

Towards Personalized Performance Feedback: Mining the Dynamics of Facial Keypoint Data in Engineering Lab Environments

Christian E. López¹ and Dr. Conrad S. Tucker^{1,2}

1 Department of Industrial and Manufacturing Engineering, the Pennsylvania State University.

2 School of Engineering Design, Technology and Professional Programs, the Pennsylvania State University.

Abstract

According to the National Academy of Engineering, the development of personalized learning is one of the grand engineering challenges of the 21st century¹. Even though affect-sensitive systems have been used for personalized learning, current systems provide feedback based on predefined relationships between affective state and performance. However, studies have shown that the affective state that correlates to good performance could vary between tasks and students. Hence, these systems can only provide accurate performance feedback once the student has completed the task at hand. In light of the limitations of current methods, this work presents a machine learning method for predicting students' performance by using the dynamics of their facial keypoint data captured while reading the instructions of a task, thus, avoiding the need to infer their affective state. A case study involving 40 students performing tasks in an engineering lab environment is used to validate the proposed method. The results reveal that the proposed model yielded an accuracy of 80%. The results indicate the importance of using students' facial keypoint data, captured while reading the instructions of a task, to predict their performance. This method could be implemented in engineering lab environments to provide real-time feedback to students and advance personalized learning.

Keywords

Personalized feedback, facial expression, facial keypoint data, affective computing.

1. Introduction

Proper feedback has the potential to improve students' performance in a wide variety of tasks^{2,3}. Research indicates a strong correlation between students' affective state and their learning performance⁴⁻⁶. Understanding students' affective state allows instructors to provide personalized assistance that can enhance students' learning experience⁷. Traditionally, instructors provide personalized assistance and real-time feedback to students based on the facial or body cues they project, as well as their performance on the task. However, students' performance on a task is usually evaluated after it is completed. This approach limits the ability to provide timely and systematic feedback to students before completing a task. Furthermore, in-person and personalized assistance might be difficult to achieve in online learning environments, where in-person interactions are challenging, or in engineering laboratories where the student to instructor ratio is high.

According to the National Academy of Engineering, the development of personalized learning systems is one of the grand engineering challenges of the 21st century¹. Researchers have shown an increasing interest in the development of systems capable of providing feedback based on students' perceived affective state with the objective to improve students' performance⁷⁻¹¹. Fig. 1 illustrates how most of the current affect-sensitive systems provide personalized intervention to

a student while using engineering equipment (e.g., band saw). First, the system captures the students' facial expression with the use of an RGB sensor (i.e., video cameras), and with the use of computer vision algorithms extract his/her facial keypoint data. This facial keypoint data is then used as input in a machine learning pipeline that infers his/her affective state (e.g., sad). Based on the student's inferred affective state, the system provides an intervention (e.g., feedback) with the objective of improving his/her performance on the task.

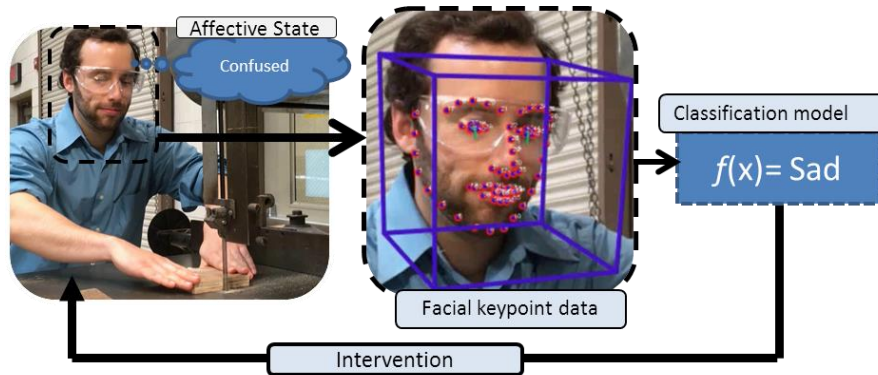


Figure 1. Representation of an affect-sensitive system in an engineering lab environment.

