



Prospective predictors of electronic nicotine delivery system initiation in tobacco naive young adults: A machine learning approach

Nkiruka C. Atuegwu^{a,*}, Eric M. Mortensen^a, Suchitra Krishnan-Sarin^b, Reinhard C. Laubenbacher^c, Mark D. Litt^d

^a Department of Medicine, University of Connecticut School of Medicine, Farmington, CT 06030, USA

^b Department of Psychiatry, Yale University School of Medicine, Connecticut Mental Health Center, 34 Park Street, New Haven, CT 06519, USA

^c Laboratory for Systems Medicine, Department of Medicine, University of Florida, Gainesville, FL 32610, USA

^d Division of Behavioral Sciences and Community Health, University of Connecticut Health Center, Farmington, CT 06030, USA

ARTICLE INFO

Keywords:

E-cigarette
Electronic nicotine delivery systems
ENDS
Machine learning
Tobacco naïve
Never tobacco users
Vaping
Young adults
Prospective predictors
PATH
Population Assessment of Tobacco and Health survey

ABSTRACT

The use of electronic nicotine delivery systems (ENDS) is increasing among young adults. However, there are few studies regarding predictors of ENDS initiation in tobacco-naïve young adults. Identifying the risk and protective factors of ENDS initiation that are specific to tobacco-naïve young adults will enable the creation of targeted policies and prevention programs. This study used machine learning (ML) to create predictive models, identify risk and protective factors for ENDS initiation for tobacco-naïve young adults, and the relationship between these predictors and the prediction of ENDS initiation. We used nationally representative data of tobacco-naïve young adults in the U.S drawn from the Population Assessment of Tobacco and Health (PATH) longitudinal cohort survey. Respondents were young adults (18–24 years) who had never used any tobacco products in Wave 4 and who completed Waves 4 and 5 interviews. ML techniques were used to create models and determine predictors at 1-year follow-up from Wave 4 data. Among the 2,746 tobacco-naïve young adults at baseline, 309 initiated ENDS use at 1-year follow-up. The top five prospective predictors of ENDS initiation were susceptibility to ENDS, increased days of physical exercise specifically designed to strengthen muscles, frequency of social media use, marijuana use and susceptibility to cigarettes. This study identified previously unreported and emerging predictors of ENDS initiation that warrant further investigation and provided comprehensive information on the predictors of ENDS initiation. Furthermore, this study showed that ML is a promising technique that can aid ENDS monitoring and prevention programs.

1. Introduction

Electronic nicotine delivery systems (ENDS), commonly known as e-cigarettes, are extremely popular among young adults. Between 2018 and 2019, ENDS use increased from 7.6 % to 9.3 % among 18–24-year-olds (QuickStats, 2019; Cornelius et al., 2020). In 2019, ENDS use was highest among young adults aged 18–24 years compared with other adult age groups, and 56.0 % of these young adults were never cigarette smokers (Cornelius et al., 2020). This use of ENDS by these young adults has major implications for their health because of the nicotine and other toxicants that they contain. ENDS use can lead to nicotine addiction, substance use problems and future addictions (Overbeek et al., 2020).

Some cross sectional studies have identified age, male sex, race,

cigarette and alcohol use, history of depressive disorder, sensation seeking, higher perceived benefits of ENDS and low adulthood socioeconomic-status, as being associated with ENDS use (Atuegwu et al., 2020; Kenne et al., 2016; Patel et al., 2016; Roys et al., 2020; Teah and Conner, 2021). There are a few studies on the prospective predictors of ENDS initiation in young adults. These studies found that ENDS use by peers, use of alcohol, marijuana, and tobacco products, exposure to ENDS marketing, perceptions of ENDS being less harmful and addicting than cigarettes, and that ENDS use could aid in smoking cessation were predictors of ENDS initiation (Agarwal et al., 2017; Choi and Forster, 2014; Cooper et al., 2022; Cooper et al., 2018; Loukas et al., 2019; Spindle et al., 2017). These studies examined a few predictors at a time in young adults with different tobacco product use status. However,

* Corresponding author at: University of Connecticut, Department of Medicine, 263 Farmington Avenue, Farmington, CT 06030, USA.

E-mail addresses: atuegwu@uchc.edu (N.C. Atuegwu), mortensen@uchc.edu (E.M. Mortensen), suchitra.krishnan-sarin@yale.edu (S. Krishnan-Sarin), Reinhard.Laubenbacher@medicine.ufl.edu (R.C. Laubenbacher), litt@uchc.edu (M.D. Litt).

<https://doi.org/10.1016/j.pmedr.2023.102148>

Received 9 October 2022; Received in revised form 11 January 2023; Accepted 10 February 2023

Available online 13 February 2023

2211-3355/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

reasons for ENDS use have been shown to be different for smokers and never cigarette smokers (Biener et al., 2015). Therefore, predictors of ENDS initiation for tobacco-naïve young adults may be different from those of young adults who have used other tobacco products. Identifying the risk and protective factors of ENDS initiation that are specific to tobacco-naïve young adults will enable the creation of targeted policies and prevention programs.

This study is among the first to use machine learning (ML) models to do a systematic study of the associations between a large number of variables and ENDS initiation in a nationally representative sample. This study adds to the existing literature on ENDS initiation by using ML techniques to provide comprehensive information on the prospective predictors and groups of tobacco-naïve young adults who are most likely to try, or initiate ENDS use in the future.

