



# Can adverse childhood experiences predict chronic health conditions? Development of trauma-informed, explainable machine learning models

Hanin B. Afzal, *University of Toronto*  
Tasfia Jahangir, *Emory University*  
Yiyang Mei, *Emory University*  
Annabelle Madden, *Columbia University*  
[Abeed Sarker](#), *Emory University*  
Sangmi Kim, *Emory University*

---

**Journal Title:** Frontiers in Public Health

**Volume:** Volume 11

**Publisher:** Frontiers | 2023, Pages 1309490-None

**Type of Work:** Article | Final Publisher PDF

**Publisher DOI:** 10.3389/fpubh.2023.1309490

**Permanent URL:** <http://pid.emory.edu/ark:/25593/w9dm7>

---

Final published version: <https://doi.org/10.3389%2Fpubh.2023.1309490>

## **Copyright information:**

© 2024 Afzal, Jahangir, Mei, Madden, Sarker and Kim.

This is an Open Access work distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

Accessed April 25, 2025 4:42 AM EDT



## OPEN ACCESS

## EDITED BY

Dabney Evans,  
Emory University, United States

## REVIEWED BY

Fuad Ismayilov,  
Azerbaijan Medical University, Azerbaijan  
Carolyn Gentle-Genitty,  
Indiana University Bloomington, United States  
Saahoon Hong,  
Indiana University, United States

## \*CORRESPONDENCE

Tasfia Jahangir  
✉ tasfia.jahangir@emory.edu

<sup>†</sup>These authors have contributed equally to this work and share first authorship

RECEIVED 08 October 2023

ACCEPTED 27 December 2023

PUBLISHED 15 January 2024

## CITATION

Afzal HB, Jahangir T, Mei Y, Madden A, Sarker A and Kim S (2024) Can adverse childhood experiences predict chronic health conditions? Development of trauma-informed, explainable machine learning models.

*Front. Public Health* 11:1309490.  
doi: 10.3389/fpubh.2023.1309490

## COPYRIGHT

© 2024 Afzal, Jahangir, Mei, Madden, Sarker and Kim. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Can adverse childhood experiences predict chronic health conditions? Development of trauma-informed, explainable machine learning models

Hanin B. Afzal<sup>1†</sup>, Tasfia Jahangir<sup>2\*†</sup>, Yiyang Mei<sup>3</sup>,  
Annabelle Madden<sup>4</sup>, Abeed Sarker<sup>5</sup> and Sangmi Kim<sup>6</sup>

<sup>1</sup>Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada,

<sup>2</sup>Department of Behavioral, Social and Health Education Sciences, Rollins School of Public Health, Emory University, Atlanta, GA, United States, <sup>3</sup>School of Law, Emory University, Atlanta, GA, United States, <sup>4</sup>Teachers College, Columbia University, New York, NY, United States, <sup>5</sup>Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States,

<sup>6</sup>Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States

**Introduction:** Decades of research have established the association between adverse childhood experiences (ACEs) and adult onset of chronic diseases, influenced by health behaviors and social determinants of health (SDoH). Machine Learning (ML) is a powerful tool for computing these complex associations and accurately predicting chronic health conditions.

**Methods:** Using the 2021 Behavioral Risk Factor Surveillance Survey, we developed several ML models—random forest, logistic regression, support vector machine, Naïve Bayes, and K-Nearest Neighbor—over data from a sample of 52,268 respondents. We predicted 13 chronic health conditions based on ACE history, health behaviors, SDoH, and demographics. We further assessed each variable's importance in outcome prediction for model interpretability. We evaluated model performance via the Area Under the Curve (AUC) score.

**Results:** With the inclusion of data on ACEs, our models outperformed or demonstrated similar accuracies to existing models in the literature that used SDoH to predict health outcomes. The most accurate models predicted diabetes, pulmonary diseases, and heart attacks. The random forest model was the most effective for diabetes (AUC = 0.784) and heart attacks (AUC = 0.732), and the logistic regression model most accurately predicted pulmonary diseases (AUC = 0.753). The strongest predictors across models were age, ever monitored blood sugar or blood pressure, count of the monitoring behaviors for blood sugar or blood pressure, BMI, time of last cholesterol check, employment status, income, count of vaccines received, health insurance status, and total ACEs. A cumulative measure of ACEs was a stronger predictor than individual ACEs.

**Discussion:** Our models can provide an interpretable, trauma-informed framework to identify and intervene with at-risk individuals early to prevent chronic health conditions and address their inequalities in the U.S.

## KEYWORDS

behavioral risk factor surveillance survey, machine learning, adverse childhood experiences, chronic diseases, health behaviors, health outcomes

## 1 Introduction

Adverse childhood experiences (ACEs) represent a critical public health issue. Defined as potentially traumatic events that occur in childhood (0–17 years old), ACEs include but are not limited to children experiencing emotional, physical, and sexual abuse, parental neglect, household instability such as parents' divorce or separation, and suicide attempts (1). According to the Centers for Disease Control and Prevention (CDC), approximately 61% of adults surveyed across 25 states reported having experienced at least one ACE before adulthood; one in six claimed that they had experienced four or more ACEs (1). Despite the widespread prevalence of ACEs, some groups are at a higher risk of ACE exposure than others. For example, Black, Hispanic, or low-income individuals show the highest prevalence of ACEs (2). Additionally, social, economic, and environmental inequities are greater in the environments of those who have endured four or more ACEs (3).

Current literature has documented that experiences of maltreatment and psychosocial stress during childhood play a significant role in shaping a wide range of chronic health conditions, which constitute physical and mental health problems that last for a prolonged period (i.e., 1 year or longer) (4). The seminal ACE Study with 17,000 adults found a clear and strong correlation between the number of negative experiences during childhood and a wide spectrum of poor health and behavioral outcomes in adult life (5). The study demonstrates a dose–response relationship between the number of ACEs and chronic diseases (e.g., ischemic heart disease, cancer, and chronic lung disease) (5). Since then, mounting evidence indicates the positive associations between ACEs and chronic health conditions, including arthritis, pulmonary disease, cancers, cardiovascular disease, stroke, pre-diabetes, diabetes, high cholesterol, and renal disease (6–28). In addition, individuals with ACEs are found to be at greater risk of experiencing poor mental health (e.g., depression, anxiety, and hallucination) (29–36).

Multiple pathways connect ACEs to chronic health conditions, including social determinants of health (SDoH) and health behaviors. Individuals with a higher number of ACEs tend to live in areas of greater poverty, fewer economic and health resources, worse food access, less green space, and more community instability (3, 37, 38). ACE survivors are also more likely to engage in harmful behaviors, such as smoking, heavy alcohol consumption, substance use, high-risk sexual behavior, interpersonal violence, excess screen time, and inadequate sleep (5, 27, 30, 39–42).

Such clustering of social and disease conditions in a specific population is well-explained by syndemic theory. A syndemic is defined as the “aggregation of two or more diseases or other health conditions in a population in which there is some level of deleterious biological or behavior interface that exacerbates the negative health effects of any or all of the diseases involved” (43). In syndemics, social conditions contribute to disease formation, accumulation, spread, and progression by increasing susceptibility and reducing immune function, particularly among marginalized populations; hence, syndemics are most likely to

emerge under conditions of health inequality (43). A syndemic can be exemplified by the interactions of ACEs, negative social conditions (i.e., SDoH), and risky health behaviors worsening the risk of various chronic health conditions (2, 3, 44). However, an accurate assessment of these complex associations can be methodologically challenging, as the involved risk factors may be highly correlated, interactive, or synergistic. In such cases, it is essential to employ models that are more flexible than linear regression, and robust at handling and computing features linked in nonlinear fashions. This need can be met by using more advanced modeling techniques such as machine learning (ML).

Most applications of Artificial Intelligence (AI) in healthcare read in categorical, numerical, or image-based data as an input; utilize algorithmic and statistical models to process the data; identify patterns; and produce a probability or classification (45–49). ML refers to the range of algorithms conducting these predictions (50). As briefly stated above, ML offers considerable benefits compared to traditional statistical modeling, as it is capable of handling complex multi-dimensional data, adapting new data as it becomes available, capturing non-linear relationships and interactions among variables more effectively, and generally accounting for noise and outliers in the data in a robust manner (50, 51). Moreover, ML can promote the P4 medicine paradigm—predictive, preventive, personalized, and participatory—an approach that proactively engages both providers and patients in early monitoring and intervention (52–54). For these reasons, there has been an exponential increase in using ML to predict the prognosis and outcome of chronic diseases.

Despite their advantages, however, health-related AI models are often impermeable black boxes: their inner workings are opaque, unintuitive, and uninterpretable to end-users. A lack of interpretability can compromise the end users' trust and confidence in model predictions, especially when the model and its outcomes influence people's decisions on their health and healthcare. In response to this growing need for transparency, explicability, and interpretability in AI models, the explainable AI (XAI) has emerged as a field. Today, XAI principles are applied for multiple purposes (e.g., reducing model bias toward certain racial or gender groups), and involve providing contextual information about the importance of variables in model decision-making (55).

Several existing studies have employed ML to predict an extensive range of chronic health conditions, such as autoimmune, cardiovascular, cerebrovascular, hepatic, metabolic, neurodegenerative, pulmonary, renal, and rheumatic diseases, as well as cancers (56–61). Most of these studies used K-nearest neighbors (KNN), support vector machines (SVM), Naïve Bayes (NB), deep neural networks, random forest (RF), and logistic regression (LR) (58, 60, 62–64). Existing classical ML models in the literature have predicted health outcomes based on SDoH with accuracies between 61 and 74% (65). It is common to combine different types and sources of data for these analyses, such as electronic medical records linked to omics data (63); clinical information linked to sociodemographic, behavioral, or anthropometric factors (58); and primary care data linked to insurance claims, cancer registries, or administrative sources (64). In terms of predictors, sociodemographic (e.g., age, sex, gender) and lifestyle factors (e.g., physical activity, lack of sleep, and use of alcohol, tobacco, and other drugs) are predominantly used for modeling chronic health conditions (58). However, only a small number of studies include ACE exposure in ML models to predict rheumatic and musculoskeletal disease (66), neurocognitive outcomes (67), and emergency department visits (68). Although a study by Ammar

Abbreviations: ACE, Adverse childhood experience; AI, Artificial intelligence; AUC, Area under the curve; BMI, Body mass index; BRFSS, Behavioral Risk Factor Surveillance Survey; CDC, Centers for Disease Control and Prevention; KNN, K-nearest neighbor; LR, Logistic regression; ML, Machine learning; NB, Naïve Bayes; RF, Random forest; SDoH, Social determinants of health; SVM, Support vector machine; XAI, Explainable Artificial Intelligence.

and Shaban-Nejad (69) proposes a proof-of-concept semantic XAI model for using ACEs and SDoH data to improve mental health surveillance, the model's accuracy and usability are yet to be evaluated. Beyond these studies, few examine the use of ACEs in tandem with SDoH and health behaviors to predict a suite of chronic health conditions. Further, none of the previous studies use large national survey data to better represent the U.S. adult population.

