BMJ Open

Predicting population health with machine learning: a scoping review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037860
Article Type:	Original research
Date Submitted by the Author:	19-Feb-2020
Complete List of Authors:	Morgenstern, Jason; McMaster University, Health Research Methods, Evidence, and Impact Buajitti, Emmalin; University of Toronto Dalla Lana School of Public Health; Institute for Clinical Evaluative Sciences, O'Neill, Meghan; University of Toronto Dalla Lana School of Public Health, Piggott, Thomas; McMaster University, Health Research Methods, Evidence, and Impact Goel, Vivek; University of Toronto Dalla Lana School of Public Health; Institute for Clinical Evaluative Sciences, Fridman, Daniel; Hospital for Sick Children Kornas, Kathy ; University of Toronto Dalla Lana School of Public Health Rosella, Laura; University of Toronto Dalla Lana School of Public Health; Institute for Clinical Evaluative Sciences,
Keywords:	PUBLIC HEALTH, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

PREDICTING POPULATION HEALTH WITH MACHINE LEARNING: A SCOPING REVIEW

AUTHORS

Jason D. Morgenstern[1], MD
Emmalin Buajitti[2,3], MPH
Meghan O'Neill[2], MPH
Thomas Piggott[1], MD, MSc
Vivek Goel[2,3], MD, MSc, PhD
Daniel Fridman[4], MPH, MA
Kathy Kornas[2], MSc
Laura C. Rosella[2,3,5,6], MHSc, PhD

AFFILIATIONS

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
3. Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada
4. The Hospital for Sick Children, Toronto, Ontario, Canada
5. Public Health Ontario, Toronto, Ontario, Canada
6. Vector Institute, Toronto, Ontario, Canada

CORRESPONDENCE TO:

Dr. Laura Rosella
Dalla Lana School of Public Health
University of Toronto
27 King's College Circle
Toronto, M5S 1A1, Canada
laura.rosella@utoronto.ca
Tel: (416) 978 6064

ABSTRACT

Objective To determine how machine learning has been applied to prediction applications in population health contexts, including which outcomes were studied, which data sources were used, and how models were developed.

Design A scoping review.

Data Sources MEDLINE, EMBASE, CINAHL, ProQuest, Scopus, Web of Science, Cochrane Library, INSPEC, and ACM Digital Library were searched on July 18th, 2018.

Eligibility criteria We included English articles published since 1980 that used machine learning to predict population health-related outcomes. We excluded studies that only used logistic regression or were restricted to a clinical context.

Data extraction and synthesis We summarized findings extracted from published reports, which included general study characteristics, aspects of model development, reporting of results, and model discussion items.

Results Of 22 618 articles found by our search, 231 were included in the review. The United States (n=71, 30.74%) and China (n=40, 17.75%) produced the most studies and cardiovascular disease (n=22, 9.52%) was the most studied outcome. The median number of observations was 5414 (interquartile range (IQR)=16 543·5) and the median number of features was 17 (IQR=31). The most commonly used data sources were health records (n=126, 54.5%) and investigator-generated (n=86, 37.2%). Many studies did not incorporate recommended guidelines on machine learning and predictive modeling. Predictive discrimination was commonly assessed using area under the receiver operator curve (n=98, 42.42%) and calibration was rarely assessed (n=22, 9.52%).

Conclusions Machine learning applications in population health have concentrated on regions and diseases well-represented in traditional data sources, infrequently using big data. Additionally, important aspects of model development were under-reported. Greater use of big data and uptake of guidelines for predictive modeling could improve the yield from machine learning applications in population health.

Registration Registered on the Open Science Framework on July 17th, 2018 (available at: https://osf.io/rnqe6/).

Strengths and limitations of this study

- Our review is one of the first syntheses of machine learning applications in population and public health.
- We used a comprehensive search strategy, including nine peer-reviewed databases, grey literature, and reference searching.
- We extracted a wide array of study characteristics, including important elements of predictive modeling reporting guidelines.
- Since both machine learning and population health have broad definitions, there may be some relevant articles that were not included.
- Given our focus on prediction, we could not address many other important intersections of machine learning and population health, such as surveillance and health promotion.

INTRODUCTION

Predictive models have a long history in clinical medicine. One well-known example is the Framingham risk score, which was first developed in 1967.[1] Such models have proliferated throughout clinical practice to inform management and interventions, including preventive approaches. More recently, researchers have developed prediction models beyond individual clinical applications, for population health uses.[2,3] While there is no universal definition of population health, it includes "the health outcomes of a group of individuals, including the distribution of such outcomes within the group."[4] Similarly to clinical medicine, population-level models can be used to identify high-risk groups, directing the implementation of preventive interventions. Additionally, population health prediction models can inform policymakers about future disease burden and help to assess the impact of public health actions. Thus far, most predictive modeling in both medicine and population health has used parametric statistical regression models. More recently, there has been increasing interest in the use of a broader range of machine learning methods for prediction tasks.[5–7]

Machine learning can be loosely defined as the study and development of algorithms that learn from data with little or no human assistance.[8] These approaches have been increasingly applied in the past two decades as a result of the enabling growth of big data reserves and computational power.[9] Recent machine learning applications to prediction in population health contexts include forecasting childhood lead poisoning,[10] yellow fever incidence,[11] and the onset of suicidal ideation.[12]

The distinction between machine learning algorithms and parametric regression models is debated.[13] Regression models tend to impose more structure on the data, requiring greater human input for the verification of distributional assumptions and incorporation of domain knowledge in choosing the input parameters.[14] Algorithms employed in machine learning often derive more structure directly from the data, making fewer distributional assumptions about the data or variables. The literature remains divided on the relative advantages of more traditional approaches compared to newer methods;[15] however, given the wide variation in applications and the data used in these examples, broad assessments of superiority are often not appropriate. Also, there are debates regarding the differences in developing and validating machine learning approaches for health applications.[15,16]

Population health applications of prediction models are relatively new compared to clinical applications; correspondingly, the role of machine learning in these applications has been far less studied and discussed in the health literature. The goals of our review are to determine how machine learning has been applied to prediction in population health, the nature of the models and data used, and how the models have been developed. We hope that our results will help to inform future research in this area, including the development of guidelines for machine learning applications in population health.

METHODS

We based our scoping review on the framework proposed by Arksey and O'Malley[17] and refined by the Joanna Briggs Institute.[18] We also followed the more recent Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews.[19] Our study protocol was registered on the Open Science Framework on July 17th, 2018 (available at: https://osf.io/rnqe6/).

Our initial goal was to scope out all machine learning applications in population health. However, the screening process identified a much larger number of publications than anticipated. Consequently, to describe the subject area comprehensively, we restricted our scope to articles predicting future outcomes.

Search Strategy

Our search strategy consisted of peer-reviewed literature databases, grey literature, and reference searches. First, we searched nine interdisciplinary, indexed databases (MEDLINE, EMBASE, CINAHL, ProQuest, Scopus, Web of Science, Cochrane Library, INSPEC, and ACM Digital Library) on July 18th, 2018. Our search was informed by consultation with a health science librarian, a machine learning textbook,[20] and a similar registered review.[15] Supplementary table A includes an example search query.

Our grey literature search included Google Scholar and Google. We developed a Google Scholar search based on terms related to 'machine learning' and 'population health', which was refined based on the relevance of initial results. The first 200 results were included in screening. A similar approach was used for the general Google search, which we restricted to the first 30 results. We examined relevant websites for publications. Results were limited to articles published on or before the date of the peer-reviewed literature search.

Finally, we searched the references of relevant reviews for additional articles. Most of these reviews were identified during screening.

Eligibility Criteria

We included articles if they used machine learning to develop a predictive model that could be applied in a population health context. Therefore, we excluded articles where the model was trained primarily on people with a pre-existing disease. We also excluded articles that were only indirectly related to population health; for example, traffic accident models that did not predict a health outcome. Studies predicting individual outcomes were included if the approach was determined to be scalable to a population level. Finally, articles using only logistic regression were excluded. See appendix A for the full eligibility criteria.

In order to manage the scope, articles were excluded if their full text could not be retrieved with our institutional licenses and if they were not written in English. Finally, articles published prior to 1980 were excluded as earlier machine learning investigators lacked comparable amounts of digitized data, software, and computational resources.

Screening Process

Initially, individual reviewers screened titles for obvious irrelevance to the review topic (JDM and EB). Examples of articles removed at this stage are outlined in appendix B. Then, we imported remaining references into Covidence systematic review management software.[21] Two reviewers screened the abstracts of remaining articles (JDM, EB, MO, and DF). Prior to evaluating full texts using all eligibility criteria, we then screened out articles that did not focus on a prediction application (JDM, EB, MO). Finally, two reviewers screened the full text of remaining articles (JDM, EB, MO). Conflicts were resolved by discussion between at least two reviewers.

Data Extraction and Synthesis

Individual authors extracted article data (JDM, EB, MO, and DF). We based our extraction items on important aspects of machine learning identified in a recent biomedical guideline[16] and on the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.[22] Major extraction categories included general study characteristics (e.g. geographic location and sample size), model development (e.g. algorithms used and type of validation), results (e.g. discrimination and calibration measures), and model discussion (e.g. practical costs of errors and implementation). See supplementary table B for a description of each extraction item.

We computed descriptive statistics for all extraction items. We also completed a narrative synthesis of discussion elements.

