# Detection of Engineering Interest in Children Through an Intelligent System Using Biometric Signals

**Gustavo Olivas Martínez, Milton Osiel Candela Leal, Juan Carlos Ocampo Alvarado, Luis Fernando Acosta Soto, Diego Mauricio Botín Sanabria, Gladys Guadalupe Díaz Armas, Aimé Judith Aguilar Herrera, Esther Aimeé Delgado Jiménez, Myriam Alanis Espinosa, Mauricio Adolfo Ramírez Moreno, Rubén Morales Menéndez, Ricardo Ambrocio Ramírez Mendoza, Jorge de Jesús Lozoya Santos**
School of Engineering and Sciences
Tecnológico de Monterrey
Monterrey, Nuevo León, México
gustavo.olivas@tec.mx, milton.candela@tec.mx, a00825409@itesm.mx, luis.acosta@tec.mx a00825786@itesm.mx, a00820245@itesm.mx, a00828691@itesm.mx, a00827948@itesm.mx ,myriam.alanis@tec.mx, mauricio.ramirezm@tec.mx, rmm@tec.mx, ricardo.ramirez@tec.mx, jorge.lozoya@tec.mx

**Jorge Murrieta González, Aleyka García Briseño,**
MachineCare Education
Monterrey, Nuevo León, México
jorge.murrieta@machinecare.com.mx, aleyka.garcia@machinecare.com.mx

## Abstract

The promotion of interest in Science, Technology, Engineering and Mathematics (STEM) areas is a priority in the current and future context of Mexico, evidenced through concrete actions, both public and private. As such, there is a need for tools to accurately measure vocational interest to effectively promote STEM education. According to Holland Occupational Themes theory, vocational interest can be conceived as an aspect of personality. There is empirical evidence supporting personality prediction using neurophysiological signals. Therefore, vocational interest may be predicted using similar data. The objective of this study is to develop an intelligent system that estimates engineering interest in children through their physiological response to engineering-related activities. The participants were 13 children between 6 and 15 years old. For each participant, 8 different electroencephalographic channels were measured, as well as the electrodermal activity, heart rate variability, facial gesticulation, and body temperature data for four 2-hour sessions of engineering-related extracurricular educational activities. Psychometric tests were also included to evaluate the children's interest in specific engineering fields according to their learning activities. The generated data was processed and used to design a machine learning algorithm. The results and implications are discussed.

## Keywords
Vocational Interest, STEM, Machine Learning, and Biometric Signals.

## 1. Introduction
Many countries are struggling to employ more people in science, technology, engineering and mathematics (STEM) industries who are trained in the design and use of technologies capable of advancing national production (Kier 2014). The Mexican government has implemented public policies with the purpose of strengthening the STEM education system. Based on this, projects have been developed in the private and public sphere to promote interest in these professional fields. To study the effects of such initiatives, the project must be able to evaluate in a standardized, efficient, and precise way the interests and capabilities of individuals in STEM areas. Furthermore, interest not only

serves to predict who will choose these careers but as established by systematic reviews, also how well they will perform (Nye 2012).

Currently, most empirical studies in this area are in the field of psychometry and there are tests, based on different models and theories, that measure interest in STEM areas (Typer 2010). Even so, as they are relatively novel evaluations, they do not present sufficient evidence to establish their convergent, discriminant, predictive and concurrent validity (Hughes 2018). Also, since they are designed for foreign cultural groups, their effectiveness in Mexican samples cannot be guaranteed without cross-cultural adaptation. In addition, its results can be affected by biases of social desirability, acquiescence, extreme response, among others.

### 1.1 Objectives

The main research objectives in which this investigation is focused are the following:

- Design and implementation of an acquisition system using Python that allows the collection of data from an electroencephalograph cap, a PPG sensor and a camera.
- Implementation of a Machine Learning algorithm that is able to analyze the collected data and produce predictions on two different aspects: a score denoting the student's level of understanding and the change in interest for a determined topic within the evaluated areas, which are Programming, 3D Design and Robotics.

## 2. Literature Review

In the field of emotion detection, neuroscientific research methodologies can provide new measurement paradigms. Despite this, we found only one article using these techniques for the study of vocational interests. Rasheed et al. evaluated electroencephalographic signals (EEG) as a predictor of a psychometric test of vocational interests in university students, obtaining initial results and concluding that further research is needed (Rasheed 2019). As such, there is inconclusive evidence and a knowledge gap regarding the use of biometric signals for the detection of vocational interest.