The current study attempts to fill these gaps by developing interpretable ML models aimed at (i) predicting 13 chronic health conditions based on demographic characteristics, ACEs, SDoH, and health behaviors among U.S. adults and (ii) explaining the relative importance of variables in predicting each of the chronic health conditions. We use data from the CDC's Behavioral Risk Factor Surveillance System (BRFSS), the world's largest continuing national health survey (70). We employ classical ML models identified in the literature as robust tools for predicting chronic health conditions: LR, Gaussian NB, SVM, RF, and KNN (58, 60, 62–64). Although neural networks are also promising for this prediction task (58), they lack interpretability and demand greater computational power and time (71, 72). Computational resources are crucial during model deployment, given the higher prevalence of ACEs in disadvantaged communities that can benefit most from the models we developed (3, 37, 38, 73–76). Accordingly, we focus on classical ML models that can be scalable and adaptable, even in low-resource settings, while empowering end-users with explainable results to aid clinical decision-making.

## 2 Materials and methods

### 2.1 Data source

We utilized a subset of the latest publicly available data from the 2021 BRFSS (70). The BRFSS is a federally sponsored telephone-based survey conducted annually among U.S. adults. In 2021, the survey was conducted with 546,569 adults in all 50 states, the District of Columbia, and three territories in the U.S. The national survey collects data on SDoH, risky health behaviors, and the use of preventive services, among many other health-related factors, to facilitate health promotion efforts (70). Survey questions related to ACE exposure belong to an optional module of the BRFSS, which was implemented in 16 states in 2021 (Alabama, Arkansas, Iowa, Kansas, Maine, Mississippi, Nevada, New Hampshire, New Jersey, New York, Ohio, Oregon, South Carolina, Virginia, and Wisconsin). As ACE exposure was the study's key predictor, our final dataset was limited to the data collected by these 16 states.

### 2.2 Inclusion criteria

Our inclusion criteria were individuals who (a) resided in any of the 16 U.S. states that administered the optional ACE module of the BRFSS, (b) answered all questions about ACEs, and (c) answered at least one of the questions regarding the pre-determined 13 chronic health conditions ( $n=86,168$ ). We excluded respondents with inconclusive responses (i.e., “Do not know/Not sure,” “Not Defined,” “Not asked,” “Yes, but female told only during pregnancy,” “Refused,” or missing answers) for any predictor and outcome variables ( $n=32,900$ ). As a result, our total sample size for analysis was 52,268 respondents.

## 2.3 Measures

The study's outcome variables included 13 chronic health conditions (Supplementary Table S1). The predictor variables included self-reported ACE exposure, SDoH, health behaviors, and demographic and anthropometric characteristics (Supplementary Table S2). Please refer to Supplementary material for the answering options of each variable.

### 2.3.1 Chronic health conditions

The outcome variables included self-reported diagnoses of 13 conditions with a well-established link to ACEs: (1) arthritis (including rheumatoid arthritis or other diseases with related symptoms, such as gout, lupus, or fibromyalgia), (2) asthma, (3) cancer (any type except skin cancer), (4) coronary heart disease (or angina), (5) depressive disorder (including depression, major depression, dysthymia, or minor depression), (6) pre-diabetes, (7) diabetes, (8) heart attack, (9) high blood pressure, (10) high cholesterol, (11) kidney disease, (12) pulmonary disease (chronic obstructive pulmonary disease, emphysema, or chronic bronchitis), and (13) stroke. These outcomes were categorized by the BRFSS as “Chronic Health Conditions” (77). Our final dataset included “Yes” and “No” responses.

### 2.3.2 ACE exposure

ACE exposure was assessed with 11 questions on ACEs and two questions on positive childhood experiences (PCEs): (1) living with someone who was depressed, mentally ill, or suicidal (Yes/No); (2–3) two questions about living with someone who was a problem drinker or alcoholic or used illicit street drugs/abused prescription medications (Yes/No); (4) living someone who served time or was sentenced to serve time in prison or other correctional facility (Yes/No); (5) having parents who were separated or divorced (Yes/No/Parents Never Married); (6–8) three questions about living with parents who were physically and verbally abusive toward each other or the respondent (1 = “Never,” 2 = “Once,” 3 = “More than once”); (9–11) three questions on being sexually abused by an adult (1 = “Never,” 2 = “Once,” 3 = “More than once”); (12) the presence of an adult who made the respondent feel safe and protected; (13) the presence of an adult who ensured that the respondent's basic needs were met. Both PCEs were evaluated on a 5-point Likert scale (1 = “Never,” 2 = “A little of the time,” 3 = “Some of the time,” 4 = “Most of the time,” 5 = “All of the time”), which were reverse-coded. Additionally, we computed two composite indices for ACE exposure: a binary variable measuring whether a respondent has experienced at least one ACE (Yes/No) and a numeric variable calculating the total number of ACEs experienced (range: 0–13).

### 2.3.3 SDoH

The eight variables on SDoH included area of residence (urban vs. rural counties), education, employment status, income, renting/home ownership status, source of health insurance, availability of a personal healthcare provider, and inability to see a medical provider due to cost. These variables were categorical and had answering options unique to each question.

### 2.3.4 Health behavior

The 13 variables included both health-promoting and deteriorating behaviors, such as exercise, smoking cigarettes, chewing tobacco, using e-cigarettes or vaping, heavy drinking, time since last cholesterol check, ever tested for HIV, monitoring blood sugar or

blood pressure (two composite variables), cancer screening (two composite variables), and vaccination status (two composite variables). Like the SDoH, these variables were categorical and had differing rating scales.

We created six composite variables to handle missing data to preserve the information without dropping respondents: count of monitoring behaviors for blood sugar or blood pressure, ever monitored blood sugar or blood pressure, count of cancer screenings, ever screened for any cancer, count of vaccines received, and ever received any vaccines. The predictors for monitoring blood sugar or blood pressure were generated from two individual variables in the dataset (i.e., tested for blood sugar or diabetes in the past 3 years and regularly checked for blood pressure at home). These two variables were recoded, to where we assigned 1 (“Yes”) if the respondent checked their blood sugar or blood pressure and 0 (“No”) otherwise. The variable for the count of monitoring blood sugar or blood pressure was the sum of these binary items (range: 0–2).

Similarly, the cancer screening predictors were generated from six variables in the dataset (i.e., CT/CAT scan for lung cancer, mammogram for breast cancer, any cervical cancer screening, PSA test for prostate cancer, colonoscopy or sigmoidoscopy for colorectal cancer, and any other screening for colorectal cancer). These six variables were also re-engineered into binary variables (1 = “Yes,” 0 = “No”). The variable for the count of cancer screenings was the sum of their answers (range: 0–6). The variable measuring whether the respondent ever screened for any cancer was coded as 1 (“Yes”) if they underwent any of the six cancer screenings and 0 (“No”) if they underwent none.

Lastly, the vaccination status predictors were generated from five variables in the dataset (i.e., flu, pneumonia, tetanus, shingles, and zoster), which were re-engineered into binary variables (1 = “Yes,” 0 = “No”). The variable for the count of vaccines received was the sum of their answers (range: 0–5). The variable measuring whether the respondent ever received any vaccines was coded as 1 (“Yes”) if they received any of the five vaccines and 0 (“No”) if they received none.

### 2.3.5 Demographic and anthropometric variables

Demographic variables included age (grouped in 13 five-year categories [1 = “18–24” to 13 = “80 ≤”]), race (White, Black, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, Multiracial, Hispanic, Other), and sex (Male/Female). Body Mass Index (BMI) was the sole anthropometric variable available in the data and was assessed in four standard categories (1 = “Underweight,” 2 = “Normal Weight,” 3 = “Overweight,” and 4 = “Obese”).

## 2.4 Preprocessing

We recoded all variables (reverse coding as needed) on a 0–N scale, such that all “Never” and “No” variables were coded as zero. As noted previously, we excluded any respondents with “Do not know/Not sure,” “Refused,” “Not asked,” “Not defined,” and missing values for the outcome and predictor variables. In addition, we excluded variables for sexual orientation, transgender status, nutrition (i.e., consumption of fruits and vegetables and salt intake), and marijuana consumption in the last 30 days due to large volumes of missing data ( $n > 26,000$  or roughly 50% of our data).

Moreover, given the data imbalance in our outcome variables (i.e., the proportion of respondents without chronic health conditions

substantially exceeding their counterparts with such conditions), we performed random under-sampling of the majority class for each outcome by retaining the data for respondents with the chronic health conditions, and randomly dropping the data from the larger group without the conditions. This approach ensured equally sized classes for the outcome data, which could reduce the risk of model bias and computational burden (see [Supplementary Figure S1](#)). Relative to other sampling methods, random under-sampling is considered an effective approach to reducing data imbalance in sufficiently large datasets while minimizing the risk of generalization error on test data (78–80).

## 3 Data analysis

### 3.1 Univariate and bivariate

We conducted descriptive analyses (i.e., counts, percentages, mean, and standard deviation [SD]) for the predictor variables. Adopting Chi-square tests, we compared respondents with vs. without missing information to investigate any significant differences in their racial and income distributions and health outcomes and ultimately prevent potential biases that might be introduced into the final dataset by deleting the missing cases.

### 3.2 ML modeling

After random under-sampling, we split the data into training and test datasets. 80% of the data was allocated for training, while the remaining 20% was reserved for testing. We built a suite of supervised ML methods, such as LR, Gaussian NB, SVM, RF, and KNN, specific to each of our target chronic health conditions.