Current affect-sensitive systems do not consider students' unique facial characteristics. These systems are often trained with data sets collected from a limited set of individuals¹². Thus, they implement general models (i.e., models trained with data of a general population) to infer a student's affective state. Hence, their capability to provide personalized feedback based on inferred affective states is limited⁹. In the example shown in Fig. 1, the student was not sad; instead, he was confused. Consequently, the intervention or feedback provided by the system was not optimal. Furthermore, these methods provide feedback based on predefined relations between students' affective state and performance. However, studies have indicated that based on tasks and student characteristics, the affective state that correlates to good performance could vary^{13,14}. Hence, an intervention given to a student i on a task t , might not be ideal for 1) the same student i on a different task k , or 2) another student j on that same task t . Finally, these systems focus on predicting individuals' affective state while performing a task¹⁴. This limits their ability to provide timely and systematic feedback to students before they start a task.

Due to the limitations of current affect-sensitive systems and the heterogeneity of students, Lopez and Tucker¹⁵ presented a method that implemented facial keypoint data to predict students' performance on a task. Nonetheless, this method only considered the average value of students' facial keypoint. Hence, information regarding the variation of these facial keypoints over time (e.g., dynamic components) was not explored. The affective computing community has revealed the value of considering the variation of facial keypoints over time since it help captures the temporal component of individuals' facial expression^{11,14,16,17}. Similarly, studies have indicated that considering the variation of facial keypoints provides additional information that could help provide personalized feedback¹⁸. Another limitation of the previous method is that it was never tested in a real engineering lab environment. Hence, more effort should be given to design and test systems capable of predicting student's performance on engineering lab environment. Systems designed to recognize student's facial expressions while reading the instruction of a task and predict their subsequent performance on that task, could provide feedback to students prior to the start of the task. Such systems have the potential to provide real-

time and personalized feedback in a wide range of environments, and potentially improve students' performance and self-confidence. In light of this gap, this work proposes a machine learning method for predicting a student's performance by using the dynamics of his/her facial keypoint data captured while reading the instructions for a task. A case study involving 40 students performing tasks in an engineering lab environment is used to validate the proposed method.

2. Method

A machine learning method for predicting students' performance on a task is presented in this work. The method uses the dynamic component of students' facial keypoint data captured while they read the instructions for a task. The (i) *Data Acquisition & Feature Extraction*, (ii) *Model Building & Tuning*, and (iii) *Model Evaluation* steps of the proposed method are illustrated in Fig. 2. For this work, students' performance is assumed to be a function of their task completion time measured in seconds. Students are classified based on this performance as (i) "below" average or (ii) "above" average. That is, if a student i takes longer than the average of all student that performed that same task t , student i is classified as A) "above"; otherwise is classified as B) "below"; for $i \in$ set of students $\{\mathbf{I}\}$ and $t \in$ set of tasks $\{\mathbf{T}\}$.

