

## RESEARCH ARTICLE



**Received:** 25-06-2020

**Accepted:** 29-07-2020

**Published:** 11-8-2020

**Editor:** Dr. Natarajan Gajendran

**Citation:** Musa RM, Suhaimi MZ, Musa A, Abdullah MR, Abdul Majeed APP, Maliki ABHM (2020) Predicting students academic performance from wellness status markers using machine learning techniques. *Indian Journal of Science and Technology* 13(29): 2947-2955. <https://doi.org/10.17485/IJST/v13i29.999>

\* **Corresponding author.**

[rabiumuazu86@gmail.com](mailto:rabiumuazu86@gmail.com)

**Funding:** None

**Competing Interests:** None

**Copyright:** © 2020 Musa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.org/))

**ISSN**

Print: 0974-6846

Electronic: 0974-5645

## Predicting students academic performance from wellness status markers using machine learning techniques

**Rabiu Muazu Musa<sup>1\*</sup>, Muhammad Zuhaili Suhaimi<sup>1</sup>, Azlina Musa<sup>1</sup>, Mohamad Razali Abdullah<sup>2</sup>, Anwar P. P. Abdul Majeed<sup>3</sup>, Ahmad Bisryi Husin Musawi Maliki<sup>4</sup>**

**1** Center for Fundamental and Continuing Education, Department of Credited Core-curriculum, Universiti Malaysia Terengganu Gong Badak, Kuala Nerus, 21030, Terengganu, Malaysia

**2** East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, 21300, Terengganu, Malaysia

**3** Innovative Manufacturing, Mechatronics and Sports Laboratory Faculty of Manufacturing & Mechatronics Engineering Technology Universiti Malaysia Pahang, Pekan, 26600, Pahang Darul Makmur, Malaysia

**4** Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, 21300, Terengganu, Malaysia

## Abstract

**Background:** A high level of wellness is vital to produce a well-balanced and competent graduate. Despite the importance of wellness to the attainment of high productivity, limited attention is directed towards predicting the academic performance of students with regards to wellness markers. This study aims to ascertain the association between wellness and academic achievement of undergraduate students. **Methods:** A total of 250 undergraduate students drawn from various faculties in one of the public universities in Malaysia participated in the study. The wellness Lifestyle inventory which evaluates an initial rating of a person's present attempt to remain healthy and assessed nine major areas namely; health-related fitness, nutrition, avoiding chemical dependency, stress management, personal hygiene and health, disease prevention, emotional well-being, personal safety and environmental health protection was used as a tool for determining the wellness status of the students whilst their CGPA was utilized as a measure for their academic achievement. K-means clustering analysis was used to group the students into high grades students (HGS) and low grades students (LGS) through their CGPA. A Logistic Regression Model (LR) is developed to classify the students based on their wellness status markers. **Findings:** An excellent classification accuracy of 99 to 100 % was obtained from the LR model for both training and testing, respectively. Moreover, analysis of variance demonstrated that the HGS and LGS differ in their effort to stay healthy with respect to certain markers  $p < 0.05$ . **Applications:** To ensure better academic grades, some wellness status elements need to be accentuated amongst undergraduate students.

**Keywords:** Wellness status; undergraduate student; academic performance; K-means clustering; logistic regression

## 1 Introduction

It is generally believed that partaking in physical exercise enhances the quality of life of an individual. However, the advent of technology coupled with changes in lifestyles, have demonstrated that fitness alone could not always be enough to lower the risk of ailments and ensure better health. Some lifestyles habit such as smoking, drinking too much alcohol, excessive stress as well as eating various foods that are high in saturated fat could pose a risk for cardiovascular disorder and other chronic diseases<sup>(1,2)</sup>. Although, many people are aware of their unhealthy lifestyle behaviors, nonetheless, they often seem to contend with life as long as they remain free from symptoms or illness. Changes in this category of people are not considered until major health problems are incurred. Therefore, it is worth to highlight that the current lifestyle habits dictate the health as well as the well-being of tomorrow. Consequently, the wellness of an individual could to a large extent determine his or her productivity in any performance-based activity<sup>(3)</sup>.

Predicting students' performances have received several attentions from different researchers globally due to its relative importance in understanding the factors that could affect the students learning hitherto their final grades. One of the commonly used methods of predicting the student's performances is the machine learning-based supervised technique. The supervised learning method of machine learning is essentially a technique used to predict numerical or categorical indicators with respect to a set of predefined parameters or variables<sup>(4)</sup>. To date, the application of machine learning methods towards the prediction of students' academic performances, dropout rates, course registration, grades as well as learning techniques has gained wider recognition for over a decade.

In a very recent study, Anusha et al.,<sup>(5)</sup> applied a different machine learning algorithms i.e k nearest neighbor (k-NN), Decision Tree, Random Forest and a conventional linear regression model to predict undergraduate computer science students' performances. A set of historical data for a total of four years which constitutes theory marks, term test marks, viva as well as practical marks for each student was used to predict the students' GPA. It was demonstrated from the results that the conventional linear regression was able to provide a good prediction of the student's GPA with an overall  $R^2$  of 94%. The authors concluded that the developed model could help the students to be fully engaged towards improving their academic performances by virtue of their historical record of performances.