Our ML analysis uses tree-based algorithms which are effective for large observational data. The ML techniques employed can accommodate multiple predictors and complex relationships, and can identify covariate patterns that predict outcomes better than prespecified models. These ML techniques also reduce errors due to statistical model misspecification and can capture non-linear relationships and interactions between predictors and the endpoint, ENDS initiation (Bzdok et al., 2018; Mooney et al., 2021; Qiu et al., 2022). Therefore, our ML analysis can identify the important predictors that may be focused on for targeted policies and prevention programs. ML has been used in ENDS research (Choi et al., 2021; Fu et al., 2021; Han et al., 2021), including in young adults (Atuegwu et al., 2020). The goal in this study was to use ML to create predictive models and identify multiple prospective predictors of ENDS initiation in tobacco-naïve young adults.

2. Method

2.1. Data and study sample

Predictors of ENDS use identified for earlier generations of ENDS users may be different from those identified today due to changing trends in attitudes, newer products, and policy interventions. Therefore, to identify current prospective predictors of ENDS initiation in tobacco-naïve young adults, we used the recent, publicly available Waves of the Population Assessment of Tobacco and Health (PATH) data: Wave 4 (12/2016–1/2018) and Wave 5 (12/2018–11/2019) (United States Department of Health Human Services et al., 2021).

The adult PATH survey is a nationally representative, longitudinal cohort study of the tobacco use patterns and health of civilian non-institutionalized adults in the US. Surveys for new Waves are conducted approximately a year after the prior Wave. Data from young adults aged 18–24 years in Wave 4 who reported never using any tobacco products, and who completed Waves 4 and 5 interviews were used for the analysis. Never tobacco users reported never use of the following tobacco products: cigarettes, ENDS, cigars, cigarillos, pipe, hookah, snus, smokeless and dissolvable tobacco. ENDS include e-cigarettes, e-cigars, e-pipes and e-hookah. This study used the publicly available data of the PATH project and was deemed exempt by the Institutional Review Board.

Young adults who have never used any tobacco product at Wave 4 and who reported any ENDS use at Wave 5 were classified as ENDS initiators. The comparison group for the outcome were those young adults who reported never using ENDS in both Waves. There were 2,746 respondents available for the analysis. The population weighted proportions of females, Whites, Hispanics and ENDS initiators were 53.8 %, 67.8 %, 21.3 % and 9.0 % respectively. Respondents' baseline descriptive statistics of sociodemographic and substance use are shown in Table 1.

2.2. Independent variables

The PATH dataset contains > 2000 variables, including

Table 1

Descriptive statistics of the baseline (Wave 4) sociodemographic and substance use stratified by ENDS initiation at Wave 5^a.

	All respondents (N = 2746)	Not ENDS initiator (N = 2437)	ENDS initiators (N = 309)
Gender			
Male	1238 (46.2)	1080 (45.6)	158 (52.0)
Female	1508 (53.8)	1357 (54.4)	151 (48.0)
Race			
White alone	1813 (67.8)	1594 (67.4)	219 (72.5)
Black alone	525 (16.1)	479 (16.4)	46 (12.7)
Other ^b	408 (16.1)	364 (16.2)	44 (14.8)
Hispanic			
Yes	790 (21.3)	695 (21.1)	95 (23.2)
No	1956 (78.7)	1742 (78.9)	214 (76.8)
Currently enrolled in a degree program			
Yes	1344 (50.3)	1173 (49.4)	171 (59.2)
No	1393 (49.5)	1257 (50.4)	136 (40.2)
Anyone who lives with you now who uses tobacco			
Cigarettes, cigars, cigarillos or filtered cigars and pipe tobacco	461 (17.2)	391 (16.7)	70 (22.2)
E-products exclusively	42 (1.6)	35 (1.5)	7 (2.1) ^c
Other tobacco products, including smokeless, snus and hookah	78 (2.5)	63 (2.3)	15 (4.2)
No one living in the home uses tobacco	2040 (73.7)	1833 (74.2)	207 (67.9)
Currently live with a spouse or romantic partner			
Yes	245 (13.3)	237 (14.2)	8 (3.6) ^c
No	1999 (71.1)	1767 (70.5)	232 (77.1)
Used marijuana in the past 12 months			
Yes	227 (7.6)	166 (6.5)	61 (19.8)
No	2518 (92.3)	2270 (93.5)	248 (80.2)
Used alcohol in the past 12 months			
Yes	1178 (47.9)	1001 (46.6)	177 (61.6)
No	1565 (51.9)	1433 (53.3)	132 (38.4)

^a Data is expressed as unweighted number (weighted percentage) of respondents.

^b Other race includes respondents who are American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander and multiracial.

^c Relative standard error (RSE) > 30 %, therefore weighted percentages may not be reliable.

demographics, tobacco and substance use, advertising exposures, risk perceptions, and peer and family influences. Details about PATH variables are available elsewhere (Hyland et al., 2017; United States Department of Health Human Services et al., 2021).

2.3. Model development and statistical analysis

After preprocessing the data (Supplementary Methods), 242 variables were available for ML analysis. The Extreme gradient boosting (XGBoost) algorithm (Chen and Guestrin, 2016) was used to create the prediction models and the SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) algorithm was used to interpret the final models (Supplementary Methods). To corroborate the findings from ML, multivariable logistic regression models for the outcome and the predictors identified using ML were created. These models adjusted for gender, race and ethnicity and accounted for the PATH complex design.

3. Results

Model performance was calculated with the Area under the Receiver

Operating Characteristic curve (AUC) and the Area under the Precision Recall Curve (AUCPR). The maximum AUC and AUCPR for the XGBoost classifier were 0.82 (mean of 0.73; standard deviation of 0.05) and 0.48 (mean of 0.30; standard deviation of 0.06) respectively. The baseline AUCPR and AUC for the analysis were 0.11 (unweighted prevalence of ENDS initiators) and 0.5 respectively. AUC and AUPRC above the baseline values indicate that the model can discriminate between ENDS initiators and non-initiators (Saito and Rehmsmeier, 2015). Descriptive statistics of influential baseline variables on the prediction of ENDS initiation selected by ML are in Supplementary Table 1.