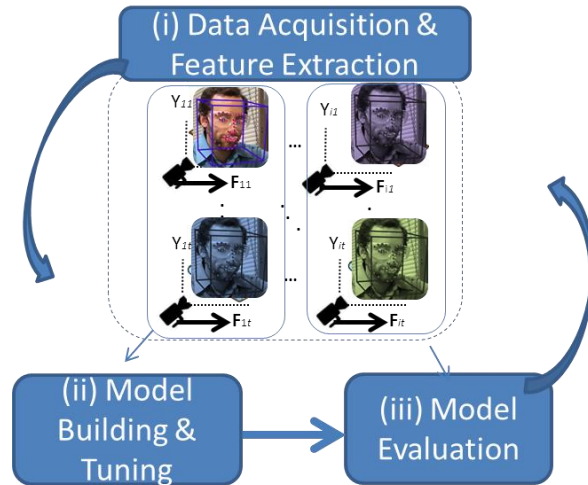


Figure 2. Outline of proposed method

2.1. Data Acquisition & Feature Extraction. The objective of this step is to capture the facial keypoint data of a student i while reading the instructions on how to perform a task t on engineering equipment, as well as the time it took them to complete the task. That is, the first objective of the data acquisition step is to capture the facial keypoint data \mathbf{F} of student i for a given task t (\mathbf{F}_{it}). To capture students' facial keypoint data, it is assumed that at least a standard resolution video (640x480 pixels) of the students' facial expression while reading the instructions of a task (\mathbf{V}_{it}), can be acquired. For each student, the video captures him/her reading the instructions of a certain task t that requires the use of engineering equipment. In this work, engineering equipment refers to equipment frequently found in engineering lab environments (e.g., band saw, drills). From the video recordings \mathbf{V}_{it} , the facial keypoint data is extracted. The facial keypoint data are given as x and y coordinates in the space of the video recording frames. Consequently, the location, size, and orientation of the facial keypoints in each frame could be characteristics of students' facial pose and location relative to the camera, and not necessarily their unique facial expressions. Therefore, the proposed method implements ordinary procrustes analysis¹⁹ on the facial keypoint data obtained from each frame of the video recordings. This is done with the objective of standardizing students' facial location and orientation while retaining their unique facial expression information. Once the facial keypoint data are normalized, the mean and standard deviation are calculated. The second objective of data acquisition step is to capture the completion time of a student i on a task t (Y_{it}). Depending on the task and engineering equipment used, a student's completion time can be manually captured by researchers, or it can be automatically captured with the use of sensors.

2.2. Model Building & Tuning. The method proposes the use of Support Vector Machine (SVM) algorithms to build a model to predict a student’s performance on a task. In the literature of affect-sensitive systems, SVM algorithms have been extensively used and shown to outperform other algorithms^{11,20,14,21}. Moreover, they are well-suited for real-time classification due to their accuracy and speed^{22,23}. The objective of an SVM algorithm is to identify the hyperplanes that classify all training vectors into their respective class. Hence, in a two-class classification problem (e.g., *below* and *above*), the objective of the SVM algorithm is to discover the function in which the margins between the two classes in the training vectors are maximized. Thus, these algorithms can be understood as optimization algorithms. Furthermore, the hyperparameter optimization step of the SVM algorithm identifies the best parameters for the model by training multiple models. The parameters to be optimized are Cost (C) and Epsilon (ϵ). The Cost (C) parameter governs the tradeoff between the model dimensions and the degree to which deviations larger than ϵ are tolerated. On the other hand, the parameter Epsilon (ϵ) controls the width of the ϵ -insensitive zone used for fitting the training vectors see Burges²⁴ and Kotsiantis²² for more details).

3.3. Model Evaluation. The accuracy of the SVM classification model is assessed with a *leave-one-out* cross-validation approach. This validation approach has been implemented in previous studies^{14,15,25} because it is appropriate for small data sets and has been shown to produce unbiased accuracy estimations²⁶. The *leave-one-out* cross-validation approach used in this work is illustrated in Fig. 3. The figure presents an example dataset composed of the (i) mean and standard deviation of the facial keypoint of the video recording, (ii) mean and standard deviation of the procrustes analysis rotation parameters, (ii) the identifiers for student i and (iv) task t , as well as (v) the performance class label of student i on task t (Y_{it}). In the first iteration of this cross-validation approach, tuple 1 is assigned to the testing set, while the remaining tuples (i.e., 2- it) are assigned to the training set. Once a model is built with the use of the training set, its classification accuracy is evaluated with the testing set. In the subsequent iterations, the process is repeated for all tuples in the dataset. The accuracy obtained in each of the iterations is measured, and the average accuracy presented.

Tuple	Facial Key point 1 μ	Facial Key point 1 σ	...	Rotation x σ	Individual (i)	Task (t)	Y_{it}
1	0.355	0.674	...	0.025	1	1	A
2	0.874	0.234	...	0.332	1	2	B
3	0.365	0.632	...	0.292	1	3	A
...
it	0.274	0.193	...	0.05	i	t	B

Figure 3. Leave-one-out cross-validation approach

3. Case Study

The method proposed in this work is implemented in a case study involving students performing tasks in an engineering lab environment. For this case study, two pieces of engineering equipment, (i) band saw and (ii) drill press, were used. Participants were required to perform tasks with each equipment. With the band saw, participants were required to cut a predefined straight line drawn in the middle of a 21cm by 19cm cardboard piece. With the drill press, participants were required to drill a predefined hole drawn in the center of a similar cardboard piece. The case study involved a total of 40 participants. All were freshman engineering students (18 to 19 years of age, 27.5% females) enrolled in EDGSN 100 *Introduction to Engineering Design* at the Pennsylvania State University. After introducing the participants to the

experimental setup and informed consent documents, they were guided toward each piece of the engineering equipment.