Kardan et al.,<sup>(6)</sup> carried out a study to investigate the students' course selection and identify the potential elements that affect students' satisfaction with regards to the online course they preferred. The data of 714 online graduate courses from 16 academic terms were used to predict the students' final number of registrations in every course after the add and drop period using an Artificial Neural Network (ANN) algorithm. It was shown from the initial finding that the network was able to provide a prediction rate of 90%. However, after some optimization, it was observed that the prediction ability of the network increased to 92%. It was concluded from the study that some influential factors specifically course characteristics, instructor characteristics, course difficulty, students workload, course grade, course type with regards to compulsory or elective, time of the course, number of clashes of the course with other courses and final examination time could predict the selection of a particular course by the students.

Amrieh et al.,<sup>(7)</sup> employed a data mining technique through which educational related data that comprised of demographic features specifically gender and nationality, academic background elements comprising of educational level, grade level and section, behavioral features including raising hands in class, visited resources, parents answering survey as well as satisfaction towards the school are utilized to forecast the academic performance of the students. Different machine learning models i.e. ANN, Naïve Bayes and Decision Tree are used to achieve the purpose of the study. Moreover, ensembles algorithms specifically bagging boosting and Random Forest were applied to improve the efficacy of the model. A strong association between the students learning behavior and their academic achievement were observed in the study. The accuracy of the model while considering the behavioral features was shown to be 22% higher improvement as opposed to the accuracy observed when the feature is eliminated. The model was shown to improve by 26% through the application of ensemble boosting technique and the model was tested against new incoming students in which an overall accuracy of 80% was obtained. It was postulated by the researchers that the method employed proved reliable in predicting student's academic performance.

Arsad et al.,<sup>(8)</sup> developed a predictive model of engineering students' performances using ANN. Features comprising of subjects selection at semester three namely; digital systems, signals and system II, material science, mathematics II as well as English I were used as the input parameters whilst the CGPA was used as the output parameter. Two separate models for undergraduate and diploma students were established. It was shown from the findings that the ANN algorithm was able to provide a good prediction of the students' academic performance with an  $R^2$  of 97% and 92% for both degree and diploma students respectively. In a different study, Ibrahim et al.,<sup>(9)</sup> predicted students' academic performance in an undergraduate program. Demographic information of the student is used as the input variable while the CGPA was utilized as the predictor variable. The authors developed three different models employing Logistic Regression, ANN as well as Neuro-Fuzzy. It was demonstrated that the Neuro-Fuzzy model outperformed all the two models. The authors concluded that the choosing parameters were effective in forecasting the student's achievement.

In an earlier study, Oladokun et al.,<sup>(10)</sup> predicted the likelihood of engineering students to be considered for admission into a university. Several indicators that could influence the performance of the students were considered such as level subject's scores, subject combination, matriculation examination scores, age on admission, parental background and gender. The aforesaid variables were trained on ANN as input variables whilst the grade of the students was used as the output parameter. It was shown from the findings that a prediction rate of up to 70% was attained. In another perspective, Romero et al.,<sup>(11)</sup> predicted computer science students' final performance through the utilization of the student's participation data in an online discussion forum. Different data mining approaches were employed in developing the model. Conventional classifiers and clustering technique were compared in their efficacy of predicting the probability of the students passing or failing a course to the data obtained from the forum usage. It was shown from the results that the application of clustering technique is effective in predicting the students' performance as compared to the other tested classifiers.

It could be observed from the various studies mentioned above that the prediction of students' performances was duly considered by several researchers. However, many of such studies focused on engineering or computer science-based students and no consideration were given to the other arts or social sciences students. Moreover, the studies mainly considered only classroom-related parameters only a few studies incorporated some demographic information of the students. It has also been noticed that no study has thus far attempted to look into the wellness markers attributed to the students. It is important to note that for students to attain a high level of academic achievement especially in a higher institution of learning, it is imperative to maintain a healthy lifestyle. The efforts of the students to pay attention to the wellness elements as well as make the necessary adjustment towards reaching a high standard of wellness, could to a large extent determine their academic performances. As such, the present study is aimed at predicting the academic performance of students from multiple programs through the utilization of their wellness status markers as well as Logistic Regression and k-means clustering algorithms.

The present paper is organized into five main headings.

1. The study participants which constitutes the number of the students involved in the study as well as their respective faculties
2. The instruments utilized for data collection.
3. The reliability as well as the power analysis carried out to establish the validity and the reliability of the study coupled with the determination of the adequate number of participants required to achieve a stable result.
4. The process of developing the k-means as well as the logistic regression algorithms are then explained. Finally,
5. The discussion of the main findings followed by a conclusion as well as recommendations is presented.

