SUPPLEMENTARY MATERIAL

Table S1: List of SDoH keywords used for searching literature.

Keywords	Related Keywords
NLP	
natural language processing	NLP, text mining, information retrieval
EHR	
electronic health records	EHR, electronic medical record, EMR
General	
behavioral determinants of health	bdoh, bdh
social determinants of health	sdh, sdoh
social and behavioral determinants of health	sbdh, sbdoh
Health behavior	
smoking	cigarette, cigarette use, tobacco, tobacco use, nicotine
alcohol use	alcohol abuse, alcohol withdrawal, alcohol consume, alcoholism, alcohol disorder
drug use	drug abuse
diet	nutrition
exercise	physical activity
sexual activity	sexual abuse
child abuse	childhood abuse, adverse childhood abuse, adverse childhood experience
abuse	-
substance use	substance abuse
Access to care	
health care	quality care, access to care, access primary care, access health care
Social and economic factors	
education	educational stress
income	financial condition, financial constraints, economic condition
work	occupation, working condition, employment, unemployment
community safety	community support
social support	social isolation, social connections, social cohesion
family support	-
food security	food insecurity
living supply	poverty
Physical environment	
housing	homeless, homelessness, living condition
air quality	water quality
public safety	-
transport	transit, transportation

Table S2: Example queries for each database

Database	Example	No results SDoH terms
ACL Anthology	*nlp + ehr + homelessness*	ACL anthology provided search results for every query, however, many search results from ACL Anthology were irrelevant.
PubMed	*(nlp OR (natural language processing) OR (text mining) OR ir OR (information retrieval)) AND (ehr OR (electronic health records) OR emr OR (electronic medical records)) AND (housing OR homeless OR homelessness OR living condition)*	food security, living supply, food insecurity, poverty, financial constraints, adverse childhood abuse, and alcohol withdrawal
Scopus	*Similar to PubMed*	food security, food insecurity, transit, transportation, air quality, water quality, sexual activity, child abuse, diet, exercise, bdh, bdoh, occupation, physical activity, social cohesion, living supply, poverty, financial constrains, living condition, financial condition, economic condition, alcohol withdrawal, and alcohol consume

Table S3: Publications for different social determinants of health. PE:Physical Environment, SEF: Social and Economic Factors, HB: Health Behaviors, T:Transportation, H:Housing issues, EN: Environment (Water/Air quality), E: Employment, ED: Education, F: Financial/Income issues, SF: Social Connection/Isolation or Family Problem, CS: Community Safety, FI: Food Insecurity, CA: Child Abuse/adverse childhood experiences, SA: Sexual activity/abuse, DS: Drug/Substance Abuse, A: Alcohol Abuse/Use, C: Cigarettes/Smoking Status, DE: Diet & Exercise, CC: Clinical care (access to care /quality of care).

Citations	PE T	PE H	PE EN	SEF E	SEF ED	SEF F	SEF SF	SEF CS	SEF FI	HB CA	HB SA	HB DS	HB A	HB C	HB DE	CC	Other
[4]												✓					
[8]											✓						
[9]		✓					✓										
[10]	✓	✓				✓	✓		✓				✓	✓			Information on clothing
[11]												✓					
[12]		✓								✓							
[16]		✓															
[17]													✓	✓	✓		
[18]		✓															
[19]		✓															
[20]		✓				✓	✓					✓	✓				Physical function, jail, nonadherence

Table S3 – *Continued from previous page*

Citations	T	H	EN	E	ED	F	SF	CS	FI	CA	SA	DS	A	C	DE	CC	Other
[21]	✓	✓		✓		✓	✓				✓					✓	legal problems, nonspecific psychosocial needs
[22]		✓															
[23]		✓				✓											Criminal Justice, Health Behaviours
[24]												✓	✓				
[25]												✓					
[26]												✓	✓				depression and suicide
[27]														✓			
[28]		✓		✓	✓												Insurance status
[29]														✓			proximal femur (hip) fracture
[30]										✓							
[31]										✓							

4

Citations	T	H	EN	E	ED	F	SF	CS	FI	CA	SA	DS	A	C	DE	CC	Other
[32]		✓									✓	✓	✓				sexual orientation and gender documentation
[33]		✓									✓	✓	✓				sexual orientation
[34]												✓					
[35]												✓	✓	✓			
[36]														✓			
[37]		✓															Social security disability
[38]												✓					
[39]												✓					
[40]		✓															
[41]													✓				
[42]		✓		✓								✓	✓	✓			
[43]												✓					
[44]												✓					
[45]												✓					
[46]														✓			