We evaluated model performance with accuracy (i.e., the rate of correct predictions) and Area under the Curve or AUC score (i.e., the probability of a model ranking a random positive observation higher than a random negative observation).

We performed hyperparameter tuning for each model on the training set to determine the most accurate predictors for each chronic health condition. Briefly, we tested a variety of optimization algorithms, penalty terms, and regularization strengths for LR; variance smoothing values for Gaussian NB; loss functions, penalty terms, and regularization strengths for SVM; the number of trees, number of features, maximum tree depth, and bootstrapping method for RF; the number of neighbors, weights, and distance metrics for KNN (see [Supplementary Table S3](#) for more details). We utilized 3-fold cross-validation and evaluated performance using validation AUC score.

### 3.3 Model interpretation

We calculated the importance of each predictor variable in predicting the occurrence of each chronic health condition using different metrics for each ML model type (81). We then examined the variable importance of the best-performing model for each chronic health condition. We performed min-max normalization on each set of variable importances, converting them to 0–1 scales. This approach allowed us to compare relative variable importance across the models. We computed variable importance for each ML model type: for LR, we referred to the coefficients of

TABLE 1 Sample characteristics.

Demographic characteristics	
<b>Age, n (%)</b>	
18–24	1,826 (3.43)
25–29	1,947 (3.66)
30–34	2,648 (4.97)
35–39	3,258 (6.12)
40–44	3,579 (6.72)
45–49	3,673 (6.9)
50–54	4,416 (8.29)
55–59	5,019 (9.42)
60–64	6,086 (11.43)
65–69	6,429 (12.07)
70–74	6,104 (11.46)
75–79	3,944 (7.4)
>80	4,339 (8.15)
<b>Race/ethnicity, n (%)</b>	
White	44,155 (82.89)
Black	4,467 (8.39)
American Indian/Alaska Native	544 (1.02)
Asian	631 (1.18)
native Hawaiian/Pacific Islander	55 (0.10)
Other	346 (0.65)
Multiracial	823 (1.55)
Hispanic	2,247 (4.22)
<b>Sex, n (%)</b>	
Male	25,226 (47.36)
Female	28,042 (52.64)
<b>Marital status, n (%)</b>	
Married	30,000 (56.32)
Divorced	6,928 (13.01)
Widowed	6,107 (11.46)
Separated	963 (1.81)
Never married	7,533 (14.14)
Unmarried couple	1,737 (3.26)
<b>Anthropometric characteristics</b>	
<b>Body mass index, n (%)</b>	
Underweight	640 (1.2)
Normal	13,933 (26.16)
Overweight	19,089 (35.84)
Obese	19,606 (36.81)
<b>Social determinants of health (SDoH)</b>	
<b>Urban or rural county, n (%)</b>	
Urban counties	43,734 (82.10)
Rural counties	9,534 (17.90)
<b>Education level, n (%)</b>	
Never attended school	21 (0.04)
Grades 1–8	572 (1.07)
Grades 9–11	1,681 (3.16)
Grades 12—GED	13,046 (24.49)
College 1–3 years	15,173 (28.48)
College >4 years	22,775 (42.76)
<b>Employment status, n (%)</b>	
Employed for wages	23,396 (43.92)
Self-employed	4,475 (8.4)

(Continued)

TABLE 1 (Continued)

Demographic characteristics	
Out of work for >1 year	1,042 (1.96)
Out of work for <1 year	868 (1.63)
Homemaker	1,715 (3.22)
Student	741 (1.39)
Retired	18,071 (33.92)
Unable to work	2,960 (5.56)
<b>Income, n (%)</b>	
< \$10,000	1,127 (2.12)
\$10,000–\$15,000	1,585 (2.98)
\$15,000–\$20,000	2,077 (3.90)
\$25,000–\$30,000	3,107 (5.83)
\$30,000–\$35,000	6,626 (12.44)
\$35,000–\$50,000	7,742 (14.53)
\$50,000–\$75,000	9,594 (18.01)
\$75,000–\$100,000	7,820 (14.68)
\$100,000–\$150,000	7,759 (14.57)
\$150,000–\$200,000	3,120 (5.86)
>\$200,000	2,711 (5.09)
<b>Rent or own home, n (%)</b>	
Own	41,687 (78.26)
Rent	10,009 (18.79)
Other arrangement	1,572 (2.95)
<b>Marital status, n (%)</b>	
Married	30,000 (56.32)
Divorced	6,928 (13.01)
Widowed	6,107 (11.46)
Separated	963 (1.81)
Never married	7,533 (14.14)
Unmarried couple	1,737 (3.26)
<b>Source of health insurance, n (%)</b>	
Employer or union plan	21,555 (40.47)
Private plan	4,342 (8.15)
Medicare	18,206 (34.18)
Medigap	59 (0.11)
Medicaid	2,500 (4.69)
Children's Health Insurance Program (CHIP)	12 (0.02)
Military-related healthcare	1,845 (3.46)
Indian Health Service	60 (0.11)
State-sponsored health plan	1,136 (2.13)
Other government program	1,442 (2.71)
No coverage	2,111 (3.96)
<b>Has personal care provider, n (%)</b>	
Yes, only one	34,089 (64.0)
More than one	14,397 (27.03)
No	4,782 (8.98)
<b>Unable to see doctor due to medical cost, n (%)</b>	
Yes	3,180 (5.97)
No	50,088 (94.03)
<b>Health behavior</b>	
<b>Regular exercise, n (%)</b>	
Yes	40,412 (75.87)
No	12,856 (24.13)
<b>Smoked at least 100 cigarettes in life, n (%)</b>	

(Continued)

TABLE 1 (Continued)

Demographic characteristics	
Yes	22,740 (42.69)
No	30,528 (57.31)
Currently use chewing tobacco, snuff or snuss, <i>n</i> (%)	
Every day	1,066 (2.0)
Some days	707 (1.33)
Not at all	51,495 (96.67)
Use-cigarettes or electronic vaping products, <i>n</i> (%)	
Every day	1,056 (1.98)
Some days	1,125 (2.11)
Not at all	43,338 (81.36)
Never used	7,749 (14.55)
Heavy drinker, <i>n</i> (%)	
No	49,893 (93.66)
Yes	3,375 (6.34)
Time since last cholesterol check, <i>n</i> (%)	
Never	3,739 (7.02)
<1 Year	39,096 (73.39)
1–2 Years	5,776 (10.84)
2–3 Years	1,803 (3.38)
3–4 Years	651 (1.22)
4–5 Years	580 (1.09)
>5 Years	1,623 (3.05)
Ever tested for HIV, <i>n</i> (%)	
Yes	17,195 (32.28)
No	36,073 (67.72)
Count of monitoring behaviors for blood sugar and blood pressure, <i>n</i> (%)	
0	33,238 (62.4)
1	18,837 (35.36)
2	1,193 (2.24)
Count of cancer screenings, <i>n</i> (%)	
0	49,771 (93.44)
1	2,354 (4.42)
2	1,095 (2.06)
3	48 (0.09)
Count of vaccines received, <i>n</i> (%)	
0	16,975 (31.87)
1	18,573 (34.87)
2	15,991 (30.02)
3	1,050 (1.97)
4	679 (1.27)
ACE exposure	
Four or more ACEs, <i>n</i> (%)	9,808 (18.41)
Total ACEs, mean (sd)	1.83 (2.27)

the predictor variables in the regression formulation (80); for Gaussian NB, we employed permutation importance that measures the decline in model performance when individual random variables are shuffled (82); for SVM, we calculated the weight vector that represents the hyperplane separating the classes in linear space (83); for RF, we examined GINI importance or mean decrease in impurity, indicating how often a specific feature is selected for splitting within the RF and, thereby, its discriminative value toward the classification (84). We performed all procedures using Python 3.8.3 run on Jupyter Notebook. We used several

open-source Python packages: *numpy*, *pandas*, *matplotlib*, *sci-kit learn*, *seaborn*, and *scipy*.

## 4 Results

### 4.1 Sample characteristics

As illustrated in Table 1, 39% of the respondents were aged 65 or older. About 83% of them self-identified as White and 8.4% as



TABLE 2 Classification model performance for all target chronic health conditions.