## 2 Materials and Methods

### 2.1 Participants

The participants of the present study comprised of a total number of 250 undergraduate students aged ( $20 \pm 0.76$ ). A probability sampling technique is applied to recruit the samples of the present study. This sampling is deemed appropriate as every individual within the study population has an equal chance of being selected<sup>(12)</sup>. The students are from Universiti Malaysia Terengganu drawn from various faculties consisting of Economics, Policy Management, Maritime Management, Accounting, Marketing, Financial Mathematics, Food Science, Computer Science, Biodiversity as well as Chemical Science. A total of 56 respondents were male students while 194 students were female. A total of 164 (65.6%) respondents were Malays, 27 (10.8%) respondents were Chinese, 43 (17.2%) respondents were Indians and 16 (6.4%) respondents were from other racial groups. Before the commencement of data collection in this study, the students were informed about the purpose of the research and informed consent was obtained. It is worth noting that the study was revised and approved by the departmental committee (UMT/PPAL/500-28 JILD 2-2019).

### 2.2 Instruments for data collection

The principal instruments for data collection were the wellness lifestyle Questionnaire (WLQ), curled from<sup>(13,14)</sup>. This instrument is considered apt in the present study due to its relative validity in measuring the wellness lifestyle of individuals within the specified age. The instrument comprises of 36 items. Responses are recorded on a 5-point Likert scale, from strongly agree (1) to strongly disagree (5). The 36 questions are grouped to contribute to 9 categories which can be further broken down into three categories of self that include "Excellence", "Good" and "Need Improvement". The categories include "health-related fitness", "nutrition", "avoiding chemical dependency", "stress management", "personal hygiene", "disease prevention", "emotional well-being", "personal safety" as well as "environmental health and protection". The CGPA of the students were collected and utilized as an indicator of their academic achievement.

### 2.3 Internal Consistency Reliability and Power analysis

A reliability testing was carried out prior to the commencement of full analysis in the present study. The internal consistency reliability was conducted to examine the consistency of the responses on the items of the instrument. A Cronbach's alpha coefficient was employed to examine the degree of consistency amongst the items and to ensure that the items are evaluating a single construct (unidimensional) and the student's responses are independent of each other<sup>(15,16)</sup>. It is worth noting that the coefficient values of the items demonstrated a satisfactory value ranging from 0.79 to 0.89. Moreover, a priori power analysis using G\*Power was conducted to determine the sample necessary to draw a meaningful conclusion in the study. A power analysis of multivariate analysis with a power equivalent to 0.95 and alpha of 0.05 suggested that a sample size of 150 respondents would be sufficient to detect a medium effect size of 0.25. Similarly, power analysis for the analysis of variance with a power equivalent to 0.80 and an alpha of 0.05 revealed that a sample size of 200 respondents would be effective to identify a medium effect size of 0.05<sup>(17)</sup>. Therefore, a sample size of 250 would be adequate to avoid the problem of Type II error in the current investigation.

### 2.4 Data Analysis

#### 2.4.1 Data Preprocessing

Before the commencement of the full analysis in this investigation, the data were checked to ensure that it is free from any typing errors or missing information. After the cleaning was completed, the Shapiro-Wilk test was used to ascertain the distribution as well as the skewness

of the data gathered. It was observed that the data was normally distributed based on the nine elements being investigated. However, the data was observed to be skewed towards female respondents. This skewness is considered normal due to the comparatively higher number of female students in the university as compared to the males. Moreover, the data is normalized by calculating the z score in order to ensure that all the features possessed the same scale. The z score is determined using the following equation:

$$x' = (x - \mu) / \sigma \quad (1)$$

Where:  $x'$  = mean whilst  $\sigma$  = standard deviation.

#### 2.4.2 The k-means Clustering Algorithm

Clustering is amongst the most common exploratory data analysis method that is often employed to obtain an intuition on the structure of a given dataset. Cluster analysis has been documented to be useful in identifying the feature's subgroups or samples based on certain observation<sup>(18,19)</sup>. A k-means clustering algorithm is essentially a type of cluster analysis that iteratively attempt to segregate a dataset into a k- predefined distinct as well as non-overlapping subgroups known as clusters such that each data point is assigned to only one group. In this process, the algorithm tries to make the inter-cluster data points as homogenous as possible whilst keeping the clusters as heterogeneous as possible. The k-means algorithm is selected in the present study due to its efficacy to cater for linearity relationship within the dataset since the time complexity of the k-means algorithm is linear i.e.  $O(n)$ . Moreover, k-means clustering is shown to be effective when prior knowledge of k is established i.e. a predetermined number of clusters<sup>(20,21)</sup>.