Due to inconsistencies during the experiment, five video recordings were excluded from the dataset. Hence, a total of 75 videos of participants were implemented. The number of frames in each video recording varied as the time taken for each participant to read the instructions of a given task differed. Sixty-eight facial keypoints were extracted from the participants' facial expression recordings using the *OpenFace facial behavior analysis toolkit* ²⁷. Fig. 1 shows a representation of the facial keypoint data of a student captured while in an engineering lab environment using the *OpenFace toolkit*. Each of the facial keypoints consisted of x and y coordinates, in which ordinary procrustes analysis was performed to normalize all the faces to a canonical orientation, scale centered at the origin, and scaled to unit variance. The participants' completion time in each of the tasks was manually recorded by the research team. The classification of a student's completion time on a given task as "*below*" or "*above*" was done by calculating average completion time of all 40 students that performed that same task. Hence, one participant could have been "*below*" average for the one task, while "*above*" for another task.

The model for predicting students' performance was generated using an SVM algorithm implemented with the R package "e1071" version 1.6-8 ²⁸. The SVM algorithm was used to predict the completion time class of a student i on a task t , based on his/her facial keypoint data captured while reading the instructions for that task. The facial keypoint data were given as x and y coordinates; hence, the mean and standard deviations for all 136 features, along with the procrustes parameters, were used as input on the classification model. The cost and epsilon parameters of the SVM algorithm were tuned using grid search approach. The model's accuracy was evaluated using a *leave-one-out* cross-validation approach. For this case study, a total of 75 tuples were collected. From these tuples, 44 instances were from student classified as "*below*", while 31 as "*above*". Each tuple was composed of a participants' class label (Y_{it}), participants' facial keypoint data predictors (F_{it}) (i.e., mean and standard deviation of x and y coordinates and Procrustes parameters), participants' identification (i), and task identification (t), as shown in Fig. 3.

4. Results and Discussion

Table 1 shows the confusion matrix for the *student-task* model proposed in this work. The results show that the model was able to correctly predict participants' completion times with an accuracy of 80% (95%CI: 69.17%-88.35%). Moreover, this accuracy was statistically significantly greater than the accuracy that could have been obtained by random chance (p -value: 7.11e-8). These results reveal that the machine learning method proposed in this work was able to accurately predict participants' performance by using their unique facial keypoint data captured while reading the instructions of a task. In contrast, the *general* model, which did not take into consideration students' facial keypoint data, provided a classification accuracy of only 58.67% (95%CI: 46.70%-69.92%). This accuracy was not statistically significantly greater than random chance (p -value: 0.549). Table 2 shows the confusion matrix for the *general* model. The difference in models' performance can be attributed to the additional information captured by the students' facial expression. These results support the authors' argument that machine learning models that take into consideration students' facial keypoint data can be implemented to predict students' performance and potentially advance personalized feedback. Furthermore, Table 1

shows that the *student-task* model tends to correctly classify the cases where the participants had below average completion times (i.e., *Below*), rather than the cases where they had above average (i.e., *Above*). In others words, if the class *Below* is considered to be the positive condition, the *Sensitivity* (or the true positive rate) of the model (97.73%) is greater than the *Specificity* (or true negative rate) of the model (54.84%). These results indicate that the model had difficulties correctly classifying students that had completion times above average.

Table 1. Confusion matrix for the student-task model

Ground truth /Predicted	Below	Above
Below	43	14
Above	1	17
Total	44	31

Table 2. Confusion matrix for the general model

Ground truth /Predicted	Below	Above
Below	44	31
Above	0	0
Total	44	31

5. Conclusions and Future Work

In recent years, researchers have started exploring how systems that capture facial expressions can infer students' affective states and be implemented in engineering environments. Nonetheless, current methods still label students' affective states into discrete emotion categories and provide feedback based on predefined relationships between performance and affective states. However, a students' affective state that correlates to good performance could vary between tasks. Additionally, this relationship of affective state and a good performance could vary between students. While a recent study presented a machine learning method to predict students' performance by using their unique facial keypoint data, bypassing the need to infer their affective states, limitations still exist. First, the method did not consider the dynamics of students' facial expression and only implemented a limited set of facial keypoints (i.e., 10). Moreover, the feasibility of implementing this method in a real engineering lab environment was never explored.