5

Citations	T	H	EN	E	ED	F	SF	CS	FI	CA	SA	DS	A	C	DE	CC	Other
[47]		✓		✓	✓	✓	✓		✓								insurance
[48]		✓		✓									✓	✓	✓		Social history (Seatbelt, Language, Abuse, Helmets), family, marital status
[49]												✓					
[50]						✓									✓		72 other categories
[51]		✓											✓	✓			Living alone, use of substance use services
[52]										✓							
[53]											✓						
[54]											✓						
[55]												✓					
[56]							✓										
[57]														✓			Other information not related to SDoH
[58]														✓			
[59]		✓				✓								✓			

Citations	T	H	EN	E	ED	F	SF	CS	FI	CA	SA	DS	A	C	DE	CC	Other
[60]														✔			
[61]														✔			
[62]								✔									
[63]		✔													✔		
[64]														✔	✔		
[65]															✔		
[66]		✔															
[67]				✔	✔												Vocational Interests and Driving Ability
[68]														✔			
[69]										✔							
[70]											✔						
[71]														✔			
[72]														✔			
[73]												✔					
[74]														✔			
[75]														✔			
[76]														✔			

7

Table S3 – *Continued from previous page*

Citations	T	H	EN	E	ED	F	SF	CS	FI	CA	SA	DS	A	C	DE	CC	Other
[77]														✓			
[78]													✓				
[79]														✓			
[80]																	Quality of clinical care
[81]					✓												
[82]													✓	✓			
[83]																	Marital status
[84]					✓												
[85]															✓		
[87]						✓									✓		falls, walking difficulty, dementia, vision impairment, absence of fecal control, severe lack of urinary control, weight loss, pressure ulcer
[88]													✓	✓			BMI, socioeconomic status

Table S4: SDoH category wise citations for rule-based and semi-automated methods for SDoH lexicon creation

SDoH Categories	Citations	
	Manual	Semi-automatic
Transportation	[10, 21]	
Housing issues	[10, 16, 20, 21, 28, 42, 47, 48, 59]	[9, 12, 23]
Environment	[63]	
Employment	[21, 28, 42, 47, 67]	
Education	[28, 47, 67]	
Financial issues	[10, 20, 21, 47]	[23]
Social issues	[10, 20, 21, 39, 47, 56, 59]	[9]
Community safety	[62]	
Food insecurity	[10, 47]	
Child abuse	[31, 69]	[12]
Sexual abuse	[20, 21]	
Drug abuse	[20, 42, 59]	
Alcohol abuse	[10, 17, 42, 48]	
Smoking status	[10, 17, 42, 48]	
Diet	[17, 42, 48, 63, 65, 85, 87]	
Clinical care	[21, 28, 42, 47, 80]	

Table S5: SDoH category wise citations for rule-based, supervised, unsupervised, and NLP tool based methods for SDoH identification

SDoH Categories	Citations			
	Rule-based	Supervised	Unsupervised	Tool based
Transportation	[10, 21]			
Housing issues	[10, 12, 21, 28]	[32, 33, 42, 47]	[48]	[9, 16]
Environment				[63]
Employment	[21, 28]	[42, 47]	[48]	[67]
Education	[28]	[47]		[67]
Financial issues	[10, 21]	[47, 81, 84]		
Social issues	[10, 21, 39, 87]		[50]	[9, 56, 59]
Community safety				[62]
Food insecurity	[10]	[47]		
Child abuse	[12, 31, 69]	[31, 52]		[52]
Sexual abuse	[21]	[32, 33]		
Drug abuse		[32, 33, 42]		[59]
Alcohol abuse	[17]	[32, 33, 42]	[48]	
Smoking status	[17]	[42]	[48]	[64]
Diet	[17, 65, 85, 87]	[42]	[48, 50]	[63, 64]
Clinical care	[21, 28]	[42, 47]		[80]

Table S6: Classifiers used for SDoH extraction with the corresponding citations.

Classifiers	Citations
SVM	[29, 30, 32, 33, 35, 36, 43, 46, 47, 57, 71, 75, 76, 78]
Random forest	[24, 30, 31, 33, 34, 40, 47, 75, 81, 82, 84, 88]
Decision trees	[75]
Logistic Regression	[31, 33, 40, 41, 45, 75, 78]
LASSO regularization	[41]
Naïve Bayes (NB)	[30, 46]
K-nearest neighbors	[47, 75]
caRT	[33]
Adaboost	[33]
SGD classifier	[78]
Passive Aggressive Classifier	[78]
XGBoost	[47]
Gradient boost classifier	[88]
n-gram based classifier	[45]
Latent Class Analysis (LCA)	[29, 49]
Topic modeling	[48]
LDA	[50]
BiLSTM-CRF	[42]
BiLSTM	[42, 47]
Convolutional Neural Networks (CNN)	[45, 46]
Gradient descent	[75]
Long short-term memory (LSTM)	[46]
Multilayer Perceptron Neural Networks (MLPNN)	[29]
Recurrent neural network (RNN)	[46]