Machine learning classification of design team members' body language patterns for real time emotional state detection

Ishan Behoora and *Conrad S. Tucker*, Industrial and Manufacturing Engineering, Engineering Design, Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

Design team interactions are one of the least understood aspects of the engineering design process. Given the integral role that designers play in the engineering design process, understanding the emotional states of individual design team members will help us quantify interpersonal interactions and how those interactions affect resulting design solutions. The methodology presented in this paper enables automated detection of individual team member's emotional states using non-wearable sensors. The methodology uses the link between body language and emotions to detect emotional states with accuracies above 98%. A case study involving human participants, enacting eight body language poses relevant to design teams, is used to illustrate the effectiveness of the methodology. This will enable researchers to further understand design team interactions.

Keywords: computational models, information processing, design activity, team work, user behavior

Engineering design is "widely considered to be the central or distinguishing activity of engineering" (Bucciarelli, 1994). Yet, it remains an insufficiently researched and understood topic (Ackroyd, 2006; Dym, Agogino, Eris, Frey, & Leifer, 2005). In particular, interactions within design teams are amongst the least understood aspects associated with the engineering design process. This is due to the dynamic, nonlinear and often loosely coupled nature of design (Edmondson & Nembhard, 2009). While there are established methods of evaluating the ideas and concepts generated by a design team (Goldschmidt & Tatsa, 2005; Liu, Chakrabarti, & Bligh, 2003; Shah, Kulkarni, & Vargas-Hernandez, 2000), the process of generating these ideas and concepts remains difficult to study. Existing approaches rely heavily on hand coding of video-recorded or observed interactions and design team member surveys (Brannick & Prince, 1997). An important aspect of the team dynamic is the interpersonal interactions between its team members. The emotions expressed by the individuals during these interactions can lead to insights about the team's dynamics. Traditional self-reported feedback of team interactions is often unreliable because it is susceptible to user

Corresponding author:
Conrad S. Tucker
ctucker4@psu.edu

www.elsevier.com/locate/destud
0142-694X *Design Studies* **39** (2015) 100–127
http://dx.doi.org/10.1016/j.destud.2015.04.003

reported biases (Barker, Pistrang, & Elliott, 2002). Additionally, this feedback is not in real time. Thus, there is a need for a system that can capture individuals' emotions in real time. Such a system would allow for a better understanding of interpersonal interactions in design teams.

Body language has been linked to emotional states by past studies (Panksepp, 1998). An individual can exhibit various body language poses, depending on whether he or she is interested, bored, frustrated, delighted etc. (Birdwhistell, 2010; Panksepp, 1998). Unfortunately, having a human observer assess body language poses exhibited by design team members can be costly and time consuming. Thus, there is a need for an automated system. Studies have been conducted that quantify body language using various sensors such as pressure sensitive chairs and motion tracking suits or by measuring other reactions such as pupil dilation (Craig, Graesser, Sullins, & Gholson, 2004; Kapoor & Picard, 2005). Unfortunately, such methods require expensive specialized, wearable hardware. To address these factors, the authors of this work propose a machine learning driven approach that utilizes off-the-shelf, non-wearable sensors to detect individuals' body language in a real time minimally-invasive manner. This approach enables researchers to quantify the emotional states of individual team members in a design team and thus, better understand the team dynamics. The methodology outlined in this work demonstrates the efficacy of non-wearable sensors and machine learning algorithms to model individuals' body language in non-design tasks with the ultimate goal of applying these methods to quantify design team interactions.

This paper is organized into four sections. In Section 1, the authors provide an overview of related literature, followed by the methodology in Section 2. Thereafter, the authors illustrate the methodology in practice with a case study in Section 3, before finally concluding in Section 4.

1 Literature review

1.1 Team dynamics and human emotions

Team dynamics is a complicated research topic. In engineering design teams in particular, the process rarely follows a linear, prescribed methodology (Stempfle & Badke-Schaub, 2002). A variety of factors such as team members' work load, time pressures, etc., are at play. Studies have shown that design team interactions are an interplay of design discussions, walkthroughs and progress evaluations (Olson, Olson, Carter, & Storrosten, 1992). Within this context, past research has shown that social aspects also interact significantly with the technical and cognitive processes of design (Cross & Cross, 1995). Additionally, conscientiousness, agreeability and emotional stability are positively related to job performance involving interpersonal interactions (Mount, Barrick, & Stewart, 1998). The success of teams is modeled using team members' behaviors as key elements (Reilly, Lynn, & Aronson, 2002). Also, the use of Myers-Briggs personality types has been studied with respect to learning in teams within the education system (Jensen, Wood, & Wood, 2003). Learning itself has also been shown to be associated with emotions (Marchand & Gutierrez, 2012; Munoz & Tucker, 2014; Shen, Wang, and Shen, 2009). In recent years, emotional state detection has been shown to be a useful tool with the potential to improve learning (Craig et al., 2004; D'Mello and Graesser 2011; Graesser et al., 2006; Shen et al., 2009). Design teams involve interactions where there is learning between team members. Therefore, the link between emotional states and learning further highlights the importance of team members' emotional states (Baker et al., 2012; Won & Bailenson, 2014).

The mood and behavior of individuals can have an impact on the entire team. The usefulness of involving all team members in the design process has been shown to have both a perceptive and productive impact (Fisher, 1993; Stasser & Dietz-Uhler, 2001; Yang, 2010). Individual team members can spread both positive and negative emotions throughout the entire team (Kelly & Barsade, 2001) and their belief structures play a role in the development of shared cognition (Klimoski & Mohammed, 1994). Additionally, non-verbal communication has been shown to have an impact in design team interactions (Le Dantec & Do, 2009). Such aspects have become even more important with the spread and effectiveness of cross-functional teams (Parker, 2003). Thus, understanding individual team member's emotional state is of importance as it can enable researchers to gain insight into team dynamics. Finally, regular assessment has been shown to improve team performance (Busseri & Palmer, 2000). A real time automated system, that captures design team members' emotional states, can help address these challenges and potentially result in significant team performance improvements.

1.2 Body language and emotional states

Body language plays a critical role in the communication process. It provides cues to detect various aspects of an individual's mental state. According to Birdwhistell (Birdwhistell, 2010), words represent only 7% of the communication process, while non-verbal communication represents 55%. Various studies have shown that body language and individuals' emotional states, are closely related (Panksepp, 1998). This linkage is usually referred as *emotional body language*. Wallbott analyzed body language cues and its link to various sets of emotions (Wallbott, 1998). Clear patterns were encountered for body movement and posture representing different emotions. A recent study showed that body poses were more effective for discriminating between intense positive and negative emotions than facial expressions (Aviezer, Trope, & Todorov, 2012). Other studies have supported the idea that intense emotions can be recognized more accurately (Gao & Maurer, 2009; Marneweck, Loftus, & Hammond, 2013). Thus, body language cues can be used to

Table 1 Emotional states and their associated body language

Emotional state	*Typical behavior*
Engagement/Interest	Slightly inclined forward, Nodding head, Head tilting, Feet pointing toward the speaker, Cross fingers, Rub palms against each other, Eye contact
Delight	Clapping hands, Laughing
Frustration	Hands on hip, Scratching hair or back part of the neck, Drumming fingers Hands clasped behind back
Boredom	Yawning frequently, Chin resting on hand, Pulling ears, Tapping hands or feet, fidgeting, Support hand on cheek, Slouching, Sitting with legs crossed Laid back and foot kicking

determine emotional states. In Table 1, typical body language poses associated with various emotional states are shown (Baker, D'Mello, Rodrigo, & Graesser, 2010; Coker & Burghoon, 1987; D'Mello & Graesser, 2009; Eastwood, Frischen, Fenske, & Smilek, 2012; Furnham & Petrova, 2010; Pease & Pease, 2008; Provine, 1986).

1.3 Automated detection of body language

Past studies have achieved high rates of accuracy in detecting emotional states such as boredom, frustration, confusion, engagement, delight and interest using sensors. Kapoor and Picard (Kapoor & Picard, 2005) used a camera and a pressure sensitive chair to classify participants' emotional states denoting interest or disinterest. The camera allowed them to track facial features as well as head gestures. Their proposed classification methods obtained a recognition rate of 86%. However, a rate of 82% was achieved when using only posture. A single-mode system using pupil tracking was proposed by Kapoor and Picard (Kapoor & Picard, 2001) for real-time detection of head nods and head shakes, which are body language poses indicating interest (Panksepp, 1998). The recognition achieved was 78.46%. Frustration has also been detected by automated systems in an intelligent system environment. Kapoor et al. (Kapoor, Burleson, & Picard, 2007) achieved a prediction accuracy of 79%. In their approach, non-verbal behavior was captured through a camera, pressure sensitive chair, mouse, and a skin conductance device. An automated emotion detection system using students' gross body language was proposed by D'Mello and Graesser (D'Mello & Graesser, 2009). They studied the states of boredom, confusion, delight, flow, and frustration. The detection accuracies were 73%, 72%, 70%, 83%, and 74% respectively while using a pressure sensitive seat to detect emotional states.