Literature reviews to date agree that Holland's vocational model is the basis for most of the research on vocational interests to date (Mikolajczak 2010). This model proposes a hexagon of occupational themes based on personality typologies since it presupposes that vocational interest is an expression or aspect of it (Holland 1997). In the last three decades significant advances have been made on personality prediction using EEG (Chi 2005, Johannisson 2016). Although without a decisive conclusion, a review by Roslan et al. found that some studies were able to differentiate extroverts from introverts using EEG (Roslan et al. 2017). While a review by Saffiera et al. concludes that event-related potentials (ERP) when performing cognitive tasks are a useful technique to determine personality traits (Saffiera et al. 2020). A study by Maksimenko et al. found a relation between EEG characteristics and 16 personality theory traits (Masksimenko et al. 2018). Lastly, Li et al. found that video EEG measurements can quantitatively predict the Big Five personality traits (Li et al. 2020). Based on the above, the aim of this study is to develop an intelligent system that estimates engineering interest in children through their physiological response to engineering-related activities.

## 3. Methods

This study analyzed and determined the possible physiological signals that indicate an achievement of three engineering education topics taught to students between 6 and 15 years old by the company MachineCare Education (MCE), the topics include Programming, 3D Design and Robotics. Specifically, the heart rate variability (HRV), electrodermal activity (EDA), temperature, facial gesticulation, and EEG signals were analyzed. Subsequently, features of such signals were contrasted with results of psychometric tests to validate how the physiological signals can predict whether students have developed the necessary skills to develop themselves, in the future, in the field of engineering.

Thirteen children participated, within the age range of 6 and 15 years, within these test subjects, a selection criterion was applied. For the inclusion criteria all participants must be current students from MCE. They should also have taken a course on the same subject for at least one period before the test. Finally, they would have agreed to attend face-to-face sessions and have signed the informed consent form for participating. The exclusion criteria states that a student cannot participate if they manifested being under medical treatment or having a psychiatric diagnosis. They also weren't allowed to participate if they are left-handed, have heart problems or if they had consumed foods or beverages with high sugar levels, or caffeine, at least 12 hours before the study.

## 4. Data Collection

For the acquisition of physiological signals, a variety of devices were used. The Empatica E4 bracelet (Empatica E4, Empatica Inc., Cambridge, MA, USA) for the acquisition of temperature, movement, EDA and blood volume pulse (BVP). The EEG signals were recorded with a mobile recording system (LiveAmp 8, Brain Products GmbH, Gilching, Germany). The LiveAmp 8 includes 8 active gel electrodes arranged according to the international 10-20 system (FP1, FP2, FC3, FC4, C3, C4, Pz). A stereo and RGB camera (RealSense Depth Camera d435i, Intel Corporation, SC, California, USA). Lastly, a psychometric instrument, based on the engineering subscale of the Questionnaire of interest in STEM areas (Kier 2014). This evaluation has 11 items and a five-point Likert scale response format, from strongly disagree to strongly agree. Likewise, this instrument was applied in electronic format to a sample with similar characteristics.

This project’s data collection methodology was divided into six phases, these are discussed below:

1. Approach: a multidisciplinary team was formed with experience in electronics, psychology, programming, and the use of biomedical engineering. Also, a protocol was drawn up, which was reviewed by Tecnológico de Monterrey’s ethics committee.
2. Data collection: The following physiological signals were obtained: temperature, brain waves (8 EEG channels), EDA, and BVP at 250, 4 and 64 Hz respectively. Additionally, one session before beginning the acquisition and one session after finishing, the subjects filled in a questionnaire to detect their vocational interest.
3. Data analysis and results: as soon as the first data acquisition session was concluded, the raw signals were filtered and analyzed in diverse ways:

a. Brain waves: Butterworth bandpass filter of order 7 between 1 and 100 Hz, and a notch filter was centered at 60 Hz. Both filters were employed to remove powerline-related noise and drift. Also, an independent component analysis (ICA) was used to remove muscular and ocular movement artifacts. An adaptive filter was applied to obtain the different frequency bands sought: theta (4-8 Hz), alpha (8-13 Hz), low beta (13-21 Hz), high beta (21-30 Hz), and gamma (30-80 Hz). The adaptive filter had two constant values: a passband ripple of 7 dB and a stopband attenuation of 10 dB.
b. Electrodermal activity: the moments in which the signal increased its value over a pre-defined threshold generating a peak in the graph was searched.
c. Blood volume pulse: it was filtered with and processed to calculate the HRV metrics in time and frequency using the HeartPy (Van Gent et al. 2019) and Neurokit2 (Makowski et al. 2020) Python libraries.
d. Temperature: How constant the temperature was during the tests was examined to detect if there were any sudden changes that may affect the electrodermal activity information capture.
e. Video: pre-existing Python libraries were used to detect the emotions that the subject presents throughout the test through their facial expressions using Python’s DeepFace library (Serengil et al. 2020).
f. Psychometric evaluation: a psychometric measurement instrument was applied to parameterize vocational interest at the beginning and end of the study.

Afterwards, data was processed for Machine Learning (ML) purposes, as on Machine Learning, situations usually have two types of variables: Source and target. Where source variables represent the variables that would be used to predict the target variables. Regarding the current situation, biometric data is provided by: EEG; Empatica; and Computer Vision (CV) algorithm that predicted emotions. While there are two target variables: The student's performance based on the business' metric, depending on the activities of the student and the lecturer's perspective; and the change of STEM interest, provided by the psychometric test.

