Prediction Modeling of Mental Well-Being Using Health Behavior Data of College Students

Hanif Abdul Rahman (✉ hanif.rahman@ubd.edu.bn)

Universiti Brunei Darussalam

Madeline Kwicklis

University of Michigan–Ann Arbor

Mohammad Ottom

University of Michigan–Ann Arbor

Areekul Amornsriwatanakul

Mahidol University

Khadizah H. Abdul-Mumin

Universiti Brunei Darussalam

Michael Rosenberg

University of Western Australia

Ivo D. Dinov

University of Michigan–Ann Arbor

Research Article

Keywords: Mental well-being, Machine learning, Algorithms, University students, Asian population, Health behaviors

Posted Date: February 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1281305/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Since the onset of the COVID-19 pandemic in early 2020, the importance of timely and effective assessment of mental well-being has increased dramatically. Due to heightened risks for developing mental illness, this trend is likely to continue during the post-pandemic period. Machine learning (ML) algorithms and artificial intelligence (AI) techniques can be harnessed for early detection, prognostication and prediction of negative psychological well-being states.

Objective: Studies using machine learning classification of mental well-being are scarce in Asian populations. This investigation aims to develop reliable machine learning classifiers based on health behavior indicators applicable to university students in South-East Asia.

Methods: Using data from a large, multi-site cross-sectional survey, this research work models mental well-being and reports on the performance of various machine learning algorithms, such as generalized linear models, k-nearest neighbor, naïve-Bayes, neural networks, random forest, recursive partitioning, bagging, and boosting. Prediction models were evaluated using various metrics such as accuracy, error rate, kappa, sensitivity, specificity, Area Under the recursive operating characteristic Curve (AUC), and Gini Index.

Results: Random forest and adaptive boosting algorithms achieved the highest accuracy of identifying negative mental well-being traits. The top five most salient features associated with predicting poor mental well-being include body mass index, number of sports activities per week, grade point average (GPA), sedentary hours, and age.

Conclusions: Based on the reported results, several specific recommendations and suggested future work are discussed. These findings may be useful to provide cost-effective support and modernize mental well-being assessment and monitoring at the individual and university level.

Introduction

Over four decades of research has linked positive mental well-being to improvements in health, development, and longevity (Agteren et al., 2021). Mental well-being can be seen as a separate, independent state from mental illness. A 10-year longitudinal study showed that improving mental well-being reduced the risk of developing mental illness by up to 8.2 times in people without mental health disorders (Keyes, Dhingra, and Simoes 2010). Long term poor psychological well-being is an important indicator in developing mental illness, viz., depression, anxiety disorders, eating disorders, and addictive behaviors (Srividya, Mohanavalli, and Bhalaji 2018). Thus, elevating mental well-being has become an essential therapeutic route to disease prevention.

Negative mental well-being typically manifests in young adults. However, some evidence showed that help-seeking behaviors start late when symptoms of mental illness have already appeared (Kessler and Bromet 2013). Help-seeking is further delayed with societal stigma, particularly in Asian cultures that

often results in under-reporting cases related to mental illness (Tan et al. 2020). Therefore, it is important to have a predictive mechanism to identify young people with negative mental well-being early to minimize the risk of developing mental health disorders (Srividya, Mohanavalli, and Bhalaji 2018; Tan et al. 2020).

The literature evidence promisingly that good physical health is a crucial factor influencing mental well-being (Kanekar and Sharma 2020; Nagy-Pénzes, Vincze, and Bíró 2020; Milne-Ives et al. 2020; Linden and Stuart 2020). Meta-analyses aggregating the results from numerous studies have revealed important links between mental disorders and physical inactivity and association with non-communicable diseases (NCDs) such as diabetes, heart disease, and multi-morbidity disorders (Stein et al. 2019). Although NCDs are usually asymptomatic in young adults, it is an added benefit to promote healthy and active behaviors early to prevent or delay the development of NCDs - as these health habits track into mid- and older- age [REF]. The health behaviors of young adults, particularly university students, provide important insight into NCD levels in the future (Secretariat 2017).

As data science techniques are no longer restricted to its predecessors of applied mathematics, statistics, and computer science, it is timely and practical for social and health sciences to utilize machine learning algorithms to address issues that have profound effects on human lives (Metcalf and Crawford 2016). In this regard, machine learning classifiers can be used to close this gap and provide more effective early detection and assessments of mental well-being in health prevention programs. Furthermore, mental well-being is directly related to the social and cultural aspects of the population in different regions (Srividya, Mohanavalli, and Bhalaji 2018). Hence, it is essential to use regional data for a prediction system that is customized for the target region. This system is particularly important for the Association of South-East Asian Nation (ASEAN) University Network-Health Promotion Network (AUN-HPN) as mental illness prevention is one of the key and immediate priority of the network. In addition, studies on mental well-being have become a priority in higher education institutions during the COVID-19 pandemic and will continue to be important in the post-pandemic era due to the heightened risk for developing serious mental health issues (Agterén et al. 2021; Liu et al. 2021). Furthermore, the Southeast and East Asia region, where ASEAN is inclusive, is the fastest growing digital market in the world with values exceeding US\$100 billion in 2019 and is expected to grow by four times that of the regional gross domestic product by 2023 (Chen & Ruddy, 2020). To incentivize the growing digital economy, higher education institutions should prioritize and ensure data infrastructure readiness and connectivity in the region for easement of research and development in the era of digital revolution. Therefore, this study aims to classify negative mental well-being based on indicators of healthy behaviors among university students in ASEAN using machine learning prediction models.