In the present study, a standard k-means clustering algorithm is applied to group the students based on their CGPA grades. The CGPA grade is set at 2 clusters ( $k = 2$ ). The evaluation process is carried out by means of a silhouette-based quality estimation technique. It is worth to highlight that a maximum number of 300 iterations with 15 re-runs were used. The initialization was set at random whilst the class centroids were estimated in order to normalize the observations from the initially aligned points to stimulate the efficacy as well as the effectiveness of the clustering<sup>(22)</sup>. Consequently, every single point is assigned to a particular cluster whilst each cluster centroid is updated simultaneously. Euclidean distance was used as a distance metric of portioning the two clusters formed i.e. high-grade students (HGS) and low-grade students (LGS). The analysis was implemented via Orange-Canvas version 3.24.0 for Windows.

#### 2.4.3 Development of the Logistic Regression Model

Logistic regression (LR) is one of the popular machine learning algorithms often employed for classification problems. LR is coined for the function used at the core of the method (the logistic function). The logistic function, also known as the sigmoid function, was primarily developed by statisticians in a view to ascertaining the properties of population growth in ecology which was deemed rapidly growing towards the threshold of the carrying capacity of the environment. The logistic function is an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1. For more detail information about the LR algorithm, the readers are encouraged to refer to the following literature<sup>(23–25)</sup>.

The LR classifier is utilized in the present study to classify the HGS and LGS based on their scores in the wellness status elements. It is worth to mention that the LR is robust in classifying two distinct sets of clusters and hence motivate the selection of the algorithm for this study<sup>(26,27)</sup>. The dataset that consists of 250 observations was split into the 70:30 ratio for training and testing purposes<sup>(26,27)</sup>. The responses on the wellness status elements (health-related fitness, nutrition, avoiding chemical dependency, stress management, personal hygiene, disease prevention, emotional well-being, personal safety as well as environmental health and protection) were used as the independent variables whilst the cluster determined i.e. HGS and LGS were treated as the dependent variables. In the training stage, the five-fold cross-validation technique was utilized. The regularization types used for the algorithm is the Ridge(L2) and the c was set at 1.

The efficacy of the model in classifying the classes of the students based on the elements investigated was evaluated through Classification accuracy (CA), Area under the curve (AUC), Precision, Recall as well as F1. The CA is importantly the ratio between the true identified observation versus the entire observation. The AUC measures the overall accuracy of the quantitative analysis; the higher the accuracy the more value is close to 1 and vice versa<sup>(28)</sup>. The precision computes the number of actual positive predictions over the entire sets of positive predictions. The Recall otherwise known as sensitivity is the ratio of the accurately predicted positive observations with respect to all the observations in true class. Conversely, the F1 is the weighted mean of both the Precision and Recall which signifies that its score considers both false positives as well as false negatives. The mathematical expressions of the said evaluation metrics are given as follows:

$$CA = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

$$PREC = TP / (TP + FP) \quad (3)$$

$$RECALL = TP / (TP + FN) \quad (4)$$

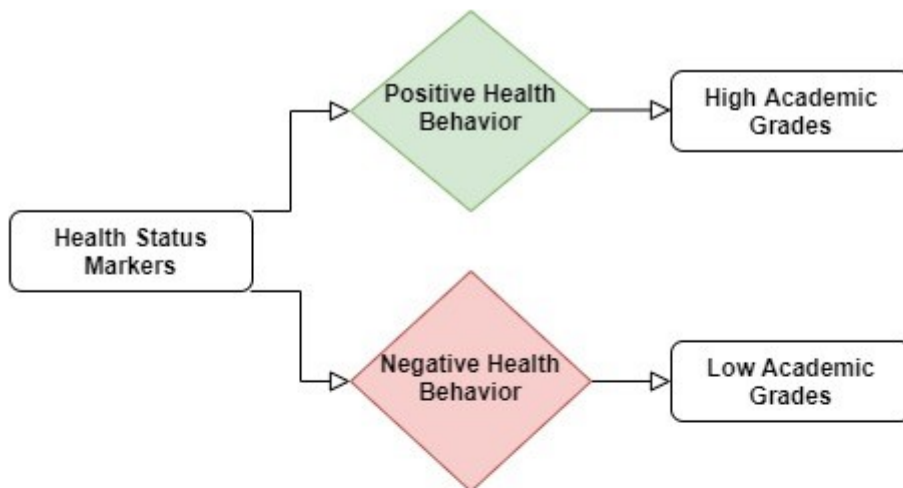
$$F1 = 2 * (Recall * Precision) / (Recall + Precision) \quad (5)$$

$$AUC = \int_a^b f(x) dx \tag{6}$$

Whereas, TP and TN stand for true positive and true negative (the number of positive and negative observations accurately predicted). The FP and FN are false positives and false negative (the number of positive and negative samples incorrectly predicted). The AUC between the two points could be determined by means of definite integral between the two points as shown in equation 6.

### 3 Results

Figure 1 represents the classes identified by the k-means clustering algorithm. The CGPA grades of the students enabled the classes formation. It could be observed from the figure that somewhat clear demarcations are formed separating the HGS and LGS.