Analyzing the biometric data: EEG and Empatica data had continuous variables and the domain from each of their features is diverse, while the CV algorithm had a categorical variable regarding the emotion detected. The proposal was to transform the domain of every feature to a continuous variable with domain between 0 and 1, and thus $0 \leq x \leq 1$.

Starting with EEG and Empatica data, a standard scaler was first applied to transform their values to mean equal to 0 and standard deviation equal to 1, based in the following equation:

$$x(t)_{(norm),\,F,\,S} = \frac{x(t)_{F,\,S} - \mu_{F,\,S}}{\sigma_{F,\,S}}$$

Where *x(t)* is the non-normalized value, while $\mu$ the mean and $\sigma$ the standard deviation of each feature *F* and student *S*. Furthermore, the sigmoid function, which has a range between 0 and 1, was applied to the normalized values, and so for each $x(t)_{norm}$ value, a new $x(t)_{sigmoid}$ value between 0 and 1 is returned. And so, the aforementioned transformation is shown in the following equation:

$$x(t)_{sigmoid} = \frac{1}{1 + e^{-x(t)_{norm}}}$$

The aforementioned pipeline was followed for the Empatica and EEG features. Although, a set of combined features were first created prior to normalization, by iterating over each feature (not combining features from distinct devices):

- $F_i$ -I: Inverse on ith feature ($\frac{1}{x(t)_{Fi}}$)
- $F_i$ -L: Natural logarithm on ith feature ($\ln(x(t)_{Fi} + \varepsilon)$)
- $F_i$ -M- Fj: Multiplication of ith feature with jth feature ($x(t)_{Fi} \times x(t)_{Fj}$)
- $F_i$ -D- Fj: Division of ith feature with jth feature ($\frac{x(t)_{Fi}}{x(t)_{Fj} + \varepsilon}$)

It could be noted that a small number in the form of an epsilon $\varepsilon$ was used to avoid errors, due to 0-sensitive functions: the natural logarithm and the division by 0. Moreover, a separate list of *combinations* was used to keep track of the combinations of ith and jth features, and so not to generate duplicate features when multiplying ith feature with jth feature and vice versa (as they produce the same number).

Moving on to the CV algorithm, an emotions' probability of distribution based on a 5-minute window was created. Additionally, an adapted version of the Laplace transformation for probabilities was employed, which removes probability 0 on all events. The transformation is widely used in bioinformatics, as the probability of mutation on a nucleobase is always possible, even if it is not found on a given chain of genes, and so the probability on a real-world scenario is not 0. The adapted version of this transformation is shown in the following equation:

$$p_{e,\,w} = \frac{f_e + 1}{n_w + m}$$

The previous equation is a transformation derived from a basic equation of probability, where probability of an event *p* is given by the frequency *f* of an event in a certain window of events. Although here the frequency is summed by 1 on the numerator and by the possible events *m* in the denominator. And so $f_e$ is the total number of occurrences of a given emotion on the 5-minute, $n_w$ is the total number of emotions detected on the window, while *m* is the number of emotions available (6). This transformation always returns a probability different from 0, and so it removes sparsity when the CV algorithm does not detect a wide range of emotions on a given window, which is mainly due to errors rather than a subject not having any emotion.

# 5. Results and Discussion

## 5.1 Numerical Results

The mean result of the Programming, Robotics, and 3D Design subscales for psychometric interest detection before the intervention were 3.04, 3.17, and 3.23 respectively. Where 0 means no interest and 4 denotes a very high interest. After the intervention, the results were 3.22, 3.46, and 3.11 in the same order. The results show a high to very-high interest in all STEM subdimensions before and after the intervention. Furthermore, the tests taken after the intervention report a higher interest in all subdomains except 3D Design. This may relate partly to sampling bias, as the examined sample is intendedly enrolled in an extracurricular STEM training. Thus, we may infer there is some amount of interest. Additionally, if it we were to isolate the stimuli, we may infer that the measured before and after change is partly due to the intervention.

Further analysis showed that children aged 6 to 11 showed a higher mean interest before the intervention in all subdomains than children aged 12 to 15. This finding holds after the intervention except in 3D Design. A probable cause may be that younger children are prone to extreme response bias, where participants report more extreme answers. For example, younger children may be answering strongly disagree to strongly agree much often than their older counterparts. The mean interest also increases when segmented by age group on all occasions, but 3D Design, supporting the possibility that the intervention may increase interest in Programming and Robotics.

Moving on to ML, a feature selection was done, which started removing features that correlated with each other more than 95%, this reduced the number of features from 10799 to 898. Afterwards, a feature selection criterion known as the Mean Decrease Impurity (MDI), measured by the GINI index and computed by a Random Forest Regression algorithm, was used to reduce the number of features. The algorithm is a randomized algorithm, and so a total of 20 iterations were done to generalize on multiple seeds and so get unbiased results, thereafter, the top 20 features with the greatest MDI were chosen.