Methods

Data

The data used in this study was extracted from an online cross-sectional survey of 15,366 university students from the ASEAN countries. The target universities consisted of 17 ASEAN University Network (AUN) member universities across seven ASEAN countries, namely, Brunei Darussalam, Indonesia, Malaysia, Philippines, Singapore, Thailand, and Vietnam.

The questionnaire was developed in several rounds of consultation meetings with experts from the AUN Health Promotion Network committee and member universities. The measurement tools used selected were widely used and validated in multiple countries (**Appendix A**). The features are extracted based on the focus of this study. Mental well-being was measured using the shortened Warwick-Edinburgh Mental Well-being Scale (WEMWBS), a reliable and valid tool for university student. WEMWBS score was dichotomized into poor well-being" (7.0-17.99) and good well-being" (≥ 18.00).

Physical activity (PA) was measured using the Global Physical Activity Questionnaire (GPAQ) version 2.0. *Low PA* is classified as those who had less than 600 Metabolic equivalents (MET)-minutes/week that resulted in a failed to comply with the conditions of minimum energy expenditure for physical activity. Number of sport activities were also collected and categorized into none, one to three, four to six, and more than six activities per week.

Health-risk behaviors were also collected including consumption of alcohol, smoking, fruits and vegetables, salts, and sugar-sweetened beverages were measured using items from existing instruments. For tobacco consumption, students who smoked daily were dichotomized into "Yes" (current smokers) and "No" (not current smokers). For alcohol consumption, students were asked if they do or do not drink alcohol. For fruit/vegetable consumption, students were asked how many servings of fruits/vegetable they usually eat each day, and consumption of ≥ 5 servings/day was considered healthy. Consumption of snacks/fast food was assessed by asking how many days per week students eat fast food. Students who consumed fast food every day were categorized into "Yes" and the remaining responses were collapsed into "No." Salt intake was assessed by asking if they added salt in their food before eating (< 1 tea spoon to ≥ 3 tea spoons). Adding ≥ 1 tea spoon or 6 gm/per day was considered excessive sodium intake. Students were also asked how many days they drank sugar-sweetened beverages. Response were handled similarly to the consumption of fast food. Participants provided demographic information including age, gender, GPA (grading system for students' academic performance), and Body Mass Index (BMI). An open-ended question regarding opinion on physical activity was asked to obtain textual data.

Ethical approval was obtained from the institutional review board of each university prior to conducting the study (See Declarations).

Data preprocessing

Data cleaning procedures were employed including removal of ineligible cases, duplicate responses, responses with more than 50% missing values (listwise deletion), and invalid questionnaire responses. A total of 15,366 remaining cases were used in the subsequent analysis. Missing data in these valid cases were handled using multiple imputation techniques - MICE (Multivariate Imputation via Chained

Equations) using 10 multiple imputations to replace missing with predicted values, using R package *mice* (Zhang 2016). The dataset with unbalanced with respect to the binary outcome of negative or poor mental well-being. To avoid potential bias in the AI/ML modeling, the dataset was re-balanced using the Synthetic Minority Oversampling TEchnique (SMOTE) (Chawla et al. 2002).

Feature selection

According to the principle of parsimony, simplicity or a simple *apriori* model often provides the best explanation of a problem, relative to more complex models because inclusion of unnecessary features creates intrinsic and extrinsic noise (Naser 2021). Accounting only for key data elements avoids model overfitting, provides better predictive accuracy and generalization, and facilitate practical application (Guan and Loew 2020). Due to limitations of different types of feature selection method, three strategies were used to validate selection of salient variables or features that will be used in the training models in this study. The first strategy was based on the Benjamini-Hochberg False Discovery Rate method that controls for expected proportion of false rejection of features in multiple significance testing (Benjamini and Hochberg 1995), which could be expressed as follows:

$$\underbrace{FDR}_{\text{False Discovery Rate}} = \underbrace{E}_{\text{expectation}} \left(\frac{\#FalsePositives}{\text{total number of selected features}} \right) \cdot \underbrace{False\ Discovery\ Proportion}$$

Second, a deterministic wrapper method based on stepwise selection, an iterative process of adding important features to a null set of features and removing worst-performing features from the list of complete features, was computed (Naser 2021). The final strategy utilized a randomized wrapper method, Boruta, which iteratively removes features that are relatively less statistically significant compared to random probes, was employed (Kursa, Rudnicki, and others 2010). Our aggregate feature-selection technique utilized the intersection of these three variable elimination strategies and generated a smaller collection of variables used in the subsequent AI modeling.