Several studies support that body language could be used to detect various emotions over time with relatively good accuracy, if compared with other single or multi-channels (D'Mello & Graesser, 2009). However, many of the proposed methodologies found in the literature require wearable sensors for each participant. While wearable sensors (e.g., eye tracking devices) are becoming less invasive (Kassner, Patera, & Bulling, 2014; Spagnolli, Guardigli, Orso, Varotto, & Gamberini, 2014; Ye et al., 2012), they still face negative social perceptions (Bodine & Gemperle, 2003; Hong, 2013). Furthermore, the constraint that each individual in a design team would need a separate wearable device presents scalability challenges in real world engineering design scenarios. The authors of this work seek to address these limitations by utilizing commercial, off-the-shelf, non-wearable sensors to capture human body language data, which is then mined for distinct body language patterns using machine learning techniques.

1.4 Modeling body poses using skeletal joint data

With the growth of motion capture technologies, the use of skeletal joint data inferred form individuals' body language poses is becoming more prevalent. Additionally, with the advent of relatively low cost infrared cameras, tracking human skeletal joints has become readily accessible (Shotton et al., 2013). For example, using skeletal data captured by the Microsoft Kinect, researchers were able to employ dynamic time warping techniques, a template matching algorithm from speech recognition, to recognize gestures made by humans, and classify them according to joint positions (Celebi, Aydin, Temiz, & Arici, 2013). In fact, human gesture and action recognition using skeletal data has proven to be a very effective approach with high accuracies achieved using various machine learning techniques such as support vector machines (Chung & Yang, 2013; Miranda et al., 2012; Wang & Lee, 2009; Xia, Chen, & Aggarwal, 2012). Researchers have also been able to illustrate real time classification of dance gestures with an average accuracy of 96.9%, in spite of noisy sensor data (Raptis, Kirovski, & Hoppe, 2011). In order to combat noise in skeletal data, processing of the raw joint position data into angular velocities between joints and their ratios has proved to be helpful in increasing accuracies (Miranda et al., 2012; Raptis et al., 2011). Evaluations of dance performances using skeletal data have also been explored (Alexiadis et al., 2011). Additionally, the non-wearable sensors have been utilized to recognize sign language. Researchers were able to use the Kinect to achieve 76% sentence verification in adults while standing and 51% sentence verification in adults while seated (Zafrulla, Brashear, Starner, Hamilton, & Presti, 2011). Another study utilized hidden Markov models with a continuous observation density and was able to achieve a recognition rate of 97% in initial results (Lang, Block, & Rojas, 2012). In other works, cues such as inferring respiratory rates (by the rising and falling of the chest) and detecting fidgeting (by detecting rapid oscillations of a person's knee), have been explored (Burba, Bolas, Krum, & Suma, 2012). Manohar and Tucker utilized skeletal joint data to predict the emergence of human threats in an audience (Manohar & Tucker, 2013). Such techniques illustrate the potential of machine learning methods using skeletal data. The authors of this work explore the use of machine learning methods using skeletal data to classify body language poses and infer individuals' emotional states within a design team.

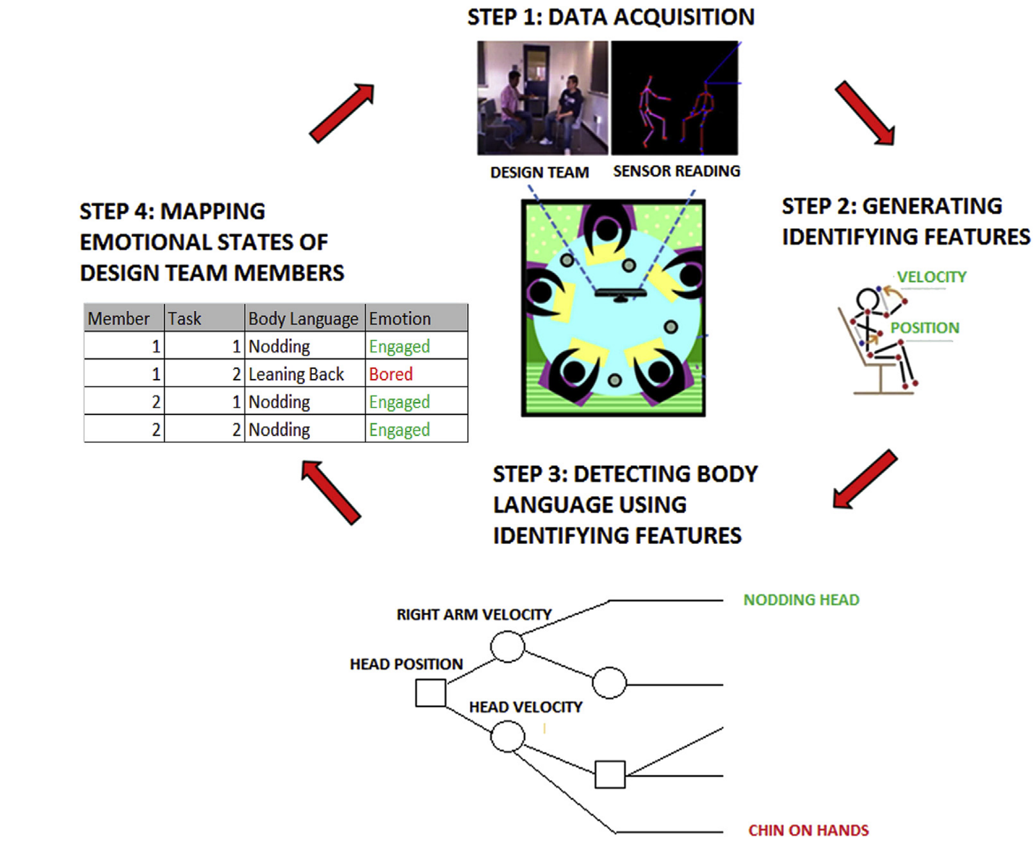

Figure 1 Overview of the methodology: using non-wearable sensors to predict emotional states

2 Methodology

The methodology presented in Figure 1 proposes the use of non-wearable sensors and machine learning algorithms to infer individuals' emotional states in a design team. The methodology demonstrates the efficacy of non-wearable sensors and machine learning algorithms to model individuals' body language in non-design tasks, with the ultimate goal of applying these methods to quantify design team interactions. This inference is based on the acquisition and classification of different body language poses that can be captured for each member of a design team using non-wearable sensors. Step 1 of the proposed methodology (Data Acquisition in Figure 1) utilizes non-wearable sensors to acquire and store skeletal joint data from a particular individual in the design team. This allows for minimally-invasive and real time acquisition of body movement data. As can be seen from Step 1 in Figure 1, existing sensors are capable of capturing multiple skeletal joint images using a single sensor device, hereby mitigating the need to a separate sensor for each design team member. For the purposes of this study, the skeletal data for one team member will be studied. The expansion towards simultaneously capturing multiple skeletal data pertaining to each team member would follow the similar steps outlined in Figure 1 and is a topic for future work. Step 2 of the methodology is the processing of the acquired skeletal joint data to generate 3D position, velocity, and acceleration values for the skeletal joints. This generates identifying features in order to classify body language poses. Step 3 of the methodology employs machine learning algorithms on the generated set of features (i.e., 3D position, velocity, and acceleration) to classify the body language exhibited by an individual into one of several mutually exclusive emotional states (e.g., engagement, frustration and boredom). Step 4 is the interpretation of the quantified body language states observed into emotional states of the individual. Within the context of a design team, this allows for real time tracking of the emotional states of the individuals comprising the design team and how those emotional states evolve over time.

