In this previous experiment, the EEG results illustrated that for the frontal asymmetrical activity, neutral thermal environment led to relatively more positive emotion. The EEG rest frontal asymmetrical activity was more correlated to the thermal acceptability metric in the questionnaire, which showed that neutral environment was perceived as the most acceptable. The EEG task frontal asymmetrical activity was also correlated to the task performance, which showed that neutral environment led to better task performance. In the current study, machine learning methods using EEG data will be investigated in details.

2.2. Machine learning-based EEG pattern recognition methods

The data were pre-processed by the EEGLAB toolbox 13.4.4b [36], which operates under the Matlab [37]. The steps recommended by the toolbox developers were used. The high-pass and low-pass thresholds were set at 3 Hz and 45 Hz respectively to remove DC offset, low-frequency artifacts (*e.g.* skin potential), and other high-frequency artifacts. As recommended by the developers, DC offset can be removed by applying high-pass filter of around 0.5 Hz. Low-frequency artifacts such as skin potential are usually <1 Hz [38]. Therefore a slightly higher margin of 3 Hz threshold was chosen to minimize the potential impacts of low frequency interferences. High-frequency artifacts such as building's electrical power system are 50/60 Hz, and movement and muscle artifacts are typically larger than 60 Hz [38]. In addition, the upper limit of the gamma range (highest of the five frequency ranges in EEG literature) is usually selected as 45 Hz [39]. Therefore the upper threshold in this study was also chosen as 45 Hz. Non-stereotyped artifacts (such as irregular head movement) were removed by manually scrolling the data. The stereotyped artifacts (such as eye blinks) were then removed by the independent component analysis (ICA), which decomposes data into maximally independent components and then rejects the components mostly related to artifacts. Afterwards the time series were segmented into epochs of 8-second length. The spectopo function in the toolbox was adopted to perform the power spectrum analysis, which computed the discrete power densities in the range of 3–45 Hz with 50% overlap. The computed data were then used for the machine learning analysis.

Three techniques were used in this study for machine learning-based EEG pattern recognition, *i.e.,* linear discriminant analysis (LDA), Naive Bayes (NB), and K-nearest neighbor (KNN). These classifiers are from the Matlab statistics toolbox [37], and all have relatively fast training and prediction speed.

2.2.1. Linear discriminant analysis classifier

The LDA classifier uses multivariate normal distribution model. The classifier is trained with labeled data by adjusting the distribution parameters. To classify the unlabeled data, the classifier minimizes the expected misclassification cost in Eq. (1).

$$\hat{y} = \arg\min_{y=1,\ldots,K} \sum_{k=1}^{K} \hat{P}(k|x) C(y|k) \tag{1}$$

where $\hat{y}$ is the predicted class; K is the total number of classes; $\hat{P}(k|x)$ is the posterior probability of class k conditional on an observation x; and $C(y|k)$ is the cost of classifying an observation as class y when its real class is k. The default values were adopted: $C(y|k)$=1 if y ~= k, and$C(y|k)$ = 0 if y = k. These suggest the cost is 1 for incorrect classification, and 0 for right classification.

The posterior probability $\hat{P}(k|x)$ is defined in Eq. (2).

$$\hat{P}(k|x) = \frac{P(x|k)P(k)}{P(x)} \tag{2}$$

where $P(k)$ is the prior probability of class k, calculated as the proportion of class k samples among training data; $P(x)$ is for normalization, *i.e.* the summation over k of $P(x|k)P(k)$; and $P(x|k)$ is the multivariate normal density defined in Eq. (3).

$$P(x|k) = \frac{1}{(2\pi|\Sigma_k|)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\} \tag{3}$$

where μ_k is the mean; Σ_k is the covariance matrix; $|\Sigma_k|$ is the determinant; and Σ_k^{-1} is the inverse matrix.

The LDA is generally accurate when the assumptions are met, *i.e.,* each class is multivariate normally distributed. Otherwise the prediction accuracy may vary. The computing memory usage is relatively low.

2.2.2. Naive Bayes classifier

The NB classifier assumes features are conditionally independent given the class, but it often works satisfactorily in practice even if the independent assumption is invalid. To train a NB classifier, the model estimates the parameters of a specific probability distribution. To classify the unknown data, this method chooses the class that each data point belongs to with the largest posterior probability. This independence assumption given the class simplifies the training process, as one-dimensional density distribution for each feature can be computed separately. In this study normal density distribution was used, *i.e.,* each feature is modeled as normally distributed. Other parameter values in the NB classifier were kept default. The prediction accuracy of this classifier is generally medium. The computing memory usage for normal distribution modeling is low.