Patient and Public Involvement Statement

There was no patient or public involvement in this study.

RESULTS

We initially retrieved 16 172 articles, after removing duplicates (figure 1). We excluded 6494 articles after title screening, 7860 after abstract screening, 1453 when screening out non-prediction articles, and 121 after full-text screening. This resulted in 231 articles being included in the final review (appendix C).

General Study Characteristics

The number of articles published in the population health prediction area that used machine learning increased dramatically after 2007 (supplementary figure A). Studies were undertaken worldwide, with the largest representation from the United States (US) (n=71, 30.74%) and China (n=40, 17.75%) (table 1). Relatively few articles came from Oceania (n=2, 0.87%), Africa (n=5, 2.16%), and the Americas outside of the US (n=13, 5.63%).

Characteristic*	**Number or Median**	**Percent or Interquartile Range**
Region		
United States	71	30.74%
Asia Excluding China	41	17.75%
China	40	17.32%
Europe	36	15.58%
Americas Excluding United States	13	5.63%
Africa	5	2.16%
Oceania	2	0.87%
Multi-region	15	6.49%
Not Reported	8	3.46%
Year published		
before 1990	1	0.4%
1990-1999	3	1.3%
2000-2004	13	5.6%
2005-2009	18	7.8%
2010-2014	70	30.3%
2015-2018	126	54.5%
Outcome level†		
Individual Risk Prediction	139	60.17%
Population Risk prediction	92	39.83%
Number of observations	5414	16 543.5

Not reported	72	31.2%
Number of features	17	31
Not reported	59	25.5%
Used any unstructured text		
Yes	24	10.4%
No	207	89.6%
Machine learning model was compared with other statistical methods	111	48.1%
Reported data pre-processing‡		
Yes	160	69.3%
No	71	30.7%
Reported method of feature selection		
Yes	164	71.0%
No	67	29.0%
Reported hyper-parameter search		
Yes	114	49.4%
No	117	50.6%
Method of Validation		
Holdout	112	48.5%
Cross-validation or bootstrap	84	36.4%
External	15	6.5%
Not reported	32	13.9%
Reported descriptive statistics§		
Yes	140	60.6%
No	91	39.4%
Discussed the practical costs of prediction errors¶		
Yes	36	15.6%

No	195	84.4%
Stated rationale for using machine learning		
Yes	179	77.5%
No	52	22.5%
Discussed model usability		
Yes	91	39.4%
No	140	60.6%
Stated model limitations		
Yes	161	69.7%
No	70	30.3%
Discussed model implementation		
Yes	184	79.7%
No	47	20.3%
Dataset Availability by Study‖		
Closed	149	64.5%
Public	42	18.2%
Closed and Public	38	16.5%
Unknown	1	0.4%

*Refer to supplementary table A for a description of each characteristic and rationales for including some elements.

†Individual risk prediction refers to studies that developed models to predict the health outcomes of individuals, while population risk prediction refers to studies that developed models to predict aggregated population-level health outcomes.

‡Whether any aspects of data cleaning or pre-processing were reported. Examples include how missing data was handled, whether log transformations were done, and if derived variables were generated.

§Included a broad array of descriptive statistics such as sample population demographics, feature distributions, and outcome distributions.

¶Whether the article discussed the relative risks of false negative and false positive results based on their predictive model in contexts where it might be used.

‖Closed refers to datasets that were not immediately available in the public domain or were not identifiable as such.

Table 1: **Summary statistics of included articles**

The median number of observations in each article was 5414 (interquartile range (IQR)=16 543.5) and the median number of features (i.e. independent variables) used was 17 (IQR=31) (table 1). Seventy-two studies (31.2%) did not report the number of observations. These studies often used data from reportable disease databases, which do not necessarily have a firm sampling frame, making ascertainment of the number of observations difficult.

Algorithms

The most frequently used machine learning algorithms were neural networks (n=95, 41.13%), followed by support vector machines (n=59, 25.54%), single tree-based methods (n=52, 22.51%), and random forests (n=48, 20.78%) (supplementary table C). About half of the articles made a comparison with statistical methods (n=111, 48.1%), which were generally logistic regression or autoregressive integrated moving average models (table 1).

Outcomes

Non-communicable disease outcomes were assessed by many articles (n=95, 41.13%), with communicable diseases (n=76, 32.90%) and non-disease outcomes (n=60, 25.97%) studied somewhat less often. The outcome most frequently predicted was cardiovascular disease (n=22, 9.52%) (figure 2). Other commonly forecasted non-communicable disease outcomes were suicidality (n=13, 5.63%), cancer (n=12, 5.19%), and perinatal health (n=12, 5.19%). Influenza (n=15, 6.49%) and dengue fever (n = 14, 6.06%) were the most predicted communicable disease outcomes. Aside from non-communicable and communicable disease, mortality (n=13, 5.63%) and healthcare utilization (n=14, 6.06%) were also frequently predicted.

Data

Data sources were usually structured (n=207, 89.6%) and closed, i.e. not publicly available (n=189, 81.8%) (table 1). The most frequently reported data sources were health records (n=126, 54.5%) and investigator-generated (e.g. cohort studies) (n=86, 37.2%) (table 2). A large proportion of studies (n=42, 18.2%) used an environmental data source (e.g. satellite imagery), mostly for prediction of infectious disease. Government databases (n=32, 13.9%) and internet-based data (n=21, 9.1%) were less frequently used. Among studies from China and the US, 80.0% and 67.6% respectively used health records data, whereas 54.5% of studies overall used these data sources (supplementary figure B).

Sources of Data Used*	**Number**	**Percent**
Environmental	42	18.2%
Geographical Information Database	12	5.2%
Meteorological/Air Quality Datasets	32	13.9%
Satellite Imagery	21	9.1%
Health Records Database	126	54.5%
Clinical Record Database†	46	19.9%
Disease Registry	2	0.9%
Population Health Survey	15	6.5%
Reportable Disease Database	42	18.2%
Other Health Records Database	30	13.0%
Government Database	32	13.9%
Census	11	4.8%

Vital Statistics	13	5.6%
Other Government Database	14	6.1%
HealthMap	3	1.3%
Private Insurance Data	9	3.9%
Private Insurance Claims	9	3.9%
Private Insurance Questionnaire	3	1.3%
Internet-based	21	9.1%
Search engine	12	5.2%
Social Media	12	5.2%
Investigator-generated‡	86	37.2%
Public Repositories§	19	8.2%
Health Organization Reports¶	5	2.2%
Not Reported	6	2.6%

*Categories are not mutually exclusive.

†Any dataset produced primarily for the purpose of delivering clinical care, such as electronic medical records and administrative healthcare databases produced by hospitals.

‡Any datasets resulting from researcher-driven studies such as randomized controlled trials, cohort studies, and case-control studies.

§Any freely available datasets such as MIMIC or the UC Irvine Machine Learning Repository.

¶Health-related reports, typically including disease burden estimates, produced by non-governmental or governmental organizations such as the World Health Organization.

Table 2: **Data sources**

Features

Biomedical and sociodemographic features were frequently used (supplementary figure C). Of these, the most commonly used were disease history (43.3%), age (48.5%), and sex/gender (41.1%). Among lifestyle features, smoking was the most frequently used (25.1%) and of environmental features, meteorology was common (17.3%). Social media posts (5.2%) and web search queries (5.2%) were not often used. See supplementary table D for more details.

Model Development and Validation

The majority of articles reported how data pre-processing (n=160, 69.3%) and feature selection (n=164, 71%) were done (table 1). Fewer authors reported how hyperparameters were selected (n=114, 49.4%). Most studies used a holdout method of validation (n=112, 48.5%), fifteen (6.5%) externally validated their models, and thirty-two (13.9%) did not report how models were validated.

Performance Metrics

Most articles reported a prediction discrimination metric (n=172, 74.46%), with fewer reporting a measure of overall model fit (n=77, 33.33%), and few reporting a measure of calibration (n=21, 9.09%) (table 3). The most common discrimination metrics employed were area under the receiver operator curve (n=98, 42.42%), accuracy (n=76, 32.90%), and recall (n=68, 29.44%). Calibration was mostly assessed with graphing methods (n=9, 3.90%) and Hosmer-Lemeshow statistics (n=8, 3.46%). Overall performance was usually measured with a form of mean error, such as root mean squared error (n=35, 15.15%).

Prediction Performance Metrics Used	**Number**	**Percent**
Any overall performance metric	77	33.33%
RMSE	35	15.15%

MSE	26	11.26%
MAE	24	10.39%
MAPE	23	9.96%
R2*	19	8.23%
Correlation	8	3.46%
AIC or BIC	8	3.46%
Other performance metric†	21	9.09%
Any discrimination metric	172	74.46%
Area under the curve‡	98	42.42%
Accuracy§	76	32.90%
Recall¶	68	29.44%
Precision‖	39	16.88%
F statistics	10	4.33%
Likelihood Ratio**	4	1.73%
Youden Index	3	1.30%
Manual or visual comparison	3	1.30%
Other discrimination metric††	4	1.73%
Any calibration metric	21	9.09%
Manual or visual comparison‡‡	9	3.90%
Hosmer-Lemeshow	8	3.46%
Observed/Expected	5	2.16%
Other calibration metric§§	3	1.30%
Any reclassification metric	6	2.60%
Net Reclassification Index	5	2.16%
Integrated Discrimination Improvement	3	1.30%

RMSE = Root Mean Squared Error; MSE = Mean Squared Error; MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

*Includes R2 and pseudo-R2 metrics.