2.1 Data acquisition

Step 1 of the proposed methodology utilizes low cost, off the shelf infrared sensors (e.g., Microsoft Kinect) to acquire data pertaining to an individual's body movement patterns. While the Kinect has been used extensively to capture body movement patters in different applications, the methodology is not limited to the use of a specific sensor hardware. For example, alternative non-wearable sensors such as Asus Xtion Live ("Xtion PRO – Overview" 2015), Primesense Carmine, now owned by Apple ("Depth Sensors Comparison – iPiSoft Wiki" 2015), etc., are capable of achieving similar data acquisition requirements and serve as practical alternatives. Non-wearable infrared sensors can approximate nodes on a human body in a minimally-invasive manner due to the absence of the wearable aspect. This data can be used to map the 3D locations of the joints of an individual's skeletal system. With each reading, the sensor collects **X, Y, Z** coordinate data for k joints, representing various points on the human skeleton as shown in Figure 2. The number of joints tracked, k, varies depending on the sensor used. The sensor collects the 3D coordinates of k skeletal joints represented by the red (in the web version) dots on the participant in Figure 2. A single data sample is illustrated with $k = 10$ in Figure 2 (the origin of the coordinate system is the sensor itself and sample readings of two joints are indicated in parenthesis). In the context of a design team, this non-wearable technique of skeletal joint data acquisition enables skeletal data to be acquired for each member of a design team. As mentioned earlier, existing sensors are capable of simultaneously capturing skeletal data for multiple individuals using a single sensor device, hereby mitigating the need for multiple sensor devices for a typical engineering design team.

The non-wearable sensors such as those discussed in this section and illustrated in Figure 2 are portable and can be connected to a desktop or mobile computing device (e.g., tablet device), depending on the location and data needs of the design team. Research teams have even created Wi-Fi enabled

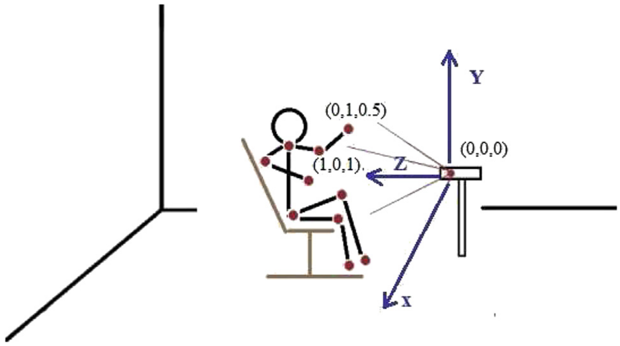

Figure 2 Skeletal joint data collection

solutions so that sensors can capture and remotely stream data to a computer in another location ("Projects | Mobile Kinect | BIG" 2015).

2.2 Raw skeletal data processing

Step 2 of the proposed methodology uses the raw data acquired by the sensors to generate identifying features that can then be used to characterize human body language. For each reading taken by the sensors in step 1, a time stamp and 3D coordinate (**X, Y, Z**) positions for each of the k joints are acquired, generating a total of 3 k features.

To effectively capture human body language (i.e., motion and temporal characteristics), multiple readings are taken by the sensor to calculate the velocity and the acceleration of each joint. The velocity of each of the k joints are subsequently used to calculate the acceleration. This is illustrated in Figure 3 above. Thus, for each tuple in the data set, there are 9 k-features whose elements represent the position, velocity and acceleration in 3 dimensions for each of the k joints. Each reading is considered an independent sample so as to minimize assumptions about body language poses.

2.3 Machine learning to quantify body language states

Machine learning classification builds a predictive model based on the available data. Typically, the process consists of getting a feature set (input) which the machine learning algorithm has to correctly label (output). In supervised machine learning, training data is provided that consists of example feature sets and their correct labels. This training data is then used to build a classification model and thereafter, for any subsequent feature sets, the machine learning algorithm can use its model to predict the label. In order to quantify the ability of machine learning algorithms to accurately classify body language data, the label/output variable consists of q known emotion states. The machine learning models aim to identify body language poses and classify each reading as one of these q poses using the r-feature set generated, where $r = 9\ k$. The r-feature set serves as input variables (features) to a classifier

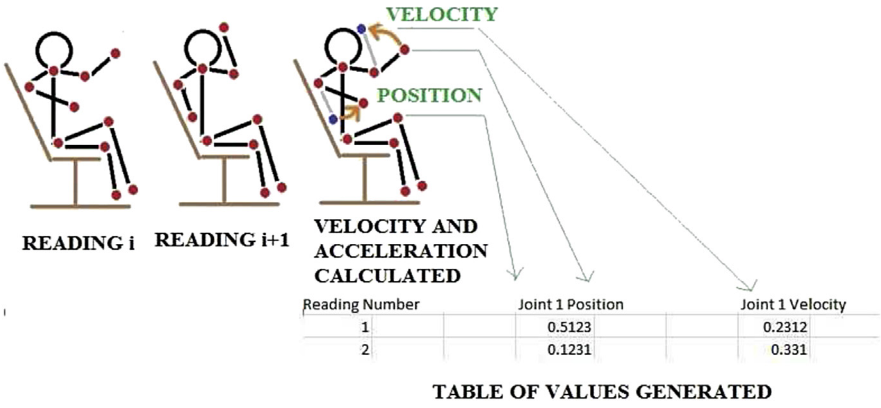

Figure 3 Data generation from sensor readings

and the output is the classification into one of the q known body language poses (label) representing an individual's emotional state. The classifier is trained using a data set of participants enacting or exhibiting known body language poses (i.e., ground truth) and can be used to classify new participants in a real time, dynamic manner. The authors explore the use of the following four machine learning classifiers. The suitability of the models for the methodology is outlined in the subsections for each classifier.

2.3.1 *C4.5*

The *C4.5* decision tree classifier is a supervised machine learning algorithm that builds a decision tree from the training set on the basis of information entropy. At each node of the tree, the algorithm chooses the features that most effectively splits its input into subsets enriched in one label or the other (Murthy, 1998). C4.5 has a good combination of error rate and low computational complexity for learning and classification (Lim, Loh, and Shih 2000; Quinlan, 2014). Additionally, decision trees have high comprehensibility. Unfortunately, Decision trees tend to perform better when dealing with discrete features, not continuous variables such as velocity and acceleration (Kotsiantis, 2007).

2.3.2 *IBK*

The IBK classifier is a supervised machine learning classifier that has low computational complexity for learning (Mitchell, 1997). The algorithm classifies a particular feature set by using the majority vote of the labels of its n closest feature sets from the labeled training data (Cover & Hart, 1967). Its main disadvantages are that it has large storage requirements and the choice of n affects the effectiveness of the classifier. The IBK classifier selects appropriate values of n based on cross-validation. However, this adds to the computational complexity (Guo, Wang, Bell, Bi, & Greer, 2003). Additionally, it is susceptible to irrelevant and noisy data and has high computational complexity for classification which is a disadvantage, as the algorithm classifies each reading (Wettschereck, Aha, & Mohri, 1997). However, such n nearest neighbor classifiers have demonstrated good performance in real life applications. In addition, their low computational complexity for learning makes them desirable (Kotsiantis, 2007).

2.3.3 Random forest

The Random Forest (RF) algorithm is an ensemble classifier that operates on the principle of generating a random forest of decision trees to classify the problem (Breiman, 2001). The method combines the existing approaches of bagging (Ho, 1995) and the random selection of features (Amit & Geman, 1997) to improve performance. Random forest provide better error rates comparable to boosting techniques such as ADABoost but is more robust to noise and avoid over fitting (Breiman, 2001). Additionally, the Random Forest algorithm has been shown to have good real world performance across a variety of domains (Verikas, Gelzinis, & Bacauskiene 2011).

2.3.4 Naïve Bayesian

Naive Bayesian is a probabilistic classifier based on Bayes' theorem with strong independence assumptions. The presence or absence of a particular feature is considered unrelated to the presence or absence of any other feature (Kotsiantis, 2007). Despite its simplified nature, Naive Bayes classifiers have worked quite well in complex, real-world situations (Domingos & Pazzani, 1997; Kononenko, 1990; Zhang & Su, 2004). Additionally, analysis has shown that there is theoretical basis for the surprising efficiency of the classifier, despite its feature independence assumptions (Zhang & Su, 2004). Another advantage of the algorithm is that it often requires a small amount of training data to perform well (Kotsiantis, 2007).