The top 20 predictors based on SHAP values are shown in Fig. 1A (Global importance) and 1B (Summary plot). SHAP dependence plots (Fig. 2) show the relationship (e.g., complex) between the variable and the prediction of the outcome. Positive SHAP values indicate increased risk of initiating ENDS while negative SHAP values indicate decreased risk of initiating ENDS.

The most influential variable was susceptibility to ENDS. As described elsewhere (Nicksic and Barnes, 2019; Pérez et al., 2021; Pierce et al., 2018; Pierce et al., 2017), susceptibility to ENDS was determined by reporting “Definitely yes”, “Probably yes” or “Probably not” to the

following questions: “Think you will use an electronic nicotine product in the next year”, “Think you will try an electronic nicotine product soon”, “Would try an electronic nicotine product if one of your best friends offered it to you” and “A little curious” “Somewhat curious” or “Very curious” to the question “Ever been curious about using an electronic nicotine product”. Non-susceptible young adults answered “Definitely not” to the first three questions and “Not at all curious” to the last question. Young adults who recorded any other response combinations were considered susceptible. Susceptibility to cigarettes and hookah were determined with similar questions. Reporting ENDS susceptibility was associated with positive SHAP values and therefore increased risk of initiating ENDS, while reporting no ENDS susceptibility was associated with negative SHAP values and therefore decreased risk of initiating ENDS (Fig. 1A and 2). We grouped the influential variables identified with SHAP (Fig. 2) into categories for clarity.

3.1. Substance use and susceptibility to tobacco

Susceptibility to ENDS, cigarettes and hookah, and using marijuana and alcohol in the past 12 months was associated with positive SHAP

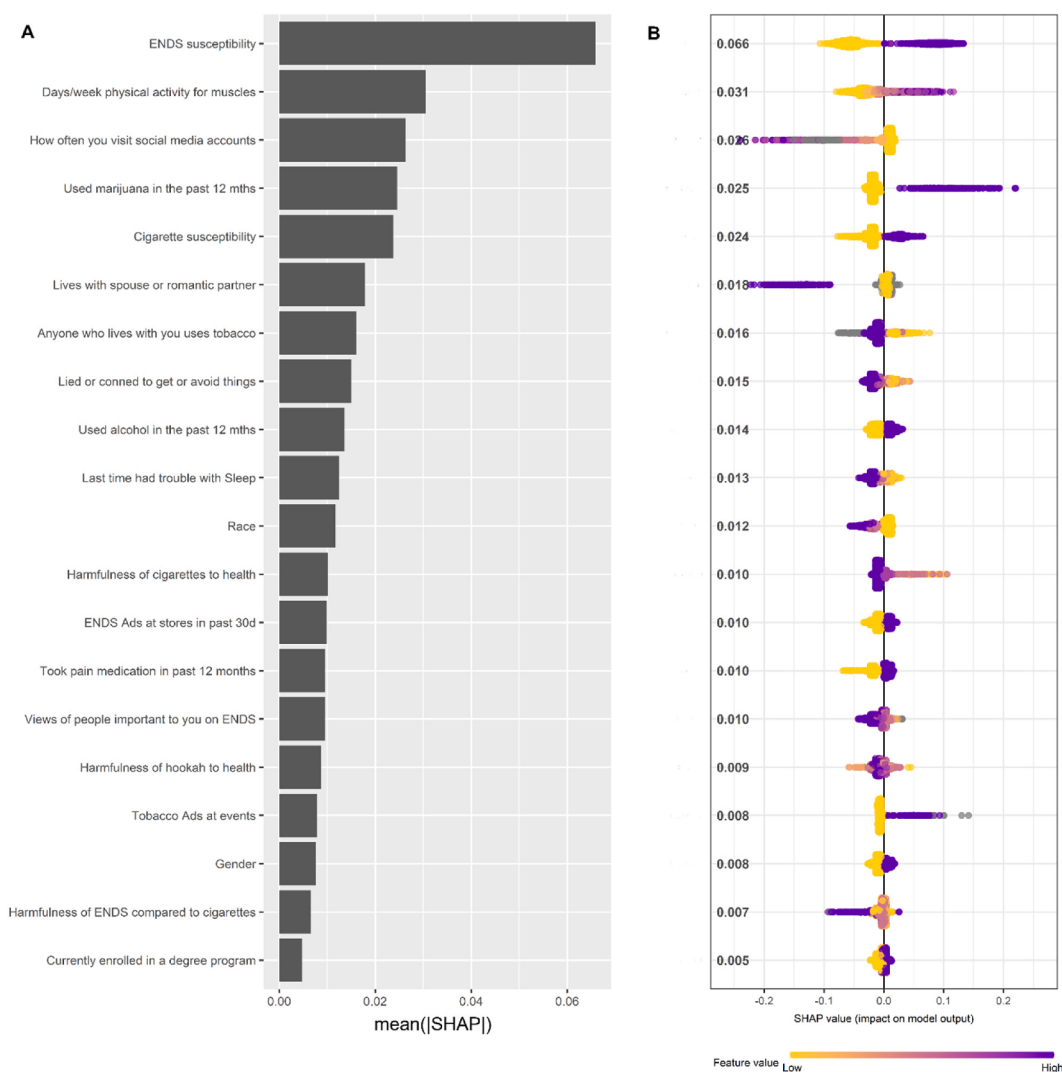


Fig. 1. SHAP global importance (A) and summary plot (B) of the prospective predictors of ENDS initiation at 1-year follow up for the top 20 variables. Global importance ranks predictors in order of their importance to the prediction of ENDS initiation. The variable importance is determined by the mean of the absolute value of the SHAP values shown in the x-axis in (A) and y-axis in (B). The higher the mean of the absolute SHAP value, the more important the variable is to the prediction of ENDS initiation. In (B) each dot is a respondent, the color of the dot represents the value of the predictor and the horizontal location of the dot shows the SHAP value. A negative SHAP value indicates that the predictor contributed to the respondent not initiating ENDS while a positive SHAP value indicates that the predictor contributed to the respondent initiating ENDS.