The previous process was done for each target variable (student's performance, change in STEM interest), and so the best features for the student's performance target variable are shown in Table 1. With a final feature selection process that used the Pearson correlation coefficient (R) with the target variable, as well as their P-value to determine whether the feature is statistically significant or not (using $\alpha = 0.05$), this criterion then removes the last 4 non-significant features, which leave us with 16 significant features for the student's performance (EEG = 12, CV = 2, Empatica = 2).

Table 1. Correlation between student's performance and their top 20 features with their greatest MDI.

| Feature | Pearson correlation coefficient (R) | P-value | Device |
|---|---|---|---|
| Alpha_FP2-M-Fatigue | 0.411913 | 0.000000 | EEG |
| Alpha_F4-M-Fatigue | 0.390876 | 0.000000 | EEG |
| Theta_FP2-M-Fatigue | 0.376956 | 0.000000 | EEG |
| Theta_PZ-L | 0.352213 | 0.000000 | EEG |
| Theta_PZ | 0.346399 | 0.000000 | EEG |
| Load-D-Fatigue | 0.334513 | 0.000000 | EEG |
| Fatigue-D-Load | 0.328633 | 0.000000 | EEG |
| Theta_FP1-M-Fatigue | 0.318455 | 0.000000 | EEG |
| LowBeta_C3-M-Theta_PZ | 0.260609 | 0.000000 | EEG |
| Surprise | 0.247494 | 0.000000 | CV |
| Gamma_FP1-D-Theta_PZ | 0.241915 | 0.000000 | EEG |
| HRV_PIP-D-HRV_C1d | 0.199660 | 0.000001 | Empatica |
| HighBeta_PZ-D-HighBeta_F3 | 0.182521 | 0.000005 | EEG |
| Happy | 0.160123 | 0.000061 | CV |
| Gamma_F3-D-Theta_PZ | 0.131813 | 0.000993 | EEG |
| HRV_MeanNN | 0.081000 | 0.043616 | Empatica |
| HRV_MeanNN-D-HRV_CVI | 0.076910 | 0.055421 | Empatica |
| Alpha_PZ-D-Fatigue | 0.048423 | 0.228217 | EEG |
| Fatigue-D-Theta_F3 | 0.015169 | 0.705970 | EEG |

| | | | |
|---|---|---|---|
| LowBeta_C4-D-HighBeta_PZ | 0.001088 | 0.978417 | EEG |

The last feature selection criterion is also applied on the top 20 best features with their greatest MDI, according to the last target variable: change in STEM interest. Results are then shown in Table 2, using the same level of significance (using $\alpha = 0.05$), the last three features are removed, which leave us with 17 best features (EEG = 15, CV = 1, Empatica = 1),

Table 2. Correlation between change in STEM interest and their top 20 features with their greatest MDI.

| Feature | Pearson correlation coefficient (R) | P-value | Device |
|---|---|---|---|
| Surprise | 0.349198 | 0.000000 | CV |
| Alpha_F4-D-Alpha_FP1 | 0.320681 | 0.000000 | EEG |
| LowBeta_F4-M-High-Beta_PZ | 0.318051 | 0.000000 | EEG |
| Fatigue-D-Alpha_F4 | 0.269225 | 0.000000 | EEG |
| Fatigue | 0.232230 | 0.000000 | EEG |
| LowBeta_PZ-D-Fatigue | 0.230479 | 0.000000 | EEG |
| Alpha_C4-D-Alpha_PZ | 0.219637 | 0.000000 | EEG |
| Fatigue-D-Alpha_C4 | 0.209250 | 0.000000 | EEG |
| Alpha_F4-D-Fatigue | 0.206663 | 0.000000 | EEG |
| Alpha_PZ-M-Alpha_F4 | 0.192964 | 0.000001 | EEG |
| Gamma_F3-M-Fatigue | 0.171491 | 0.000017 | EEG |
| Alpha_PZ-I | 0.170675 | 0.000019 | EEG |
| Alpha_C4-M-Load | 0.123173 | 0.002105 | EEG |
| HRV-MedianNN | 0.107745 | 0.007201 | Empatica |
| Alpha_PZ-M-LowBeta_FP1 | 0.103809 | 0.009634 | EEG |
| Alpha_PZ-M-LowBeta_C3 | 0.086749 | 0.030657 | EEG |
| Fatigue-D-Load | 0.085656 | 0.032830 | EEG |
| Theta_C4-D-Theta_C3 | 0.035904 | 0.371752 | EEG |
| Alpha_PZ-D-Alpha_C4 | 0.023034 | 0.566692 | EEG |
| HRV_MeanNN-D-HRV_MedianNN | 0.005457 | 0.892040 | Empatica |

### 5.2 Graphical Results

The previously described, best statistically significant features, are then displayed using two separate bar plots for each target variable, which are shown in the Figure 1. It can be seen that the EEG is the biometric device from which most significant features are drawn (14 for student's performance, 15 for change in STEM interest), although this device had less features than the Empatica device (38 and 64 features before feature generation). On the other hand, the probability of surprise and HRV_MeanNN are two, non-EEG features, present in both target variables.