†Includes penalty error, Total Sum of Squares, proportional reduction in error, overall prediction error, specific prediction error, Nash-Sutcliffe, Root Mean Squared Percentage Error (2), mean relative absolute error, Analysis of Variance F-stat, 2LogLikelihood, relative efficiency, deviance, Ljung-Box test, mean absolute deviation, standard error, Mean Percentage Error, Brier score, and log score.

‡Includes c-statistic, s-index, area under ROC / AUC.

§Includes accuracy, misclassification, and error rate.

¶Includes sensitivity, specificity, true/false positive, and true/false negative.

||Includes positive predictive value, negative predictive value, and precision.

**Includes positive/negative LR.

††Includes G-means (2), k-statistic, Matthews correlation coefficient.

‡‡Includes calibration plots.

§§Includes mean bias (from Bland-Altman plot), calibration factoring, and Calibration statistic.

Table 3: **Prediction Performance Metrics**

Study Discussion

Most articles included some discussion of their rationale for using machine learning (n=179, 77.5%), limitations of their study (n=161, 69.7%), and how the model might be implemented (n=184, 79.7%) (table 1). Few discussed model usability (n=91, 39.4%) and only a small number discussed the costs of prediction errors in real-world contexts (n=36, 15.6%). See appendix D for a narrative synthesis of discussion reporting items.

DISCUSSION

Our results show that machine learning is increasingly being applied to make predictions related to population health. Nearly half of the included studies were conducted in the US or China. Both countries produce the greatest number of scientific publications in general;[23] however, they also likely benefited from robust health data infrastructures. The US has rapidly digitized much of its healthcare system, resulting in large electronic medical records (EMRs) linked with government data through public-private partnerships, including processes to make these data available to researchers.[24,25] Both the US and China made greater use of health records and less use of investigator-generated data relative to other regions, which may have made machine learning projects more tractable. They also used more internet-based data, which typically includes many observations and is high-dimensional, making it amenable to machine learning methods. Other countries with substantial EMR-use and government database linkage such as Finland, Singapore, and Denmark[26] likely have untapped potential for machine learning research. We noted that studies from Oceania, Africa, and the Americas (outside of the US) were limited. This may be partly due to less availability of traditional sources of structured health data. However, given that machine learning methods can incorporate non-traditional data sources, there is the potential to expand use of these methods even when structured health data is unavailable.

We found that a wide range of population health outcomes have been the focus of machine learning prediction models. However, relative to morbidity and mortality, multiple outcome categories like cancer, human immunodeficiency virus, dementia, gastroenteritis, pneumococcal disease, perinatal health, tuberculosis, and malaria appear understudied.[27] Many of these conditions are most prevalent in regions with decreased access to traditional health data, perhaps stymieing research. If machine learning methods are used to leverage novel data sources for research in these regions, it could enable greater study of neglected diseases.

Most investigators did not analyze a large number of observations and features. We observed a high reliance on investigator-generated data, which likely made it difficult to achieve high sample sizes or high dimensional data. The use of smaller datasets may affect the performance of studied models, as machine learning algorithms generally require a high number of observations relative to features.[28] Additionally, most studies focused on features typical of clinical prediction models, such as biomedical factors and limited aspects of broader socioeconomic or environmental determinants of health. We also observed infrequent use of unstructured data and wearable data for prediction purposes. A reliance on small datasets and traditional numbers and types of features is unlikely to fully leverage any benefits of machine learning. This may be contributing to the small differences frequently seen between parametric regression and machine learning model performance. Greater use of linked population-level databases, large EMRs, internet data, and unstructured features would likely improve these approaches.

Based on the elements of model development that we studied, adherence to existing machine learning[16] and prediction model[22] guidelines appears limited. Most articles did not report their method of hyper-parameter selection, discuss practical costs of prediction errors, or consider model usability, which are needed for transparency and model assessment. Many studies did not report the number of features included, method of validation, method of feature selection, or any performance metric. Given these issues, it would be difficult or impossible to compare many of these machine learning models with existing approaches. However, we acknowledge that existing guidelines were not available when many included studies were published. Future work should apply existing guidance,[16] including from TRIPOD,[22] and anticipate the forthcoming TRIPOD-ML statement.[29]

Lastly, we noted that included studies rarely assessed predictive performance in terms of calibration, which refers to a model's ability to accurately predict the absolute probability of outcomes.[30] In contrast, discrimination measures of predictive performance quantify a model's ability to correctly rank-order individuals. Many traditional machine learning tasks, such as image recognition, often have a high signal to noise ratio. In these cases, discrimination may be a suitable lone performance metric, as the algorithm can achieve near perfect performance. Conversely, health outcomes tend to be more stochastic. As a result, accurate prediction of probabilities is more important.[30] Models can have good predictive discrimination, but poor calibration, making them less useful in practice, particularly for population health applications. A further issue is that many measures of discrimination, such as accuracy and recall, artificially impose a threshold for calling events. Thresholds should ideally be ascertained by decision-makers based on their cost-utility curves.[30] Overall, applications of machine learning in population health would benefit from greater use of calibration performance metrics.

A strength of our study is that we addressed an understudied area, the intersection of machine learning and population health. Additionally, prediction is an application with untapped potential in population health, and where machine learning has the potential to make significant improvements. Our study also employed a comprehensive search strategy, including numerous multidisciplinary peer-reviewed databases, alongside a grey literature search. Furthermore, we applied insights from the field of clinical prediction modeling to population health and machine learning. Finally, given the focus on prediction, we were able to take a comprehensive approach to data extraction and synthesis.

In terms of limitations, concentrating on prediction prevented us from exploring applications of machine learning to other important aspects of population health, such as disease surveillance. These should be the focus of future research. Our review was also limited by including only English articles and articles with available full text, which may have introduced selection bias. Lastly, the two main concepts underlying our review, machine learning and population health, are not universally defined. As a result, we may have excluded articles that may be relevant to these fields.

This was the first scoping review specifically focused on machine learning prediction in population health applications. Predictive modeling in population health can help to inform preventive interventions, anticipate future disease burden, and assess the impact of health policies and programs. Advances in machine learning offer opportunities to improve these models, particularly when incorporating big data. This is still a nascent field, but based on our findings more research in Oceania, Africa, and South America would be particularly beneficial. Diseases such as malaria, tuberculosis, and dementia should also be further studied. Additionally, future machine learning projects could incorporate larger datasets and more non-traditional features. Greater use of resources such as HealthMap, social media, web search patterns, remote sensing, and WHO reports would enable more work in regions without formal data sources and enrich research in others. Another largely untapped prospect is using machine learning and high-dimensional data to incorporate richer representations of the social determinants of health. Opportunities should continue to grow as governments increasingly digitize their health service records and link databases to both health and non-health data. Overall, as applications of machine learning in population health develop, adherence to existing guidance[16,22,29] will improve our ability to assess and advance machine learning applications. Finally, it will be important to evaluate the impact of prediction models on decisions made in population health and the practice of public health.

CONTRIBUTORS

JDM contributed to the design of the study and led the literature search, article screening, data extraction, analysis, and writing of the manuscript. EB contributed to the design of the study, the literature search, article screening, data extraction, and analysis. MO contributed to article screening, data extraction, analysis, and writing of the manuscript. DF contributed to article screening and data extraction. KK contributed to the design of the study. LCR led the design of the study. All authors interpreted study results and contributed to drafting of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Catherine Bornbaum for her assistance with the initial design of the study.

FUNDING

This work was funded by the Canadian Institutes of Health Research. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

COMPETING INTERESTS

None declared.

PATIENT CONSENT FOR PUBLICATION

Not required.

ETHICS APPROVAL

Not required as only prior published research was included in the review.

DATA AVAILABILITY

The full data extraction table used for this review will be made publicly available after publication with no end date on Mendeley Data (DOI: 10.17632/7rrz9xrp2j.1).

REFERENCES

1 Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chronic Dis* 1967;**20**:511–24. doi:10.1016/0021-9681(67)90082-3

2 Nunes MB, McPherson M, Kommers P, *et al.* Proceedings of the International Association for Development of the Information Society (IADIS) International Conference on E-Learning (Lisbon, Portugal, July 20-22, 2017). 2017;:251.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/2013525439?accountid=12347

3 Manuel DG, Tuna M, Bennett C, *et al.* Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ* 2018;**190**:E871–82. doi:10.1503/cmaj.170914

4 Kindig D, Stoddart G. What is population health? *Am J Public Health* 2003;**93**:380–3.http://www.ncbi.nlm.nih.gov/pubmed/12604476 (accessed 24 Jun 2018).

5 Panch T, Pearson-Stuttard J, Greaves F, *et al.* Artificial intelligence: opportunities and risks for public health. *Lancet Digit Heal* 2019;**1**:e13–4. doi:10.1016/s2589-7500(19)30002-0

6 Aldridge RW. Research and training recommendations for public health data science. *Lancet Public Heal* 2019;**4**:e373. doi:10.1016/S2468-2667(19)30112-4

7 Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annu Rev Public Health* 2018;**39**:95–112. doi:https://dx.doi.org/10.1146/annurev-publhealth-040617-014208

8 Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev* 1959;**3**:210–29. doi:10.1147/rd.33.0210

9 Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. 2nd ed. Upper Saddle River, New Jersey: : Prentice Hall 2003.