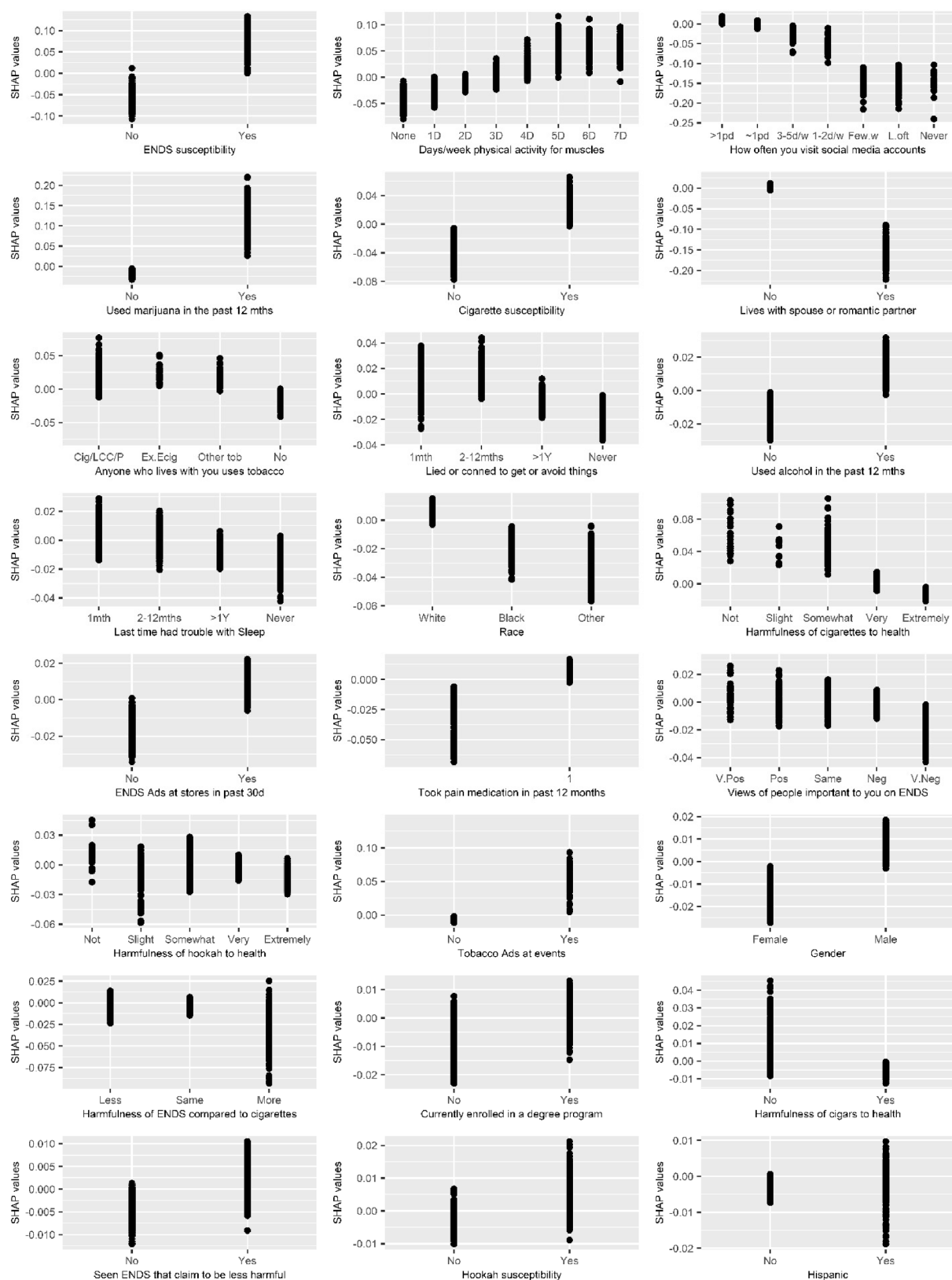


Fig. 2. SHAP dependence plots for the predictors of ENDS initiation. The x and y-axes are the value and SHAP values of the predictors respectively. Positive SHAP values indicate increased risk of ENDS initiation and negative SHAP values indicate decreased risk of ENDS initiation.

values. Also, reporting taking pain medication in the past 12 months was associated with positive SHAP values.

3.2. Physical activity

The SHAP values for physical exercise specifically designed to strengthen muscles (not including cardio) were negative for no days of strength training and increased with increasing days of strength training with positive values occurring at ≥ 4 days a week.

3.3. Internalizing and externalizing problems

Responding “never” to: “lied or conned to get things or avoid doing something” (externalizing problem) and “sleep trouble” (internalizing problems) (Conway et al., 2018), were associated with negative SHAP values.

3.4. Media use and exposure to advertisement

The SHAP values for frequency of use of social media were positive for reporting use of social media more than once a day and decreased with decreasing frequency of social media use. Noticing ENDS being advertised at gas stations and stores, noticing tobacco products being advertised at events, and seeing any advertisements claiming that ENDS were less harmful than cigarettes was associated with positive SHAP values.