### Graphical Abstract

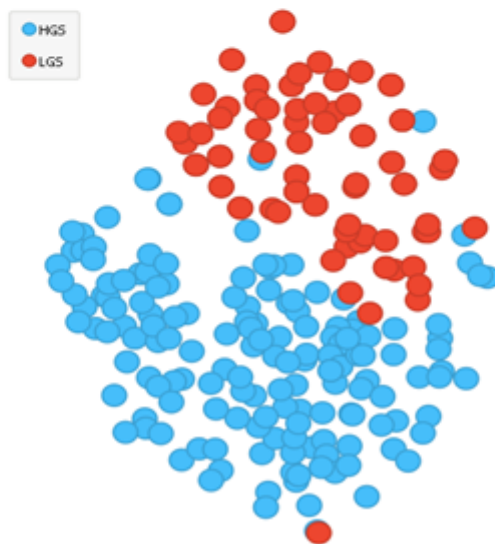


Fig 1. Clusters forms through k-means algorithm

Table 1 demonstrates the performance of the model developed in the present study. From the table, the classification accuracy (CA), the area under the curve (AUC), the Precision, Recall as well as the F1 parameters are shown for both the training and test dataset. The CA for

both the training and tests are 99 and 100% respectively whilst the AUC which is essentially the measure of separability of the model that ranges from 0 (no separation ability) to 1 high (separation ability). A precision of 97% and 100% for training and testing is displayed which revealed the ratio of correctly predicted positive observations to the total predicted positive observations. A Recall or otherwise sensitivity of the model which measures the ratio of correctly predicted positive observations to all observations in actual class is shown to be 97 and 100% for training and testing. Moreover, the F1 score which is the weighted average of Precision and Recall that takes both false positives and false negatives into account is found to be also 97 and 100% for training and testing respectively. Overall, based on the model parameters evaluated in the present study, it is tempting to postulate that the performance of the model developed is rather excellent and therefore, the wellness related markers considered in the study could potentially predict the CGPA grades level of the undergraduate students.

**Table 1.** Performance of the model parameters

Model Parameters	CA	AUC	Precision	Recall	F1
Training	99%	1	97%	97%	97%
Test	100%	100%	100%	100%	100%

Figure 2 shows the confusion metrics of the model developed. It could be seen that two misclassifications are observed from the HGS groups whilst 3 misclassifications are accrued from the LGS group during the training phase of the model i.e. (a). However, no single misclassification was transpired at the testing phase of the model (b) which further accentuated the robustness of the model in classifying the grouping of the students excellently.



**Fig 2.** Confusion metrics of the developed model

**Table 2.** Descriptive statistics and analysis of variance for the study variables

Wellness Variables	HGS		LGS		Sig.
	Mean	Std Dev+	Mean	Std Dev+	
CGPA	3.350	0.267	3.247	0.357	0.001*
Health R. Fitness	9.656	2.342	14.452	2.008	0.108
Nutrition	11.127	2.823	14.505	2.273	0.024*
Avoiding Chemical Dependence	18.510	1.693	15.054	2.891	0.001*
Stress Management	14.586	2.562	14.71	2.434	0.595
Personal Hygiene and Health	14.911	2.642	13.527	1.827	0.001*
Disease Prevention	13.669	3.249	15.043	2.216	0.001*
Emotional Well-being	16.389	2.518	14.925	2.369	0.526
Personal Safety	17.535	1.910	14.269	2.193	0.130
Environmental Health Protection	15.363	3.880	14.269	2.081	0.001*

\*P<0.05

Table 2 highlights the descriptive statistics and analysis of variance for the variables examined in the current study. It could be observed that the CGPA of the students is statistically different between the groups. Moreover, it is demonstrated that HGS is better in avoiding chemical

dependency, maintaining personal hygiene and health, and environmental health protection whereas the LGS are good in controlling nutrition and diseases prevention  $p < 0.05$ . Conversely, no significant difference was observed between the groups in health-related fitness, stress management, emotional well-being as well as personal safety.

## 4 Discussion

The purpose of the present study was to examine the wellness status of undergraduate students and to classify the students based on their scores on the wellness status markers as well as their academic achievement. The present study employed a non-conventional approach in assessing the link between wellness status of undergraduate students and the corresponding academic performance. It is worth to mention that the current study is amongst the first study to consider applying a non-conventional technique in evaluating the wellness status and academic performance of students. This study is deemed timely due to the relative importance of maintaining healthy life especially amongst undergraduate students who are often anticipated to excel in their studies irrespective of challenges posed by the school system as well as a learning process.

The findings from the present study revealed that HGS is characterized by avoiding chemical dependency, sustaining personal hygiene and health as well as a good habit of environmental health protection (Table 2). Wellness is a framework within which individuals form a productive and enjoyable life, much of which is developed within the scope of their education process. Avoiding chemical dependency is the well-being of individuals through self-care and personal security which is closely related to the students' understanding of their roles within the university, where they realize that they are responsible for taking care of themselves while dealing with stress in their work and other related claims<sup>(29)</sup>. Similarly, it has been documented that a good understanding of factors associated to the students' burnout and stress could assist the school administration and the other relevant stakeholders in mounting students-wellness initiatives that could aid in promoting empathy, self-care as well as overall health awareness of the students<sup>(30)</sup>.