10 Potash E, Brew J, Loewi A, *et al.* Predictive modeling for public health: Preventing childhood lead poisoning. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2015;**2015-August**:2039–47. doi:10.1145/2783258.2788629

11 Shearer FM, Longbottom J, Browne AJ, *et al.* Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Heal* 2018;**6**:e270–8. doi:10.1016/S2214-109X(18)30024-X

12 De Choudhury M, Kiciman E, Dredze M, *et al.* Discovering shifts to suicidal ideation from mental health content in social media. In: *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery 2016. 2098–110. doi:10.1145/2858036.2858207

13 K.G.M. M, J.A.H. DG, W. B, *et al.* Checklist for data extraction and critical appraisal for systematic reviews of prediction modelling studies: The charms checklist. *Eur J Epidemiol* 2015;**30**:904. doi:http://dx.doi.org/10.1007/s10654-015-0072-z

14 Breiman L. Statistical modeling: The two cultures. *Stat Sci* 2001;**16**:199–215. doi:10.1214/ss/1009213726

15 Christodoulou E, Ma J, Collins GS, *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;**110**:12–22. doi:10.1016/j.jclinepi.2019.02.004

16 Luo W, Phung D, Tran T, *et al.* Guidelines for developing and reporting machine learning

predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res* 2016;**18**:e323. doi:10.2196/jmir.5870

17 Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;**8**:19–32. doi:10.1080/1364557032000119616

18 Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual 2015 - Methodology for JBI Scoping Reviews. Adelaide: 2015. doi:10.1017/CBO9781107415324.004

19 Tricco AC, Lillie E, Zarin W, *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* Published Online First: 4 September 2018. doi:10.7326/M18-0850

20 Hastie T, Tibshirani R, Witten D, *et al. An Introduction to Statistical Learning: With Applications in R*. New York, NY: : Springer 2013.

21 Veritas Health Innovation. Covidence systematic review software. www.covidence.org

22 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;**13**:1. doi:https://dx.doi.org/10.1186/s12916-014-0241-z

23 Country outputs | Nature Index. https://www.natureindex.com/country-outputs/generate/All/global/All/score (accessed 15 Nov 2019).

24 Hecht J. The future of electronic health records. *Nature* 2019;**573**:S114–6. doi:10.1038/d41586-019-02876-y

25 Gliklich RE, Dreyer NA, Leavy MB. Public-Private Partnerships. 2014.

26 OECD. Health Data Governance: Privacy, Monitoring and Research. Paris: 2015. http://dx.doi.org/10.1787/9789264244566-en

27 Naghavi M, Wang H, Lozano R, *et al.* Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;**385**:117–71. doi:10.1016/S0140-6736(14)61682-2

28 van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a

simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;**14**:137. doi:10.1186/1471-2288-14-137

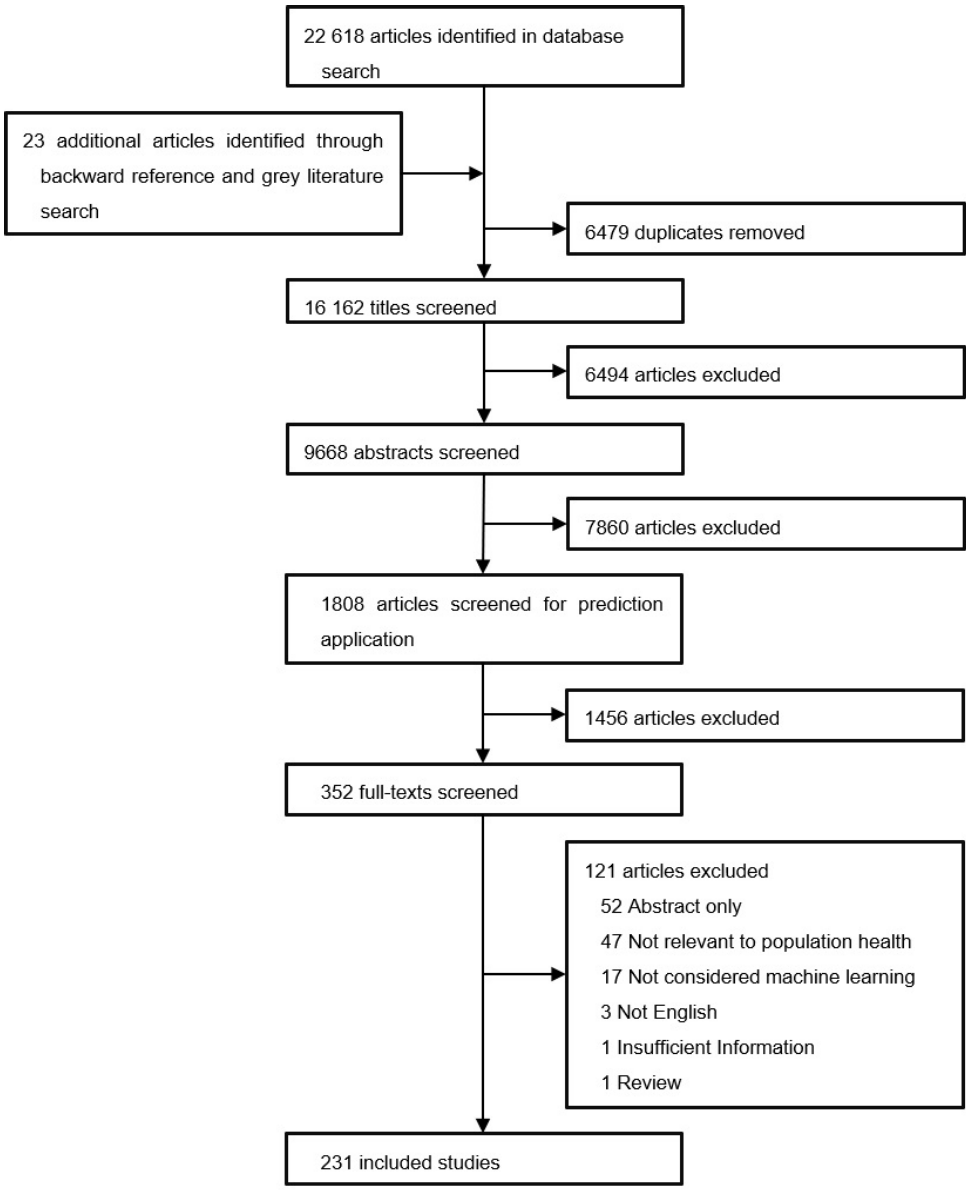
29 Collins GS, M Moons KG. Reporting of artificial intelligence prediction models. Published Online First: 2019. doi:10.1016/S0140-6736(19)30235-1

30 Steyerberg EW. *Clinical prediction models*. New York: : Springer 2009.

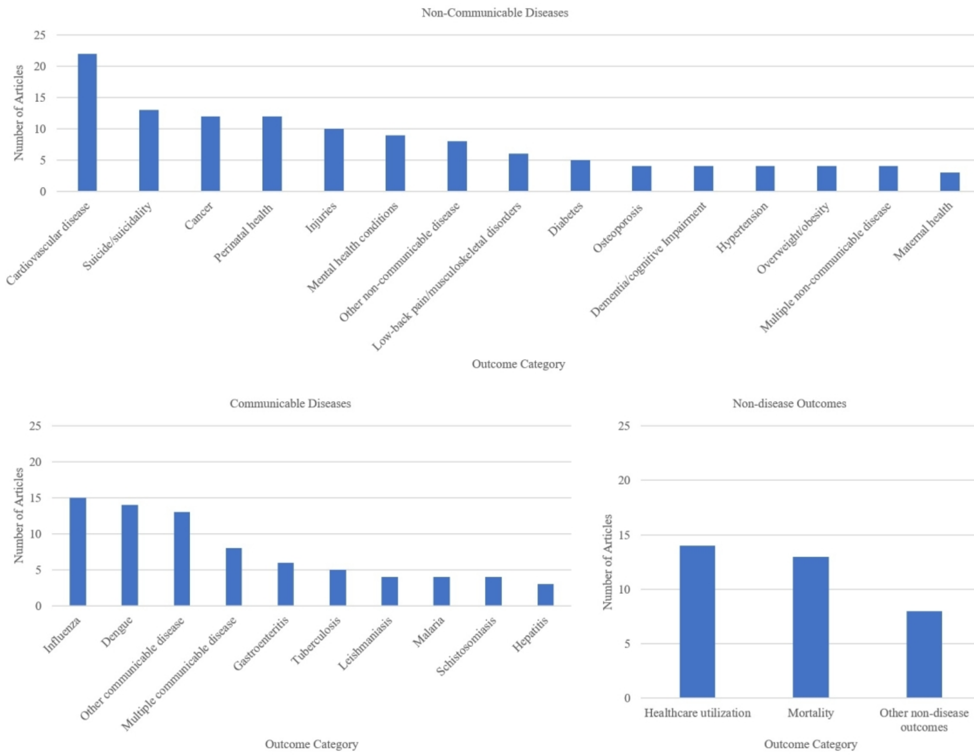
FIGURES LEGENDS

Figure 1: PRISMA flowchart of article screening process.

Figure 2: Number of articles by outcome.