The authors' motivation behind choosing these four classifiers is that the classifiers have been shown to have good real world performance across a wide variety of classification areas (Kotsiantis, 2007). The relative pros and cons of various popular classifiers has been studied (Kotsiantis, 2007), making them well suited for direct comparison. If we assume four stars represents the best performance for a metric and one star represents the worst performance (Table 2), we can illustrate the advantages and disadvantages of various machine learning classifiers. Support vector machines (SVMs) have high accuracy across a multitude of applications and are thus assigned a score of four stars for model accuracy. The C4.5 decision tree classifier is typically not as accurate as SVMs and is assigned a score of two stars, which is still better than the performance of the Naïve Bayesian classifier which performs the worst and is assigned a score of one star. However, SVMs suffer from slow training speed, when compared with the other classifiers. The performance variations of these classifiers across various metrics is summarized in Table 2 below. Random forest provides us with high accuracy while maintaining performance in the other three metrics comparable to decision tree classifiers (Verikas et al., 2011) and

Table 2 Comparison of classifiers used (Kotsiantis, 2007), where** represents the best and * represents the worst performance**

	Decision tree based algorithms (C4.5, random forest)	*SVM*	*IBk*	*Naïve Bayes*
Model accuracy	**	****	**	*
Training speed	***	*	****	****
Classification speed	****	****	*	****
Tolerance to noise	**	**	*	***

are thus chosen in place of SVMs for the methodology. However, if training is performed offline, while classification is performed online, SVMs may also be a suitable candidate machine learning algorithm for this design problem. In this work, it is assumed that training may have to be performed online in cases involving concept drift, a phenomenon in data stream mining where there is a shift in the target variable, resulting in a diminished accuracy of the data mining models (Li, Hu, & Wu 2008; Tucker and Kim 2011). For example, if there becomes a new societal gesture involving the hand on cheek that shifts for representing the emotional state of boredom, then one would want the machine learning model to process this new incoming data stream and retrain the model online in a timely and efficient manner.

2.4 Emotion detection and feedback

Finally, once the machine learning algorithms have classified the readings into the q known body language poses, the model uses the known links between body language poses and emotions to predict the individual's emotional state. This step links the body language exhibited by a team member to the emotional state of the team member in real time. This results in knowledge of the emotional states of the team member during the entire design process and thus enables fine grained analysis of the emotional states exhibited. Furthermore, this knowledge, when used in conjunction with other design team metrics such as productivity and design solution, will reveal the extent of these correlations. Additionally, feedback from the results can be incorporated to improve the classification stage of the methodology. This iterative cycle of data collection, processing, classification and feedback for improvement, provides a continuously improving methodology for non-wearable real time emotion detection. An example is illustrated in Figure 4 below.

3 Case study

A case study demonstrating the feasibility of the proposed methodology was conducted using participants in an engineering design laboratory. This study was performed in accordance with IRB guidelines as per Penn State's IRB 40258: "A Dynamic Pattern Recognition Framework for Mining and Predicting Emerging Threats". For the case study, each participant is recorded enacting various body language poses which were then classified using the machine learning algorithms discussed in Section 2.3. The enacted poses and the

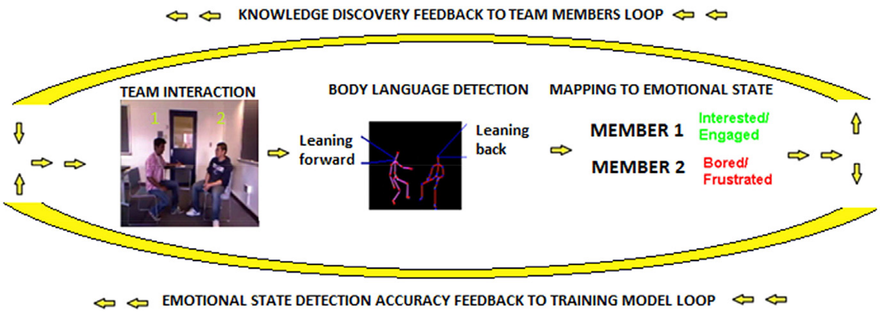

Figure 4 Example of emotion state detection

emotions they represent were chosen to be from those expected during design team interactions and are listed in Table 3 below.

For example, the IDEO shopping cart documentary showcases a design team deeply engaged in the brainstorming process of design (ABC Nightline – IDEO Shopping Cart 2009). The IDEO design team exhibits many of the body language poses utilized in this study, which have also been shown to be relevant to emotional states (Aviezer et al., 2012). Some of the body language poses seen in the shopping cart video include *tilted head* (@1:42 min), *nodding head* (@1:56 min) and *chin on hand* (@4:37 min). Individually customized emotional state models can be generated for each team member, provided that each team member has provided ground truth data to validate the predicted emotional states. The following design scenario outlines the potential for the proposed research.

3.1 Design scenario

A design team is focused on creating a design solution (Y). While there exists well established methods for evaluating design concepts and solutions, the design team would like to understand the design team interactions (I) that influence these resulting design solutions. E.g., what combination of emotions expressed by designers within a team influence the quality of design solution (Y)? Prior to the design challenge, characteristics (X) of each member of a design team can be acquired such as:

Table 3 Enacted body language and associated emotions

Body language	*Emotion*
Tilted head	Engagement/Interest
Nodding head	Engagement/Interest
Hands behind back	Frustration
Scratching back of the neck	Frustration
Pulling of ears	Boredom
Chin on hand	Boredom
Hand on cheek	Boredom
Neutral	Neutral

- Design expertise
- Personality type (introvert, extrovert, etc.)
- Demographics (age, gender, etc.)

Given this data, researchers can answer a wide range of questions pertaining to a design team, based on prior design team interaction models (I). For example, what characteristics (X) of a design team, correlate with desired/undesired design team interactions (I) such as engagement or boredom? Beyond a deeper understanding of the composition of a design team, designers may want to understand what aspects of design team interactions (I), result in the most successful or creative design solutions (Y). The ability to quantify correlations between design team interactions and either the design team characteristics (X) or resulting design solutions (Y) require an accurate method of quantifying design team interactions (I). In the case study presented in this work, a design team would utilize the non-invasive sensors as a means of generating a baseline model of design team interactions in a non-invasive manner. The validity of the non-wearable approach to modeling design team interactions could then be evaluated based on either survey feedback from each of the design team members about their emotional state during a given design process or through more quantitative methods such as the use of wearable eye tracking devices that are becoming less invasive (Kassner, Patera, & Bulling 2014; Spagnolli et al. 2014; Ye et al., 2012). Therefore given the proposed non-wearable system for capturing and inferring body language poses, designers will be able to explore both the design team characteristics (X) and the design team solutions (Y). However, before such correlations can be explored, the validity of utilizing non-wearable approaches to capturing and mining design team interactions must be established. The case study presented in this work demonstrates the feasibility of capturing and mining individual body postures with high accuracy. From this model, designers will be able to explore research questions pertaining to either their design team characteristics or design solutions.

3.2 *Data acquisition*

For the experiment, the Microsoft Kinect multimodal sensor was used to capture 3D skeletal images in a minimally-invasive manner. The MS Kinect is an off-the-shelf, low cost sensor. The Kinect is capable of tracking 20 joints in the human skeleton via its infrared sensor which captures data at the rate of 30 Hz and at a resolution of 640×480. The horizontal and vertical fields of view of the sensor are 57° and 43° respectively (Breiman, 2001). Additionally, while the sensor's accuracy decreases with distance, the maximum error reaches 1.5 inches. These factors make it suitable for our study (Khoshelham & Elberink, 2012). Although, the authors utilized the Kinect for their case study, other devices with similar capabilities such as PrimeSense are available as an alternative.

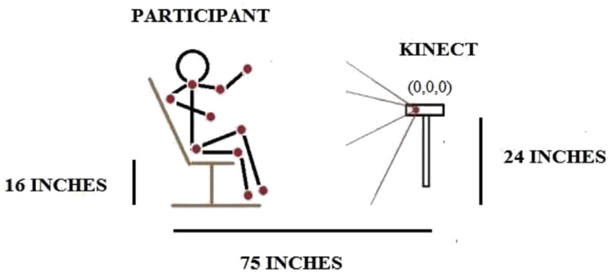

Figure 5 Experimental setup

During the experiment, the Kinect was configured at an elevation of 2 feet above the floor as seen in Figure 5. For each participant, data was collected while they were seated at a distance of 75 inches from the Kinect on a chair that was 16 inches above the ground level. This is illustrated in Figure 5 below.

For our study, each participant was then instructed to enact each of the 8 body language poses for 12 s. For poses involving motion such as nodding, the participants repeated the movements. For gesture learning, human variations can often be accounted for by using as few as 3 participants (Gillian, Knapp, & O'Modhrain 2011). The training data requires the class variable to have mutually exclusive postures so that once a model is generated, test data can be assigned to one of the mutually exclusive classes. The presence of more than one individual enacting a body posture introduces variability in the manner in which the same body posture can be enacted, hereby making the resulting model more robust to unseen test data. For our study, we recorded 4 participants, with each participant generating approximately 150 reading for each pose. The focus is the number of independent labeled samples obtained for the classifier. Each such reading, captured once every 33 ms, represents an independent sample. Thus for the study, the authors obtained 4072 labeled samples split evenly among the 8 poses. Examples of the skeletal joint data collected during the poses are visualized in Figure 6. As can be seen from Figure 6, the left image presents skeletal joint data of a person sitting, while the right image in Figure 6 presents skeletal joint data of a person sitting and pulling their ear. The positions of the head, elbow and knee joints are included in Figures 6 and 7 for clarity and reference.