Substance abuse and dependence may arise as a result of multiple factors faced by students, including genetic vulnerability, environmental stress, social pressures, individual personality characteristics, and psychiatric problems. Students provide myriad justifications for using drugs such as exploring the feelings, pressure from their peers as well as the need to develop a sense of belonging to the group<sup>(31)</sup>. Nonetheless, the students might likely use drugs to navigate obstacles emanating from schools, occupations, family as well as friends. The feelings of sadness, anxiety or depression could also motivate others to use drugs to overcome the situations. These issues can escalate rapidly particularly amongst college students. It is pertinent to highlight that preventing any element of addiction could be more advantageous as well as cost-effective as opposed to treatment measures. Consequently, regular dissemination of information to the students about the negative effects of any drug abuse could assist in guiding the students to make the necessary adjustment that could lead to a brighter future. It is obvious from the findings of the present study that the HGS are informed about the risk of chemical dependency which translated to their high score in the variable as compared to the LGS.

It was shown from the study findings that the HGS is attributed to better personal hygiene and health practices as well as a good habit of environmental health protection control. Personal hygiene is one of the pillars of healthy living of an individual. It constitutes all the personal elements that could affect the health as well as the well-being of an individual. Personal hygiene encompasses regular bathing, washing clothing and hands, taking care of nails, feet, teeth, personal appearance as well as inculcating the habits of cleanness<sup>(32,33)</sup>. A good practice of personal hygiene among students is non-trivial owing to the fact that a large number of students spend most of their time in public places, such as schools, colleges, or universities in direct contact with other people. The transmission of communicable disease to students may lead to the abscondment from school, which could consequently affect their academic performance<sup>(32)</sup>.

It was apparent from the finding of the present study that the HGS is attributed to a certain degree of environmental health protection habits. Environmental problems are a global phenomenon that threatens to destabilize that lives of living creatures. Problems such as global warming, ozone layer destruction, natural damage, loss of biodiversity as well as environmental pollution could seriously militate against the lives of both present and future generation<sup>(29,34)</sup>. It should be noted that environmental problems could strongly be influenced by environmental awareness, values as well as attitude of the people within the environment. This has prompted the effort of many countries to embark on assessing the environmental awareness of their populace with the aim of introducing intervention to mitigate the lingering issues of environmental hazards<sup>(35)</sup>. It is, therefore, vital for the student to be aware of environmental problems and actively involved in protecting the environment in order to safeguard the health of the school community and the population at large.

The LGS group were found to be better in controlling nutrition and diseases prevention. This is not surprising because high academic performance students are mostly focused on achieving higher academic grades even at the expense of keeping a healthy lifestyle. It was reported in the previous study that a good number of higher academic performance students often neglect taking a nutritious breakfast<sup>(36)</sup>. Furthermore, the preceding researchers stressed that among the factors which contributed to this problem are students' time constraints, no appetite, not liking to eat in the morning or sleeping too late to get up early<sup>(37)</sup>. On the other hand, the finding from the current study has indicated that both the HGS and LGS groups are similar in wellness attributes of health-related fitness, stress management, emotional well-being as well as personal safety. This finding has pointed out that the students are informed about the need to avoid stress, exercise regularly as well as inculcate the habit of personal safety regardless of their academic grades. This finding is also congruent with the previously reported data that demonstrated students who participate in physical activities are more incline to exhibit a favorable attitude towards both behavioral as well as cognitive abilities<sup>(38)</sup>.

## 5 Conclusion

To enhance students' academic performance, higher institutions of learning have consistently engaged the stakeholders in modifying programs, introducing new concept as well as initiating a myriad support system to improve academic performance. However, despite these efforts, students still have both satisfactory and unsatisfactory progress. This study examined the link between wellness status variables and academic performance of undergraduate students. The findings from the study demonstrated that students with high academic performance are attributed to some wellness qualities that consist of avoiding chemical dependency, sustaining personal hygiene and health as well as a good habit of environmental health protection whilst they fell short in nutritional health coupled with disease prevention. It is also apparent from the study that students are knowledgeable in the aspect of maintaining regular exercise, emotional well-being and personal safety. The overall findings of the present study accentuated the association between students' academic performance and wellness status. The findings of this study provide significant implications for the understanding of undergraduate students' wellness and its practices. The results from the study could serve as a foundation for the university to understand students' overall wellness and its connection with academic performance, in order to be able to design appropriate activities and programs to cater to their needs. Moreover, it is worth to mention that the nonconventional approach employed in the present study serves useful in systematically highlighting the relationship between the student's wellness and their academic performance. It is, therefore, recommended that the approach employed in the current study should be extended to some specific health care practices related to students' performances by considering a larger sample including both private and public institutions.