Training Machine Learning Classifiers

Classification is a supervised machine learning technique that group records into sets of homologous observations associated with particular classes. Different classifiers or classification algorithms are available. In this study, six different classifiers were trained including generalized linear model (glm), k-nearest neighbor (knn), naïve-Bayes (nb), neural network (nnet), random forest (rf), and Recursive partitioning (RPART).

The generalized linear model, specifically, logistic regression, is a linear probabilistic classifier. It takes in the probability values for binary classification, in this case, positive (0) and negative (0) mental well-being, and estimate class probabilities directly using the logit transform function (Myers and Montgomery 1997).

Naïve-Bayes predicts class membership probabilities based on the Bayes theorem and naive assumption that all features are equally important and independent (Dinov 2018). Bayes conditional probability could be expressed as:

$$\text{Posterior Probability} = \frac{\text{likelihood} \times \text{Prior Probability}}{\text{Marginal Likelihood}}.$$

Essentially, the probability of class level L given an observation, represented as a set of *independent features* $F_1, F_2, \dots, F_n$. Then the posterior probability that the observation is in class L is equal to:

$$P(C_L | F_1, \dots, F_n) = \frac{P(C_L) \prod_{i=1}^n P(F_i | C_L)}{\prod_{i=1}^n P(F_i)},$$

where the denominator, $\prod_{i=1}^n P(F_i)$, is a scaling factor that represents the marginal probability of observing all features jointly.

For a given case $X = (F_1, F_2, \dots, F_n)$, i.e., given vector of *features*, the naive Bayes classifier assigns

the **most likely class** $\hat{C}$ by calculating $\frac{P(C_L) \prod_{i=1}^n P(F_i | C_L)}{\prod_{i=1}^n P(F_i)}$ for all class labels L , and then assigning

the class $\hat{C}$ corresponding to the maximum posterior probability. Analytically, $\hat{C}$ is defined by:

$$\hat{C} = \operatorname{argmax}_L \frac{P(C_L) \prod_{i=1}^n P(F_i | C_L)}{\prod_{i=1}^n P(F_i)}.$$

As the denominator is static for L , the posterior probability above is maximized when the numerator is maximized, i.e., $\hat{C} = \operatorname{argmax}_L P(C_L) \prod_{i=1}^n P(F_i | C_L)$.

Artificial neural networks, or simply neural nets, simulate the underlying intelligence of the human brain by using a synthetic network of interconnected neurons (nodes) to train the model. The features are weighted by importance and the sum is passed according to an activation function, and generate an output (y) at the end of the process (Dinov 2018). A typical output could be expressed as:

$$y(x) = f \left(\sum_{i=1}^n w_i x_i + w_o b \right).$$

Random forest classifier is a randomized ensemble of decision trees that recursively partition the dataset into roughly homogeneous or close to homogeneous terminal nodes. It may contain hundreds to thousands of trees that are grown by bootstrapping samples of the original data. The final decision is

obtained when the tree branching process terminates and provides the expected forecasting results given the series of events in the tree (Dinov 2018; Nguyen, Wang, and Nguyen 2013).

Recursive partitioning (RPART) is another decision tree classification technique that works well with variables with definite ordering and unequal distances. The tree is built similarly as random forest with a resultant complex model. However, RPART procedure also trims back the full tree into nested terminals based on cross-validation. The final model of the sub-tree provides the decision with the 'best' or lowest estimated cross-validation error (Therneau, Atkinson, and others 1997).

The caret package was used for automated parameter tuning with *repeatedcv* method set at 15-fold cross-validation re-sampling that was repeated with 10 iterations (Kuhn 2009).

In this study, random forest outperformed other machine learners. However, general decision trees might overfit model to noise in the training dataset. To overcome this, we implemented bootstrap aggregation (bagging) and boosting to reduce variance and bias, respectively.

Bagging decreases the variance in the prediction model by essentially generating additional data for training original dataset using bootstrapping methods. Boosting reduces bias in parameter estimation by sub-setting the original data to produce a series of models and boost their performance (in this case, measured by accuracy) by combining them together (Dinov 2018).

Model performance metrics

Classification model performance could not be evaluated with a single metric, therefore, a number of metrics were used to assess model performance including Accuracy, Error rate, Kappa, Sensitivity, Specificity, Area Under the Receiver Operating Characteristics Curve (AUC), and Gini Index.

In binary classification, accuracy is calculated using the 2×2 confusion matrix, which can be expressed as:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{\text{Total number of observations}}.$$

Where, **True Positive**(TP) is the number of observations that correctly classified as "yes" or "success." **True Negative**(TN) is the number of observations that correctly classified as "no" or "failure." **False Positive**(FP) is the number of observations that incorrectly classified as "yes" or "success." **False Negative**(FN) is the number of observations that incorrectly classified as "no" or "failure" (Dinov 2018).