Chronic disease	ML model	Validation accuracy [95% CI]	Test accuracy [95% CI]	Test AUC score [95% CI]
Arthritis <i>n</i> = 39,184	LR	0.696 [0.691, 0.7]	0.697 [0.693, 0.702]	0.697 [0.695, 0.7]
	NB	0.682 [0.678, 0.687]	0.688 [0.683, 0.693]	0.688 [0.685, 0.691]
	SVM	0.691 [0.686, 0.696]	0.697 [0.692, 0.701]	0.697 [0.694, 0.699]
	KNN	0.667 [0.662, 0.671]	0.677 [0.673, 0.682]	0.677 [0.675, 0.68]
	<b>RF (Best Predictor)</b>	<b>0.697 [0.693, 0.702]</b>	<b>0.701 [0.697, 0.706]</b>	<b>0.701 [0.699, 0.704]</b>
Asthma <i>n</i> = 14,268	LR	0.615 [0.607, 0.623]	0.615 [0.607, 0.623]	0.615 [0.61, 0.619]
	NB	0.572 [0.564, 0.58]	0.578 [0.57, 0.587]	0.578 [0.574, 0.583]
	SVM	0.615 [0.607, 0.623]	0.616 [0.608, 0.624]	0.616 [0.612, 0.621]
	KNN	0.569 [0.56, 0.577]	0.571 [0.563, 0.58]	0.571 [0.567, 0.576]
	<b>RF (Best Predictor)</b>	<b>0.61 [0.602, 0.618]</b>	<b>0.627 [0.619, 0.635]</b>	<b>0.627 [0.622, 0.631]</b>
Cancer <i>n</i> = 11,726	LR	0.675 [0.667, 0.684]	0.668 [0.66, 0.677]	0.668 [0.664, 0.673]
	NB	0.652 [0.644, 0.661]	0.661 [0.653, 0.67]	0.661 [0.656, 0.666]
	SVM	0.673 [0.665, 0.682]	0.671 [0.663, 0.68]	0.671 [0.667, 0.676]
	KNN	0.647 [0.639, 0.656]	0.659 [0.65, 0.668]	0.659 [0.654, 0.664]
	<b>RF (Best Predictor)</b>	<b>0.675 [0.666, 0.683]</b>	<b>0.687 [0.678, 0.695]</b>	<b>0.687 [0.682, 0.691]</b>
Coronary heart disease <i>n</i> = 6,554	LR	0.732 [0.721, 0.743]	0.715 [0.705, 0.726]	0.716 [0.709, 0.722]
	NB	0.704 [0.693, 0.715]	0.694 [0.683, 0.705]	0.694 [0.688, 0.701]
	<b>SVM (Best Predictor)</b>	<b>0.732 [0.721, 0.742]</b>	<b>0.725 [0.715, 0.736]</b>	<b>0.725 [0.719, 0.732]</b>
	KNN	0.7 [0.689, 0.711]	0.678 [0.667, 0.689]	0.678 [0.672, 0.685]
	RF	0.734 [0.724, 0.745]	0.719 [0.708, 0.729]	0.719 [0.712, 0.725]
Depressive disorder <i>n</i> = 21,288	<b>LR (Best Predictor)</b>	<b>0.708 [0.702, 0.714]</b>	<b>0.705 [0.699, 0.711]</b>	<b>0.705 [0.702, 0.709]</b>
	NB	0.648 [0.642, 0.655]	0.642 [0.636, 0.649]	0.642 [0.639, 0.646]
	SVM	0.705 [0.699, 0.711]	0.7 [0.694, 0.706]	0.7 [0.697, 0.704]
	KNN	0.665 [0.659, 0.671]	0.665 [0.659, 0.672]	0.665 [0.662, 0.669]
	RF	0.709 [0.703, 0.715]	0.702 [0.696, 0.709]	0.702 [0.699, 0.706]
Diabetes <i>n</i> = 15,504	LR	0.784 [0.778, 0.791]	0.772 [0.765, 0.778]	0.772 [0.768, 0.775]
	NB	0.764 [0.758, 0.771]	0.751 [0.744, 0.758]	0.751 [0.747, 0.755]
	SVM	0.783 [0.777, 0.79]	0.774 [0.768, 0.781]	0.774 [0.771, 0.778]
	KNN	0.737 [0.73, 0.744]	0.729 [0.722, 0.736]	0.729 [0.725, 0.733]
	<b>RF (Best Predictor)</b>	<b>0.79 [0.784, 0.797]</b>	<b>0.784 [0.778, 0.791]</b>	<b>0.784 [0.781, 0.788]</b>
Heart attack <i>n</i> = 6,236	LR	0.724 [0.713, 0.735]	0.725 [0.714, 0.736]	0.725 [0.719, 0.731]
	NB	0.695 [0.684, 0.707]	0.698 [0.687, 0.709]	0.698 [0.691, 0.704]
	SVM	0.722 [0.711, 0.733]	0.725 [0.714, 0.736]	0.725 [0.719, 0.731]
	KNN	0.689 [0.678, 0.701]	0.691 [0.679, 0.702]	0.691 [0.684, 0.697]
	<b>RF (Best Predictor)</b>	<b>0.728 [0.717, 0.739]</b>	<b>0.732 [0.721, 0.743]</b>	<b>0.732 [0.726, 0.739]</b>
High blood pressure <i>n</i> = 45,976	LR	0.707 [0.703, 0.711]	0.712 [0.708, 0.717]	0.712 [0.71, 0.715]
	NB	0.669 [0.665, 0.673]	0.669 [0.665, 0.673]	0.669 [0.667, 0.672]
	SVM	0.707 [0.703, 0.711]	0.711 [0.707, 0.715]	0.711 [0.709, 0.714]
	KNN	0.676 [0.672, 0.68]	0.679 [0.674, 0.683]	0.679 [0.676, 0.681]
	<b>RF (Best Predictor)</b>	<b>0.713 [0.709, 0.717]</b>	<b>0.716 [0.711, 0.72]</b>	<b>0.716 [0.713, 0.718]</b>
High cholesterol <i>n</i> = 49,526	LR	0.656 [0.652, 0.66]	0.657 [0.653, 0.662]	0.657 [0.655, 0.66]
	NB	0.612 [0.607, 0.616]	0.613 [0.608, 0.617]	0.613 [0.61, 0.615]
	SVM	0.66 [0.656, 0.664]	0.661 [0.657, 0.665]	0.661 [0.658, 0.663]
	KNN	0.609 [0.605, 0.613]	0.612 [0.608, 0.617]	0.612 [0.61, 0.615]
	<b>RF (Best Predictor)</b>	<b>0.67 [0.666, 0.675]</b>	<b>0.671 [0.667, 0.675]</b>	<b>0.671 [0.668, 0.673]</b>

(Continued)

TABLE 2 (Continued)

Chronic disease	ML model	Validation accuracy [95% CI]	Test accuracy [95% CI]	Test AUC score [95% CI]
Kidney disease <i>n</i> = 4,432	LR	0.692 [0.678, 0.706]	0.677 [0.663, 0.69]	0.677 [0.669, 0.685]
	NB	0.679 [0.665, 0.692]	0.653 [0.639, 0.667]	0.653 [0.645, 0.661]
	SVM	0.691 [0.678, 0.705]	0.669 [0.655, 0.683]	0.669 [0.661, 0.677]
	KNN	0.674 [0.66, 0.688]	0.633 [0.619, 0.647]	0.633 [0.625, 0.641]
	<b>RF (Best Predictor)</b>	<b>0.698 [0.684, 0.711]</b>	<b>0.681 [0.667, 0.695]</b>	<b>0.681 [0.673, 0.689]</b>
Pre-diabetes <i>n</i> = 51,060	LR	0.711 [0.707, 0.715]	0.717 [0.714, 0.721]	0.714 [0.712, 0.717]
	NB	0.694 [0.69, 0.698]	0.701 [0.697, 0.705]	0.696 [0.694, 0.699]
	SVM	0.708 [0.704, 0.712]	0.715 [0.711, 0.719]	0.711 [0.709, 0.713]
	KNN	0.671 [0.667, 0.675]	0.679 [0.675, 0.684]	0.674 [0.671, 0.676]
	<b>RF (Best Predictor)</b>	<b>0.724 [0.72, 0.728]</b>	<b>0.728 [0.724, 0.732]</b>	<b>0.726 [0.724, 0.728]</b>
Pulmonary disease <i>n</i> = 8,890	<b>LR (Best Predictor)</b>	<b>0.745 [0.736, 0.754]</b>	<b>0.753 [0.744, 0.762]</b>	<b>0.753 [0.748, 0.758]</b>
	NB	0.711 [0.701, 0.72]	0.715 [0.706, 0.725]	0.715 [0.71, 0.721]
	SVM	0.745 [0.736, 0.754]	0.749 [0.74, 0.758]	0.749 [0.744, 0.754]
	KNN	0.706 [0.697, 0.716]	0.711 [0.701, 0.72]	0.711 [0.706, 0.716]
	RF	0.756 [0.747, 0.765]	0.744 [0.734, 0.753]	0.744 [0.738, 0.749]
Stroke <i>n</i> = 4,488	LR	0.714 [0.701, 0.728]	0.714 [0.701, 0.727]	0.714 [0.706, 0.721]
	NB	0.706 [0.693, 0.72]	0.688 [0.675, 0.702]	0.688 [0.68, 0.696]
	SVM	0.712 [0.698, 0.725]	0.71 [0.697, 0.724]	0.71 [0.703, 0.718]
	KNN	0.698 [0.685, 0.712]	0.682 [0.668, 0.695]	0.682 [0.674, 0.689]
	<b>RF (Best Predictor)</b>	<b>0.718 [0.705, 0.731]</b>	<b>0.715 [0.702, 0.728]</b>	<b>0.715 [0.707, 0.722]</b>

Bold values indicate best-performing models for each chronic disease.

Black. Slightly over 50% were female and married, respectively. Over 40% of the respondents completed 4 years of college education or more and were employed, respectively, while 78% owned a home.

In terms of BMI, 35.8% of the respondents were overweight, and 36.8% were obese. For healthcare access, 64% reported having a personal provider, while 40.5% reported having an employer or union-sponsored insurance. Nevertheless, 94% reported that they could not see a doctor in the past 12 months due to cost. Regarding health behaviors, a majority of the respondents reported exercising in the past month (75.9%) and never using chewing tobacco (96.7%) and electronic cigarettes/vaping products (81.4%). Also, 57.3% had smoked less than 100 cigarettes in their lifetime, and around 6% were involved in heavy drinking.

73.4% of the respondents checked their cholesterol last time less than a year ago. On the other hand, a majority reported never having been tested for HIV (67.7%), not monitoring blood sugar or blood pressure (62.4%), and not screening for cancer (93.4%). Nearly one in three respondents received at least one vaccine. Lastly, the mean number of ACEs was 1.83 (SD = 2.27), and 18.4% of the respondents encountered four or more ACEs.

## 4.2 Analysis of missing data

There was no significant difference in the racial distribution of the missing and non-missing cases (data not shown). However, we found a significant difference in the income distribution between the two

groups, wherein respondents with missing data were more likely to be in a higher-income group earning \$75,000 or more. Regarding chronic health conditions, we found significant differences only in high blood pressure and arthritis, whereby those with missing data were more likely to experience these conditions. However, we do not expect the removal of missing data on high blood pressure and arthritis to impact model performance, as we performed under-sampling to ensure balanced distributions of classes for each outcome variable.

## 4.3 Model performance

With the inclusion of data on ACEs, our ML models achieved higher or similar accuracy and AUC scores compared to existing models in the literature that predicted health outcomes based on SDoH (65) (Table 2). Nine of the 13 models obtained test accuracies above 70% and test AUC scores above 0.7. The top-performing models were those predicting diabetes (78.4% accuracy, 0.784 AUC), pulmonary disease (75.3% accuracy, 0.753 AUC), and heart attack (73.2% accuracy, 0.732 AUC).

Training a single iteration of each model took an average of 38 s. Validation and model selection involved training a single iteration of each algorithm for every combination of the hyperparameters that were tested. This process determined the optimal performance for each model.