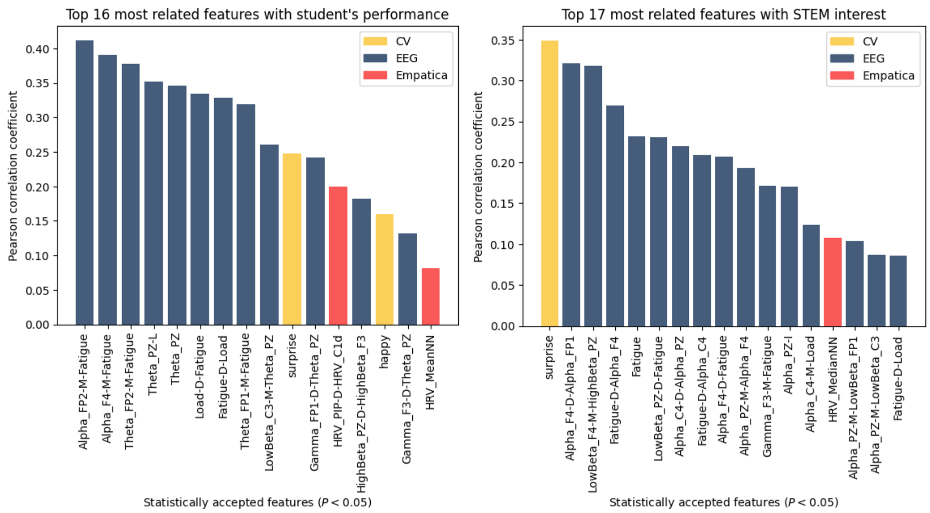


Figure 1. Statistically accepted features and their correlation with the target variable

## 5.3 Proposed Improvements

In addition to the best features used for each target variable, the student's performance model used an additional categorical feature that indicated the subject of the current lecture (Robotics, Programming, 3D Design), given that four lectures were used, this generated 4 new features (16 previous + 4 information features = 20 total features).

These features were only included in the student's performance model, as it represented improvement using the coefficient of determination (R2) as performance metric and Multiple Linear Regression (MLR) as model: Increase (+ 0.3014 R2) for student's performance and decrease (- 0.0859 R2) for change in STEM interest.

Additional improvements that could be implemented to increase the performance of the generated models are the inclusion of more children into the experiment, as these would then increase the data available for training. For the current study, only the median value on a 5-minute window was used (EEG and Empatica features), although more features could be generated using additional statistical and mathematical features, such as: mean, mode, standard deviation, maximum, minimum, interquartile range, skewness, kurtosis and many more.

## 5.4 Validation

Using the most significant features for each target variable, three regression models were fitted in order to know which model performed best according to each target variable:

- Multiple Linear Regression (MLR)
- Random Forest Regressor (RF)
- Gradient Boosting Regressor (GBR)

A different model was created for each target variable, and the validation methodology used was Leave-One-Out (LOO) using lectures, this methodology uses all but one lecture to train the ML model, while the removed lecture is used for testing.

So, for each student and lecture, three new models were trained to predict the score on the removed lecture's samples. Afterwards, the mean of predicted scores was calculated to get the overall score of that lecture, which was saved on an array for later comparison. A student could have multiple lectures and so the other lectures from the student from which the removed lecture was from, were also removed, in order to reduce bias on using lectures from the same student on training and testing.

Using the predicted scores, the coefficient of determination ($R^2$) was calculated to evaluate each model's predictions with respect to the ground truth. Afterwards, the continuous prediction is encoded to its categorical feature (either student's performance or whether the change of interest is positive or negative).

The aforementioned encoding was employed to calculate an additional categorical measure: accuracy, defined as the number of correctly predicted labels divided by the number of total samples. Now then, with the combination of both continuous and categorical metrics, the best model for each target variable was selected.

Results based on the evaluation metrics are displayed in Table 3. The difference between accuracies (19.44%) may be due to the fact that the interest on STEM subjects is drawn from a standardized, psychometric test, while the student's performance is based on subjectivity of the tutor and subject's grades.