Whereas, error rate is the proportion of mis-classified observations calculated using:

$$errorrate = \frac{FP + FN}{TP + TN + FP + FN} = \frac{FP + FN}{\text{Total number of observations}} = 1 - accuracy.$$

The accuracy and error rate and accuracy add up to 1. Therefore, a 95% accuracy means 5% error rate (Dinov 2018).

Kappa statistic measures the possibility of a correct prediction by chance alone and evaluate *the agreement* between the expected truth and the machine learning prediction. When kappa = 1, there is a perfect agreement between a computed prediction and an expected prediction (typically random, by-chance, prediction). Kappa statistics can be expressed as (Dinov 2018):

$$kappa = \frac{P(a) - P(e)}{1 - P(e)}.$$

where $P(a)$ and $P(e)$ simply denote the probability of actual and expected agreement between the classifier and the true values.

A common interpretation of the Kappa statistics includes (Dinov 2018):

- *Poor* agreement: less than 0.20
- *Fair* agreement: 0.20-0.40
- *Moderate* agreement: 0.40-0.60
- *Good* agreement: 0.60-0.80
- *Very good* agreement: 0.80-1

Sensitivity is a statistic that indicates the true positive rate measures the proportion of “success” observations that are correctly classified (Dinov 2018). This can be expressed as:

$$sensitivity = \frac{TP}{TP + FN}.$$

On the other hand, specificity is a statistic that indicates the true negative rate measures the proportion of “failure” observations that are correctly classified (Dinov 2018). This can be expressed as:

$$specificity = \frac{TN}{TN + FP}.$$

The Receiver Operating Characteristic (ROC) curve plots the trade-off between classification of true positive (sensitivity) and avoiding false positives (specificity). The area under this curve serves as a proxy of classifier performance and is normally interpreted as (Dinov 2018):

- Outstanding: 0.9-1.0
- Excellent/good: 0.8-0.9
- Acceptable/fair: 0.7-0.8
- Poor: 0.6-0.7
- No discrimination: 0.5-0.6

The Gini index is based on variable importance measure and evaluate information gain by calculating the estimated class probabilities (Dinov 2018). This can be expressed as:

$$GI = \sum_k p_k(1 - p_k) = 1 - \sum_k p_k^2$$

where k is the number of classes.

Results

The cleaned and preprocessed dataset comprises n=15,366 cases with k=20 features. The majority of respondents were from Vietnam (33.3%), followed by Indonesia (28.8%) and Thailand (25.6%). Approximately half of the respondents were female (52.6%), were 19-21 years old (66.3%), and had normal BMI (61.5%). Over half of the respondents achieved a moderate GPA of 3.3-3.9 out of 5 (69.2%) and lived off-campus (65.2%). The highest prevalence of health-risk behaviors was consumption of sugar-sweetened beverages (82.0%), followed by snacks/fast food (65.2%), low consumption of fruits and vegetables (47.8%) and high salt intake (46.0%). Insufficient physical activity levels (less than 600 MET-min/week) were observed among 39.7% of respondents. A negative or poor mental well-being level was observed among 16.7% of respondents, whereas 13.4% drank alcohol, and 8.9% smoked. See **Appendix B** for basic descriptive statistics of the survey.

Feature importance

Figure 1 illustrated ten features that are salient to the prediction model of mental well-being. This is corroborated by the error plot, variable importance plot (accuracy) and Gini index (Figure 2). The ten salient indicators for mental well-being rank order by importance based on health behaviors comprised of body mass index, number of sports activity per week, grade point average (GPA), sedentary hours, age, gender, salt intake, fruits and vegetables consumption, hours of sleep, and achieved recommended physical activity levels.

Model evaluation

The dataset was randomly partitioned into the training set (80%) and the testing set (20%). The training dataset was used to build the classifier models using different classification algorithms including generalized linear model, k-nearest neighbors, naïve-Bayes, neural net, random forest, and recursive partitioning. The performance of the trained classifiers were then evaluated using accuracy and kappa statistics. Figure 3 illustrates the results of the trained model performance. The overall performance effectiveness of a classifier indicated using accuracy and kappa statistics showed that random forest (accuracy = 0.921, kappa = 0.788) was the best classifier, followed by k-nearest neighbor (accuracy = 0.775, kappa = 0.554) and naïve-Bayes (accuracy = 0.723, kappa = 0.433).

The trained model classifiers were then applied to the testing data set to evaluate how well they predict poor mental well-being. Table 1 shows the model evaluation on testing data. With tuning parameters using *repeatedcv* method set at 10-folded cross-validation re-sampling repeated with 5 iterations showed that random forest clearly outperforms other classifiers (AUC=0.966). Model optimization using bagging

(AUC=0.677) did not improve the selected random forest classifier. However, boosting (AUC=0.959) also performed similarly. Adding complementary unstructured-text information to the structured data elements did not significantly improve the performance of the random forest classifier (AUC=0.951). Such data augmentation adds more than 20 text-derived structured data elements to the standard survey features, which runs counter to the principle of parsimony.