## Acknowledgement

The authors wish to thank the participants for their cooperation and keenness throughout the data collection process. The authors have no conflicting interests to declare.

## References

- 1) Olufemi AJ. Assessment of Wellness Status Among a Multi-Ethnic Based Adult Sample. *International Journal of Physical Education, Fitness and Sports*. 2016;5(2):05–08. Available from: <https://dx.doi.org/10.26524/1622>.
- 2) Santana CCA, Hill JO, Azevedo LB, Gunnarsdottir T, Prado WL. The association between obesity and academic performance in youth: a systematic review. *Obesity Reviews*. 2017;18(10):1191–1199. Available from: <https://dx.doi.org/10.1111/obr.12582>.
- 3) Musa RM, Abdullah MR, Maliki ABHM, Kosni NA, Haque M. The Application of Principal Components Analysis to Recognize Essential Physical Fitness Components among Youth Development Archers of Terengganu, Malaysia. *Indian Journal of Science and Technology*. 2016;9(44). Available from: <https://dx.doi.org/10.17485/ijst/2016/v9i44/97045>.
- 4) Musa RM, Majeed APPA, Abdullah MR, Nasir AFA, Hassan MHA, Razman MAM. Technical and tactical performance indicators discriminating winning and losing team in elite Asian beach soccer tournament. *PLOS ONE*. 2019;14(6):e0219138–e0219138. Available from: <https://dx.doi.org/10.1371/journal.pone.0219138>.
- 5) Anusha M, Karthik K, Rani P, Srikanth P, V. Prediction of student performance using machine learning. *International Journal of Engineering and Advanced Technology*. 2019;8(6):247–255. Available from: <https://doi.org/10.35940/ijeat.E7520.088619>.
- 6) Kardan AA, Sadeghi H, Ghidary SS, Sani MRF. Prediction of student course selection in online higher education institutes using neural network. *Computers & Education*. 2013;65:1–11. Available from: <https://dx.doi.org/10.1016/j.compedu.2013.01.015>.
- 7) Amrieh EA, Hamtini T, Aljarah I. Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*. 2016;9(8):119–136. Available from: <https://dx.doi.org/10.14257/ijdta.2016.9.8.13>.
- 8) Arsad PM, Buniyamin N, Manan J. Prediction of engineering students' academic performance using artificial neural network and linear regression: A comparison. *2013 IEEE 5th International Conference on Engineering Education: Aligning Engineering Education with Industrial Needs for Nation Development, ICEED 2013*;2014:43–48. Available from: <https://doi.org/10.1109/ICEED.2013.6908300>.
- 9) Ibrahim Z, Rusli D. Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision tree And Linear Regression. In: and others, editor. Proceedings of the 21st Annual SAS Malaysia Forum. 2007;p. 1–6. Available from: <https://www.researchgate.net/publication/228894873>.
- 10) Oladokun V, Adebajo A, Charles-Owaba O. Predicting students academic performance using artificial neural network: a case study of an engineering course. *The Pacific Journal of Science and Technology*. 2008;9(1):72–79.
- 11) Romero C, López MI, Luna JM, Ventura S. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*. 2013;68:458–472. Available from: <https://dx.doi.org/10.1016/j.compedu.2013.06.009>.
- 12) R RM, N A, M MABH, M MR, A KL, H J. Unsupervised Pattern Recognition of Physical Fitness Related Performance Parameters among Terengganu Youth Female Field Hockey Players. *International Journal on Advanced Science, Engineering and Information Technology*. 2017;7(1):100–100. Available from: <https://dx.doi.org/10.18517/ijaseit.7.1.1145>.
- 13) Fitness LP, Program P. Your Complete Solution for Health. 2014. Available from: <https://clubsolutionsmagazine.com/2020/01/fully-integrated.Accessed28>.
- 14) Hoeger W, Hoeger S. Lifetime Physical Fitness and Wellness: A Personalized Program: Cengage Learning. 2014.
- 15) Musa RM, Haque M. Academic performance of pre-clinical and clinical medical students' of East Coast Malaysian Peninsula: A cross-sectional and descriptive study that stimulates their life. *Journal of Applied Pharmaceutical Science*. 2017;7(6):169–175. Available from: <https://doi.org/10.7324/JAPS.2017.70624>.
- 16) Nunnally JC. Psychometric Theory 3E. and others, editor;Tata McGraw-Hill Education. 1994.
- 17) Cohen J, Primer P. A power primer. *Psychological bulletin*. 1992;112(1):155–159. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19565683>. Accessed September 29, 2018.
- 18) Taha Z, Musa RM, Majeed APPA, Alim MM, Abdullah MR. The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach. *Human Movement Science*. 2018;57(1):184–193. Available from: <https://dx.doi.org/10.1016/j.humov.2017.12.008>.
- 19) Taha Z, Musa RM, Majeed APPA, Alim MM, Abdullah MR. The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach. *Human Movement Science*. 2018;57:184–193. Available from: <https://dx.doi.org/10.1016/j.humov.2017.12.008>.
- 20) Rastrollo-Guerrero JL, Gómez-Pulido JA, Durán-Domínguez A. Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Applied Sciences*. 2020;10(3):1042–1042. Available from: <https://dx.doi.org/10.3390/app10031042>.
- 21) Fard MM, Thonet T, Gaussier E. Deep k-Means: Jointly clustering with k-Means and learning representations. *Pattern Recognition Letters*. 2020;138:185–192. Available from: <https://dx.doi.org/10.1016/j.patrec.2020.07.028>.
- 22) Zhu J, Jiang Z, Evangelidis GD, Zhang C, Pang S, Li Z. Efficient registration of multi-view point sets by K-means clustering. *Information Sciences*. 2019;488:205–218. Available from: <https://dx.doi.org/10.1016/j.ins.2019.03.024>.
- 23) Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*. 2002;35(5–6):352–359. Available from: [https://dx.doi.org/10.1016/s1532-0464\(03\)00034-0](https://dx.doi.org/10.1016/s1532-0464(03)00034-0).
- 24) Walker E. Regression Modeling Strategies. *Technometrics*. 2003;45(2):170–170. Available from: <https://doi.org/10.1198/tech.2003.s158>.
- 25) Musa RM, Majeed APPA, Taha Z, Abdullah MR, Maliki ABHM, Kosni NA. The application of Artificial Neural Network and k-Nearest Neighbour classification models in the scouting of high-performance archers from a selected fitness and motor skill performance parameters. *Science & Sports*. 2019;34(4):e241–e249. Available from: <https://dx.doi.org/10.1016/j.scispo.2019.02.006>.
- 26) Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression;vol. 398. and others, editor;John Wiley & Sons. 2013.