Table 3. Performance metrics for each target variable.

| Student's performance | | | Change in STEM interest | | |
|---|---|---|---|---|---|
| Model | Coefficient of determination ($R^2$) | Accuracy (%) | Model | Coefficient of determination ($R^2$) | Accuracy (%) |
| MLR | 0.4477 | 58.33 | MLR | 0.0457 | 66.67 |
| GBM | 0.3553 | 52.78 | GBM | 0.4264 | 75.00 |
| RF | 0.3244 | 55.56 | RF | 0.4504 | 80.56 |

The detailed predicted scores and ground truth is further represented in Figure 2. It can then be seen that the LOO validation does not perform well on predicting scores that are near to maximum and minimum values from the current population, this was a limitation from our data because class imbalance was present in extreme values, as few participants performed either excellent or quite poorly. Although, the LOO validation accurately predicted when the score of a given lecture was near to the population's mean, as more similar data could be used to generate the given prediction.

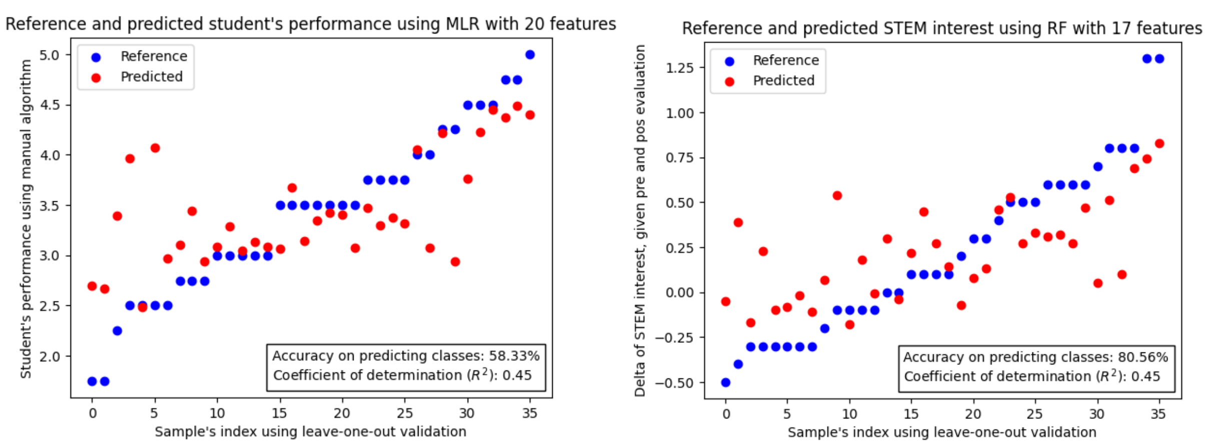


Figure 2. Predictions (red dots) and ground truth (blue dots) using the best model for each target variable: (Left) Student's performance, (Right) Change in STEM interest

## 6. Conclusion

This investigation proposes a methodology through which a machine learning algorithm is to be trained for the purpose of vocational interest detection in children and adolescents. Said proposal stands based on past investigations demonstrating the relationships between physiologic parameters and cognitive variables such as cognitive load and level of focus, as well as stress and excitement. It is intended that with this investigation, the interpretation given to Machine Learning results can be refined to point out novel investigation frontiers regarding vocational interest detection which can be employed by a single user, educational institutions, governments, and enterprises. The main potential of these new measurement methodologies based on biometry lies in the promptness in which the capabilities as well as the interest can be assessed, without any cultural biases. This creates a powerful instrument that can be used to improve the way society works, and transform it into a healthier, non-cultural biased structure that may positively impact the economic development by increasing the chances of having prepared and passion-oriented professionals working in areas where they can be fully prepared to thrive in.

## References


M. W. Kier, M. R. Blanchard, J. W. Osborne, and J. L. Albert, "The development of the stem career interest survey (stem-cis)," *Research in Science Education*, vol. 44, pp. 461–481, Jun2014.

C. D. Nye, R. Su, J. Rounds, and F. Drasgow, "Vocational interests and performance: A quantitative summary of over 60 years of research," *Perspectives on Psychological Science*, vol. 7, no. 4, pp. 384–403, 2012. PMID: 26168474.

T. Tyler-Wood, G. Knezek, and R. Christensen, "Instruments for assessing interest in stem content and careers," *Journal of Technology and Teacher Education*, vol. 18, 01 2010.

D. J. Hughes, *Psychometric Validity*, ch. 24, pp. 751–779. John Wiley & Sons, Ltd, 2018.

S. Rasheed, H. M. Al-Barhamtoshy, H. B. Saifaddin, and W. Shalash, "Assessment of vocational types and eeg analysis using holland test questionnaire," in *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control*, AIRC '19, (NewYork, NY, USA), p. 53–60, Association for Computing Machinery, 2019.

M. Mikolajczak, K. Bodarwé, O. Laloyaux, M. Hansenne, and D. Nelis, "Association between frontal eeg asymmetries and emotional intelligence among adults," *Personality and Individual Differences*, vol. 48, no. 2, pp. 177–181, 2010.

J. L. Holland, *Making vocational choices: A theory of vocational personalities and work environments*. Odessa and
FL: Psychological Assessment Resources, 3 ed., 1997.