Table 1
Model evaluation metrics of machine learning classifiers on testing data set.

Classifier	Accuracy	Kappa	Sensitivity	Specificity	AUC
Random Forest	0.901	0.801	0.980	0.815	0.966
Random forest + text	0.881	0.759	0.965	0.782	0.951
Adaptive boosting	0.893	0.785	0.951	0.828	0.959
k-nearest neighbor	0.795	0.593	0.711	0.887	0.886
Naïve-Bayes	0.702	0.399	0.775	0.621	0.678
Bagging (Bootstrap Aggregation)	0.672	0.349	0.617	0.735	0.677
Recursive partitioning	0.633	0.268	0.607	0.662	0.665
Neural Network	0.615	0.231	0.597	0.633	0.674
Generalized linear model	0.603	0.201	0.673	0.527	0.653

As the top performing classifier (random forest) represents an implicit (black box) model, Figure 4 illustrates an example of a single decision tree from the aggregate forest model that illustrates one explicit classification strategy for predicting poor mental well-being.

Discussion

Main findings

Mental well-being is an important indicator for mental health and this study developed prediction models using machine learning classifiers to predict mental well-being status among university students in ASEAN. This is particularly important, with no end in sight of the pandemic due to the emergent of different COVID-19 variants, social and physical distancing and isolation will be the way of life in the new normal that indubitably increase risk of developing serious mental illness in the future.

In the present study, the prediction models that produced high accuracy were achieved by random forest, random forest with text predictors, and adaptive boosting. Models using the additional text-derived features did not improve the model performance. This could be explained by the non-specific nature of the open-ended survey question regarding physical health. Future studies could examine this aspect

closely using more sophisticated methods of natural language processing (NLP), deep learning, and language syntax techniques to transform the unstructured text into quantitative data elements (Dinov, 2018). Such advanced machine learning strategies could enhance the contribution of the textual content in the forecasting of mental well-being. Nevertheless, studies on mental well-being using various classification techniques among samples in the Asian population are scarce. A few studies also reported that random forest or decision tree-based algorithms were some of the best techniques for forecasting mental health (Srividya, Mohanavalli, and Bhalaji 2018; Edla et al. 2018; Tao et al. 2021). A recent systematic review and meta-analysis has revealed that psychological interventions' efficacy was generally low to moderate and largely concentrated in clinical mental health settings (Agteren et al. 2021). The disparity of evidence for general population can be reduced using better prediction models. The prediction models used in this study may aid in health promotion strategies and research designed to improve mental well-being.

In addition, this project also identified salient features for assessing mental well-being, namely, body mass index, number of sports activities per week, grade point average (GPA), sedentary hours, age, gender, salt intake, fruits and vegetables consumption, hours of sleep, and achieved recommended physical activity levels. Physical health and dietary status have been well-documented as strong predictors of psychological well-being (Liu et al. 2021; Terebessy et al. 2016). Furthermore, interventions that targets physical health outcomes have also shown to benefit mental well-being (Dale, Brassington, and King 2014).

Limitations

The key strength of the present study is the use of large data set from multiple sites in South-East Asia to objectively and rank order significant variables of predicting mental well-being. The utilization of more than one feature selection methods, machine learning classifiers, and model evaluation metrics reduces the errors and biases of the results. Despite the strengths, several limitations should be noted for the current study. The data was collected using a cross-sectional survey design and is not able to draw causal inferences. Even though the survey consisted of items from widely used, validated questionnaires, self-reporting bias and likelihood of under-reporting are still present. However, this survey took place during the COVID-19 pandemic and would continue to complement and benefit future studies post-pandemic.

Recommendations and future work

The ASEAN university network (AUN) has a central role in coordinating effective use of available digital infrastructure and research activities, which currently focused on particular individual universities. Here are three-point recommendations, from general to specific, that provides feasible and practical solutions particularly for its health promotion network. *Firstly*, the continuity of data collection to train the machine learning algorithms are vital. An appropriate central data collection, processing, and analytical centre using existing infrastructure, particularly from resource-rich AUN member universities, need to be identified and established. With agreed upon data collection, storage, and sharing policies and

regulations, it encourages active participation and contribution of university students' health data, even from resource deprived institutions. The systematic and long-term collection of health data is crucial for answering critical research inquiry, encourage innovative interventions, and experimentation of data created, used, and shared among higher education institutions.