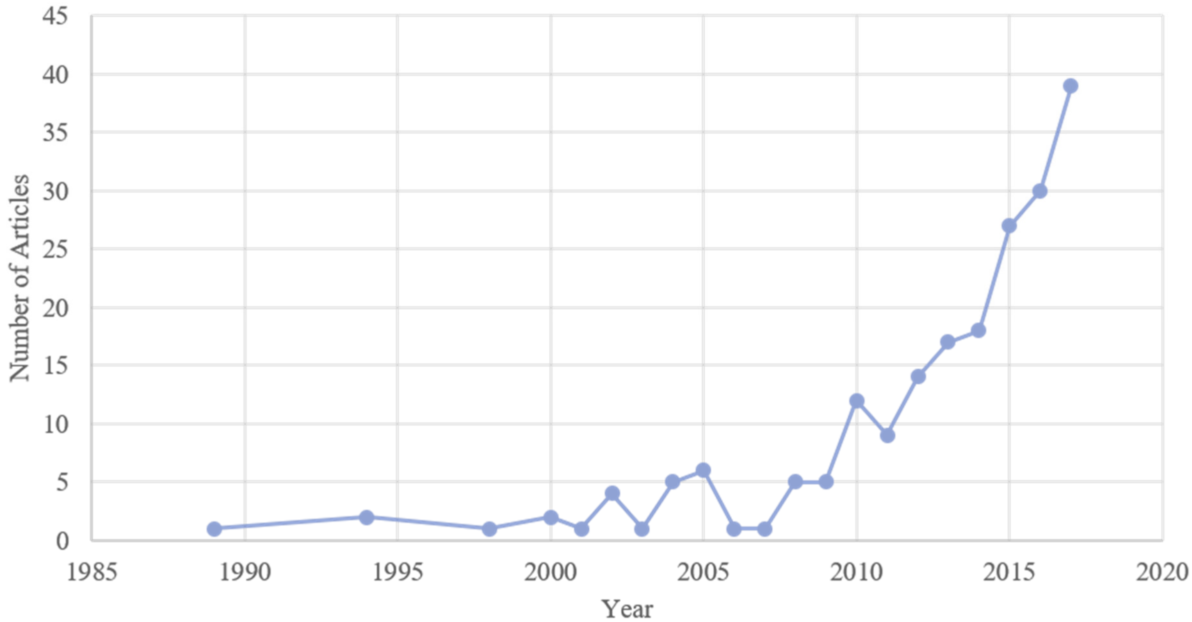


144x180mm (144 x 144 DPI)

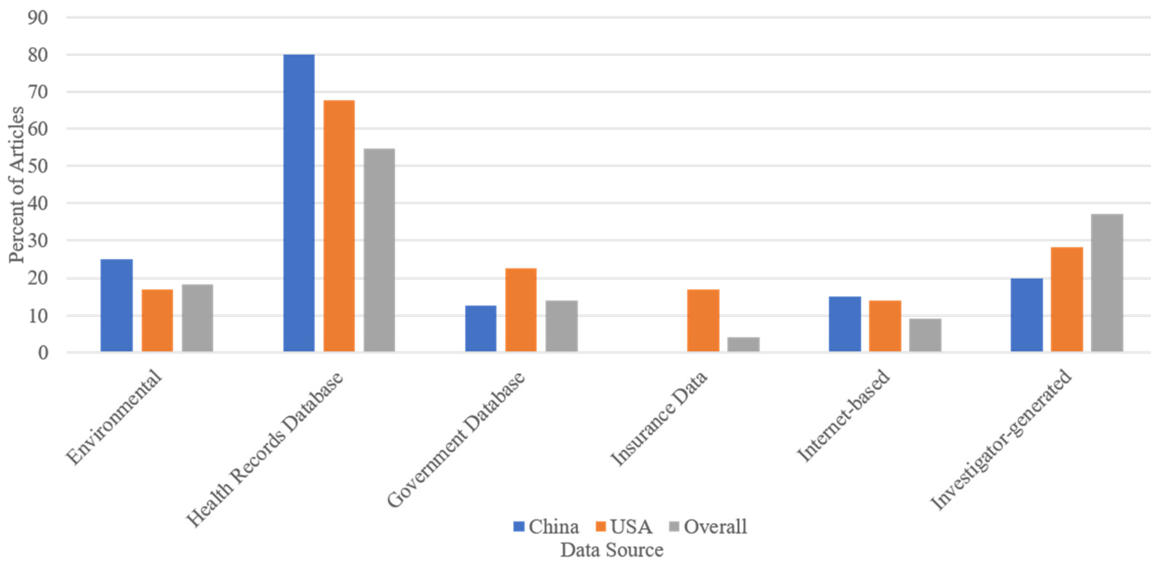


298x230mm (144 x 144 DPI)

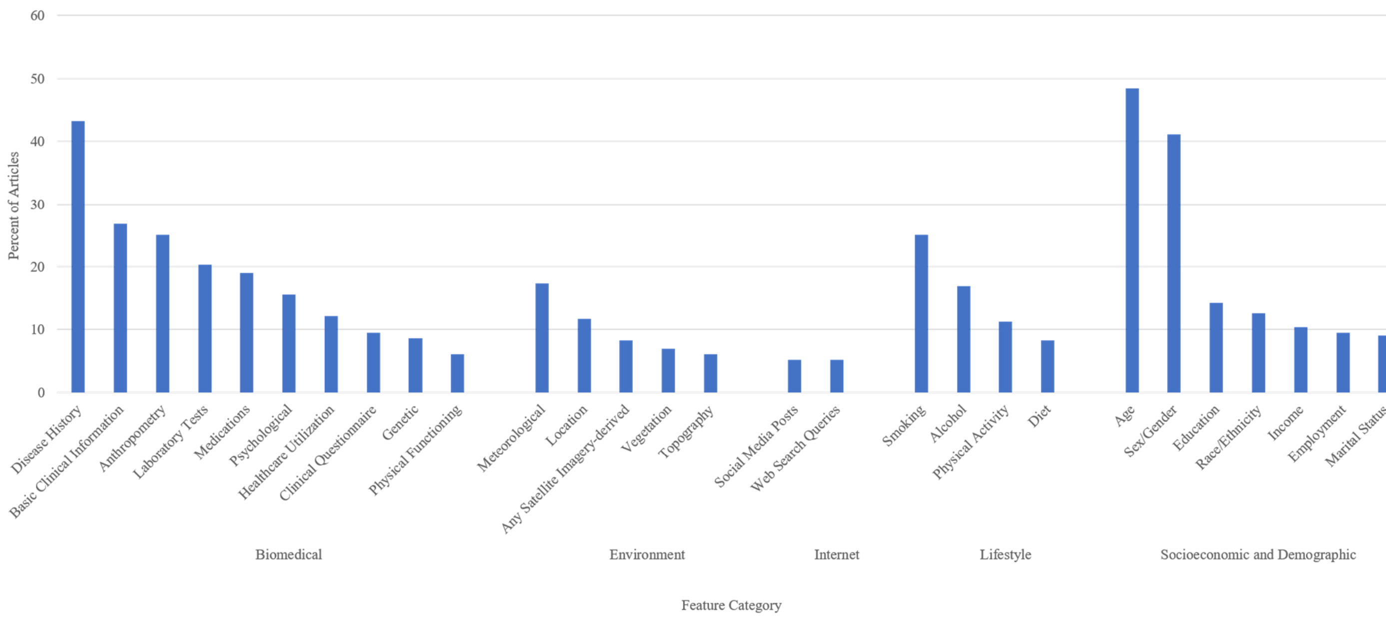
Supplementary Figure A: Number of articles by publication year. Articles from 2018 are not plotted in this figure because the review does not include all studies published that year.



Supplementary Figure B: Data Source by selected region.



Supplementary Figure C: Most commonly used feature categories.



Supplementary Table A: MEDLINE search query[1]

Machine Learning Terms	Population Health Terms
1. Exp Artificial Intelligence/	24. Exp Population Health/
2. Exp "neural networks (computer)"/	25. Exp Population Surveillance/
3. Support vector machine*.kf,tw	26. Exp Health Equity/
4. Neural net*.kf,tw	27. Health status/
5. Perceptron*.kf,tw	28. Health status disparities/
6. Deep learning.kf,tw	29. Public health systems research/
7. Random forest*.kf,tw	30. "Social determinants of health"/
8. Lasso*.kf,tw	31. Health surveys/
9. Gaussian mixture*.kf,tw	32. Health status indicators/
10. Bayesian network*.kf,tw	33. "global burden of disease"/
11. Classification tree*.kf,tw	34. Global health/
12. Regression tree*.kf,tw	35. Environmental health/
13. Relevance vector machine*.kf,tw	36. Harm reduction/
14. Nearest neighbo*.kf,tw	37. Public health informatics/
15. Probability estimation tree*.kf,tw	38. Community medicine/
16. Elastic net*.kf,tw	39. Public health/
18. Naive bayes.kf,tw	40. Epidemiology/
19. Genetic algorithm*.kf,tw	41. Preventive medicine/
20. Artificial intelligence.kf,tw	42. Occupational medicine/
21. Machine learning.kf,tw	43. Environmental medicine/
22. Statistical learning.kf,tw	44. Public health practice/
23. /or 1-22	45. Preventive health services/
	46. Health promotion/
	47. public health.kf,tw
	48. population health.kf,tw
	49. health promot*.kf,tw
	50. population surveillance.kf,tw
	51. health surveillance.kf,tw
	52. health equity.kf,tw
	53. preventive medicine.kf,tw
	54. health protection.kf,tw
	55. disease prevention.kf,tw
	56. social determinant* of health.kf,tw
	57. health determinant*.kf,tw
	58. determinant* of health.kf,tw
	59. occupational medicine.kf,tw
	60. community medicine.kf,tw
	61. epidemiolog*.kf,tw
	62. health status*.kf,tw
	63. global health.kf,tw

	64. environmental health.kf,tw
	65. harm reduction.kf,tw
	66. environmental medicine.kf,tw
	67. /or 24-66
	68. 23 and 67

[1]Limited to articles published in 1980 or after.