S. E. Chi, C. B. Park, S. L. Lim, E. H. Park, Y. H. Lee, K. H. Lee, E. J. Kim, and H. T. Kim, "Eeg and personality dimensions: A consideration based on the brain oscillatory systems," *Personality and Individual Differences*, vol. 39, no. 3, pp. 669–681, 2005.

T. Johannisson, "Correlations between personality traits and specific groups of alpha waves in the human EEG,"
*PeerJ*, vol. 4, p. e2245, jul 2016.

N. S. Roslan, L. I. Izhar, I. Faye, M. N. M. Saad, S. Sivapalan, and M. A. Rahman, "Review of eeg and erp studies of extraversion personality for baseline and cognitive tasks," *Personality and Individual Differences*, vol. 119, pp. 323–332, 2017.

C. Saffiera, R. Hassan, and A. R. Ismail, "Preventive and curative personality profiling based on eeg, erp, and big five personality traits: A literature review," *Journal of Computational and Theoretical Nanoscience*, vol. 17, pp. 531–545, 02 2020.

V. A. Maksimenko, A. E. Runnova, M. O. Zhuravlev, P. Protasov, R. Kulanin, M. V. Khramova,A. N. Pisarchik, and A. E. Hramov, "Human personality reflects spatio-temporal and time-frequency eeg structure," *PLOSONE*, vol. 13, pp. 1–22, 09 2018.

W. Li, X. Hu, X. Long, L. Tang, J. Chen, F. Wang, and D. Zhang, "Eeg responses to emotional videos can quantitatively predict big-five personality traits," *Neurocomputing*, vol. 415, pp. 368–381, 2020.

Empatica Inc., "Empatica E4," 2021.15. Brain Products Inc., "Brain Products GmbH," 2021.

P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behavior*, vol. 66, pp. 368–378, 2019.

D. Makowski, T. Pham, Z. J. Lau, J. Brammer, F. Lespinasse, C. Schölzel, S. Chen, and H. Pham, "Neurokit2: A python toolbox for neurophysiological signal processing," 07 2020.

S. I. Serengil and A. Ozpinar, "Lightface: A hybrid deep face recognition framework," in 2020 *Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–5, 2020


## Biography

**Gustavo Olivas** Martínez currently attends his undergraduate studies in Mechatronics Engineering at Tecnológico de Monterrey, Nuevo León, México. His interests consist of control systems, programming, 3D design and mechanics. He has worked both in academic and industrial projects. He collaborated as a research engineer for the SIEMENS factory in Nuevo León for a project which aimed to optimize internal processes for cost reductions. His most recent collaboration was in a joint investigation between MachineCare Education and Tecnológico de Monterrey whose objective is to find a pattern in biometric signals for interest detection using Machine Learning. There he performed tasks as a software developer to design an acquisition and analysis system for most of the collected biometric data.

**Milton O. Candela** Leal is a Sophomore Undergraduate Student in Biomedical Engineering at Tecnológico de Monterrey, Nuevo León, México. He has experience in Python, MATLAB, and R. His research interests are Data

Science, Machine Learning and Deep Learning applied in biomedical engineering context such as Brain-Computer Interfaces, Neurorobotics and Health Informatics. He has collaborated in many IUCRC BRAIN TEC initiatives that involve the creation of Artificial Intelligence models, based on wearables and biometric data.

**Juan C. Ocampo** Alvarado is pursuing a BSc in Clinical Psychology at Instituto Tecnológico y de Estudios Superiores de Monterrey. He is a research assistant for IUCRC BRAIN initiative at that university and at Universidad Regiomontana's Center for Intercultural Studies of the Northeast, leading a 2-year study on the Mexican prison crisis. He has collaborated with the Neuroscience Institute of Guayaquil, University of Texas at Austin, and the Organization of American States. He has co-authored peer-reviewed articles on disability inclusion, prison violence, and neuroeducation.

**Luis F. Acosta** Soto is a Biomedical Engineer with a minor in biomedical microtechnologies. He is recipient of the Academic Excellence scholarship for undergraduate studies, and a co author of a Microfluidic MEMS published article, and a book chapter on fluorescent BIOMEMS. He holds the position of research assistant at Writing Lab, an institution focused on the professional development of academic writing and investigation at the Instituto Tecnologico de Estudios Superiores de Monterrey. Further, he has worked on neuroscience-related projects; his main areas of interest comprehend the neurosciences and signal processing technologies in the biomedical field.

**Diego M. Botín** Sanabria is a Mechatronics Engineering bachelor student in Tecnologico de Monterrey. He is Project Manager Jr. for the Digital Twins and Digital Cab research programs and technical lead for this work. Diego has current collaboration with the University of Technology Sydney and Macquerie University, Australia. Diego is the Chief Engineer for his university's scuderia: Tec Racing and has experience in C, MATLAB/Simulink, SolidWorks and CarSim. His research interests are systems modeling and simulation, vehicle dynamics, digital twins, and motorsports engineering.