2.2.3. K-nearest neighbor classifier

The KNN classifier categorizes the unknown data according to their distance to the training data, which can be simple but still effective. From the majority vote of the k-nearest neighbors' classes, unknown data can be classified. Many different types of distance metrics are available, and the Euclidean distance was used in this study. Other parameter values in the KNN classifier were kept default. Given mx (1-by-n) vectorsx_1, x_2, ..., x_{mx} and my (1-by-n) observation vectorsy_1, y_2, ..., y_{my}, the Euclidean distance between the vector x_s and y_t is defined in Eq. (4).

$$\text{Euclidean distance}: d_{st}^2 = (x_s - y_t)(x_s - y_t)' \quad (4)$$

The prediction accuracy is generally good for low dimensions, but may be poor for high dimensions. The computing memory usage is relatively high.

2.2.4. User-dependent classifier and selection of features

User-dependent approach was adopted in the current paper, *i.e.,* a separate classifier was adopted for each participant. Related research [21,23,40] often used this method, as EEG inter-person diversities are usually significant [41]. The feature selection approach in the current study followed previous works [21,25]: the mean power density in each 1 Hz bin between 3 and 45 Hz (42 bins) was adopted. As the headset has 14 channels, 588 features were available for selection.

2.3. Method to compare different classifiers' performance

The performances of LDA, NB and KNN classifiers were compared by data classification rates. The classification rate was calculated as follows: for each participant, 1) half of his/her data were randomly selected for training, and the remaining half were kept for classification; 2) features were chosen based on his/her training data by ANOVA with $p < 0.05$; 3) his/her classifiers were trained with training data; and 4) his/her classifiers were utilized to classify the classification data. The classification rate was computed as the percentage of correctly classified data. The classification rates of different classifiers were then compared.

2.4. Method to find linear continuous features for class interpolation

A potential way to interpolate different classes is to interpolate the features. In this study, linear continuous features were searched, which can be used for further class interpolation. The neutral condition was interpolated from the experimental data at cool and warm conditions, and then the experimental data at neutral condition were used to check the performance of this interpolation. The method was as follow: for each individual feature of each human subject, 1) the interpolated mean value of neutral condition was obtained as the average of the mean values at cool condition and warm condition; 2) three cases of interpolated standard deviation of neutral condition were considered: mean, minimum, and maximum of the standard deviations at cool condition and warm condition; 3) the interpolated data at neutral condition were then generated by the normal random numbers function using the interpolated mean value and the interpolated standard deviation; 4) a LDA classifier was trained by the experimental data at cool and warm conditions and the interpolated data at neutral condition; 5) the experimental data at neutral condition were then classified by this trained LDA classifier. Finally if this classification rate passed a satisfactory threshold, then this feature was deemed as linear continuous. As each individual feature was examined here for its respective interpolation performance, the classification rate threshold was considered satisfactory as long as it was well above random rate (*i.e.* 33%), because the models here were univariate rather than multivariate.

It should be noted that higher classification rate indicates that the mean values under the three thermal conditions should be different enough from each other, and the standard deviations should be small enough. In addition, after the linear continuous features were found, these features were also tested for extrapolation cases. Two cases were tested: extrapolated cool condition and extrapolated warm condition. The procedures were similar to that of interpolation.

3. Results and discussion

3.1. Environmental measurements and human subjects' thermal sensation votes

Temperature, relative humidity (RH) and air velocity were recorded by air velocity meters (Velocicalc air velocity meter 9545, TSI Inc.). CO_2 was recorded by CO_2 meters (model CM-0018, CO_2 Meter Inc.). Data were recorded at seat level near participants, and the sampling rate was 1/60 s^{-1}, *i.e.,* 1-min interval. As shown in Table 1, the experimental temperatures were close to the targeted temperatures.

For thermal sensation votes, 23 °C resulted in slightly cool, 26 °C resulted in neutral, and 29 °C resulted in slightly warm according to the notation in ASHRAE standard 55 [3] (1 is cold, 4 is neutral, and 7 is hot in the current study), and the pairwise p-values were all significant ($p < 0.002$). The results were in accordance with the ASHRAE standard 55 [3].

3.2. Performance of different classifiers

The results of LDA, NB and KNN classifiers are shown in Fig. 3 (rest condition) and Fig. 4 (task condition). The full-set features after selection by ANOVA varied among participants and were between 400 and 500. The average number of training data was 36 for rest condition and 50 for task condition. For each scenario (*e.g.,* LDA classifier in rest condition): 1) each human subject's full-set features classification rate was computed; 2) then the features were dropped randomly one at a time and the corresponding classification rates were computed; 3) finally the average classification rates among participants were computed. The 1/3 rate (random rate) was given to the final zero feature cases. The standard deviations among participants were also plotted. The purpose of gradually dropping features was to investigate if features were

Table 1
Environmental measurements near human subjects and thermal sensation votes [20].