Three of the top five models employed RF (diabetes, heart attack, and prediabetes), whereas LR (pulmonary disease) and SVM

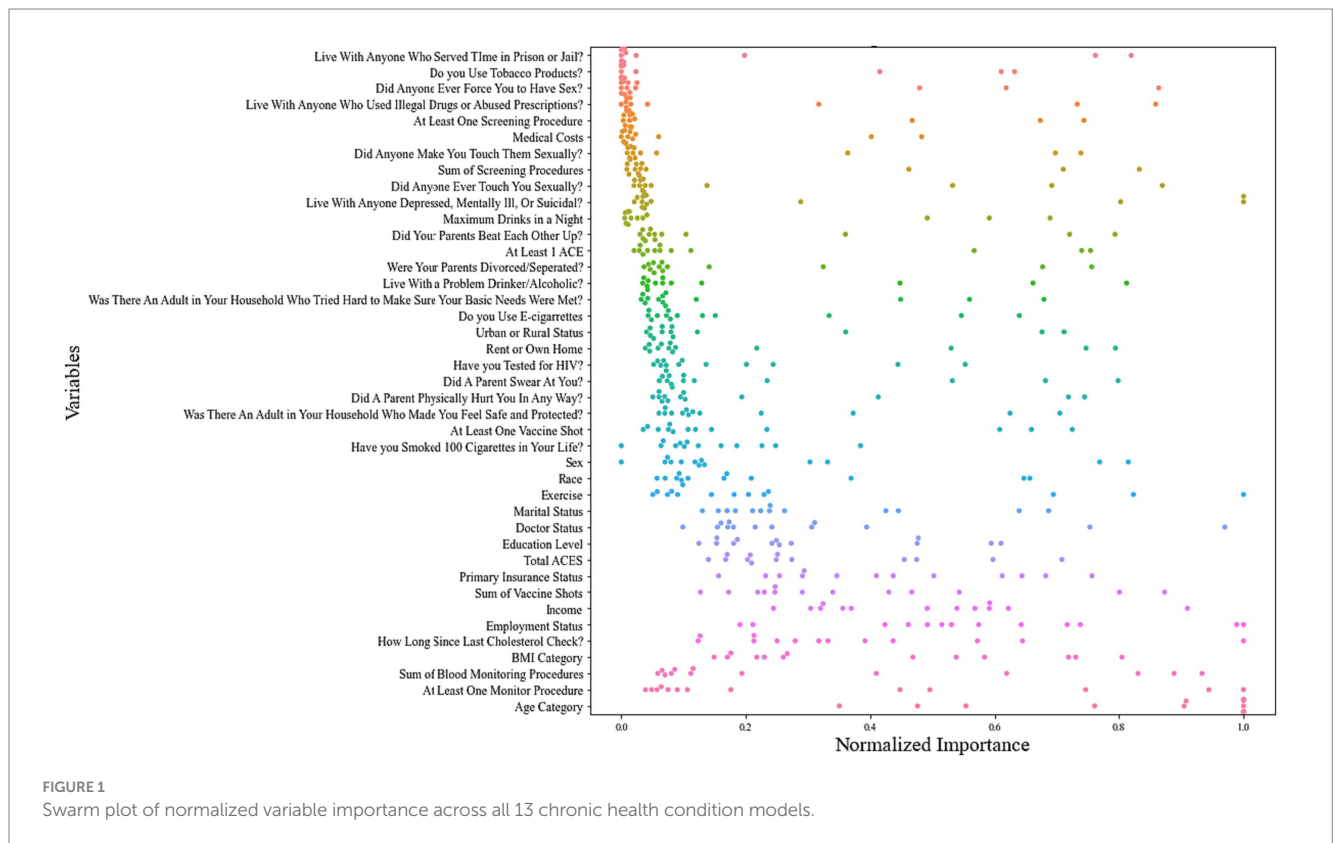
TABLE 3 Most predictive variables for each best-performing classification model.

Disease	Best performing model	Top 5 predictor variables
Arthritis	Random forest	1. Age
		2. Employment status
		3. Income
		4. Source of health insurance
		5. Count of vaccines received
Asthma	Random forest	1. Age
		2. Income
		3. Total ACEs
		4. BMI category
		5. Employment status
Cancer	Random forest	1. Age
		2. Employment status
		3. Source of health insurance
		4. Income
		5. Count of vaccines received
Coronary heart disease	Support vector machine	1. Sex
		2. ACE—lived with mentally ill/suicidal person
		3. Has personal health care provider available
		4. Smoking cigarettes
		5. Age
Depressive disorder	Logistic regression	1. ACE—lived with mentally ill/suicidal person
		2. Inability to see medical provider due to cost
		3. Sex
		4. Exercise
		5. Ever received any vaccines
Diabetes	Random forest	1. Age
		2. Ever monitored blood sugar or blood pressure
		3. Count of blood monitoring behaviors for blood sugar or blood pressure
		4. BMI
		5. Time since last cholesterol check
Heart attack	Random forest	1. Age
		2. Employment status
		3. Income
		4. Source of health insurance
		5. Sex
High blood pressure	Random forest	1. Age
		2. BMI
		3. Employment status
		4. Time since last cholesterol check
		5. Income
High cholesterol	Random forest	1. Time since last cholesterol check
		2. Age
		3. Income
		4. Employment status
		5. Source of health insurance

(Continued)

TABLE 3 (Continued)

Disease	Best performing model	Top 5 predictor variables
Kidney disease	Random forest	1. Employment status
		2. Age
		3. Source of health insurance
		4. Income
		5. Count of vaccines received
Pre-diabetes	Random forest	1. Ever monitored blood sugar or blood pressure
		2. Count of blood monitoring behaviors for blood sugar or blood pressure
		3. Age
		4. Income
		5. Time since last cholesterol check
Pulmonary disease	Logistic regression	1. Smoking cigarettes
		2. Exercise
		3. Inability to see medical provider due to cost
		4. Have you tested for HIV?
		5. ACE—lived with mentally ill/suicidal person
Stroke	Random forest	1. Age
		2. Employment status
		3. Income
		4. Source of health insurance
		5. Time since last cholesterol check



(coronary heart disease) were used in the other two. Overall, RF performed best for 10 of the 13 chronic health conditions: diabetes, heart attack, prediabetes, high blood pressure, stroke, arthritis, cancer, kidney disease, high cholesterol, and asthma. The linear model (i.e., LR) performed best only for two chronic health conditions.

## 4.4 Model interpretation

Age and SDoH, such as income, employment, and health insurance, were among the top five strongest variables to predict each chronic health condition (Table 3). ACEs, either cumulatively or individually, were also identified as an important variable for asthma, coronary heart disease, depressive disorder, and pulmonary disease. When individually examined, living with a mentally ill/suicidal person during childhood was the only ACE predictive of these health conditions (except asthma). Specifically, living with a mentally ill/suicidal person seemed to play the most critical role in the depressive disorder and coronary heart disease models and was listed as their first and second most important predictor, respectively. Supplementary Figure S2 outlines the variable importance of all models.

Normalized variable importance revealed the 10 most predictive variables across a total of 65 models (5 ML models  $\times$  13 chronic health conditions): age, ever monitored blood sugar or blood pressure, count of monitoring behaviors for blood sugar or blood pressure, BMI, time since last cholesterol check, employment status, income, count of vaccines received, primary insurance status, and the total number of ACEs (Figure 1).

## 5 Discussion

Our study developed explainable ML models using large national survey data to predict 13 chronic health conditions prevalent among U.S. adults. We found that non-linear models, particularly RF, outperformed the linear model in predicting chronic health conditions. In addition, our ML models cast light on the most predictive features of each condition. Among these, ACEs and SDoH such as income, employment, and health insurance, were robust predictors of multiple chronic health conditions. Additionally, cumulative ACEs were a stronger predictor than individual ACEs across chronic health conditions. Our models achieved comparable or superior performance to classical ML-based health outcome prediction models that previously used SDoH as predictors (65). Our findings not only align with previous studies linking ACEs to chronic health conditions (6–28, 30, 31), but also expand upon them by employing ML to factor in complex interactions between ACEs and other socioeconomic and behavioral factors to predict chronic health conditions. Our primary focus on ACEs and relevant socioeconomic and behavioral factors can distinguish the current study from others. While previous studies have documented excellent performance of classical ML models (e.g., RF, gradient boost, SVM, LR, KNN, decision trees, and NB) to predict chronic health conditions, they commonly focused on biomedical predictors such as clinical, biomarker, and genetics data (58, 60, 63, 64).

Our study, which emphasizes the role of ACEs and their cumulative impact, highlights the significance of predictive values of total ACEs in shaping chronic diseases. ACEs were among the top five predictors for four chronic health conditions: asthma, coronary heart disease, depressive disorder, and pulmonary disease. Living with a mentally ill/suicidal person during childhood was particularly predictive of coronary heart disease, depressive disorder, and pulmonary disease. These results are supported by Gallagher and colleagues, who found that living with a severely mentally ill person is associated with poorer subjective health, activity limitations, and higher utilization of physician visits than living with non-mentally ill household members (85). Beyond this single ACE, the total number of ACEs was a stronger predictor than individual ACEs across all the best-performing models of the 13 chronic health conditions. Notably, the total number of ACEs was among the top five predictors for the asthma model, which aligns with findings from the existing literature on the dose–response relationship between ACEs and asthma (12). The composite measure may more accurately represent how ACEs operate: not arbitrarily, but rather in clusters, especially among historically marginalized populations (2, 86–88). This finding underscores the significance of cumulative ACEs on an individual's likelihood of developing chronic health conditions. Although we demonstrated a strong association between ACEs and chronic health conditions by comparing various base learners, including LR, Gaussian NB, SVM, RF, and KNN, future work is guaranteed to improve prediction accuracy. For example, we may employ stacked ensemble algorithms (e.g., XGBoost), which has been reported to improve classification with imbalanced data (89–91); this may enhance performance while requiring smaller degrees of undersampling, thereby allowing the use of a larger volume of data. Additionally, we may perform more extensive iterations of training and validation using a wider range of hyperparameters.