**Gladys G. Díaz** Armas received her bachelor's degree in Biomedical Engineering from Tecnológico de Monterrey, Mexico, in 2021. She was part of the Emerging Leaders in the Americas Program as a visiting student at the University of British Columbia, Canada, in 2019. During her studies, she collaborated as a research assistant in various national and international groups dedicated to distinct areas of bioengineering, including tissue engineering, biosensors, and neurosciences. She currently works as a biology research analyst in the private sector. Her research interests include the use of machine learning and point-of-care devices for the diagnosis and treatment of chronic diseases.

**Aime J. Aguilar** Herrera is currently a biomedical engineering student at the Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM) and Community Manager at Campus City and IUCRC BRAIN TEC. She studied technical degree in clinical laboratory at the Center for Industrial Technology and Services Baccalaureate 189 (CBTis 189). Her areas of interest include neuroengineering, bioinstrumentation, and bio-signal processing. She is currently working on research projects in biomechanical evaluation using machine learning algorithms and signal processing to evaluate fatigue in university students.

**Esther A. Delgado** Jiménez studied technical degree in Business Administration at the Instituto Internacional de Estudios Superiores (IIES) in Reynosa. Currently, she is a biomedical engineering student at the Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM) and project manager of two projects that are under development at Campus City and BRAIN. Nowadays, she is working on research projects in biomechanical evaluation using machine learning algorithms and signal processing to evaluate fatigue in university students. Her areas of interest include bioinstrumentation, biomechanics, and biomedical signal processing.

**Myriam Alanis** Espinosa received a Ph.D. in biomedical engineering from the Center of Research and Advanced Studies (CINVESTAV) in 2019 and the M.Sc. in Electronic Systems from Tecnologico de Monterrey in 2010. She is the co-founder and Director of Research and Technology of the startup named SATI where she is in charge of R&D of products and services of health monitoring. Currently, she is the program director of biomedical engineering at Tecnologico de Monterrey, Guadalajara Campus. Her research interests are biomedical signal processing, neuroengineering, Virtual Reality, Internet of Medical Things, and machine learning.

**Mauricio A. Ramírez** Moreno received his PhD in Biomedical Engineering in 2019 in Cinvestav Monterrey (Mexico). In 2019, he joined the School of Engineering and Sciences at Tecnologico de Monterrey. His main research interests include Brain-Computer Interfaces, neuroengineering, robotics, biomechanics, smart cities and machine learning; and has published five journal papers, three conference papers, and one book chapter in the fields of smart cities and neuroengineering. He is currently a postdoctoral researcher, and the Program Manager of Campus City Smart Mobility and the IUCRC BRAIN TEC initiatives.

**Rubén Morales** Menendez has a PhD in Artificial Intelligence (2003). Dr. Morales has taught more than 5,000 students through 30 different courses at Tecnologico de Monterrey. His research interests include machine learning approaches for control systems, fault diagnosis and estimation of dynamic systems, fault control of vehicle dynamics, and active learning techniques for engineering education. He has more than 250 research and has also graduated 6 PhD and 25 MsC students. He is currently the Dean of Graduate Studies at the School of Engineering and Sciences.

**Ricardo A. Ramirez** Mendoza received his Ph.D. degree in Automation and Production from Grenoble Institute of Technology, France (1997). He has published over 500 papers in journals, conferences, etc, and has mentored over 40 graduate students, who occupy leading positions in academia & industry. His research lines include Automotive Control, Active Control, Vehicle Dynamic Control, Mechanical Vibrations, Brain-Computer-Interface, and Biomedical Signal Analysis. He is Dean of Research and Graduate Studies at the School of Engineering and Sciences, and Professor of Mechatronics & Mechanical Engineering in Tecnologico de Monterrey.

**Jorge de-J. Lozoya** Santos received his PhD Degree in Mechatronics and Advances Materials (2013) from Tecnologico de Monterrey. Jorge has current collaborations with Politecnico di Milano, Italy; Institute Polytechnique de Grenoble, France; Universita degli Studi di Modena and Reggio Emilia, Italia among others. He has more than 25 international conferences, 10 indexed journals and 4 patent applications. His research interests are intelligent transportation systems, modeling and control of systems, applied automatic control and automotive systems. He is Research Professor in Tecnologico de Monterrey, School of Engineering and Sciences.

**Jorge Murrieta** González ended his studies in the Electronics Engineering area in 2000 at the Tecnológico de Monterrey. He has more than 17 years of experience in the Test System Integration Industry. He also worked for 9 years with Cygnus Automated Test & Control Solutions. Then, with his strong entrepreneurship skills, founded Machine Care Industrial Services as a company which supplies prompt support and maintenance of machinery.

**Aleyka García** Briseño is in charge of detecting new talents that may be added to Machine Care Education sector, where they teach the children different areas of engineering. She performed critical tasks for this investigation both to confirm the psychometric findings as well as to coordinate the help and availability of the selected case subjects.