Targeted	23 °C		26 °C		29 °C	
Monitored temperature and RH	Temp (°C)	RH (%)	Temp (°C)	RH (%)	Temp (°C)	RH (%)
	22.6 ± 0.2	70.1 ± 0.6	25.8 ± 0.1	61.3 ± 0.6	29.1 ± 0.4	50.5 ± 1.1
Thermal sensation votes*	3.09 ± 1.02		4.19 ± 0.73		5.09 ± 0.68	
Monitored CO_2 and air velocity	CO_2 < 1000 ppm; air-velocity < 0.1 m/s					

*7-point scale: 1 is cold, 4 is neutral, and 7 is hot.

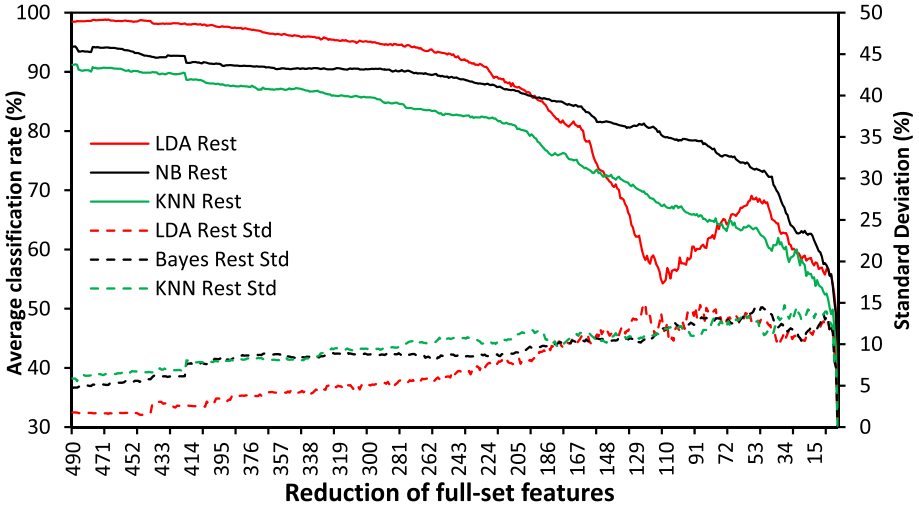

Fig. 3. Comparison of different classifiers for rest condition. Classification rate averaged across human subjects for full-set features: a) LDA rest: 98%; b) NB rest: 92%; c) KNN rest: 89%.

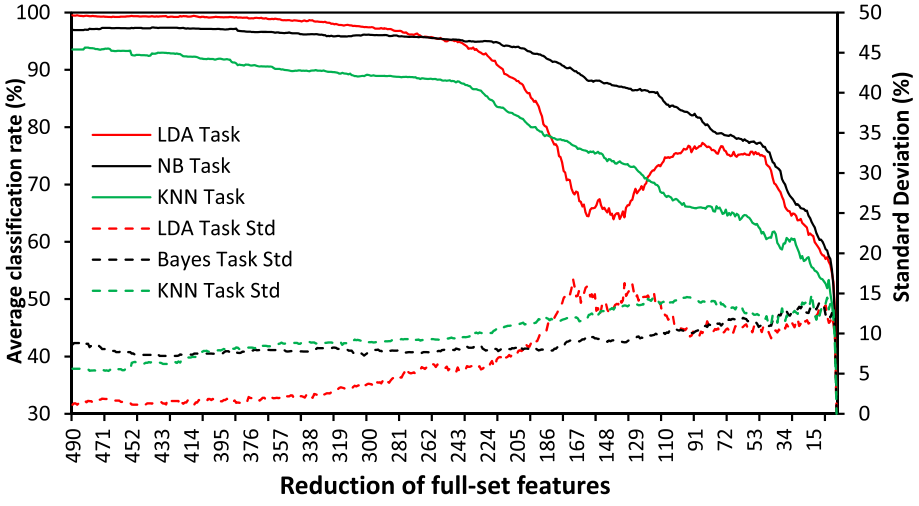

Fig. 4. Comparison of different classifiers for task condition. Classification rate averaged across human subjects for full-set features: a) LDA task: 99%; b) NB task: 97%; c) KNN task: 92%.

reduced/lost, how the classification rate would change accordingly. The results could also give more insights to the balance between efficiency and accuracy, as less number of features might be more efficient for implementation in practice.