3.5. Demographics, social influence, and beliefs about tobacco products

Living with a spouse/romantic partner, reporting being of Black or Other race, female gender, believing that cigarettes, hookah and cigars were harmful to health, not living with a tobacco user, and having important people with very negative views on ENDS were associated with negative SHAP values. In contrast, reporting current enrollment in a degree program was associated with positive SHAP values.

3.6. Multivariable logistic regression models

The odds ratios for most of the predictors identified with ML showed statistically significant associations with ENDS initiation (Supplementary Table 2). We however found no statistically significant association between race, gender and ENDS initiation. A possible explanation may be that the relationship between gender, race and ENDS initiation is more complex than our prespecified logistic regression models. However, our ML approach may have been able to capture this relationship because of its ability to consider non-linear relationships between the predictors and prediction of the outcome.

3.7. Variable interactions

The magnitude of the SHAP values for the pairwise interactions between variables in the model are shown in Supplementary Fig. 1. The most notable interaction occurred between “susceptibility to ENDS” and “susceptible to cigarettes”. Respondents with both of these positive risk factors had even greater predicted risk of initiating ENDS (Supplementary Fig. 2A and C). We also identified some other pairwise variable interactions. The interaction between use of marijuana in the past 12 months and reporting susceptibility to ENDS showed an increased predicted risk of ENDS initiation for young adults who reported no susceptibility to ENDS but had used marijuana (Supplementary Fig. 2D and 2F). The interaction between physical activity designed to strengthen muscles and ENDS susceptibility showed a greater predicted risk of ENDS initiation for young adults who were susceptible to ENDS and who also frequently engaged in physical exercise to strengthen muscles (Supplementary Fig. 2G and 2I).

4. Discussion

ML techniques were used to create predictive models and identify multiple risk and protective factors of ENDS initiation from a large number of baseline variables at 1-year follow-up in a nationally representative sample of tobacco-naïve young adults. Several predictors of ENDS initiation were identified, and these were ranked in order of their importance to the prediction. The predictors identified with ML were also corroborated with multivariable logistic regression models.

Some of the predictors identified with ML aligned with those previously reported in the literature. Similar to our results from the ML analysis, other investigators also noted that respondents who reported race as Black or Other, and gender as female were less likely to initiate ENDS (Kenne et al., 2016). Also, ever use of alcohol and marijuana, living with a tobacco user, and increased exposure to ENDS and other tobacco product marketing have been shown to be predictors of ENDS initiation in adults (Cooper et al., 2022; Nicksic et al., 2017; Villanti et al., 2015).

The most influential predictor of ENDS initiation was reporting any susceptibility to ENDS. While susceptibility to ENDS has been shown to be associated with ENDS initiation in youth (Bold et al., 2020), to the best of our knowledge this has not been reported in tobacco-naïve young adults. Susceptibility to cigarettes was also identified as a predictor of ENDS initiation. This suggests that tobacco-naïve young adults, who are curious about smoking or have the intent to smoke, are also more likely to initiate ENDS. This could have implications for policies that have strict bans on ENDS while leaving cigarettes accessible to these young adults. Perhaps, some of these young adults may move on to cigarettes if no ENDS are available. Conversely, ENDS may also lead to future use of cigarettes (Bold et al., 2018). Furthermore, our analysis found that tobacco-naïve young adults who were susceptible to both ENDS and cigarettes had further increased risk of initiating ENDS. Intervention strategies to reduce ENDS initiation should identify and target young adults who are susceptible to ENDS as well as other tobacco products. Advertisements in media, and schools, might preempt nicotine use by appealing directly to those who are curious but not using nicotine products.

We found that young adults who used marijuana in the past 12 months were more likely to initiate ENDS. We also found that marijuana use in the past 12 months was associated with increased risk of ENDS initiation among young adults who reported no susceptibility to ENDS. These preliminary findings indicate that while reporting no susceptibility to ENDS was protective against initiating ENDS, the use of marijuana by these young adults may have increased the risk of ENDS initiation for these young adults. Additional studies are needed to understand the association between marijuana use and ENDS initiation especially for young adults who report no susceptibility to ENDS.

We found that young adults who used social media accounts more than once a day had increased risk of initiating ENDS, and that this risk decreases as the frequency of use of social media decreases. A possible explanation may be that the use of social media exposes young adults to ENDS and tobacco advertisement as well as other tobacco related content from peers, all of which could increase susceptibility to ENDS (Yang et al., 2021). Efforts to regulate ENDS and tobacco related messages on social media, and the placement of anti-ENDS messages on social media sites may reduce ENDS initiation.

The current study extends the literature on ENDS initiation by identifying ≥ 4 days a week of physical activity designed to strengthen muscles as a predictor of ENDS initiation. We also found further increased risk of ENDS initiation for young adults who were susceptible to ENDS use and who frequently engaged in exercise to strengthen muscles. It is plausible that young adults who are trying to gain muscle mass and improve their physique are using nicotine supplied via ENDS to suppress appetite and reduce body fat. Nicotine has been shown to reduce body fat and decrease food intake (Calarco and Picciotto, 2020). Other explanations could be that young adults who are drawn to the

'novelty' of muscle building may also be drawn to the 'novelty' of ENDS use. Also, these young adults may be subject to more pressure in gyms to use ENDS. Further studies are required to understand the association between frequency of muscle building exercises and ENDS initiation including the interaction with ENDS susceptibility, and to validate if this may be an emerging group of young adults who are vulnerable to ENDS use.