Supplementary Table B: Data Extraction Field Descriptions

Data Extraction Field	**Description**
Title	The article titles.
First Author	The last name and first initial of the first listed author of each article
Year of Publication	The year of publication noted for each article.
Outcome level	One of two categories: 1. *Population risk prediction:* the aggregated outcome of a whole population was predicted 2. *Individual risk prediction*: outcomes of individual participants were predicted
Outcome	Selected from the following, which are not mutually exclusive, as some articles predicted multiple outcomes: **Non-communicable Disease** 1. *Cardiovascular disease*: any disease characterized by atherosclerosis and resulting ischemia, including myocardial infarction and stroke 2. *Suicide/suicidality* 3. *Cancer* 4. *Perinatal health*: including pre-term birth, fetal alcohol spectrum disorder, congenital heart disease, growth failure, and neural tube defects 5. *Mental health conditions* 6. *Osteoporosis* 7. *Low-back pain and other musculoskeletal disorders* 8. *Diabetes* 9. *Dementia and cognitive Impairment* 10. *Hypertension* 11. *Injuries*: including fractures, falls, traffic injury, and foreign body injuries 12. *Overweight and obesity* 13. *Maternal health*: including fertility, pregnancy risk, and severe maternal morbidity 14. *Multiple non-communicable disease* 15. *Other non-communicable disease*: including liver disorders, Crohn's disease, glaucoma, dental caries, and lead poisoning **Communicable Disease** 16. *Influenza* 17. *Dengue* 18. *Gastroenteritis* 19. *Tuberculosis* 20. *Leishmaniasis* 21. *Malaria* 22. *Schistosomiasis* 23. *Hepatitis*: of viral origin 24. *Multiple communicable disease* 25. *Other communicable disease*: including zika, hand food and mouth disease, leptospirosis, yellow fever, West Nile, and typhoid fever **Non-disease Outcomes** 26. *Mortality* 27. *Healthcare utilization* 28. *Other non-disease outcomes*: including health behaviours, vitamin d status, and wellness score
Region	Categorized based on Organisation for Economic Cooperation and Development (OECD) region except for the United States and China, which were given their own categories due to the high number of publications. One of the following: *1. Africa* *2. Americas except for the United States* *3. Asia except for China* *4. China* *5. Europe* *6. Oceania* *7. United States* *8. Multi-region* *9. Other/Unknown*
Study Setting	One of two categories:

	1. *Clinical*: when data was collected in any type of clinical setting 2. *Community*: when data was collected in a community setting
Data Source Categories	Selected from the following categories, which were not mutually exclusive, and often more than one was used: 1. *Geographical Information Database*: any dataset containing basic map-based spatial information such as distances and topography 2. *Meteorological/Air Quality Datasets* 3. *Satellite Imagery*: examples include the moderate resolution imaging spectroradiometer (MODIS) and the Shuttle Radar Topography Mission (SRTM) 4. *Clinical Record Database*: any dataset produced primarily for the purpose of delivering clinical care, such as electronic medical records and administrative healthcare databases produced by hospitals 5. *Disease Registry*: a dataset maintained to monitor and/or provide care for a specific disease 6. *Population Health Survey*: a regular epidemiological survey administered periodically to assess the health of populations 7. *Reportable Disease Database*: a dataset containing reports of diseases for which it is mandatory for healthcare providers to report 8. *Other Health Records Database*: any other health records dataset not encompassed in other categories, including various surveillance systems 9. *Census* 10. *Vital Statistics*: information regularly collected by governments regarding births and deaths 11. *Other Government Database*: other governmental datasets including socioeconomic and demographic information 12. *HealthMap*: a public health surveillance system using natural language processing to analyze informal data sources such as online news, individual reports, expert-curated discussions 13. *Private Insurance Claims:* including medical, hospital, and prescription drug claims 14. *Private Insurance Questionnaires* 15. *Internet Search*: including the number of searches of certain key terms and meta data such as the location of the searches 16. *Social Media*: both posts and metadata 17. *Investigator-generated*: any datasets resulting from researcher-driven studies such as randomized controlled trials, cohort studies, and case-control studies 18. *Public Repositories*: any freely available datasets such as MIMIC 19. *Health Organization Reports*: health-related reports, typically including disease burden estimates, produced by non-governmental or governmental organizations such as the World Health Organization 20. *Not Reported*
Feature Categories	Selected from the following categories, which were not mutually exclusive, as often more than one category was used (if more than one instance of a feature category was found in an article it was only counted once): **Biomedical** 1. *Anthropometry*: measurements of the human body such as height and weight 2. *Basic Clinical Information*: information typically collected during a brief physician encounter such as a focused medical history and physical examination, including blood pressure 3. *Basic Medical Tests*: any test requiring somewhat specialized equipment such as an electrocardiogram 4. *Clinical Questionnaire*: a standardized questionnaire administered in a clinical context such as the Montreal Cognitive Assessment or Patient Health Questionnaire-9 5. *Disease History*: information regarding present and/or past diagnoses of an individual 6. *Genetic* 7. *Healthcare Utilization* 8. *Instrumental Activities of Daily Living*: features relating to an individual's daily functioning in areas such as cooking and shopping 9. *Laboratory Tests*: any features derived from human specimens requiring specialized equipment for analysis, such as hematological and microbiological results 10. *Medical Imaging* 11. *Medications* 12. *Physical Functioning*: features including the presence of any physical disabilities or the status of activities of daily living 13. *Prenatal*: relevant aspects of the period before birth such as the use of prenatal vitamins or the results of routine lab results

	14. *Psychological*: features including mood or anxiety symptoms 15. *Self-Reported Health Status* **Internet-based** 16. *Social Media Images* 17. *Social Media Location*: either aggregated or individual 18. *Social Media Metadata*: any information other than the content of social media posts, such as the frequency of general posts and time of posting 19. *Social Media Posts*: social media post content 20. *Social Network*: the interconnections among individuals in a social media platform 21. *Web Search Metadata*: any aspects of web searches other than their content 22. *Web Search Queries*: the content of web search queries either individual or aggregated **Lifestyle** 23. *Alcohol* 24. *Diet* 25. *Physical Activity* 26. *Sleep* 27. *Smoking* 28. *Unspecified* 29. *Other Substance-use* 30. *Other Lifestyle* **Environment** 31. *Air Quality* 32. *Any Satellite Imagery-derived* 33. *Biodiversity and Domestic Animals* 34. *Satellite-based Built Environment* 35. *Other Built Environment* 36. *Connectivity*: the ease of access to large urban centers and/or general services 37. *Electrical Lighting (satellite-based)* 38. *General Environmental Exposures (not included in other categories)* 39. *Hazard*: characteristics of an external hazard such as the presence of lighting on a roadway 40. *Satellite-based Land-use* 41. *Other Land-use* 42. *Location* 43. *Meteorological* 44. *Surface Water Distribution/Flooding (satellite-based)* 45. *Satellite-based Topography* 46. *Other Topography* 47. *Vector/Reservoir Characteristics*: including mosquito surveillance numbers and the population of non-human primates in the case of yellow fever 48. *Vegetation (satellite-based)*: such as the normalized difference vegetation index (NDVI) 49. *Water Composition* 50. *Other Satellite Imagery-derived* **51. *Population Disease or Healthcare Statistics*** **Socioeconomic and Demographic** 52. *Adverse Adult Experiences/Trauma* 53. *Adverse Childhood Experiences* 54. *Age* 55. *Antisocial Behaviour* 56. *Economy Makeup*: such as the number of individuals working in various types of occupations 57. *Education* 58. *Electricity* 59. *Employment* 60. *Garbage Collection* 61. *Healthcare System*: such as the availability of universal, public healthcare 62. *Household Characteristics*: the number of individuals in the household and their ages 63. *Housing Structure*: aspects of the physical structure of housing such as the number of units and age of the building 64. *Human Development Index* 65. *Immigration Status* 66. *Income* 67. *Income Inequality* 68. *Language* 69. *Legal System* 70. *Literacy*