For the LDA classifier with full set of features, the classification rates were above 95% for both rest state and task state. For the rest state, the average rate was around 90% when the number of features decreased to around 200, and a saddle point emerged when the number of features decreased to around 100. The rate then increased until the number of features decreased to around 50, and finally the rate decreased. The task state had similar pattern with different turning points. The task average rate was at its saddle point near 150 features, and then increased again until around 80 features, and then finally decreased. The results showed that the LDA classifier can satisfactorily classify various thermal comfort states (rest and task conditions) under three temperature levels in two feature intervals, and the first feature interval corresponded to higher rate. Furthermore for the LDA, more number of features did not necessarily imply higher rate, as a saddle point existed between the two intervals. The main reason behind the saddle point should be under-sampling of training data and inter-dependence of features. Theoretically the classification errors

should decrease with increasing number of features, but this may not be true in practice because the number of training data is often much less than the number of features, which is known as under-sampling problem. Some researchers have put forward methods to optimize the number of features as a function of sample size for different classification methods [42], or put forward feature reduction methods such as uncorrelated LDA to cope with potential under-sampling problems that could arise [43]. Still, the results in Figs. 3 and 4 showed that as long as the number of features was in the two intervals, *i.e.,* either sufficiently large or close to the number of training data, the classification rates were generally satisfactory.

For the NB classifier, the full-set features classification rates were lower than those of the LDA classifier for both rest and task conditions. This slightly worse performance suggested that the features were not entirely independent of each other, and therefore for full-set features the covariance approach of LDA was more appropriate than the NB's approach of independent feature treatment. Unlike the LDA classifier, for NB classifier the trend of dropping features was monotonic. Again, since the main difference between LDA and NB classifiers in this study is their treatments of inter-dependent characteristics of features, the pattern difference between LDA and NB classifiers further showed that the non-monotonic decreasing behavior of LDA classifier should be caused by the feature inter-dependence characteristics coupled with under-sampling problem. For the NB classifier, the classification rate dropped below 90% at around 270 features for the rest condition and at around 170 features for the task state, and the task state had better classification performance than the rest condition throughout the feature reduction process. For both the rest state and task state, the NB classifier had lower classification rates and larger standard deviations among human subjects than the LDA classifier in the range from the full-set features to the vicinity of the saddle point of LDA classifier.

For the KNN classifier, the full-set features classification rates were lower than those of the NB classifier for both the rest state and task state. The rate dropped below 90% at around 450 features for the rest state and at around 350 features for the task state. The trend of dropping features was also monotonic. For both the rest state and task state, the KNN classifier had lower classification rates than the NB classifier throughout the feature reduction process. This suggested that the probability-based approach is more appropriate than the simple geometric distance-based approach. The standard deviations across participants were similar for both the KNN and NB classifiers.

In general, these three classifiers can classify occupants' real-time thermal comfort states in certain range of features for both rest and task conditions. The potential for energy saving as a result is to adopt time varying set points in ACMV system when real-time changes in thermal comfort can require less energy input.

3.3. Linear continuous features for class interpolation and extrapolation

The results of interpolation for rest state and task state are shown in Figs. 5 and 6, respectively. Three cases of standard deviations were used for interpolation: the mean, maximum and minimum standard deviations of cold condition and warm condition. These three cases give the possible range of interpolated standard deviation, and are denoted as "Meanstd", "Maxstd" and "Minstd". The horizontal axis is the selection threshold, which means that a particular feature will be considered as linear continuous for that threshold only if its classification rate passes that threshold. In this study, the selection threshold ranged from 33% to 80%. The 33% is the random rate to correctly classify a thermal comfort state and therefore was chosen as the bottom line. The 80% was regarded as a good classification rate for the individual feature's respective univariate model. The variation of selection thresholds in this range was to investigate the number and performance of features that could be suitable for linear interpolation when different selection criteria were chosen: the higher the threshold, the more stringent the criteria. The left vertical axis labels the average classification rate of features that passes a selection threshold, and the right vertical axis labels the average number of these features. All these average values were across human subjects. The error bars show the deviations across participants, and were shown for the "MeanStd" case only, as the three standard deviation cases were similar.

The figure shows that for the rest state and task state, the average number of features passing 33% threshold was around 60–70, and the average classification rate of these features was around 60%. Higher selection threshold led to fewer selected features but better average performance of these features, *i.e.,* higher average classification rates. For instance, for both the rest state and task state, the average number of features passing 67% threshold was 20–30 and the average classification rate of these features was around 80%. The number and the average classification rates of linear continuous features were not very sensitive to the way of interpolating standard deviations, as "Meanstd", "Maxstd" and "Minstd" cases were very close. For the average classification rates of the selected features, the deviations among human subjects (*i.e.,* the error bars) decreased with higher selection threshold for both the rest state and task state. For the number of the selected features, the deviations among human subjects did not decrease significantly with higher selection threshold for the rest condition, but decreased significantly for the task condition. The features that did not pass the pre-defined threshold may have large standard deviations or less mean value differences among three temperature conditions. As the range of temperature conditions was quite large in this study, linear interpolation may perform even better in narrower temperature intervals, *i.e.,* more linear continuous features could be found with higher average classification rates.