Additionally, we found that young adults who used pain medication in the past 12 months were more likely to initiate ENDS. Perhaps these young adults may be using ENDS for some pain relief due to the analgesic effects of nicotine (Ditre et al., 2016). It is important to understand the association between pain and ENDS initiation especially in tobacco-naïve young adults. We also found that young adults who were currently enrolled in a degree program were more likely to initiate ENDS. This is in contrast with a study that showed that non-college young adults had higher prevalence of ENDS use (Buu et al., 2020a,b). The difference may be due to the data used. The earlier study used Wave 1 of PATH (2013–2014). The current study, using a later PATH Wave, may be capturing a more recent trend in ENDS use among college students. A recent report by the American College Health Association showed that ENDS are the most commonly used tobacco product among young adults transitioning to and on college campuses (American College Health Association, 2021). Similarly, the Monitoring The Future (MTF) study showed a 6 % to 19 % increase from 2017 to 2020 in past 30-day nicotine vaping for college students (Schulenberg et al., 2021), and a higher annual prevalence of nicotine vaping among college than non-college students from 2018 to 2020. Strategies such as targeted campaigns, promoting awareness of ENDS use with student health providers, and promoting tobacco-free campus policies may reduce ENDS initiation in colleges (American College Health Association, 2021; Escoto et al., 2021).

Our analysis showed that young adults who did not report internalizing and externalizing problems were less likely to initiate ENDS. Internalizing and externalizing problems are indicative of mental health problems (Riehm et al., 2019). This contrasts with a study that showed no association between reporting internalizing or externalizing problems and ENDS initiation in young adults (Cooper et al., 2022). This difference may be due to the sample of young adults used. Our analysis focused on tobacco-naïve young adults. Other studies however have shown that reporting higher numbers of internalizing and externalizing problems was associated with higher risks of initiating ENDS in youth (Buu et al., 2020a,b; Cooper et al., 2022). We also found that young adults who lived with spouses/romantic partners were less likely to initiate ENDS. This may be because these young adults were less likely to engage in risky behaviors such as ENDS use. Intimate relationships have been shown to have a significant effect on health behaviors (Umberson et al., 2010). Our analysis found that young adults who had negative views on tobacco, did not live with tobacco users, and had a social network of people who were not supportive of ENDS use, were less likely to initiate ENDS. Youth and young adults who had friends who smoked cigarettes or had used ENDS were more likely to initiate ENDS themselves (Jayakumar et al., 2020). ENDS prevention programs should identify young adults with internalizing and externalizing problems for targeted interventions, should emphasize the negative effects of tobacco products, and should encourage establishing a social network of non-tobacco users.

A strength of the current study was the use of respondents who were nationally representative of tobacco-naïve young adults in the U.S. Additionally, ML removed the biases in preselected models of ENDS initiation. We did not prespecify the predictors and models of ENDS initiation, rather we used ML to simultaneously consider many variables, identify predictors including those previously unknown, and identify nonlinear relationships and interactions between the predictors and the prediction of ENDS initiation. This decreased the possibility of statistical misclassification, and the omission of predictors. We also used multivariable logistic regression models to independently corroborate

the findings from ML. Furthermore, our findings showed the predictors that may have the most impact on ENDS initiation in tobacco-naïve young adults and may provide possible areas that public health officials could concentrate on to decrease ENDS initiation in tobacco-naïve young adults.

This study has some limitations. The data are based on self-report and may have recall bias. We cannot establish causality from our analysis. We could not include some potential predictors such as the age of the respondents, "sensation seeking" and a diagnosis of depression in the model because they were not available in the public adult PATH survey. There may be other confounders that can affect the identified predictors. While we feel that the sample size of young adults is sufficient to carry out the ML analysis, a larger sample size may increase the accuracy of the models. We created the ML models on unweighted PATH data, because the ML algorithms could not account for the PATH complex design. However, our findings are still relevant and were corroborated with logistic regression models that incorporated the PATH complex design and weights. There is also a possibility of spurious findings. We reduced the possibility of spuriousness because the ML algorithms considered all the variables and their interactions simultaneously and not serially. Also, the use of repeated CV and different ML algorithms reduced the dependence of the results on one ML or iteration of a ML algorithm, which reduced the likelihood of spurious findings. Furthermore, our findings were corroborated with logistic regression models. Respondents who initiate ENDS use may become experimental or frequent users. Future work should create ML models for these types of ENDS use.

In conclusion, we showed that ML can identify predictors and the complex relationships between these predictors and ENDS initiation. This indicates that ML may be a promising technique for ENDS monitoring and prevention programs, and may identify emerging predictors of ENDS use. Our findings if validated may have implications for policies and programs designed to prevent ENDS initiation in young adults.

Funding

Dr. Atuegwu receives support from Robert E. Leet and Clara Guthrie Patterson Trust Mentored Research grant.

CRediT authorship contribution statement

Nkiruka C. Atuegwu: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Eric M. Mortensen:** Writing – review & editing. **Suchitra Krishnan-Sarin:** Writing – review & editing. **Reinhard C. Lauenbacher:** Writing – review & editing. **Mark D. Litt:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pmedr.2023.102148>.

References

- Agarwal, D., Loukas, A., Perry, C.L., 2017. Examining College Students' Social Environment, Normative Beliefs, and Attitudes in Subsequent Initiation of Electronic Nicotine Delivery Systems. *Health Educ. Behav.* 45, 532–559.
- American College Health Association, 2021. Addressing E-cigarette Use and Vaping on College and University Campuses.