The body language of scratching the back of one's neck (left image in Figure 7) and the tilting of one's head (right image in Figure 7) are visually distinguishable, as seen in Figure 7.

3.3 Data processing

Once the data collection is complete, an r-feature set is generated, as described in the methodology. For the Kinect, 20 joints can be tracked, $k = 20$.

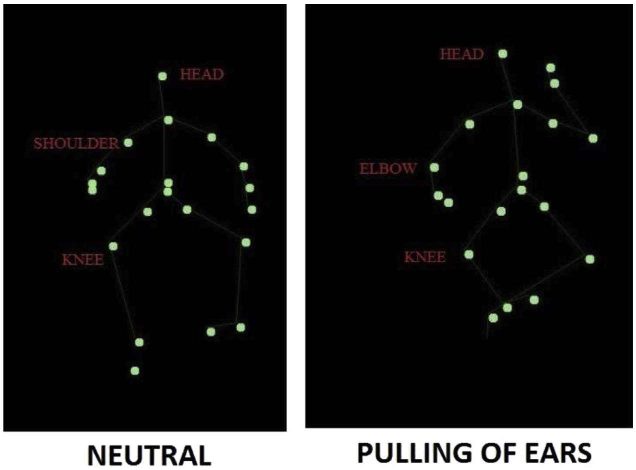

Figure 6 Pulling of ears

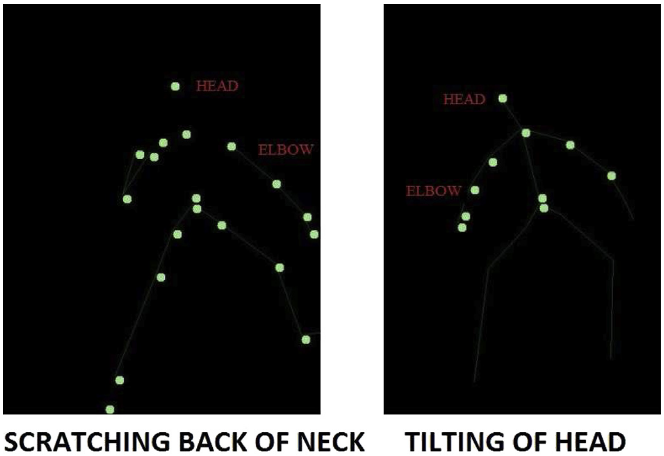

Figure 7 Scratching the back of the neck and tilting of head

1. Only readings where the 3D positions of the all 20 joints were measured by the sensor were considered, with the rest discarded. This results in 60-features for each data sample acquired by the Kinect (i.e., 20 features pertaining to the X coordinate, 20 features pertaining to the Y coordinate and 20 features pertaining to the Z coordinate). Each reading represents an independent sample.
2. As per the methodology, the velocity and acceleration of each node are also generated in X, Y and Z coordinate space, resulting in 120 additional features. In total, 60 features relate to the 3D position coordinates, 60 features relate to the velocity, 60 features relate to the acceleration.
3. A data set was created from the readings generated by the participants enacting the 8 body language poses. These were each labeled, based on the body language pose that the participants were enacting.

4. Given a combination of the input features (180 features representing XYZ coordinate data of position, velocity and acceleration), the goal is to classify input data into one of the eight 8 body language poses.

Waikato Environment for Knowledge Analysis (WEKA) (Hall et al. 2009) software was used to execute the four machine learning techniques described in the methodology. For the purposes of this study the default parameters of the WEKA software's machine learning algorithms were used.

3.4 Machine learning classification of body language

In order for the methodology to be viable, the accuracy and robustness of the model needs to be demonstrated. Given the static nature of the body language poses and the focus on individual emotion state models for each team member, the authors performed 10-fold cross validation (CV) and used the four machine learning algorithms to mine the data. The original training data for the entire study contained 4072 tuples that are split into training and test data sets, based on the 10-fold CV. CV involves partitioning the collected data set into two subsets; training the predictive model on one subset called the training data set, and testing/validating the model performance using the other, called the *test data* set. Typically, the split used is 90% assigned as the training and 10% as the test data set for 10 fold CV (Kohavi, 1995). For scenarios where generalizations of emotional states are needed with minimum retaining of the model for each individual, the accuracy of classification models should be evaluated, based on a leave-one-out sampling method (Refaeilzadeh, Tang, & Liu 2009).

3.4.1 Classification results

The results in Table 4 represent the accuracies of the classifiers in predicting the body language poses exhibited, given a set of input movements (i.e., based on the 180 features representing XYZ position, velocity and acceleration). Analysis of the C4.5 classifier revealed the decision tree's size was 87. This indicates there is a significant decrease in the dimensionality of the feature space. Additional data and tests can be used to select an optimal set of features. Next, the authors explored how accuracies changed as data from additional participants were added. This is illustrated in Table 4 and Figure 8. The accuracies reported in Tables 4, 5 and 7, represent average accuracies over the 10-fold CV. The resulting predictive errors reported for each algorithm therefore

Table 4 Classifier accuracy variation across number of participants

Number of Participants	*C4.5*	*Random forest*	*IBk*	*Naïve Bayesian*
1	99.59%	100%	99.92%	98.44%
2	98.25%	99.85%	98.97%	81.45%
3	98.36%	99.94%	99.04%	62.78%
4	98.40%	99.85%	99.51%	53.44%

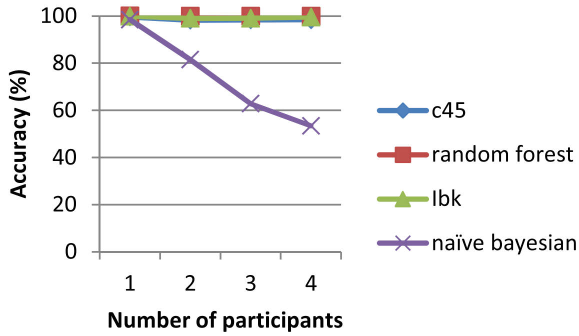

Figure 8 Accuracy across number of participants

Table 5 Accuracy percentage variation depending on attributes used

Features	*C4.5*	*Random forest*	*IBk*	*Naïve Bayesian*
Position	98.79%	99.95%	99.92%	60.16%
+Velocity	98.55%	99.90%	99.87%	53.33%
+Acceleration	98.40%	99.85%	99.50%	53.17%

represent the average of the 10 models generated during the 10-fold cross validation step. Table 6 outlines the source of the 10-fold CV model errors for each of the models, including the Naïve Bayesian algorithm, given that it has lower performance, compared to the other algorithms, especially as the number of participants increases.

Table 6 Confusion matrix for each of the machine learning classifiers

C4.5 Decision Tree

a	b	c	d	e	f	g	h	<-- classified as
436	0	0	2	1	2	0	1	a = neutral
0	596	0	1	0	0	1	0	b = nod
0	1	514	3	0	1	1	1	c = tilt
2	2	0	460	2	1	1	5	d = scratch
0	1	2	4	501	0	2	0	e = hback
1	2	0	2	0	495	3	3	f = hmouth
0	1	0	2	1	3	483	1	g = hcheek
1	1	1	3	0	1	2	522	h = ears

IBk

a	b	c	d	e	f	g	h	<-- classified as
441	1	0	0	0	0	0	0	a = neutral
1	597	0	0	0	0	0	0	b = nod
1	0	520	0	0	0	0	0	c = tilt
0	1	1	469	0	1	0	1	d = scratch
0	0	0	0	510	0	0	0	e = hback
6	4	0	0	0	495	0	1	f = hmouth
0	0	0	0	0	1	490	0	g = hcheek
0	0	0	0	0	1	0	530	h = ears

Naïve Bayes

a	b	c	d	e	f	g	h	<-- classified as
351	3	6	2	2	6	6	66	a = neutral
203	185	81	3	8	21	16	81	b = nod
118	6	295	3	4	7	18	70	c = tilt
2	0	1	160	26	15	32	237	d = scratch
71	4	27	6	380	2	5	15	e = hback
3	2	3	9	11	361	28	89	f = hmouth
90	0	0	12	27	121	191	50	g = hcheek
18	0	0	18	111	87	45	252	h = ears

Random Forest

a	b	c	d	e	f	g	h	<-- classified as
442	0	0	0	0	0	0	0	a = neutral
0	598	0	0	0	0	0	0	b = nod
0	0	521	0	0	0	0	0	c = tilt
0	0	2	471	0	0	0	0	d = scratch
0	0	1	0	509	0	0	0	e = hback
0	1	2	0	0	503	0	0	f = hmouth
0	0	0	0	0	0	491	0	g = hcheek
0	0	0	0	0	0	0	531	h = ears

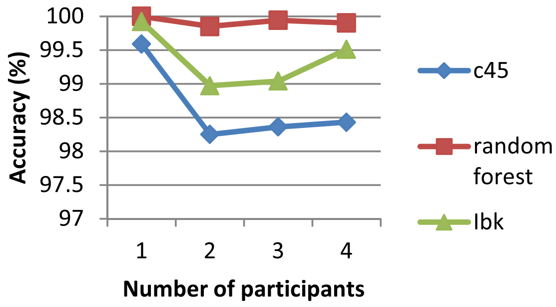

Figure 9 Accuracy variation among top 3 classifiers

Figure 8 illustrates that the classifiers (except for the Naïve Bayesian classifier) maintain high accuracy, as the number of participants increase. This indicates the Naïve Bayesian classifier may not be well suited for this classification problem. The variations across the three effective classifiers are illustrated in Figure 9 below.