In light of these limitations, this work presented a machine learning method for predicting a student's performance. The dynamics of a student's facial keypoints while reading the instructions of a task were used as input for the proposed model. In this work, the feasibility of the proposed method was tested with students performing tasks in a real engineering lab environment. The results of this work show that with the use of widely available sensors (i.e., video cameras) and open source toolkits, the dynamics of students' facial keypoint data can be captured and used to successfully model and predict their performance on a task. The machine learning model proposed in this work yielded a classification accuracy of 80%, which was statistically significantly greater than random chance. In contrast, the model that did not take into consideration students' facial expression provided a classification accuracy of only 58.67%, which was not statistically significantly greater than random chance. The results from this work highlight the potential of capturing the dynamics of students' facial expression while reading the instructions of a task to predict their subsequent performance on that task. Nevertheless, there are several areas for future work that could improve the method proposed. For example, future work should explore testing other machine learning algorithms such as neural networks and compare their accuracy and efficiency. Similarly, other tasks and performance metrics should be explored. Nonetheless, this work provides initial groundwork for systems that implement students' facial keypoint data to predict their performance. This work has the potential to advance systems intended to provide real-time feedback to students and bypass the need to predict their affective states.

Acknowledgment

This research is funded in part by NSF NRI # 1527148. Any opinions, findings, or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

References

1. Vest C. Context and challenge for twenty-first century engineering education. *J Eng Educ.* 2008;97(3):235-236.
2. Lauche K. Job design for good design practice. *Des Stud.* 2005;26(2):191-213. doi:10.1016/j.destud.2004.09.002.
3. Egan PF, Leduc PR. Utilizing emergent levels to facilitate complex systems design: Demonstrated in a synthetic biology domain. *ASME 2013 Int Des Eng Tech Conf Comput Inf Eng Conf.* 2013;V03AT03A045-V03AT03A045.
4. Ben Ammar M, Neji M, Alimi AM, Gouardères G. The Affective Tutoring System. *Expert Syst Appl.* 2010;37(4):3013-3023. doi:10.1016/j.eswa.2009.09.031.
5. Marchand GC, Gutierrez AP. The Role of Emotion in the Learning Process: Comparisons Between Online and Face-to-Face Learning Settings. *Internet High Educ.* 2012;15(3):150–160.
6. Tian F, Gao P, Li L, et al. Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems. *Knowledge-Based Syst.* 2014;55:148-164. doi:10.1016/j.knosys.2013.10.019.
7. Wu CH, Huang YM, Hwang JP. Review of affective computing in education/learning: Trends and challenges. *Br J Educ Technol.* 2015;doi: 10.1111/(bjet.12324). doi:10.1111/bjet.12324.
8. Lin HCK, Wu CH, Hsueh YP. The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability. *Comput Human Behav.* 2014;41:514-522. doi:10.1016/j.chb.2014.09.052.
9. Mondragon AL, Nkambou R, Poirier P. Evaluating the Effectiveness of an Affective Tutoring Agent in Specialized Education. *Eur Conf Technol Enhanc Learn.* 2016;446-452. doi:10.1007/978-3-319-45153-4_7.
10. Bahreini K, Nadolski R, Westera W. Towards real-time speech emotion recognition for affective e-learning. *Educ Inf Technol.* 2016;21(5):1367-1386. doi:10.1007/s10639-015-9388-2.
11. Calvo RA, D'Mello S. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans Affect Comput.* 2010;1(1):18-37. doi:10.1109/T-

AFFC.2010.1.