	71. *Marital Status* 72. *Occupational Risk*: including risk factors for low-back pain such as prolonged sitting or injury from repetitive movements 73. *Parental*: including disciplinary styles and the amount of time spent at home and number of parent-child activities 74. *Peer Group*: behaviours of peer group 75. *Political Stability* 76. *Population and Population Density* 77. *Population Growth* 78. *Race/Ethnicity* 79. *Religion* 80. *Sanitation*: availability of sewage systems 81. *Sex/Gender* 82. *Social Support* 83. *Unspecified* 84. *Vehicle Ownership*: at population level 85. *Water Supply Quality* 86. *Wealth* 87. *Other Socioeconomic and Demographic* 88. ***Other Features*** 89. ***Not Reported***
Number of Datasets Used	The number of distinct datasets used regardless of the number of sources.
Dataset Availability	Selected from the following categories: 1. *Public*: all the datasets used by article authors were publicly available 2. *Closed*: all the datasets were not publicly available or appeared not to be available 3. *Closed and Public*: the datasets used were a mix of available and not available
Any Unstructured Text Used	Natural human language was included in the model as a feature with no initial ordinal/nominal structure imposed.
Number of Observations	The number of individuals or other units of observations (such as countries) included in the predictive model. If multiple subsets of the data and/or distinct datasets were used for different models, the largest number was used.
Machine Learning Algorithm Type	The algorithm type used to build the predictive model, with multiple types often used in the same article. Algorithms were only counted once when used in each article, even if used to build multiple different models in the same article. Selected from the following categories: 1. *Neural Networks*: includes deep learning/deep neural networks as well as other simpler neural networks 2. *Support Vector Machine* 3. *Single Tree-based Methods*: includes classification trees, regression trees, and decision trees 4. *Random Forest* 5. *Least Absolute Shrinkage and Selection Operator (LASSO)* 6. *Bayesian Networks*: includes naïve bayes 7. *Feature Selection Methods*: includes k-means clustering and genetic algorithms; these were often used as a pre-processing step and in a few cases this was the only use of machine learning (i.e. a machine learning model was not used to build the predictive model itself) 8. *Boosted Tree-based Methods*: includes gradient boosting and boosted trees 9. *K-Nearest Neighbour* 10. *Elastic Net* 11. *Ridge Regression* 12. *Other*: includes association rule learning, single task learning, multitask learning, rough set classifier, associative classification, bagging, partial least squares discriminant analysis, Just-Add-Data Bio Tool, super learner, particle swarm optimization, ant colony optimization, Isomap, PCA, Disease State Index, Stacking, kernel conditional density estimation, stepwise deletion, conditional random fields, contrast mining, grammatical evolution, Learning from Examples Using ROugh Sets, AUtoregression with exogenous outputs, and natural language processing
Compared with Other Statistical Methods	Whether the machine learning method's predictive performance was compared with a traditional parametric statistical regression model such as logistic regression (yes/no).
Reported Data Pre-processing	Whether any aspects of data cleaning or pre-processing were reported (yes/no). Examples include how missing data was handled, whether log transformations were done, and if derived variables were generated. Missing data and all model development processes have been identified as important to report by TRIPOD.[1]
Reported Method of Feature Selection	Whether the method of feature selection was reported (yes/no). When there is a high number of features initially, this is usually done using algorithmic, domain knowledge-

	informed, or mixed approaches. Feature selection is an important element of reporting as identified by TRIPOD.[1]
Number of Features	The number of features included in the final prediction model after feature selection. If multiple models were used in one article, the largest number of features was chosen.
Reported Hyper-parameter Search	Whether the process for determining the hyper-parameters of the machine learning model, such as the number of features used to build each tree in a random forest, was reported (yes/no). This is an important aspect of model development[2], and thus considered an important element to report by the TRIPOD statement.[1]
Method of Validation	How the authors validated the predictive performance of their model, selected from one of the following categories: 1. *Holdout*: the dataset was divided into two parts; one part was used to train the model and the other was used to test the model 2. *Cross-validation and bootstrap*: the dataset was either divided into more than two parts and repeatedly trained and tested on different parts of the dataset or random sampling with replacement was used to train the model 3. *External*: the model was tested on a completely separate dataset
Reported Descriptive Statistics	Whether the article reported any descriptive statistics regarding their sample (yes/no). We considered a broad array of descriptive statistics including sample population demographics, feature distributions, and outcome distributions. These are all important reporting elements according to TRIPOD.[1]
Calibration Metrics	The types of calibration predictive performance metrics used to evaluate models, which could be more than one. Calibration refers to a model's ability to accurately predict absolute probabilities of the outcome occurring.[3] One or more of the following categories was selected if a calibration metric was used: 1. *Manual or visual comparison*: includes calibration plots 2. *Hosmer-Lemeshow* 3. *Observed/Expected*: is a ratio or comparison of observed and predicted/expected probabilities 4. *Other calibration metric*: includes mean bias (from Bland-Altman plot), calibration factoring, calibration statistic
Discrimination Metrics	The types of discrimination predictive performance metrics used to evaluate models, which could be more than one. Discrimination refers to a model's ability to correctly rank-order individuals according to their likelihood of developing the outcome.[3] One or more of the following categories was selected if a discrimination metric was used: 1. *Area under the curve*: meaning receiver operator curve; also includes c-statistic and s-index 2. *Accuracy*: includes accuracy, misclassification, and error rate 3. *Recall*: includes sensitivity, specificity, true/false positive, and true/false negative 4. *Precision*: includes positive predictive value, negative predictive value, and precision 5. *F statistics* 6. *Likelihood Ratio*: includes both positive and negative likelihood ratios 7. *Youden Index* 8. *Manual or visual comparison* 9. *Other discrimination metric*: includes G-means, k-statistic, and Matthews correlation coefficient
Overall Goodness of Fit Metrics	The types of overall goodness of fit performance metrics used to evaluate models, which could be more than one. Overall goodness of fit refers to a model's predictions' concordance with observed outcomes. One or more of the following categories was selected if an overall performance metric was used: 1. *Root mean squared error* 2. *Mean squared error* 3. *Mean absolute error* 4. *Mean absolute percentage error* 5. R^2: includes pseudo-R2s 6. *Correlation* 7. *Akaike Information Criterion or Bayesian Information Criterion* 8. *Other performance metric*: includes penalty error, total sum of squares, proportional reduction in error, overall prediction error, specific prediction error, Nash-Sutcliffe, root mean squared percentage error, mean relative absolute error, analysis of variance F-stat, -2LogLikelihood, relative efficiency, deviance, Ljung-Box test, mean absolute deviation, standard error, Brier score, log score, and mean percentage error
Did Machine Learning Models Outperform Traditional Methods?	Whether the machine learning-based predictive models outperformed the statistical parametric regression models based on the performance metrics supplied by the authors (yes/no). However, this should not be taken to mean that the difference in model performance was reliable or valid. Often, important performance metrics and essential aspects of model development were not reported, making accurate comparisons difficult.

Discussed the Practical Costs of Prediction Errors	Whether the article discussed the relative risks of false negative and false positive results based on their predictive model in contexts where it might be used (yes/no). These costs are important for determining the usefulness and application of predictive models.[3]
Stated Rationale for Using Machine Learning	Whether the article stated any reasons for using a machine learning approach instead of a statistical parametric regression approach (yes/no).
Rationale for Using Machine Learning - Free Text	Reviewers included article quotations and summaries in this section to capture different rationales for using machine learning. Reviewers attempted to only extract free text regarding each specific type of rationale once
Discussed Model Usability	Whether the article discussed any aspect of how the model could be practically used in a relevant context (yes/no).
Stated Model Limitations	Whether the article discussed any potential limitations of the research (yes/no).
Model limitations - Free Text	Reviewers included article quotations and summaries in this section to capture different reported limitations. Reviewers attempted to only extract free text regarding each specific type of limitation once.
Discussed Model Implementation	Whether the article included discussion of any consequences of model implementation such as potential clinical, population-health, and policy-level impacts (yes/no).
Model Implementation - Free Text	Reviewers included article quotations and summaries in this section to capture different reported consequences of model implementation. Reviewers attempted to only extract free text regarding each specific type of implementation impact once.

Supplementary Table C: Types of machine learning algorithms used.

Types of Algorithms	Number	Percent
Neural Networks[*]	95	41.13%
Support Vector Machine	59	25.54%
Single tree-based methods[†]	52	22.51%
Random Forest	48	20.78%
LASSO	25	10.82%
Bayesian Networks[‡]	23	9.96%
Feature selection methods[§]	20	8.66%
Boosted tree-based methods[¶]	19	8.23%
K-Nearest Neighbour	19	8.23%
Elastic Net	9	3.90%
Ridge regression	5	2.16%
Other[‖]	22	9.52%

[*]Includes deep neural networks.
[†]Includes CART, decision trees.
[‡]Includes naive bayes.
[§]Includes cluster methods (e.g. k-means clustering) and genetic algorithms.
[¶]Includes gradient boosting and boosted trees.
[‖]Including (all algorithms used once unless otherwise specified) association rule learning (n=3), single task learning, multitask learning, rough set classifier, associative classification, bagging, partial least squares discriminant analysis, Just-Add-Data Bio Tool, super learner, particle swarm optimization, ant colony optimization, isomap, principal components analysis, disease state Index, stacking, kernel conditional density estimation, stepwise deletion, conditional random fields, contrast mining, grammatical evolution, Learning from Examples Using ROugh Sets, AUtoregression with exogenous outputs, and natural language processing (n=2).

Supplementary Table D: Detailed feature categories included in studies.