Figure 9 illustrates that while a participant's body state can be classified effectively, there is a drop off when trying to classify multiple people due to variations in human body movements. However, the accuracy increases as the number of participants increases, indicating that acquiring more data pertaining to a wider range of individuals body movement, may lead to higher accuracies. For now, the methodology is focused on modeling individual team members' emotional states, although future work will explore generalized emotional state patterns existing within a design team.

3.4.2 Variation across features used

To understand the effect of using position values as features versus the use of velocity and acceleration values for better classification, the authors conducted tests, with the resulting accuracies depicted in Table 5 and Figure 10 below. The data used to train the classifiers were only position data, followed by both position and velocity data and finally all three position, velocity and

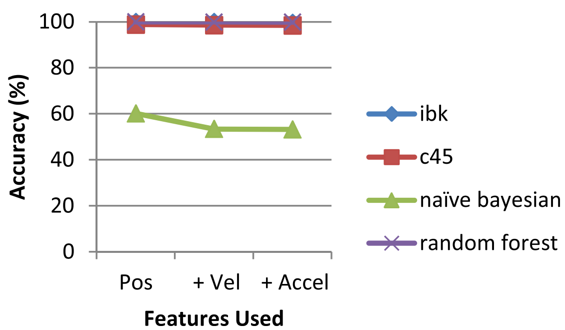

Figure 10 Variation of accuracy depending on attributes used

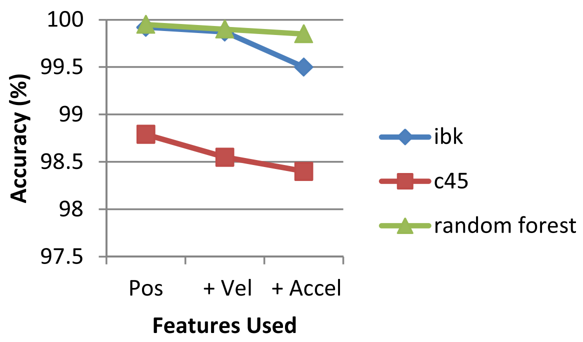

Figure 11 Accuracy variations across top 3 classifiers

acceleration data. The three are indicated in Figure 10 as "Pos", "+Vel" and "+Accel" respectively.

Figure 10 indicates that the use of velocity and acceleration as additional features other than position data, is actually counterproductive and leads to a loss of accuracy. Classifiers such as the Naïve Bayesian reveal a larger decrease in accuracy, when velocity and acceleration are included. The remaining classifiers also show the same trend as is seen in Figure 11.

3.4.3 Source of errors

To understand the source of errors, the confusion matrix for each of the algorithms was investigated. The confusion matrix is given below in Table 6, where the rows represent the actual body language depicted and the columns for each row provide the resulting distribution of classification by each of the algorithms.

Table 6 highlights the fact that the Naïve Bayes classifier suffers most in cases with slight variations in body language, compared to the other classifiers. For example, challenges in classification accuracy can be observed between a neutral pose and when the participant is nodding his/her head or when the participant is scratching the back of his/her neck versus pulling his/her ear. The similar skeletal data readings are between nodding of head and neutral pose illustrated below in Figure 12.

The similar skeletal data readings between pulling of the ears and scratching the back of the neck are shown below in Figure 13. This illustrates the difficulty in differentiation between the two poses and their respective associated emotions of frustration and boredom. However, with improvements in sensor accuracy and more data, such challenges can be addressed within the presented methodology. This would allow for more fine grained detection of emotional states and hence help us better understand design team dynamics.

For completeness, the authors performed a sensitivity analysis on the Naïve Bayes algorithm to determine how the accuracy of the model varied with

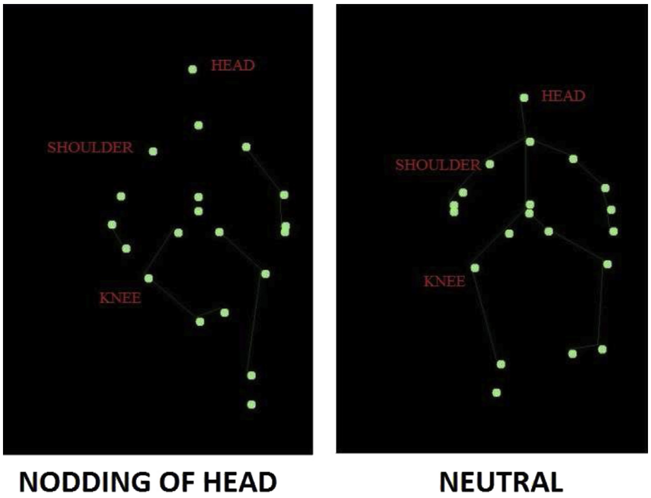

Figure 12 Nodding of head and neutral pose are similar

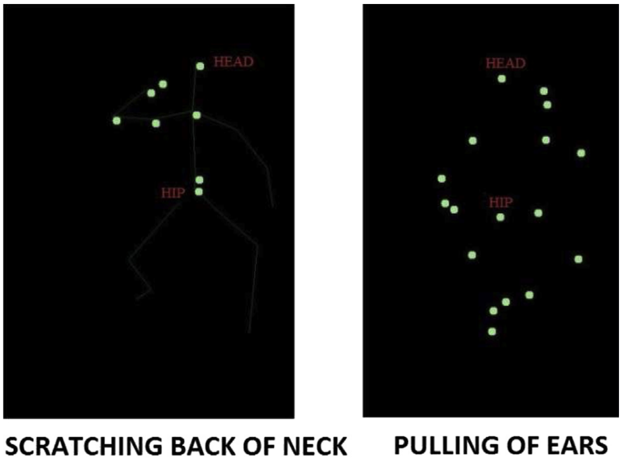

Figure 13 Scratching the back of the neck and pulling one's ears are similar

the number of folds used to train the data and the number of folds used to test the data. As can be seen in Figure 14, the stability of the algorithm is relatively consistent, independent of the number of folds used to train the model. While the first scenario of utilizing 2/3rd of the 4072 tuples to train the data, with the remaining 1/3 tuples to test the data has the highest predictive accuracy, the difference between accuracies still does not compare to the other algorithms such as C4.5, IBK or Random Forest. In the scientific literature, 10-fold CV has been shown to generate a consistent and stable approximation of an algorithm's true predictive power (Tibshirani, Hastie, Narasimhan, & Chu, 2002), (Ambroise, & McLachlan, 2002) and is employed in this work as a baseline to evaluate the predictive accuracies of each of the models.

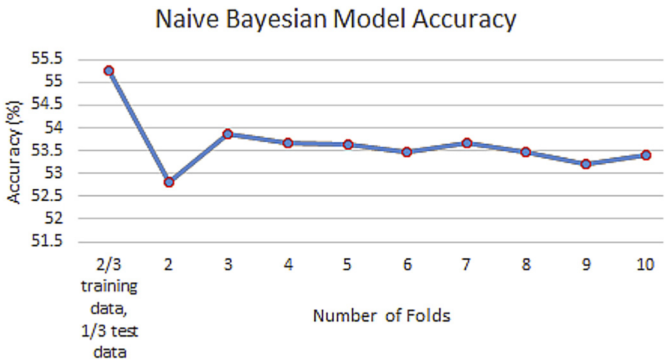

Figure 14 Variation of accuracy depending on the number of folds

3.4.4 Variation across number of joints tracked

To understand the effect of using fewer or larger number of joints on classification, the authors conducted tests, with the resulting accuracies are depicted in Table 7 and Figure 15 below.