The extrapolation performance of these selected linear continuous features was further calculated and is shown in Fig. 7. The suffix "Warm" or "Cold" means that this particular thermal condition was extrapolated. Only "Meanstd" cases were calculated, as the previous interpolation results already showed that three standard deviation interpolation methods were close. The error bars again denote the deviations across human subjects and were plotted at sparser intervals. As can be seen, the classification rates were in the 10–40% range, which was close to the 33% random rate and therefore not viable. The deviations among human subjects were not sensitive to selection threshold. The results suggested that the linear continuous features may perform relatively well for interpolation, but they may not necessarily perform well for extrapolation. This may due to the reason that the selected features were not perfectly linear, and therefore extrapolation would magnify the errors especially for the large range of temperature conditions in this study. This also suggested that interpolation is more reliable.

4. Conclusions and outlook

In this study, machine learning techniques in conjunction with passive EEG (electroencephalogram) measurement were explored to classify occupants' real-time thermal comfort states under different indoor temperatures. The performances of different classifiers were compared. The method to select linear continuous features for class interpolation was also investigated. Major conclusions include:

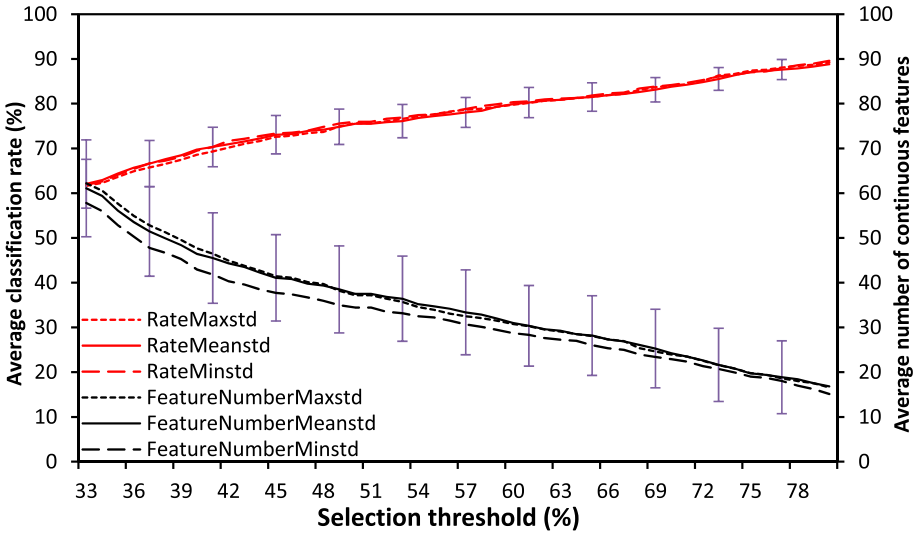

Fig. 5. Rest condition interpolations. Each line denotes the average values across all human subjects. The error bars denote the standard deviations across all human subjects, and are plotted for the "Meanstd" case only.

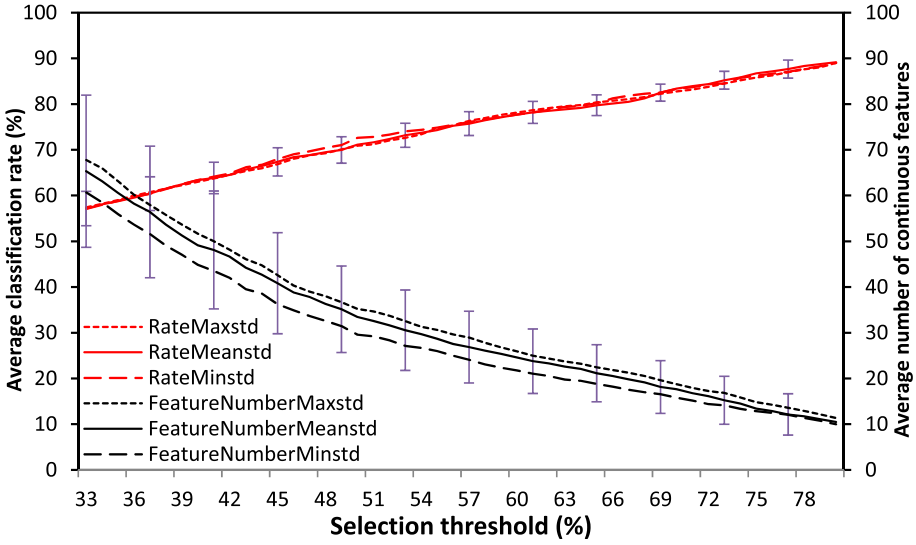

Fig. 6. Task condition interpolations. Each line denotes the average values across all human subjects. The error bars denote the standard deviations across all human subjects, and are plotted for the "Meanstd" case only.