Secondly, the selected data centre that has been setup could be populated with data from affordable technological survey tools, which provide practical, cost-effective, and long-term avenue for monitoring overall health of university students including mental well-being (Woodward et al. 2020). Traditional assessment tools took considerable amount of time to complete due to collection of too many features and collection period over long period before they provide useful insights (Areán, Ly, and Andersson 2016; Woodward et al. 2020). The use of technological survey such as mHealth mobile apps or integration into student evaluation tools that collects only few salient features over short period of time, could have potential for more effective assessment and monitoring. Working in conjunction with the diagnostic assessment by university counselors/ psychiatrists/ care professionals, the collected data could provide precise early detection and early intervention (Lederman et al. 2014; Woodward et al. 2020). However, ethical collection and storage of student information will necessarily require additional steps to ensure privacy and security as well as aligning with code of professional practice such as maintaining patient-physician confidentiality (Aguilera 2015).

Finally, an agreed upon schema of data that will be collected and displayed as active indicators for monitoring and assessment of the current state of university students' health and well-being status on a digital dashboard that is accessible by students and stakeholders alike. The dashboard will provide a sense of digital connectiveness among students in the ASEAN university network – the fundamental goal of ASEAN. The data schema could be updated from time to time. At the moment, this study has provided a rank ordered importance features of predicting mental well-being, which is essential particularly in resource deprived institutions to allocate resources appropriately and as necessary. Among the significant features, the top five variables viz., body mass index, number of sports activity per week, grade point average (GPA), sedentary hours, and age. They appeared to have large effects on the accuracy of the classifiers and therefore should be prioritized when developing and implementing psychological well-being monitoring and promotion interventions. Future research is needed to continue improving precision of the prediction models. Given that level mental well-being varies widely by social and cultural context, country or culture-specific models should be developed. In addition, resource rich institutions could consider reducing subjectivity of the data by incorporating salient objective measures such as real-time biofeedback information from physiological sensing technology that collects electroencephalogram (EEG) activity, electrocardiogram (ECG) fluctuations, heart rate, breathing rate, temperature, speech intonation and so on (Woodward et al. 2020). These variables are very useful in the next step of prediction model development using deep learning algorithms. Although objective measures typically will increase the cost of assessment and monitoring, university and related policymakers must decide on the trade-off between comprehensiveness and conciseness. More importantly, to make this a reality, the AUN-Health Promotion Network is in a central position to coordinate and collect long term monitoring data for all member institutions in the ASEAN region (Sivarajah et al. 2017). The continuous collection of data is

critical for the organization to modernize and adopt data analytics to improve operational efficiency and achieve strategic goals.

Declarations

Ethical approval and consent to participate

The study protocols have been performed in accordance with the Declaration of Helsinki and received ethical approval from the institutional review board of each participating universities: Mahidol University Central Institutional Review Board (MU-CIRB 2020/089.0704), University Research and Ethics Committee of Universiti Brunei Darussalam (UBD/OAVCR/UREC/OCTOBER2020-05), Universitas Gadjah Mada (KE/FK/1066/EC/2020), Universitas Indonesia (KET126/UN2.F1/ETIK/PPM.00.02/2021), Universiti Putra Malaysia (JKEUPM-2020-156), University of Malaya (UM.TNC2/UMREC_1326), Ateneo de Manila University (AdMUREC-20-010), Nanyang Technological University (IRB-2021-03-027), Naresuan University (P29936/63), Thammasat University (075/2563), Burapha University (HS035/2563), Chiang Mai University (AMSEC-63EX-019), Walailak University (WUEC-20-122-01), Mahasarakham University (266/2020), Vietnam National University, Ho Chi Minh City (01/2020/IRB-VN01.017-MEDVNU). Universitas Airlangga and King Mongkut's University of Technology North Bangkok used MU-CIRB. Participants in this study gave their informed online consent by clicking "I agree to participate" before completing the survey.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to restrictions on intellectual property regulations of the funding organization.

Competing interest

All authors do not have conflict of interest to declare.

Funding

This study was financially supported by Thai Health Promotion Foundation through Children and Youth Physical Activity Studies (Ref: 61-00-1814), and Centre of Advanced Research, Universiti Brunei Darussalam (UBD/RSCH/1.10/FICBF(b)/2019/005). Partial funding was also provided by NSF grants (1916425, 1734853, 1636840) and NIH grants (UL1 TR002240, R01 CA233487, R01 MH121079, R01 MH126137, T32 GM141746).

The sponsors were not involved with or had any roles regarding the conduct of this study and publication.

Author contributions

HAR, MO, IDD contributed to the conception or design of the paper. HAR conducted the data analysis. All authors contributed to data interpretation, and drafting/editing the manuscript. All authors were involved in revising the manuscript, providing critical comments, and agreed to be accountable for all aspects of the work and any issues related to the accuracy or integrity of any part of the work.