- 27) Musa RM, Majeed A, Kosni NA, Abdullah MR. Machine Learning in Team Sports: Performance Analysis and Talent Identification in Beach Soccer & Sepak-Takraw. and others, editor;Springer Nature. 2020.
- 28) Macarini LAB, Cechinel C, Machado MFB, Ramos VFC, Munoz R. Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems. *Applied Sciences*. 2019;9(24):5523–5523. Available from: <https://dx.doi.org/10.3390/app9245523>.
- 29) Moos RH. Evaluating Treatment Environments: The Quality of Psychiatric and Substance Abuse Programs. and others, editor;Taylor and Francis. 2018. Available from: <https://doi.org/10.4324/9781351291804>.
- 30) Confino JA. Where Are We on the Path to Law Student Well-Being?: Report on the ABA CoLAP Law Student Assistance Committee Law School Wellness Survey. *Journal of Legal Education*. 2019;68(3).
- 31) Pope KS, Vasquez MJT. Ethics in Psychotherapy and Counseling: A Practical Guide. and others, editor;John Wiley & Sons. 2016.
- 32) White C, Kolble R, Carlson R, Lipson N, Dolan M, Ali Y, et al. The effect of hand hygiene on illness rate among students in university residence halls. *American Journal of Infection Control*. 2003;31(6):364–370. Available from: [https://dx.doi.org/10.1016/s0196-6553\(03\)00041-5](https://dx.doi.org/10.1016/s0196-6553(03)00041-5).
- 33) Qasem J, Al-Rifaai J, Haddad AA. Personal hygiene among college students in Kuwait: A Health promotion perspective. *Journal of Education and Health Promotion*. 2018;7(1):92–92. Available from: [https://dx.doi.org/10.4103/jehp.jehp\\_158\\_17](https://dx.doi.org/10.4103/jehp.jehp_158_17).
- 34) Lai CS. A Study of Fifth Graders' Environmental Learning Outcomes in Taipei. *International Journal of Research in Education and Science*. 2018;4(1):252–262.
- 35) Ramadhan S, Sukma E, Indriyani V. Environmental education and disaster mitigation through language learning. *IOP Conference Series: Earth and Environmental Science*. 2019;314:012054–012054. Available from: <https://dx.doi.org/10.1088/1755-1315/314/1/012054>.
- 36) Gan WY, Nasir MTM, Zalilah MS, Hazizi AS. Disordered eating behaviors, depression, anxiety and stress among Malaysian university students. *College Student Journal*. 2011;45(2):296–310.
- 37) Damit N, Rahman HA, Ahmad SR. Factors associated with skipping breakfast among Inner Mongolia Medical students in China. *Pakistan Journal of Nutrition*. 2009;18(9):165–174.
- 38) Reina R, Hutzler Y, Iniguez-Santiago MC, Moreno-Murcia JA. Student Attitudes Toward Inclusion in Physical Education: The Impact of Ability Beliefs, Gender, and Previous Experiences. *Adapted Physical Activity Quarterly*. 2019;36(1):132–149. Available from: <https://dx.doi.org/10.1123/apaq.2017-0146>.