- Atuegwu, N.C., Oncken, C., Laubenbacher, R.C., Perez, M.F., Mortensen, E.M., 2020. Factors Associated with E-Cigarette Use in U.S. Young Adult Never Smokers of Conventional Cigarettes: A Machine Learning Approach. *Int. J. Environ. Res. Public Health* 17, 7271.
- Biener, L., Song, E., Sutfin, E.L., Spangler, J., Wolfson, M., 2015. Electronic Cigarette Trial and Use among Young Adults: Reasons for Trial and Cessation of Vaping. *Int. J. Environ. Res. Public Health* 12, 16019–16026.
- Bold, K.W., Kong, G., Camenga, D.R., Simon, P., Cavallo, D.A., Morean, M.E., Krishnan-Sarin, S., 2018. Trajectories of E-Cigarette and Conventional Cigarette Use Among Youth. *Pediatrics* 141, e20171832.
- Bold, K.W., Buta, E., Simon, P., Kong, G., Morean, M., Camenga, D., Krishnan-Sarin, S., 2020. Using Latent Class Analysis to Examine Susceptibility to Various Tobacco Products Among Adolescents. *Nicotine Tobacco Res.* 22, 2059–2065.
- Buu, A., Hu, Y.-H., Wong, S.-W., Lin, H.-C., 2020a. Internalizing and Externalizing Problems as Risk Factors for Initiation and Progression of E-cigarette and Combustible Cigarette Use in the US Youth Population. *Int. J. Ment. Heal. Addict.*
- Buu, A., Hu, Y.H., Wong, S.W., Lin, H.C., 2020b. Comparing American college and noncollege young adults on e-cigarette use patterns including polysubstance use and reasons for using e-cigarettes. *J. Am. Coll. Health* 68, 610–666.
- Bzdok, D., Altman, N., Krzywinski, M., 2018. Statistics versus machine learning. *Nat. Methods* 15, 233–324.
- Calarco, C.A., Picciotto, M.R., 2020. Nicotinic Acetylcholine Receptor Signaling in the Hypothalamus: Mechanisms Related to Nicotine's Effects on Food Intake. *Nicotine Tobacco Res.* 22, 152–163.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, California, USA, pp. 785–794.
- Choi, K., Forster, J.L., 2014. Beliefs and Experimentation with Electronic Cigarettes: A Prospective Analysis Among Young Adults. *Am. J. Prev. Med.* 46, 175–218.
- Choi, J., Jung, H.-T., Ferrell, A., Woo, S., Haddad, L., 2021. Machine Learning-Based Nicotine Addiction Prediction Models for Youth E-Cigarette and Waterpipe (Hookah) Users. *J. Clin. Med.* 10, 972.
- Conway, K.P., Green, V.R., Kasza, K.A., Silveira, M.L., Borek, N., Kimmel, H.L., Sargent, J.D., Stanton, C.A., Lambert, E., et al., 2018. Co-occurrence of tobacco product use, substance use, and mental health problems among youth: Findings from wave 1 (2013–2014) of the population assessment of tobacco and health (PATH) study. *Addict. Behav.* 76, 208–217.
- Cooper, M., Loukas, A., Case, K.R., Marti, C.N., Perry, C.L., 2018. A longitudinal study of risk perceptions and e-cigarette initiation among college students: Interactions with smoking status. *Drug Alcohol Depend.* 186, 257–263.
- Cooper, M., Day, H.R., Ren, C., Oniyide, O., Corey, C.G., Ambrose, B.K., Michael Cummings, K., Sargent, J., Niaura, R., et al., 2022. Correlates of tobacco product initiation among youth and young adults between waves 1–4 of the population assessment of tobacco and health (PATH) study (2013–2018). *Addict. Behav.* 134, 107396.
- Cornelius, M.E., Wang, T.W., Jamal, A., Loretan, C.G., Neff, L.J., 2020. Tobacco Product Use Among Adults - United States, 2019. *MMWR Morb. Mortal. Wkly Rep.* 69, 1736–1742.
- Ditre, J.W., Heckman, B.W., Zale, E.L., Kosiba, J.D., Maisto, S.A., 2016. Acute analgesic effects of nicotine and tobacco in humans: a meta-analysis. *Pain* 157, 1373–1381.
- Escoto, A., Watkins, S.L., Welter, T., Beecher, S., 2021. Developing a targeted e-cigarette health communication campaign for college students. *Addict. Behav.* 117, 106841.
- Fu, R., Shi, J., Chaiton, M., Leventhal, A.M., Unger, J.B., Barrington-Trimis, J.L., 2021. A Machine Learning Approach to Identify Predictors of Frequent Vaping and Vulnerable Californian Youth Subgroups. *Nicotine Tob. Res.*
- Han, D.-H., Lee, S.H., Lee, S., Seo, D.-C., 2021. Identifying emerging predictors for adolescent electronic nicotine delivery systems use: A machine learning analysis of the Population Assessment of Tobacco and Health Study. *Prev. Med.* 145, 106418.
- Hyland, A., Ambrose, B.K., Conway, K.P., Borek, N., Lambert, E., Carusi, C., Taylor, K., Crosse, S., Fong, G.T., et al., 2017. Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. *Tob. Control* 26, 371–438.
- Jayakumar, N., O'Connor, S., Diemert, L., Schwartz, R., 2020. Predictors of E-Cigarette Initiation: Findings From the Youth and Young Adult Panel Study. *Tobacco Use Insights*, 13:1179173X20977486.
- Kenne, D.R., Mix, D., Banks, M., Fischbein, R., 2016. Electronic cigarette initiation and correlates of use among never, former, and current tobacco cigarette smoking college students. *J. Subst. Abus.* 21, 491–544.