Acknowledgement

The authors would like to thank all the students who participated in the survey. We also would like to express our gratitude to the AUN-HPN research team for collecting data in each participating university. They include:

1. Denny Agustiningsih, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Indonesia
2. Surasak Chaiyasong, Faculty of Pharmacy, Mahasarakham University, Thailand
3. Michael Chia, Physical Education & Sports Science, National Institute of Education
4. Nanyang Technological University, Singapore.
5. Supat Chupradit, Department of Occupational Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Thailand.
6. Le Quang Huy, School of Medicine, Vietnam National University, Ho Chi Minh City, Vietnam
7. Katiya Ivanovitch, Faculty of Public Health, Thammasat University (Rangsit Campus), Thailand
8. Ira Nurmala, Health Promotion and Behavior, Faculty of Public Health, Universitas Airlangga, Indonesia
9. Hazreen B Abdul Majid, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya 50603 Kuala Lumpur, Malaysia
10. Ahmad Iqmer Nashriq Mohd Nazan, Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 434000 UPM Serdang, Malaysia
11. Yuvadee Rodjarkpai, Faculty of Public Health, Burapha University, Thailand.
12. Henrietta Teresa O. de la Cruz, Office of Health Services, Ateneo de Manila University and Faculty, Ateneo School of Medicine and Public Health, Philippines
13. Trias Mahmudiono, Health Promotion and Behavior, Faculty of Public Health, Universitas Airlangga, Indonesia
14. Krissachai Sriboonma, King Mongkut's University of Technology North Bangkok, Thailand.
15. Supaporn Sudnongbua, Faculty of Public Health, Naresuan University, Thailand
16. Dhanasari Vidiawati, Head of Makara Satellite Clinic of Universitas Indonesia, Academic staff of Faculty of Medicine, Universitas Indonesia
17. Nani Cahyani, Academic staff of Faculty of Medicine, Universitas Indonesia

18. Apichai Wattanapisit, School of Medicine, Walailak University, Nakhon Si Thammarat, 80160, Thailand.
19. Assistant Professor Dr Sukanya Charoenwattana, Burapha University, Chonburi, Thailand.
20. Professor Josip Car, Professor of Digital Health Sciences, Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.
21. Associate Professor Moon-Ho Ringo Ho, School of Social Sciences, Nanyang Technological University, Singapore.

References

1. Agteren, Joep van, Matthew Iasiello, Laura Lo, Jonathan Bartholomaeus, Zoe Kopsaftis, Marissa Carey, and Michael Kyrios. 2021. "A Systematic Review and Meta-Analysis of Psychological Interventions to Improve Mental Wellbeing." *Nature Human Behaviour*, 1–22.
2. Aguilera, Adrian. 2015. "Digital Technology and Mental Health Interventions: Opportunities and Challenges." *Arbor* 191 (771): a210–10.
3. Areán, Patricia A, Kien Hoa Ly, and Gerhard Andersson. 2016. "Mobile Technology for Mental Health Assessment." *Dialogues in Clinical Neuroscience* 18 (2): 163.
4. Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300.
5. Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. "SMOTE: Synthetic Minority over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57.
6. Chen, L., & Ruddy, L. (2020). Improving Digital Connectivity: Policy Priority for ASEAN Digital Transformation. *ERIA Policy Brief*, (7), 1–4.
7. Dale, Hannah, Lindsay Brassington, and Kristel King. 2014. "The Impact of Healthy Lifestyle Interventions on Mental Health and Wellbeing: A Systematic Review." *Mental Health Review Journal*.
8. Dinov, Ivo D. 2018. *Data Science and Predictive Analytics*. Springer.
9. Edla, Damodar Reddy, Kunal Mangalorekar, Gauri Dhavalikar, and Shubham Dodia. 2018. "Classification of EEG Data for Human Mental State Analysis Using Random Forest Classifier." *Procedia Computer Science* 132: 1523–32.
10. Guan, Shuyue, and Murray Loew. 2020. "Analysis of Generalizability of Deep Neural Networks Based on the Complexity of Decision Boundary." In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 101–6. IEEE.
11. Kanekar, Amar, and Manoj Sharma. 2020. "COVID-19 and Mental Well-Being: Guidance on the Application of Behavioral and Positive Well-Being Strategies." In *Healthcare*, 8:336. 3. Multidisciplinary Digital Publishing Institute.
12. Kessler, Ronald C, and Evelyn J Bromet. 2013. "The Epidemiology of Depression Across Cultures." *Annual Review of Public Health* 34: 119–38.