The results in Table 7 and Figure 15 indicate that using the sensor's full capabilities and tracking 20 joints does not always increase performance and that the accuracy gain is not always positive. For example, the Naïve Bayesian classifier shows a larger improvement, with an initial increase in the number of joints, followed by a reduction in the predictive accuracy. Similar results can be observed by the other classifiers as seen in Figure 16.

Table 7 Accuracy percentage depending on joints used

Joints tracked	*C4.5*	*Random forest*	*IBk*	*Naïve Bayesian*
6	98.62%	99.83%	99.48%	38.73%
13	97.94%	99.83%	99.46%	56.16%
20	98.43%	99.9%	99.51%	53.44%

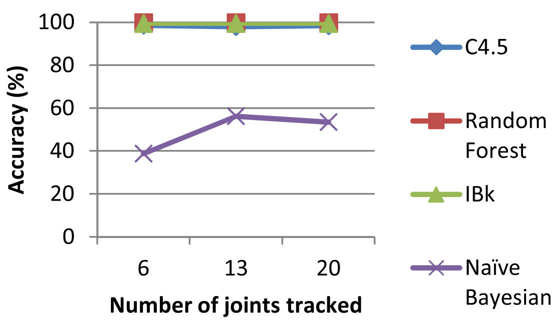

Figure 15 Variation of accuracy depending on joints tracked

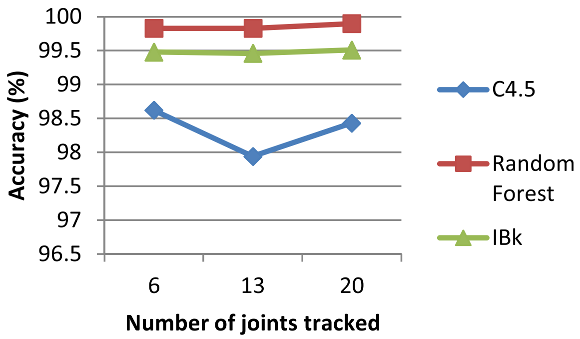

Figure 16 Accuracy variation among top 3 classifiers

4 Conclusion

In this paper, a machine learning driven methodology is proposed to quantify emotional states of individuals in a design team using low-cost, non-wearable sensor hardware. The authors achieve this by exploiting the known link between body language and emotional states. The methodology demonstrates the efficacy of non-wearable sensors and machine learning algorithms to model individuals' body language in non-design tasks with the ultimate goal of applying these methods to quantifying design team interactions. The performance of the proposed methodology is evaluated using a case study containing the classification of 8 body language poses relevant to design team interactions. The authors identify certain classifiers as effective for the methodology used. Additionally, the effect of using various attributes such as position, velocity and acceleration for classification was studied in detail. The results reveal that the impact of velocity and acceleration data may not enhance the predictive accuracy of the proposed model. Also, the effect of the number of joints tracked was investigated, indicating that the selection of which joints to track needs to be explored further.

The proposed methodology can be improved by incorporating measures to identify a larger range of body language cues that are more subtle. For now, the methodology is limited to modeling individual team members' emotional states, although future work will explore generalized emotional state patterns existing across design team members. For this to occur, predictive models will have to be evaluated based on a "leave one out" method in order to demonstrate the generalizability of detected emotional states on other team members. Beyond exploring the generalizability to multiple team members, the placement of sensors and their directionality in a given room can be optimized to reduce costs and improve detection. Additionally, with improving sensor technology, the methodology can be extended to track more subtle body language. Thus for future work, other representations of data on velocity and acceleration and related attributes such as angular velocity and angular acceleration should be explored, as acceleration and velocity may have a negative impact on accuracy. In addition, the selection of which joints to track needs to be investigated further. The optimal selection of joints will avoid over or under sampling of joint locations and thus improve accuracy and decrease computational costs. This will result in faster and more accurate evaluation of emotional states of individual team members. Beyond different representations of the feature space, the class variable can be updated to reflect a broad set of emotional states (e.g., low engagement, medium engagement and high engagement, as opposed to a single state of engagement). Moreover, this methodology opens the door for fine grained analysis of design team interactions, their emotional states, their productivity and links therein during the design process. Lastly, incorporating other means of quantifying design team interactions and studying them with conjunction with the emotional state detection of individual members may lead to a better understanding of design teams.

Acknowledgments

The authors would like to acknowledge the National Science Foundation (NSF I/UCRC grant # 1067885) and Penn State University's Center for Online Innovation in Learning (COIL) for funding this research. The authors of this work would also like to acknowledge David Munoz Soto for assisting with relevant literature pertaining to human body movements and engagement states.

References

ABC Nightline – IDEO Shopping Cart. 2009. https://www.youtube.com/watch?v=M66ZU2PCIcM&feature=youtube_gdata_player.

Ackroyd, S. (2006). *The oxford handbook of work and organization*. Oxford Handbooks Online.

Alexiadis, D. S., Kelly, P., Daras, P., O'Connor, Noel E., Boubekeur, T., & Moussa, M. B. (2011). Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proceedings of the 19th ACM international conference on multimedia* (pp. 659–662). ACM.

Ambroise, C., & McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences, 99*(10), 6562–6566.

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation, 9*(7), 1545–1588.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science, 338*(6111), 1225–1229.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241.

Baker, D., Ryan, S. J., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., et al. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In Stamper. (Ed.), *Proceedings of the 5th international conference on educational*

data mining, Chania, Greece, June 19, 21 (pp. 126). International Educational Data Mining Society, ISBN 978-1-74210-276-4.
Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology*. London: Wiley.
Birdwhistell, R. L. (2010). *Kinesics and context: Essays on body motion communication*. University of Pennsylvania Press.
Bodine, K., & Gemperle, F. (2003). Effects of functionality on perceived comfort of wearables. In *2012 16th international symposium on wearable computers*. IEEE Computer Society. 57–57.
Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. *Team Performance Assessment and Measurement. Theory, Methods, and Applications, 3*.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Bucciarelli, L. L. (1994). *Designing engineers*. MIT Press.
Burba, N., Bolas, M., Krum, D. M., & Suma, E. A. (2012). Unobtrusive measurement of subtle nonverbal behaviors with the Microsoft Kinect. *Virtual reality short papers and posters (VRW), 2012 IEEE*. IEEE 1–4.
Busseri, M. A., & Palmer, J. M. (2000). Improving teamwork: the effect of self-assessment on construction design teams. *Design Studies, 21*(3), 223–238.
Celebi, S., Aydin, A. S., Temiz, T. T., & Arici, T. (2013). *Gesture recognition using Skeleton data with weighted dynamic time warping*. In: *VISAPP, 1* 620–625.
Chung, H., & Yang, H.-D. (2013). Conditional random field-based gesture recognition with depth information. *Optical Engineering, 52*(1). 017201–017201.
Coker, D. A., & Burghoon, J. (1987). The nature of conversational involvement and nonverbal encoding patterns. *Human Communication Research, 13*(4), 463–494.
Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on, 13*(1), 21–27.
Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with autotutor. *Journal of Educational Media, 29*(3), 241–250.
Cross, N., & Cross, A. C. (1995). Observations of teamwork and social processes in design. *Design Studies, 16*(2), 143–170.
D'Mello, S., & Graesser, A. (2009). Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence, 23*(2), 123–150.
D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. ahead-of-p (ahead-of-p). *Cognition & Emotion* 1–10. http://dx.doi.org/10.1080/02699931.2011.613668.
Depth Sensors Comparison – iPiSoft Wiki." 2015. Accessed February 14 http://wiki.ipisoft.com/Depth_Sensors_Comparison.
Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*(2–3), 103–130.
Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education, 94*(1), 103–120.
Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind defining boredom in terms of attention. *Perspectives on Psychological Science, 7*(5), 482–495.
Edmondson, A. C., & Nembhard, I. M. (2009). Product development and learning in project teams: the challenges are the benefits*. *Journal of Product Innovation Management, 26*(2), 123–138.

Fisher, K. (1993). *Leading self-directed work teams, a guide to developing new team leadersip skills*. New York: McGraw-HillFisherLeading Self-Directed Work Teams: A Guide to Developing New Team Leadership skills1993.

Furnham, A., & Petrova, E. (2010). *Body language in business: Decoding the signals*. Palgrave Macmillan.

Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of Experimental Child Psychology, 102*(4), 503–521.

Gillian, N., Knapp, R. B., & O'Modhrain, S. (2011). Recognition of multivariate temporal musical gestures using N-dimensional dynamic time warping. In *Proc of the 11th int'l conference on new interfaces for musical expression*.