12. Metaxas D, Zhang S. A review of motion analysis methods for human Nonverbal Communication Computing. *Imavis*. 2013;31(6-7):421-433. doi:10.1016/j.imavis.2013.03.005.
13. Hu Q, Bezawada S, Gray A, Tucker C, Brick T. Exploring the Link Between Task Complexity and Students' Affective States During Engineering Laboratory Activities. In: *ASME 2016 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.* ; 2016:V003T04A019-V003T04A019.
14. Bezawada S, Hu Q, Gray A, Brick T, Tucker C. Automatic Facial Feature Extraction for Predicting Designers' Comfort With Engineering Equipment During Prototype Creation. *J Mech Des*. 2017;139(2):21102. doi:10.1115/1.4035428.
15. Lopez, C. E., & Tucker CS. From Mining Affective States to Mining Facial Keypoint Data: The Quest Towards Personalized Feedback. In: *In ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* ; 2017:V001T02A039-V001T02A039.
16. Krumhuber EG, Kappas A, Manstead ASR. Effects of Dynamic Aspects of Facial Expressions: A Review. *Emot Rev*. 2013;5(January):41-46. doi:10.1177/1754073912451349.
17. Littlewort G, Bartlett MS, Fasel I, Susskind J, Movellan J. Dynamics of facial expression extracted automatically from video. *Image Vis Comput*. 2006;24(6):615-625. doi:10.1016/j.imavis.2005.09.011.
18. CE Lopez CT. When to Provide Feedback? Exploring Human-Co-Robot Interactions in Engineering Environments. In: *ASEE Annual Conference & Exposition.* ; 2017.
19. Ross A. Procrustes Analysis. *Food Qual Prefer*. 1994;5(1-2):115-120. doi:10.1016/0950-3293(94)90016-7.
20. D'Mello SK, Kory J. A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Comput Surv*. 2015;47(3):A43. doi:10.1145/2682899.
21. Cohen I, Sebe N, Garg A, Chen LS, Huang TS. Facial expression recognition from video sequences: temporal and static modeling. *Comput Vis Image Underst*. 2003;91(1-2):160-187. doi:10.1016/S1077-3142(03)00081-X.
22. Kotsiantis SB. Supervised machine learning: A review of classification techniques. *Informatica*. 2007;31:249-268. doi:10.1115/1.1559160.
23. Behoora I, Tucker CS. Machine learning classification of design team members' body language patterns for real time emotional state detection. *Des Stud*. 2015;39:100-127. doi:10.1016/j.destud.2015.04.003.

24. Burges CCJC. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min Knowl Discov.* 1998;2(2):121-167. doi:10.1023/A:1009715923555.
25. Elisseeff a, Pontil M. Leave-one-out error and stability of learning algorithms with applications. *Adv Learn Theory Methods, Model Appl NATO Sci Ser III Comput Syst Sci Vol 190.* 2003:111-130. <http://www0.cs.ucl.ac.uk/staff/M.Pontil/reading/natoasi.pdf>.
26. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Int Jt Conf Artif Intell.* 1995;14(12):1137-1143. doi:10.1067/mod.2000.109031.
27. Baltrusaitis T, Robinson P, Morency LP. OpenFace: An open source facial behavior analysis toolkit. In: *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016.* ; 2016. doi:10.1109/WACV.2016.7477553.
28. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch F. E1071: Misc. Functions of the Department of Statistics (e1071). *R Packag version 16-2.* 2014:TU Wien, Wien, Austria.

Christian E. López

Christian E. López, is currently a Ph.D. student at Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at the Pennsylvania State University. He has worked as an Industrial Engineer in both the Service and Manufacturing sectors before pursuing his Ph.D. His current research focuses on the design and optimization of intelligent systems technologies through the acquisition, integration, and mining of large-scale, disparate data. He is currently working on a project that ambitions to design a system capable of providing students customized motivational stimuli and performance feedback based on their affective states.

Dr. Conrad S. Tucker

Dr. Tucker holds a joint appointment as Associate Professor of Engineering Design and Industrial Engineering at The Pennsylvania State University. He is also an affiliate faculty in Computer Science and Engineering. He is the director of the Design Analysis Technology Advancement (D.A.T.A) Laboratory. His research interests are in formalizing system design processes under the paradigm of knowledge discovery, optimization, data mining, and informatics. His research interests include applications in complex systems design and operation, product portfolio/family design, and sustainable system design optimization in the areas of engineering education, energy generation systems, consumer electronics, environment, and national security.