On a relative scale, the models for diabetes, pulmonary disease, and heart attacks performed particularly well, whereas models predicting kidney disease, high cholesterol, and asthma exhibited lower performance. Such discrepancies may be attributable to the varying importance of different variables in predicting distinct chronic health conditions. Similarly, Battineni and colleagues report that their ML models used different sets of variables to predict various chronic diseases in different populations, demonstrating no “gold standard” for ML methods to predict chronic diseases, including how to select and prioritize predictors (56). Despite improved interpretability, this unclarity could still compromise ML models' transparency and trustworthiness. To partially address the issue, future research could compare different sets of predictors across domains, ML models, and strategies for interpretability to analyze the commonalities and variations in model output.

In addition to the ACEs discussed above, the following were the most predictive variables across all models of chronic health conditions: age, ever monitored blood sugar or blood pressure, count of monitoring behaviors for blood sugar or blood pressure, BMI, time since last cholesterol check, employment status, income, vaccine count, and primary health insurance status. Previous literature has revealed that chronic health conditions are indeed associated with age (92); self-management (93, 94); BMI (95); employment, income, and wealth (96–98); immunization (99–103); health insurance (104, 105).

Our study findings undergird the pivotal role of preventing ACEs and socioeconomic inequalities in chronic disease prevention at the

population level. Our ML models could enable data-driven screening for various chronic health conditions to identify high-risk individuals, explain the most influential underlying factors, and develop personalized prevention strategies.

Despite the strengths and contributions of our study, some limitations must be acknowledged. First, we analyzed self-reported data, which could have introduced biases (e.g., recall bias, social desirability bias, or misinterpretation of the questions), potentially affecting the accuracy and reliability of the developed models. However, such reporting biases are inherent in survey data and not unique to the BRFSS. In addition, the prevalence estimates in the BRFSS data are known to be consistent with comparable national surveys (i.e., National Health Interview Survey, National Health and Nutrition Examination Survey) (106, 107). More objective measures, such as biomarkers, should be analyzed to predict chronic health conditions more accurately in the future.

Second, our final dataset comprised mostly White and middle-income respondents. Consequently, the developed models may not predict chronic health conditions among disadvantaged populations at higher risk of ACEs (e.g., Black, Hispanic, or low-income individuals) as accurately as among more privileged populations (e.g., White or affluent individuals). Future studies are needed to develop ML models optimized for subpopulations, compare their performance to models with a pooled population, and consider potential differences in important variables or magnitudes in prediction. Stratification by subpopulation could partially mitigate the system-wide bias in collecting and processing data among different populations.

Third, our random sampling method to create an artificially balanced dataset for model training may misrepresent model performance. Random under-sampling increases the possibility that the model underperforms with “real-world” data, as the inflated proportion of positive cases in the training data may introduce greater false positives in real-world data. However, relative to other sampling methods, random under-sampling minimizes the risk of generalization error on test data (78–80).

Fourth, we encountered some hurdles with data availability. For instance, there were no core questions in the BRFSS regarding transportation, food security, and green space, which are crucial SDoH. Relatedly, other variables that represent determinants of health were not factored into our models due to insufficient data: sexual orientation, transgender status, nutrition, and marijuana consumption. Furthermore, we were unable to predict specific types of cancer, joint conditions, and pulmonary disease due to unavailable data.

Lastly, our ML models were trained and tested on unweighted data due to a lack of computing resources to model the weighted data. Hence, our unweighted ML models are limited in generalizability, and their performance is likely inflated to some degree compared to weighted models (108). With these limitations in data collection and modeling, our findings should be interpreted with caution. Our models should be viewed as supplementary tools for screening and decision-making, rather than a standalone, definitive prediction system for chronic health conditions.

## 6 Conclusion

To our knowledge, this is the first study to employ interpretable ML methods to model the syndemic interactions of ACEs, SDoH, health behaviors, and chronic health conditions using extensive data

from a large national health survey in the U.S. Our findings highlighted the significance of preventing ACEs and mitigating their cumulative impact on chronic health conditions later in life. This study serves as an initial step toward developing a data-driven screening tool to identify U.S. adults at high risk of chronic health conditions, aiding in prevention and early intervention efforts. Our models also offer an interpretable and trauma-informed framework, aimed at reducing the persistent inequalities associated with early trauma and chronic health conditions among U.S. adults. Acknowledging the insights from Battineni et al., we underscore the importance of continuous validation and testing of our models to ensure their reliability and practical utility in multiple settings with different patient characteristics. ML models are bound to the data they train; therefore, the model parameters we have developed can be used as a baseline, upon which future research can develop contextualized models that will be re-fitted to other datasets of new patient populations to predict their chronic health conditions more accurately.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: [https://www.cdc.gov/brfss/annual\\_data/annual\\_2021.html](https://www.cdc.gov/brfss/annual_data/annual_2021.html).

## Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

## Author contributions

HA: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft. TJ: Conceptualization, Data curation, Methodology, Project administration, Software, Writing – original draft. YM: Visualization, Writing – original draft. AM: Visualization, Writing – original draft. AS: Validation, Writing – review & editing. SK: Supervision, Validation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Institute of Nursing Research of the National Institutes of Health under Award Number K01NR019651.

## Acknowledgments

We thank Juan Rodriguez and Devin Lucas for their feedback during the initial stages of the project.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Centers for Disease Control and Prevention. Fast facts: Preventing adverse childhood experiences [internet]. (2023). Available at: <https://www.cdc.gov/violenceprevention/aces/fastfact.html> (Accessed August 6, 2023).
- Goldstein E, Topitzes J, Miller-Cribbs J, Brown RL. Influence of race/ethnicity and income on the link between adverse childhood experiences and child flourishing. *Pediatr Res.* (2021) 89:1861–9. doi: 10.1038/s41390-020-01188-6
- Skindzielewski K, Forke CM, Sarwer DB, Noll JG, Wheeler DC, Henry KA, et al. The intersection of adverse childhood experiences and neighborhood determinants of health: an exploratory spatial analysis. *Psychol Trauma Theory Res Pract Policy.* (2022) 1–8. doi: 10.1037/tra0001320
- Centers for Disease Control and Prevention. About chronic diseases [internet]. (2022). Available at: <https://www.cdc.gov/chronicdisease/about/index.htm> (Accessed August 6, 2023).
- Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. *Am J Prev Med.* (1998) 14:245–58. doi: 10.1016/S0749-3797(98)00017-8
- Rubinstein TB, Bullock DR, Ardalan K, Mowrey WB, Brown NM, Bauman LJ, et al. Adverse childhood experiences are associated with childhood-onset arthritis in a National Sample of US youth: an analysis of the 2016 National Survey of Children's health. *J Pediatr.* (2020) 226:243–250.e2. doi: 10.1016/j.jpeds.2020.06.046
- Luiz APL, Antico H d A, Skare TL, Boldt ABW, Nisihara R. Adverse childhood experience and rheumatic diseases. *Clin Rheumatol.* (2018) 37:2863–7. doi: 10.1007/s10067-018-4200-5
- Sonagra M, Jones J, McGill M, Gmuca S. Exploring the intersection of adverse childhood experiences, pediatric chronic pain, and rheumatic disease. *Pediatr Rheumatol.* (2022) 20:14. doi: 10.1186/s12969-022-00674-x
- Baiden P, Panisch LS, Onyeaka HK, LaBrenz CA, Kim Y. Association of childhood physical and sexual abuse with arthritis in adulthood: findings from a population-based study. *Prev Med Rep.* (2021) 23:101463. doi: 10.1016/j.pmedr.2021.101463
- Ospina MB, Serrano-Lomelin JA, Amjad S, Hicks A, Giesbrecht GF. Latent factors of adverse childhood experiences and adult-onset asthma. *J Dev Orig Health Dis.* (2021) 12:50–7. doi: 10.1017/S2040174419000886
- Pape K, Cowell W, Sejbaek CS, Andersson NW, Svanes C, Kolstad HA, et al. Adverse childhood experiences and asthma: trajectories in a national cohort. *Thorax.* (2021) 76:547–53. doi: 10.1136/thoraxjnl-2020-214528
- Exley D, Norman A, Hyland M. Adverse childhood experience and asthma onset: a systematic review. *Eur Respir Rev.* (2015) 24:299–305. doi: 10.1183/16000617.00004114
- Lietzén R, Suominen S, Sillanmäki L, Virtanen P, Virtanen M, Vahtera J. Multiple adverse childhood experiences and asthma onset in adulthood: role of adulthood risk factors as mediators. *J Psychosom Res.* (2021) 143:110388. doi: 10.1016/j.jpsychores.2021.110388
- Holman DM, Ports KA, Buchanan ND, Hawkins NA, Merrick MT, Metzler M, et al. The association between adverse childhood experiences and risk of Cancer in adulthood: a systematic review of the literature. *Pediatrics.* (2016) 138:S81–91. doi: 10.1542/peds.2015-4268L
- Ports KA, Holman DM, Guinn AS, Pampati S, Dyer KE, Merrick MT, et al. Adverse childhood experiences and the presence of Cancer risk factors in adulthood: a scoping review of the literature from 2005 to 2015. *J Pediatr Nurs.* (2019) 44:81–96. doi: 10.1016/j.pedn.2018.10.009

## Author disclaimer

This work was prepared under Contract HT0014-19-C-0004 with DHA Contracting Office (CO-NCR) HT0014 and, therefore, is defined as U.S. Government work under Title 17 U.S.C.§101. Per Title 17 U.S.C.§105, copyright protection is not available for any work of the U.S. Government. For more information, please contact [dha.tbicoeinfo@health.mil](mailto:dha.tbicoeinfo@health.mil).