Goldschmidt, G., & Tatsa, D. (2005). How good are good ideas? Correlates of design creativity. *Design Studies, 26*(6), 593–611.

Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. (2006). Detection of emotions during learning with autotutor. In *Proceedings of the 28th annual meetings of the cognitive science society* (pp. 285–290). Citeseer.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. *On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE*. Springer 986–996.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter, 11*(1), 10–18.

Ho, T. K. (1995). Random Decision Forests. In*Document analysis and recognition, 1995, proceedings of the third international conference on, 1* (pp. 278–282). IEEE.

Hong, J. (2013). Considering privacy issues in the context of google glass. *Commun. ACM, 56*(11), 10–11.

Jensen, D., Wood, J., & Wood, K. (2003). Hands-on activities, interactive multimedia and improved team dynamics for enhancing mechanical engineering curricula. *International Journal of Engineering Education, 19*(6), 874–884.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic Prediction of Frustration. *International Journal of Human-Computer Studies, 65*(8), 724–736.

Kapoor, A., & Picard, R. W. (2001). A real-time head nod and shake detector. In *Proceedings of the 2001 workshop on perceptive user interfaces* (pp. 1–5). ACM.

Kapoor, A., & Picard, R. W. (2005). Multimodal Affect Recognition in Learning Environments. In *Proceedings of the 13th annual ACM international conference on multimedia* (pp. 677–682). ACM.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication* (pp. 1151–1160). ACM.

Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes, 86*(1), 99–130.

Khoshelham, K., & Elberink, S. O. (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors, 12*(2), 1437–1454.

Klimoski, R., & Mohammed, S. (1994). Team mental model: construct or metaphor? *Journal of Management, 20*(2), 403–437.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. In: *Ijcai, 14* 1137–1145.

Kononenko, I. (1990). Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition. *Current Trends in Knowledge Adquisition* 190–197.

Kotsiantis, S. B. (2007). *Supervised machine learning: A review of classification techniques*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.95.9683.

Lang, S., Block, M., & Rojas, R. (2012). Sign language recognition using kinect. *Artificial intelligence and soft computing*. Springer 394–402.

Le Dantec, C. A., & Do, E. Yi.-L. (2009). The mechanisms of value transfer in design meetings. *Design Studies, 30*(2), 119–137.

Li, P., Hu, X., & Wu, X. (2008). Mining concept-drifting data streams with multiple semi-random decision trees. *Advanced data mining and applications*. Springer 733–740.

Lim, T.-S., Loh, W.-Y., & Shih, Yu.-S. (2000). A comparison of prediction accuracy, Complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning, 40*(3), 203–228.

Liu, Y.-C., Chakrabarti, A., & Bligh, T. (2003). Towards an 'ideal' approach for concept generation. *Design Studies, 24*(4), 341–355.

Manohar, G., & Tucker, C. (2013). A privacy preserving data mining methodology for dynamically predicting emerging human threats. In *ASME 2013 international design engineering technical conferences and computers and information in engineering conference*. American Society of Mechanical Engineers. V02AT02A069–V02AT02A069.

Marchand, G. C., & Gutierrez, A. P. (2012). The role of emotion in the learning process: comparisons between online and face-to-face learning settings. *The Internet and Higher Education, 15*(3), 150–160.

Marneweck, M., Loftus, A., & Hammond, G. (2013). Psychophysical measures of sensitivity to facial expression of emotion. *Frontiers in Psychology, 4*.

Miranda, L., Vieira, T., Martinez, D., Lewiner, T., Vieira, A. W., & Campos, M. F. M. (2012). Real-time gesture recognition from depth data through key poses learning and decision forests. In *Graphics, patterns and images (SIBGRAPI), 2012 25th SIBGRAPI conference on* (pp. 268–275). IEEE.

Mitchell, T. M. (1997). *Machine learning. 1997*. Burr Ridge, IL: McGraw Hill 45.

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*(2–3), 145–165.

Munoz, D., & Tucker, C. (August 17–20 2014). Assessing students' emotional states: an approach to identify lectures that provide an optimal learning experience. In *Proceedings of the 2014 ASME IDETC/CIE, DETC2014-34782*. Buffalo, NY.

Murthy, S. K. (1998). Automatic construction of decision trees from data: a multi-disciplinary survey. *Data Mining and Knowledge Discovery, 2*(4), 345–389.

Olson, G. M., Olson, J. S., Carter, M. R., & Storrosten, M. (1992). Small group design meetings: an analysis of collaboration. *Human–Computer Interaction, 7*(4), 347–374.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Parker, G. M. (2003). *Cross-functional teams: Working with allies, enemies, and other strangers*. John Wiley & Sons.

Pease, B., & Pease, A. (2008). *The definitive book of body language*. Bantam.

Provine, R. R. (1986). Yawning as a stereotyped action pattern and releasing stimulus. *Ethology, 72*(2), 109–122.

Projects | Mobile Kinect | BIG." 2015. Accessed February 16. http://big.cs.bris.ac.uk/projects/mobile-kinect.
Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. Elsevier.
Raptis, M., Kirovski, D., & Hoppe, H. (2011). Real-time classification of dance gestures from skeleton animation. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics symposium on computer animation* (pp. 147–156). ACM.
Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. *Encyclopedia of database systems*. Springer 532–538.
Reilly, R. R., Lynn, G. S., & Aronson, Z. H. (2002). The role of personality in new product development team performance. *Journal of Engineering and Technology Management, 19*(1), 39–58.
Shah, J. J., Kulkarni, S. V., & Vargas-Hernandez, N. (2000). Evaluation of idea generation methods for conceptual design: effectiveness metrics and design of experiments. *Journal of Mechanical Design, 122*(4), 377–384.
Shen, L., Wang, M., & Shen, R. (2009). Affective e-learning: using 'emotional' data to improve learning in pervasive learning environment. *Educational Technology & Society, 12*(2), 176–189.
Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., et al. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM, 56*(1), 116–124.
Spagnolli, A., Guardigli, E., Orso, V., Varotto, A., & Gamberini, L. (2014). Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. *Symbiotic interaction*. Springer 87–98.
Stasser, G., & Dietz-Uhler, B. (2001). Collective choice, judgment, and problem solving. In M. A. Hogg, & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: Group processes* (pp. 31–55). Blackwell Publishers Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9780470998458.ch2/summary.
Stempfle, J., & Badke-Schaub, P. (2002). Thinking in design teams – an analysis of team communication. *Design Studies, 23*(5), 473–496. http://dx.doi.org/10.1016/S0142-694X(02)00004-2.
Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences2, 99*(10), 6567–6572.
Tucker, C. S., & Kim, H. M. (2011). Trend mining for predictive product design. *Journal of Mechanical Design, 133*(11), 111008.
Verikas, A., Gelzinis, A., & Bacauskiene, M. (2011). Mining data with random forests: a survey and results of new tests. *Pattern Recognition, 44*(2), 330–349. http://dx.doi.org/10.1016/j.patcog.2010.08.011.
Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879–896. http://dx.doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W.
Wang, J.-Y., & Lee, H.-M. (2009). Recognition of human actions using motion capture data and support vector machine. In*WRI world congress on software engineering, 2009 WCSE'09, 1* (pp. 234–238). http://dx.doi.org/10.1109/WCSE.2009.354.
Wettschereck, D., Aha, D. W., & Mohri, T. (1997). A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review, 11*(1–5), 273–314. http://dx.doi.org/10.1023/A:1006593614256.
Won, A. S., & Bailenson, J. N. (2014). Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *Affective Computing, IEEE Transactions on, 5*(2), 112–125. http://dx.doi.org/10.1109/TAFFC.2014.2329304.
Xia, L., Chen, C.-C., & Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3D joints. In *2012 IEEE computer society*

conference on computer vision and pattern recognition workshops (CVPRW) (pp. 20–27). http://dx.doi.org/10.1109/CVPRW.2012.6239233.
Xtion PRO – Overview." 2015. Accessed February 14. http://www.asus.com/Multimedia/Xtion_PRO/.
Yang, M. C. (2010). Consensus and single leader decision-making in teams using structured design methods. *Design Studies, 31*(4), 345–362. http://dx.doi.org/10.1016/j.destud.2010.03.002.
Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D., et al. (2012). Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (pp. 699–704). ACM.
Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., & Presti, P. (2011). American sign language recognition with the kinect. In *Proceedings of the 13th international conference on multimodal interfaces* (pp. 279–286). ACM.
Zhang, H., & Su, J. (2004). Naive Bayesian classifiers for ranking. In J.-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi (Eds.), *Machine learning: ECML 2004. Lecture notes in computer science 3201* (pp. 501–512). Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-30115-8_46.