- For the full-set features, the performances of different classifiers were satisfactory, with classification rates all above 90%. The LDA classifier had the best performance. The second best was the NB classifier, and the relatively worst was the KNN classifier.
- The patterns of feature reduction were different among classifiers. The LDA classifier had a saddle point at around 100–150 features, while the NB and KNN classifiers decreased monotonically. From the full-set features to the location of LDA classifier's saddle point, the LDA classifier had the highest rate, and the NB classifier was intermediate, and the KNN classifier was the worst. In this range, LDA classifier also had the smallest standard deviations among human subjects, while the NB and KNN classifiers had similarly larger standard deviations.
- The linear continuous features were selected by interpolation method. For both the rest state and task state, the average number of features passing 33% threshold was around 60–70, and the average classification rate of these features was around 60%. Higher selection threshold led to less selected features but higher average performance of these features, *i.e.,* higher average classification rates. The numbers and the average rates of linear continuous features were not very sensitive to the way of interpolating standard deviations.
- The linear continuous features were also checked by extrapolation, and the classification rates were mainly concentrated in the 10–40% range, which was close to the 33% random rate and therefore not viable.

In general, the EEG based machine learning methods can passively classify occupants' real-time thermal comfort states, and accordingly, have the potential for energy saving through adopting time varying set points when real-time changes in thermal comfort can require less energy input. In addition, the linear interpolation method provides a way to interpolate data that are not experimentally available. This method should work even better in narrower temperature intervals, which could be achieved with more experimental thermal conditions. The results of data extrapolation were not satisfactory, and therefore, in future studies, lower/higher tem-

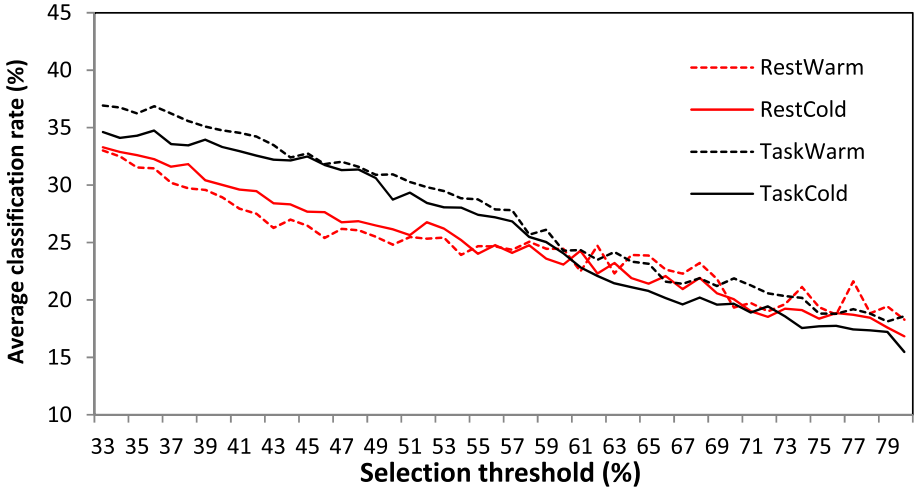

Fig. 7. Rest and task conditions extrapolation of "Meanstd" case. Each line denotes the average classification rates across all human subjects. The error bars denote the standard deviations across all human subjects. The suffix "Warm" or "Cold" means that this particular thermal condition was extrapolated.

perature levels beyond the range of those in the current study should be experimentally investigated, as it could be common in other countries and regions. For the classifiers in particular, this study established user-dependent models, in which inter-person diversities are bypassed. In the current context, one of the inter-person diversities is the different thermal sensations due to personal differences such as gender and metabolic rate, which can be seen by the thermal sensation variations in Table 1. Some people have thermal sensation votes shifted either upwards or downwards from the group means. Future studies should take this inter-person difference into account, and can also construct user-independent classifier by recruiting more human subjects to control for these inter-person diversities (*e.g.*, gender, age) [25,44–45].