- Loukas, A., Paddock, E.M., Li, X., Harrell, M.B., Pasch, K.E., Perry, C.L., 2019. Electronic Nicotine Delivery Systems Marketing and Initiation Among Youth and Young Adults. *Pediatrics* 144, e20183601.
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, NIPS.
- Mooney, S.J., Keil, A.P., Westreich, D.J., 2021. Thirteen Questions About Using Machine Learning in Causal Research (You Won't Believe the Answer to Number 10!). *Am. J. Epidemiol.* 190, 1476–1482.
- Nicksic, N.E., Barnes, A.J., 2019. Is susceptibility to E-cigarettes among youth associated with tobacco and other substance use behaviors one year later? Results from the PATH study. *Prev. Med.* 121, 109–114.
- Nicksic, N.E., Snell, L.M., Rudy, A.K., Cobb, C.O., Barnes, A.J., 2017. Tobacco Marketing, E-cigarette Susceptibility, and Perceptions among Adults. *Am. J. Health Behav.* 41, 579–590.
- Overbeek, D.L., Kass, A.P., Chiel, L.E., Boyer, E.W., Casey, A.M.H., 2020. A review of toxic effects of electronic cigarettes/vaping in adolescents and young adults. *Crit. Rev. Toxicol.* 50, 531–558.
- Patel, D., Davis, K.C., Cox, S., Bradfield, B., King, B.A., Shafer, P., Caraballo, R., Bunnell, R., 2016. Reasons for current E-cigarette use among U.S. adults. *Prev. Med.* 93, 14–20.
- Pérez, A., Blustein, M.A., Kuk, A.E., Chen, B., Sterling, K.L., Harrell, M.B., 2021. Age of Onset of Susceptibility to Different Tobacco Products Among Non-Susceptible US Young Adults: Findings from the Population Assessment of Tobacco and Health Study Waves 2–4 (2014–2017). *Tob. Use Insights*, 14:1179173x211065643.
- Pierce, J.P., Sargent, J.D., White, M.M., Borek, N., Portnoy, D.B., Green, V.R., Kaufman, A.R., Stanton, C.A., Bansal-Travers, M., et al., 2017. Receptivity to tobacco advertising and susceptibility to tobacco products. *Pediatrics* 139.
- Pierce, J.P., Sargent, J.D., Portnoy, D.B., White, M., Noble, M., Kealey, S., Borek, N., Carusi, C., Choi, K., et al., 2018. Association Between Receptivity to Tobacco Advertising and Progression to Tobacco Use in Youth and Young Adults in the PATH Study. *JAMA Pediatr.* 172, 444–451.
- Qiu, W., Chen, H., Dincer, A.B., Lundberg, S., Kaeberlein, M., Lee, S.-I., 2022. Interpretable machine learning prediction of all-cause mortality. *Commun. Med.* 2, 125.
2019. QuickStats: Percentage of Adults Aged 18–24 Years Who Currently Smoke Cigarettes* or Who Currently Use Electronic Cigarettes,(†) by Year - National Health Interview Survey, United States, 2014–2018(§). *MMWR Morb Mortal Wkly Rep* 68: 870.
- Riehm, K.E., Young, A.S., Feder, K.A., Krawczyk, N., Tormohlen, K.N., Pacek, L.R., Mojtabai, R., Crum, R.M., 2019. Mental Health Problems and Initiation of E-cigarette and Combustible Cigarette Use. *Pediatrics* 144, e20182935.
- Roys, M.R., Peltier, M.R., Stewart, S.A., Waters, A.F., Waldo, K.M., Copeland, A.L., 2020. The association between problematic alcohol use, risk perceptions, and e-cigarette use. *Am. J. Drug Alcohol Abuse* 46, 224–231.
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10, e0118432.
- Schulenberg, J.E., Patrick, M.E., Johnston, L.D., O'Malley, P.M., Bachman, J.G., Miech, R.A., 2021. Monitoring the Future National Survey Results on Drug Use, 1975–2020. Volume II, College Students & Adults Ages 19–60. Institute for Social Research.
- Spindle, T.R., Hiler, M.M., Cooke, M.E., Eissenberg, T., Kendler, K.S., Dick, D.M., 2017. Electronic cigarette use and uptake of cigarette smoking: A longitudinal examination of U.S. college students. *Addict. Behav.* 67, 66–72.
- Teah, G.E., Conner, T.S., 2021. Psychological and Demographic Predictors of Vaping and Vaping Susceptibility in Young Adults. *Front. Psychol.* 12.
- Umberson, D., Crosnoe, R., Reczek, C., 2010. Social Relationships and Health Behavior Across Life Course. *Annu. Rev. Sociol.* 36, 139–157.
- United States Department of Health Human Services, National Institutes of Health, National Institute on Drug Abuse, United States Department of Health Human Services, Food Drug Administration, Center for Tobacco Products, 2021. Population Assessment of Tobacco and Health (PATH) Study [United States] Public-Use Files: PATH Study Public Use Files User Guide. Inter-university Consortium for Political and Social Research [distributor].
- Villanti, A.C., Rath, J.M., Williams, V.F., Pearson, J.L., Richardson, A., Abrams, D.B., Niaura, R.S., Vallone, D.M., 2015. Impact of Exposure to Electronic Cigarette Advertising on Susceptibility and Trial of Electronic Cigarettes and Cigarettes in US Young Adults: A Randomized Controlled Trial. *Nicotine Tob. Res.* 18, 1331–2139.
- Yang, Q., Clendennen, S.L., Loukas, A., 2021. How Does Social Media Exposure and Engagement Influence College Students' Use of ENDS Products? A Cross-lagged Longitudinal Study. *Health Commun.* 1–10.