13. Keyes, Corey LM, Satvinder S Dhir, and Eduardo J Simoes. 2010. "Change in Level of Positive Mental Health as a Predictor of Future Risk of Mental Illness." *American Journal of Public Health* 100 (12): 2366–71.
14. Kuhn, Max. 2009. "The Caret Package." *Journal of Statistical Software* 28 (5).
15. Kursa, Miron B, Witold R Rudnicki, and others. 2010. "Feature Selection with the Boruta Package." *J Stat Softw* 36 (11): 1–13.
16. Lederman, Reeva, Greg Wadley, John Gleeson, Sarah Bendall, and Mario Álvarez-Jiménez. 2014. "Moderated Online Social Therapy: Designing and Evaluating Technology for Mental Health." *ACM Transactions on Computer-Human Interaction (TOCHI)* 21 (1): 1–26.
17. Linden, Brooke, and Heather Stuart. 2020. "Post-Secondary Stress and Mental Well-Being: A Scoping Review of the Academic Literature." *Canadian Journal of Community Mental Health* 39 (1): 1–32.
18. Liu, Chang, Melinda McCabe, Andrew Dawson, Chad Cyrzon, Shruthi Shankar, Nardin Gerges, Sebastian Kellett-Renzella, Yann Chye, and Kim Cornish. 2021. "Identifying Predictors of University Students' Wellbeing During the COVID-19 Pandemic—a Data-Driven Approach." *International Journal of Environmental Research and Public Health* 18 (13): 6730.
19. Metcalf, Jacob, and Kate Crawford. 2016. "Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide." *Big Data & Society* 3 (1): 2053951716650211.
20. Milne-Ives, Madison, Ching Lam, Caroline De Cock, Michelle Helena Van Velthoven, and Edward Meinert. 2020. "Mobile Apps for Health Behavior Change in Physical Activity, Diet, Drug and Alcohol Use, and Mental Health: Systematic Review." *JMIR mHealth and uHealth* 8 (3): e17046.
21. Myers, Raymond H, and Douglas C Montgomery. 1997. "A Tutorial on Generalized Linear Models." *Journal of Quality Technology* 29 (3): 274–91.
22. Nagy-Pénzes, Gabriella, Ferenc Vincze, and Éva Bíró. 2020. "Contributing Factors in Adolescents' Mental Well-Being—the Role of Socioeconomic Status, Social Support, and Health Behavior." *Sustainability* 12 (22): 9597.
23. Naser, MZ. 2021. "Mapping Functions: A Physics-Guided, Data-Driven and Algorithm-Agnostic Machine Learning Approach to Discover Causal and Descriptive Expressions of Engineering Phenomena." *Measurement* 185: 110098.
24. Nguyen, Cuong, Yong Wang, and Ha Nam Nguyen. 2013. "Random Forest Classifier Combined with Feature Selection for Breast Cancer Diagnosis and Prognostic."
25. Secretariat, ASEAN. 2017. "First ASEAN Youth Development Index." Jakarta: ASEAN Secretariat.
26. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
27. Srividya, M, S Mohanavalli, and N Bhalaji. 2018. "Behavioral Modeling for Mental Health Using Machine Learning Algorithms." *Journal of Medical Systems* 42 (5): 1–12.
28. Stein, Dan J, Corina Benjet, Oye Gureje, Crick Lund, Kate M Scott, Vladimir Poznyak, and Mark Van Ommeren. 2019. "Integrating Mental Health with Other Non-Communicable Diseases." *Bmj* 364.

29. Tan, Gregory Tee Hng, Shazana Shahwan, Chong Min Janrius Goh, Wei Jie Ong, Ker-Chiah Wei, Swapna Kamal Verma, Siow Ann Chong, and Mythily Subramaniam. 2020. "Mental Illness Stigma's Reasons and Determinants (MISReaD) Among Singapore's Lay Public—a Qualitative Inquiry." *BMC Psychiatry* 20 (1): 1–13.
30. Tao, Xiaohui, Oliver Chi, Patrick J Delaney, Lin Li, and Jiajin Huang. 2021. "Detecting Depression Using an Ensemble Classifier Based on Quality of Life Scales." *Brain Informatics* 8 (1): 1–15.
31. Terebessy, András, Edit Czeglédi, Bettina Claudia Balla, Ferenc Horváth, and Péter Balázs. 2016. "Medical Students' Health Behaviour and Self-Reported Mental Health Status by Their Country of Origin: A Cross-Sectional Study." *BMC Psychiatry* 16 (1): 1–9.
32. Therneau, Terry M, Elizabeth J Atkinson, and others. 1997. "An Introduction to Recursive Partitioning Using the RPART Routines." Technical report Mayo Foundation.
33. Woodward, Kieran, Eiman Kanjo, David Brown, Thomas M McGinnity, Becky Inkster, Donald MacIntyre, and Thanasis Tsanas. 2020. "Beyond Mobile Apps: A Survey of Technologies for Mental Well-Being." *IEEE Transactions on Affective Computing*.
34. Zhang, Zhongheng. 2016. "Multiple Imputation with Multivariate Imputation by Chained Equation (MICE) Package." *Annals of Translational Medicine* 4 (2).

Appendix

Appendices A & B are not available with this version.

Figures

Figure 1

Accuracy (left) and Error (right) plots of random forest classification of mental well-being.

Figure 2

Variable importance plots of random forest classification of salient features (top) of mental well-being using accuracy (left) and Gini index (right) as evaluation metrics.

Figure 3

Evaluation of performance of trained machine learning classifiers.

Figure 4

An example of a decision tree with a recursive partitioning algorithm for predicting negative mental well-being. The blue colour nodes represent higher correlation or chance to be classified as poor mental well-being. In contrast, the green nodes represent lower probability of being classified as poor mental well-being. The percentage in each refers to the proportion of cases remaining after each split – starting with 100% from the top of the hierarchy.