Furthermore, in most cases more features can generally lead to better classification accuracy. In this study the issue of efficiency was not considered, as the training and classification speed for the selected models is relatively fast. However, this could be an issue in practice, and the balance between efficiency and accuracy for each person could be further addressed by identifying the more independent features and delete some of the features that are highly correlated to them. Another issue could be the non-stereotyped artifacts, which may be difficult to remove in practice. More advanced algorithms could be established to filter these artifacts. Longer time duration for thermal comfort state classification could also be adopted to reduce these artifacts' potential impact in practice, as they occur less often as compared to stereotyped artifacts such as eye-blinks. As for wearing the portable devices to collect EEG data, the current state-of-the-art devices could still be a hassle to building occupants especially for long-term wearing. Further improvement on the portable and contactless devices for EEG data collection is necessary. This study focused on indoor thermal environment, and for the potential future studies, these methods can also be applied to other aspects of indoor environment, such as indoor air quality [46], which is also related to ACMV system and therefore has energy saving potentials.

CRediT authorship contribution statement

Xin Shan: Methodology, Data curation, Investigation, Writing - original draft. **En-Hua Yang:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The financial support from the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) program is gratefully acknowledged.

References

[1] UN Environment and International Energy Agency, Towards a zero-emission, efficient, and resilient buildings and construction sector, Global Status Report 2017 (2017).
[2] U.S. Energy Information Administration, International Energy Outlook 2013 (2013).
[3] Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE Standard: 55–2013, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2013.
[4] J. Van Hoof, Forty years of Fanger's model of thermal comfort: Comfort for all?, Indoor Air 18 (3) (2008) 182–201.
[5] H. Ning, Z. Wang, Y. Ji, Thermal history and adaptation: Does a long-term indoor thermal exposure impact human thermal adaptability?, Appl. Energy 183 (2016) 22–30.
[6] A. Uğursal, C.H. Culp, The effect of temperature, metabolic rate and dynamic localized airflow on thermal comfort, Appl. Energy 111 (2013) 64–73.
[7] G.S. Brager, R.J. de Dear, Thermal adaptation in the built environment: a literature review, Energy Build. 27 (1) (1998) 83–96.
[8] G. Jendritzky, R.J. de Dear, Adaptation and Thermal Environment, in: K.L. Ebi, I. Burton, G. McGregor (Eds.), Biometeorology for Adaptation to Climate Variability and Change, Springer, 2009, pp. 9–32.
[9] S.C. Sekhar, Thermal comfort in air-conditioned buildings in hot and humid climates–why are we not getting it right?, Indoor Air 26 (1) (2016) 138–152.
[10] L. Zagreus, C. Huizenga, E. Arens, D. Lehrer, Listening to the occupants: a Web-based indoor environmental quality survey, Indoor Air 14 (s8) (2004) 65–74.
[11] F. Jazizadeh, A. Ghahramani, B. Becerik-Gerber, T. Kichkaylo, M. Orosz, Human building interaction framework for personalized thermal comfort-driven systems in office buildings, J. Comput. Civ. Eng. 28 (1) (2014) 2–16.
[12] A. Ghahramani, C. Tang, B. Becerik-Gerber, An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling, Build. Environ. 92 (2015) 86–96.
[13] L. Lan, P. Wargocki, D.P. Wyon, Z. Lian, Effects of thermal discomfort in an office on perceived air quality, SBS symptoms, physiological responses, and human performance, Indoor Air 21 (5) (2011) 376–390.
[14] K.W. Tham, H.C. Willem, Room air temperature affects occupants' physiology, perceptions and mental alertness, Build. Environ. 45 (1) (2010) 40–44.