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1309490/full#supplementary-material>

- Hu Z, Kaminga AC, Yang J, Liu J, Xu H. Adverse childhood experiences and risk of cancer during adulthood: a systematic review and meta-analysis. *Child Abuse Negl.* (2021) 117:105088. doi: 10.1016/j.chiabu.2021.105088
- Kelly-Irving M, Lepage B, Dedieu D, Lacey R, Cable N, Bartley M, et al. Childhood adversity as a risk for cancer: findings from the 1958 British birth cohort study. *BMC Public Health.* (2013) 13:767. doi: 10.1186/1471-2458-13-767
- Su S, Wang X, Pollock JS, Treiber FA, Xu X, Snieder H, et al. Adverse childhood experiences and blood pressure trajectories from childhood to young adulthood: the Georgia stress and heart study. *Circulation.* (2015) 131:1674–81. doi: 10.1161/CIRCULATIONAHA.114.013104
- Rodriguez-Miguelez P, Looney J, Blackburn M, Thomas J, Pollock JS, Harris RA. The link between childhood adversity and cardiovascular disease risk: role of cerebral and systemic vasculature. *Function.* (2022) 3:zqac029. doi: 10.1093/function/zqac029
- Godoy LC, Frankfurter C, Cooper M, Lay C, Maunder R, Farkouh ME. Association of Adverse Childhood Experiences with Cardiovascular Disease Later in life: a review. *JAMA Cardiol.* (2021) 6:228–35. doi: 10.1001/jamacardio.2020.6050
- Campbell JA, Mendez CE, Garacci E, Walker RJ, Wagner N, Egede LE. The differential impact of adverse childhood experiences in the development of pre-diabetes in a longitudinal cohort of US adults. *J Diabetes Complicat.* (2018) 32:1018–24. doi: 10.1016/j.jdiacomp.2018.09.006
- Zhang Y, Yin Y, Zhang X, Ye J, Zhang J. Association of adverse childhood experiences with diabetes: a systematic review and meta-analysis. *J Diabetes Complicat.* (2022) 36:108289. doi: 10.1016/j.jdiacomp.2022.108289
- Kellum CE, Kemp KM, Mrug S, Pollock JS, Seifert ME, Feig DI. Adverse childhood experiences are associated with vascular changes in adolescents that are risk factors for future cardiovascular disease. *Pediatr Nephrol.* (2023) 38:2155–63. doi: 10.1007/s00467-022-05853-2
- Allen H, Wright BJ, Vartanian K, Dulacki K, Li HF. Examining the prevalence of adverse childhood experiences and associated cardiovascular disease risk factors among low-income uninsured adults. *Circ Cardiovasc Qual Outcomes.* (2019) 12:e004391. doi: 10.1161/CIRCOUTCOMES.117.004391
- O'Leary E, Millar SR, Perry IJ, Phillips CM. Association of adverse childhood experiences with lipid profiles and atherogenic risk indices in a middle-to-older aged population. *SSM Popul Health.* (2023) 22:101393. doi: 10.1016/j.ssmph.2023.101393
- Ozieh MN, Garacci E, Campbell JA, Walker RJ, Egede LE. Adverse childhood experiences and decreased renal function: impact on all-cause mortality in U.S. adults. *Am J Prev Med.* (2020) 59:e49–57. doi: 10.1016/j.amepre.2020.04.005
- Anda RF, Brown DW, Dube SR, Bremner JD, Felitti VJ, Giles WH. Adverse childhood experiences and chronic obstructive pulmonary disease in adults. *Am J Prev Med.* (2008) 34:396–403. doi: 10.1016/j.amepre.2008.02.002
- Lopes S, Hallak JEC, Machado de Sousa JP, Osório FLL. Adverse childhood experiences and chronic lung diseases in adulthood: a systematic review and meta-analysis. *Eur J Psychotraumatol.* (2020) 11:1720336. doi: 10.1080/20008198.2020.1720336
- Benjet C, Borges G, Medina-Mora ME. Chronic childhood adversity and onset of psychopathology during three life stages: childhood, adolescence and adulthood. *J Psychiatr Res.* (2010) 44:732–40. doi: 10.1016/j.jpsychires.2010.01.004
- Hughes K, Lowey H, Quigg Z, Bellis MA. Relationships between adverse childhood experiences and adult mental well-being: results from an English national household survey. *BMC Public Health.* (2016) 16:222. doi: 10.1186/s12889-016-2906-3
- van Duin L, Bevaart F, Zijlmans J, Luijckx MJA, Doreleijers TAH, Wiersma AI, et al. The role of adverse childhood experiences and mental health care use in

- psychological dysfunction of male multi-problem young adults. *Eur Child Adolesc Psychiatry*. (2019) 28:1065–78. doi: 10.1007/s00787-018-1263-4
32. Nelson CA, Bhutta ZA, Harris NB, Danese A, Samara M. Adversity in childhood is linked to mental and physical health throughout life. *BMJ*. (2020) 371:m3048. doi: 10.1136/bmj.m3048
33. Kim KN, Ha B, Seog W, Hwang IU. Long-term exposure to air pollution and the blood lipid levels of healthy young men. *Environ Int*. (2022) 161:107119. doi: 10.1016/j.envint.2022.107119
34. LeMasters K, Bates LM, Chung EO, Gallis JA, Hagaman A, Scherer E, et al. Adverse childhood experiences and depression among women in rural Pakistan. *BMC Public Health*. (2021) 21:400. doi: 10.1186/s12889-021-10409-4
35. Li S, Wang R, Thomas E, Jiang Z, Jin Z, Li R, et al. Patterns of adverse childhood experiences and depressive symptom trajectories in young adults: a longitudinal study of college students in China. *Front Psychiatry*. (2022) 13:918092. doi: 10.3389/fpsy.2022.918092
36. Chapman DP, Whitfield CL, Felitti VJ, Dube SR, Edwards VJ, Anda RF. Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord*. (2004) 82:217–25. doi: 10.1016/j.jad.2003.12.013
37. Walsh D, McCartney G, Smith M, Armour G. Relationship between childhood socioeconomic position and adverse childhood experiences (ACEs): a systematic review. *J Epidemiol Community Health*. (2019) 73:1087–93. doi: 10.1136/jech-2019-212738
38. Schweer-Collins M, Lanier P. Health care access and quality among children exposed to adversity: implications for universal screening of adverse childhood experiences. *Matern Child Health J*. (2021) 25:1903–12. doi: 10.1007/s10995-021-03270-9
39. Dube SR, Anda RF, Felitti VJ, Chapman DP, Williamson DF, Giles WH. Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the adverse childhood experiences study. *JAMA*. (2001) 286:3089–96. doi: 10.1001/jama.286.24.3089
40. Santos M, Burton ET, Cadieux A, Gaffka B, Shaffer L, Cook JL, et al. Adverse childhood experiences, health behaviors, and associations with obesity among youth in the United States. *Behav Med*. (2022) 49:381–91. doi: 10.1080/08964289.2022.2077294
41. Maurya C, Maurya P. Adverse childhood experiences and health risk behaviours among adolescents and young adults: evidence from India. *BMC Public Health*. (2023) 23:536. doi: 10.1186/s12889-023-15416-1
42. Garrido EF, Weiler LM, Taussig HN. Adverse childhood experiences and health-risk behaviors in vulnerable early adolescents. *J Early Adolesc*. (2018) 38:661–80. doi: 10.1177/0272431616687671
43. Singer M, Bulled N, Ostrach B, Mendenhall E. Syndemics and the biosocial conception of health. *Lancet (London England)*. (2017) 389:941–50. doi: 10.1016/S0140-6736(17)30003-X
44. Camacho S, Clark HS. The social determinants of adverse childhood experiences: an intersectional analysis of place, access to resources, and compounding effects. *Int J Environ Res Public Health*. (2022) 19:10670. doi: 10.3390/ijerph191710670
45. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. (2019) 6:94–8. doi: 10.7861/futurehosp.6-2-94
46. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *SVN*. (2017) 2:230–43. doi: 10.1136/svn-2017-000101
47. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak*. (2021) 21:125. doi: 10.1186/s12911-021-01488-9
48. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. Practical guidance on artificial intelligence for health-care data. *Lancet Digit Health*. (2019) 1:e157–9. doi: 10.1016/S2589-7500(19)30084-6
49. Alanazi A. Using machine learning for healthcare challenges and opportunities. *Inform Med Unlocked*. (2022) 30:100924. doi: 10.1016/j.imu.2022.100924
50. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci*. (2021) 2:160. doi: 10.1007/s42979-021-00592-x
51. Kirk D, Kok E, Tufano M, Tekinerdogan B, Feskens EJM, Camps G. Machine learning in nutrition research. *Adv Nutr*. (2022) 13:2573–89. doi: 10.1093/advances/nmac103
52. Collatuzzo G, Boffetta P. Application of P4 (predictive, preventive, personalized, participatory) approach to occupational medicine. *Med Lav*. (2022) 113:e2022009. doi: 10.23749/mdl.v113i1.12622
53. Auffray C, Charron D, Hood L. Predictive, preventive, personalized and participatory medicine: back to the future. *Genome Med*. (2010) 2:57. doi: 10.1186/gm178
54. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol*. (2011) 8:184–7. doi: 10.1038/nrclinonc.2010.227
55. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf Fusion*. (2023) 99:101805. doi: 10.1016/j.inffus.2023.101805
56. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *J Pers Med*. (2020) 10:21. doi: 10.3390/jpm10020021
57. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput*. (2023) 14:8459–86. doi: 10.1007/s12652-021-03612-z
58. Delpino FM, Costa ÂK, Farias SR, Chiavegatto Filho ADP, Arcêncio RA, Nunes BP. Machine learning for predicting chronic diseases: a systematic review. *Public Health*. (2022) 205:14–25. doi: 10.1016/j.puhe.2022.01.007
59. Rashid J, Batool S, Kim J, Wasif Nisar M, Hussain A, Juneja S, et al. An augmented artificial intelligence approach for chronic diseases prediction. *Front Public Health*. (2022) 10:860396. doi: 10.3389/fpubh.2022.860396
60. Siddegowda CJ, Devi AJ. A literature review on prediction of chronic diseases using machine learning techniques. *Int J Manag Technol Soc Sci*. (2022):28–49. doi: 10.47992/IJMTS.2581.6012.0209
61. Subramanian M, Wojtuszczyz A, Favre L, Boughorbel S, Shan J, Letaief KB, et al. Precision medicine in the era of artificial intelligence: implications in chronic disease management. *J Transl Med*. (2020) 18:472. doi: 10.1186/s12967-020-02658-5
62. Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol*. (2020) 122:56–69. doi: 10.1016/j.jclinepi.2020.03.002
63. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *Npj Digit Med*. (2020) 3:30. doi: 10.1038/s41746-020-0229-3
64. Abdulazeem H, Whitelaw S, Schauburger G, Klug SJ. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data [preprint]. *Primary Care Res*. (2022). doi: 10.1101/2022.08.25.22279229
65. Kino S, Hsu YT, Shiba K, Chien YS, Mita C, Kawachi I, et al. A scoping review on the use of machine learning in research on social determinants of health: trends and research prospects. *SSM Popul Health*. (2021) 15:100836. doi: 10.1016/j.ssmph.2021.100836
66. Vera Cruz G, Bucourt E, Réveillère C, Martailé V, Joncker-Vannier I, Goupille P, et al. Machine learning reveals the most important psychological and social variables predicting the differential diagnosis of rheumatic and musculoskeletal diseases. *Rheumatol Int*. (2022) 42:1053–62. doi: 10.1007/s00296-021-04916-1
67. Gladieux M, Gimness N, Rodriguez B, Liu J. Adverse childhood experiences (ACEs) and environmental exposures on neurocognitive outcomes in children: empirical evidence, potential mechanisms, and implications. *Toxics*. (2023) 11:259. doi: 10.3390/toxics11030259
68. Bhattarai S, Gupta A, Ali E, Ali M, Riad M, Adhikari P, et al. Can big data and machine learning improve our understanding of acute respiratory distress syndrome? *Cureus*. (2021) 13:e13529. doi: 10.7759/cureus.13529
69. Ammar N, Shaban-Nejad A. Explainable artificial intelligence recommendation system by leveraging the semantics of adverse childhood experiences: proof-of-concept prototype development. *JMIR Med Inform*. (2020) 8:e18752. doi: 10.2196/18752
70. Centers for Disease Control and Prevention. About BRFSS [internet]. (2023). Available at: <https://www.cdc.gov/brfss/index.html> (Accessed August 6, 2023).
71. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. (2021) 8:53. doi: 10.1186/s40537-021-00444-8
72. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. *Npj Digit Med*. (2019) 2:1–5. doi: 10.1038/s41746-019-0122-0
73. Blair A, Marryat L, Frank J. How community resources mitigate the association between household poverty and the incidence of adverse childhood experiences. *Int J Public Health*. (2019) 64:1059–68. doi: 10.1007/s00038-019-01258-5
74. Royer MF, Ojinnaka CO, Zhang X, Thornton AG, Blackhorse K, Bruening M. Food insecurity and adverse childhood experiences: a systematic review. *Nutr Rev*. (2022) 80:2089–99. doi: 10.1093/nutrit/nuac029
75. Rigolon A, Browning M, Jennings V. Inequities in the quality of urban park systems: an environmental justice investigation of cities in the United States. *Landsc Urban Plan*. (2018) 178:156–69. doi: 10.1016/j.landurbplan.2018.05.026
76. Schroeder K, Forke CM, Noll JG, Wheeler DC, Henry KA, Sarwer DB. The association between adverse childhood experiences, neighborhood greenspace, and body mass index: a cross-sectional study. *Prev Med Rep*. (2022) 29:101915. doi: 10.1016/j.pmedr.2022.101915
77. Centers for Disease Control and Prevention. *LLCP 2021 codebook report: Behavioral risk factor surveillance system [internet]*. Atlanta, Georgia: (2022) Available at: [https://www.cdc.gov/brfss/annual\\_data/2021/pdf/codebook21\\_llcp-v2-508.pdf](https://www.cdc.gov/brfss/annual_data/2021/pdf/codebook21_llcp-v2-508.pdf).
78. Bauder RA, Khoshgoftaar TM. The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. *Health Inf Sci Syst*. (2018) 6:9. doi: 10.1007/s13755-018-0051-3