[15] J.R. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: A review of algorithms and modeling techniques, Appl. Energy 235 (2019) 1072–1089.
[16] A. Ghahramani, G. Castro, S.A. Karvigh, B. Becerik-Gerber, Towards unsupervised learning of thermal comfort using infrared thermography, Appl. Energy 211 (2018) 41–49.
[17] A. Aryal, B. Becerik-Gerber, A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor, Build. Environ. 160 (2019) 106223.
[18] Y. Yao, Z. Lian, W. Liu, Q. Shen, Experimental study on physiological responses and thermal comfort under various ambient temperatures, Physiol. Behav. 93 (1–2) (2008) 310–321.
[19] L. Lan, Z. Lian, L. Pan, The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings, Appl. Ergon. 42 (1) (2010) 29–36.
[20] X. Shan, E.-H. Yang, J. Zhou, V.W.-C. Chang, Human-building interaction under various indoor temperatures through neural-signal electroencephalogram (EEG) methods, Build. Env. 129 (2018) 46–53.
[21] E.T. Esfahani, V. Sundararajan, Using brain-computer interface to detect human satisfaction in human-robot interaction, Int. J. Humanoid Rob. 08 (01) (2011) 87–101.
[22] C.A. Frantzidis, C. Bratsas, C.L. Papadelis, E. Konstantinidis, C. Pappas, P.D. Bamidis, Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli, IEEE Trans. Inf Technol. Biomed. 14 (3) (2010) 589–597.
[23] R.W. Picard, E. Vyzas, J. Healey, Toward machine emotional intelligence: analysis of affective physiological state, IEEE Trans. Pattern Anal. Mach. Intell. 23 (10) (2001) 1175–1191.
[24] A. Gevins, M.E. Smith, H. Leong, L. McEvoy, S. Whitfield, R. Du, G. Rush, Monitoring working memory load during computer-based tasks with EEG pattern recognition methods, Hum. Factors 40 (1) (1998) 79–91.
[25] C. Berka, D.J. Levendowski, M.M. Cvetinovic, M.M. Petrovic, G. Davis, M.N. Lumicao, V.T. Zivkovic, M.V. Popovic, R. Olmstead, Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset, Int. J. Hum.-Comput. Interact. 17 (2) (2004) 151–170.
[26] A.S. Ashrae, Standard 90.1-2004, Energy Standard For Buildings Except Low Rise Residential Buildings, American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc, 2004.
[27] J.A. Russell, Affective space is bipolar, J. Pers. Soc. Psychol. 37 (1979) 345–356.
[28] P.J. Lang, M.M. Bradley, B.N. Cuthbert, International affective picture system (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-8), University of Florida, Gainesville, FL, 2008.
[29] M.M. Bradley, P.J. Lang, The International Affective Digitized Sounds: Affective ratings of sounds and instruction manual (Technical Report No. B-3), University of Florida, Gainesville, FL, 2007.
[30] G.J. Raw, M.S. Roys, C. Whitehead, D. Tong, Questionnaire design for sick building syndrome: An empirical comparison of options, Environ. Int. 22 (1) (1996) 61–72.
[31] P. Wargocki, D.P. Wyon, Y.K. Baik, G. Clausen, P.O. Fanger, Perceived air quality, sick building syndrome (SBS) symptoms and productivity in an office with two different pollution loads, Indoor Air 9 (3) (1999) 165–179.
[32] N. Gong, K.W. Tham, A.K. Melikov, D.P. Wyon, S.C. Sekhar, K.W. Cheong, The acceptable air velocity range for local air movement in the tropics, HVAC&R Res. 12 (4) (2006) 1065–1076.
[33] A.K. Melikov, B. Krejciríková, J. Kaczmarczyk, M. Duszyk, T. Sakoi, Human response to local convective and radiant cooling in a warm environment, HVAC&R Res. 19 (8) (2013) 1023–1032.
[34] B.H. Kantowitz, H.L. Roediger, D.G. Elmes, Experimental Psychology, ninth ed., Wadsworth Cengage Learning, Belmont, CA, 2009.
[35] R.L. Solso, O.H. Maclin, M.K. MacLin, Cognitive Psychology, eighth ed., Pearson/Allyn and Bacon, Boston, 2008.
[36] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (2004) 9–21.
[37] Matlab R2013b, The MathWorks, Inc., Natick, Massachusetts, United States.
[38] C.L. Dickter, P.D. Kieffaber, EEG Methods for the Psychological Sciences, SAGE Publications, London, 2014.
[39] S. Sanei, J.A. Chambers, EEG Signal Processing, John Wiley & Sons, West Sussex, England, 2007.
[40] J. Healey, R.W. Picard, Digital processing of affective signals, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. 6 (1998) 3749–3752.
[41] S. Hamann, T. Canli, Individual differences in emotion processing, Curr. Opin. Neurobiol. 14 (2) (2004) 233–238.
[42] J. Hua, Z. Xiong, J. Lowey, E. Suh, E.R. Dougherty, Optimal number of features as a function of sample size for various classification rules, Bioinformatics 21 (8) (2005) 1509–1515.
[43] J. Ye, T. Li, T. Xiong, R. Janardan, Using uncorrelated discriminant analysis for tissue classification with gene expression data, IEEE/ACM Trans. Comput. Biol. Bioinf. 1 (4) (2004) 181–190.
[44] C.D. Katsis, N. Katertsidis, G. Ganiatsas, D.I. Fotiadis, Toward emotion recognition in car-racing drivers: a biosignal processing approach, IEEE Trans. Syst. Man. Cybern. A, Syst. Hum. 38 (3) (2008) 502–512.
[45] K.H. Kim, S.W. Bang, S.R. Kim, Emotion recognition system using short-term monitoring of physiological signals, Med. Biol. Eng. Compu. 42 (2004) 419–427.
[46] X. Shan, E.-H. Yang, J. Zhou, V.W.-C. Chang, Neural-signal electroencephalogram (EEG) methods to improve human-building interaction under different indoor air quality, Energy Build. 197 (2019) 188–195.