79. Hasanin T, Khoshgoftaar TM, Leevy JL, Bauder RA. Severely imbalanced big data challenges: investigating data sampling approaches. *J Big Data*. (2019) 6:107. doi: 10.1186/s40537-019-0274-4
80. Zuech R, Hancock J, Khoshgoftaar TM. Detecting web attacks using random undersampling and ensemble learners. *J Big Data*. (2021) 8:75. doi: 10.1186/s40537-021-00460-8
81. Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. *SN Appl Sci*. (2021) 3:272. doi: 10.1007/s42452-021-04148-9
82. scikit-learn. scikit-learn. Permutation feature importance. Available at: [https://scikit-learn/stable/modules/permutation\\_importance.html](https://scikit-learn/stable/modules/permutation_importance.html) (Accessed August 6, 2023).
83. Xing H, Ha MH, Hu BG, Tian DZ. Linear feature-weighted support vector machine. *Fuzzy Inform Eng*. (2009) 1:289–305. doi: 10.1007/s12543-009-0022-0
84. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. (2009) 10:213. doi: 10.1186/1471-2105-10-213
85. Gallagher SK, Mechanic D. Living with the mentally ill: effects on the health and functioning of other household members. *Soc Sci Med* 1982. (1996) 42:1691–701.
86. Sanderson M, Mouton CP, Cook M, Liu J, Blot WJ, Hargreaves MK. Adverse childhood experiences and chronic disease risk in the southern community cohort study. *J Health Care Poor Underserved*. (2021) 32:1384–402. doi: 10.1353/hpu.2021.0139
87. Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C, et al. The effect of multiple adverse childhood experiences on health: a systematic review and meta-analysis. *Lancet Public Health*. (2017) 2:e356–66. doi: 10.1016/S2468-2667(17)30118-4
88. Tan M, Mao P. Type and dose-response effect of adverse childhood experiences in predicting depression: a systematic review and meta-analysis. *Child Abuse Negl*. (2023) 139:106091. doi: 10.1016/j.chiabu.2023.106091
89. Cai T. Breast Cancer diagnosis using imbalanced learning and ensemble method. *Appl Comput Math*. (2018) 7:146. doi: 10.11648/j.acm.20180703.20
90. Zheng H, Sherazi SWA, Lee JY. A stacking ensemble prediction model for the occurrences of major adverse cardiovascular events in patients with acute coronary syndrome on imbalanced data. *IEEE Access*. (2021) 9:113692–704. doi: 10.1109/ACCESS.2021.3099795
91. Zhang Y, Huang Q, Ma X, Yang Z, Jiang J. Using multi-features and ensemble learning method for imbalanced malware classification. In: *2016 IEEE Trustcom/BigDataSE/ISPA [internet]*. Tianjin: IEEE; (2016), 965–973. Available at: <http://ieeexplore.ieee.org/document/7847046/> (Accessed December 15, 2023).
92. Bland JS. Age as a modifiable risk factor for chronic disease. *Integr Med Clin J*. (2018) 17:16–9.
93. McBain H, Shipley M, Newman S. The impact of self-monitoring in chronic illness on healthcare utilisation: a systematic review of reviews. *BMC Health Serv Res*. (2015) 15:565. doi: 10.1186/s12913-015-1221-5
94. Song M, Lipman TH. Concept analysis: self-monitoring in type 2 diabetes mellitus. *Int J Nurs Stud*. (2008) 45:1700–10. doi: 10.1016/j.ijnurstu.2008.04.005
95. Larsson SC, Burgess S. Causal role of high body mass index in multiple chronic diseases: a systematic review and meta-analysis of Mendelian randomization studies. *BMC Med*. (2021) 19:320. doi: 10.1186/s12916-021-02188-x
96. Campbell DJ, Ronskley PE, Manns BJ, Tonelli M, Sanmartin C, Weaver RG, et al. The Association of Income with health behavior change and disease monitoring among patients with chronic disease. *PLoS One*. (2014) 9:e94007. doi: 10.1371/journal.pone.0094007
97. Palencia-Sánchez F, Riaño-Casallas M. *Employments condition and chronic disease in the labor market: A systematic literature review [internet]*. Rochester, NY; (2022). Available at: <https://papers.ssrn.com/abstract=4228722> (Accessed August 6, 2023).
98. National Research Council, Institute of Medicine, board on population health and public health practice, division of behavioral and social sciences and education, Committee on population, panel on understanding cross-National Health Differences among High-Income Countries. *U.S. health in international perspective: Shorter lives, poorer health*. Washington DC, United States: National Academies Press (2013). 421 p.
99. Madjid M, Aboshady I, Awan I, Litovsky S, Casscells SW. Influenza and cardiovascular disease: is there a causal relationship? *Tex Heart Inst J*. (2004) 31:4–13.
100. Morris A. Heart-lung interaction via infection. *Ann Am Thorac Soc*. (2014) 11:S52–6. doi: 10.1513/AnnalsATS.201306-157MG
101. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *Eur Respir J*. (2018) 51:1701794. doi: 10.1183/13993003.01794-2017
102. Behrouzi B, Bhatt DL, Cannon CP, Vardeny O, Lee DS, Solomon SD, et al. Association of influenza vaccination with cardiovascular risk: a Meta-analysis. *JAMA Netw Open*. (2022) 5:e228873. doi: 10.1001/jamanetworkopen.2022.8873
103. Siscovick DS. Influenza vaccination and the risk of primary cardiac arrest. *Am J Epidemiol*. (2000) 152:674–7. doi: 10.1093/aje/152.7.674
104. Christopher AS, McCormick D, Woolhandler S, Himmelstein DU, Bor DH, Wilper AP. Access to care and chronic disease outcomes among Medicaid-insured persons versus the uninsured. *Am J Public Health*. (2016) 106:63–9. doi: 10.2105/AJPH.2015.302925
105. Torres H, Poorman E, Tadepalli U, Schoettler C, Fung CH, Mushero N, et al. Coverage and access for Americans with chronic disease under the affordable care act. *Ann Intern Med*. (2017) 166:472–9. doi: 10.7326/M16-1256
106. Li C, Balluz LS, Ford ES, Okoro CA, Zhao G, Pierannunzi C. A comparison of prevalence estimates for selected health indicators and chronic diseases or conditions from the behavioral risk factor surveillance system, the National Health Interview Survey, and the National Health and nutrition examination survey, 2007–2008. *Prev Med*. (2012) 54:381–7. doi: 10.1016/j.ypmed.2012.04.003
107. Centers for Disease Control and Prevention. BRFSS data quality, validity, and reliability [internet]. (2017). Available at: [https://www.cdc.gov/brfss/publications/data\\_qvr.htm](https://www.cdc.gov/brfss/publications/data_qvr.htm) (Accessed August 7, 2023).
108. MacNell N, Feinstein L, Wilkerson J, Salo PM, Molsberry SA, Fessler MB, et al. Implementing machine learning methods with complex survey data: lessons learned on the impacts of accounting sampling weights in gradient boosting. *PLoS One*. (2023) 18:e0280387. doi: 10.1371/journal.pone.0280387