Feature Category	Number of Articles	Percent
Biomedical	141	61.04
Anthropometry	58	25.11
Basic Clinical Information	62	26.84
Basic Medical Tests	10	4.33
Clinical Questionnaire	22	9.52
Disease History	100	43.29
Genetic	20	8.66
Healthcare Utilization	28	12.12
Instrumental Activities of Daily Living	6	2.60
Laboratory Tests	47	20.35
Medical Imaging	10	4.33
Medications	44	19.05
Physical Functioning	14	6.06
Prenatal	10	4.33
Psychological	36	15.58
Self-Reported Health Status	7	3.03
Internet-based	21	9.09
Social Media Images	1	0.43
Social Media Location	5	2.16
Social Media Metadata	4	1.73
Social Media Posts	12	5.19
Social Network	3	1.30
Web Search Metadata	1	0.43
Web Search Queries	12	5.19
Lifestyle	81	35.06
Alcohol	39	16.88
Diet	19	8.23
Physical Activity	26	11.26
Sleep	11	4.76
Smoking	58	25.11
Unspecified	4	1.73
Other Substance-use	13	5.63
Other Lifestyle	13	5.63
Environment	82	35.50
Air Quality	5	2.16
Any Satellite Imagery-derived	19	8.23
Biodiversity and Domestic Animals	2	0.87
Built Environment	8	3.46
Satellite	4	1.73
Other	4	1.73

Connectivity	4	1.73
Electrical Lighting[1]	1	0.43
General Environmental Exposures (not included in other categories)	9	3.90
Hazard	10	4.33
Land-use	2	0.87
Satellite	1	0.43
Other	1	0.43
Location	27	11.69
Meteorological	40	17.32
Surface Water Distribution/Flooding[1]	6	2.60
Topography	14	6.06
Satellite	12	5.19
Other	2	0.87
Vector/Reservoir Characteristics	9	3.90
Vegetation[1]	16	6.93
Water Composition	1	0.43
Other Satellite Imagery-derived	7	3.03
Population-level Disease or Healthcare Statistics	38	16.45
Socioeconomic and Demographic Factors	150	64.94
Adverse Adult Experiences/Trauma	5	2.16
Adverse Childhood Experiences	4	1.73
Age	112	48.48
Antisocial Behaviour	2	0.87
Economy Makeup	1	0.43
Education	33	14.29
Electricity	2	0.87
Employment	22	9.52
Garbage Collection	1	0.43
Healthcare System	5	2.16
Household Characteristics	10	4.33
Housing Structure	4	1.73
Human Development Index	1	0.43
Immigration Status	5	2.16
Income	24	10.39
Income Inequality	3	1.30
Language	2	0.87
Legal System	1	0.43
Literacy	2	0.87
Marital Status	21	9.09
Occupational Risk	10	4.33
Parental	3	1.30
Peer Group	1	0.43
Political Stability	1	0.43

Population and Population Density	11	4.76
Population Growth	2	0.87
Race/Ethnicity	29	12.55
Religion	3	1.30
Sanitation	5	2.16
Sex/Gender	95	41.13
Social Support	10	4.33
Unspecified	6	2.60
Vehicle Ownership	2	0.87
Water Supply Quality	5	2.16
Wealth	2	0.87
Other Socioeconomic and Demographic	29	12.55
Other Features	17	7.36
Not Reported	1	0.43

[1]See supplementary table B for greater detail regarding feature categories
[2]Satellite-derived

Appendix A: Eligibility Criteria

The following types of articles were excluded:

- Reviews;
- Focused on a methodological development;
- Only included an abstract;
- Only used linear regression, logistic regression, generalized additive models, or other approaches not considered machine learning for the purpose of this review;
- Only applied models to diagnosis, treatment decisions, or prognosis of individuals who already had a disease;
- Only related to logistics, human resources, finance, or management involved in provision of public health services;
- Focused on occupational health, traffic accidents, or environmental monitoring, with no direct link to population health outcomes;
- Used smart home or home monitoring systems;
- Used advanced imaging or other expensive predictors that would be difficult or unsafe to scale to a population level;
- Focused on clinical decision support systems;
- Predicted adverse drug effects, except vaccines.

Appendix B: Examples of article titles removed during title screening

1. Improved classification of mangroves health status using hyperspectral remote sensing data
2. Diesel engine and propulsion diagnostics of a mini-cruise ship by using artificial neural networks
3. Relationship between benthic macroinvertebrate bio-indices and physicochemical parameters of water: A tool for water resources managers
4. Adaptive one-switch row-column scanning
5. Development of a distributed bearing health monitoring and assessing system
6. Neural networks based sensor validation and recovery methodology for advanced aircraft engines
7. Mining images in publicly-available cameras for homeland security
8. The human pulvinar and attentional processing of visual distractors
9. Text classification techniques in oil industry applications
10. Research on acoustic mechanical fault diagnosis method of high voltage circuit breaker based on improved EEMD

Appendix C: All studies included in the review.

1 Achrekar H, Gandhe A, Lazarus R, *et al. Predicting Flu Trends using Twitter data*. 2011. doi:10.1109/INFCOMW.2011.5928903

2 Adamou M, Antoniou G, Greasidou E, *et al.* Mining Free-Text Medical Notes for Suicide Risk Assessment. Proc. 10th Hell. Conf. Artif. Intell. 2018. doi:10.1145/3200947.3201020

3 Adams LJ, Bello G, Dumancas GG. Development and Application of a Genetic Algorithm for Variable Optimization and Predictive Modeling of Five-Year Mortality Using Questionnaire Data. *Bioinform Biol Insights* 2015;**9**:31–41. doi:https://dx.doi.org/10.4137/BBI.S29469

4 Agarwal A, Baechle C, Behara RS, *et al.* Multi-method approach to wellness predictive modeling. *J Big Data* 2016;**3**:1–23. doi:http://dx.doi.org/10.1186/s40537-016-0049-0

5 Agarwal V, Zhang L, Zhu J, *et al.* Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. *J Med Internet Res* 2016;**18**:e251.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=27655225

6 Agopian AJ, Lupo PJ, Tinker SC, *et al.* Working towards a risk prediction model for neural tube defects. *Birth Defects Res A Clin Mol Teratol* 2012;**94**:141–6. doi:https://dx.doi.org/10.1002/bdra.22883

7 Ahn C, Hwang Y, Park SK. Predictors of all-cause mortality among 514,866 participants from the Korean National Health Screening Cohort. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0185458

8 Aichele S, Rabbitt P, Ghisletta P. Illness and intelligence are comparatively strong predictors of individual differences in depressive symptoms following middle age. *Aging Ment Health* 2017;:1–10. doi:https://dx.doi.org/10.1080/13607863.2017.1394440

9 Akbulut A, Ertugrul E, Topcu V. Fetal health status prediction based on maternal clinical history using machine learning techniques. *Comput Methods Programs Biomed* 2018;**163**:87–100. doi:http://dx.doi.org/10.1016/j.cmpb.2018.06.010

10 Akhavan P, Karimi M, Pahlavani P, *et al.* Risk mapping of Cutaneous Leishmaniasis via a fuzzy C Means-based Neuro-Fuzzy inference system. 2014;**40**:19–23. doi:10.5194/isprsarchives-XL-2-W3-19-2014

11 Alby S, Shivakumar BL. A prediction model for type 2 diabetes risk among Indian women. *ARPN J Eng Appl Sci* 2016;**11**:2037–43.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959387072&partnerID=40&md5=0fde9764a6290488b1c3472e2bbb5f7c NS -

12 Allen T, Murray KA, Zambrana-Torrelio C, *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun* 2017;**8**:1124. doi:https://dx.doi.org/10.1038/s41467-017-00923-8

13 Allore H, Tinetti ME, Araujo KLB, *et al.* A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and Social and Productive Activities scores. *J Clin Epidemiol* 2005;**58**:154–61.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15680749

14 Al-Mallah MH, Elshawi R, Ahmed AM, *et al.* Using Machine Learning to Define the Association between Cardiorespiratory Fitness and All-Cause Mortality (from the Henry Ford Exercise Testing Project). *Am J Cardiol* 2017;**120**:2078–84. doi:10.1016/j.amjcard.2017.08.029

15 Almeida AS, Werneck GL. Prediction of high-risk areas for visceral leishmaniasis using socioeconomic indicators and remote sensing data. *Int J Health Geogr* 2014;**13**. doi:10.1186/1476-072X-13-13

16 Alves EB, Costa CHN, de Carvalho FAA, *et al.* Risk Profiles for Leishmania infantum Infection in Brazil. *Am J Trop Med Hyg* 2016;**94**:1276–81. doi:10.4269/ajtmh.15-0513

17 Amini P, Ahmadinia H, Poorolajal J, *et al.* Evaluating the High Risk Groups for Suicide: A Comparison of Logistic Regression, Support Vector Machine, Decision Tree and Artificial Neural Network. *Iran J Public Health* 2016;**45**:1179–87.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5149472/pdf/IJPH-45-1179.pdf NS -

18 Amini P, Maroufizadeh S, Samani RO, *et al.* Factors Associated with Macrosomia among Singleton Live-births: A Comparison between Logistic Regression, Random Forest and Artificial Neural Network Methods. *Epidemiol Biostat Public Heal* 2016;**13**. doi:10.2427/11985

19 Anand A, Shakti D. Prediction of diabetes based on personal lifestyle indicators. 2015;:673–6. doi:10.1109/NGCT.2015.7375206

20 Anderson RT, Balkrishnan R, Camacho F. Risk classification of Medicare HMO enrollee cost levels using a decision-tree approach. *Am J Manag Care* 2004;**10**:89–98.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15011809

21 Asensio-Cuesta S, Diego-Mas JA, Alcaide-Marzal J. Applying generalised feedforward neural networks to

classifying industrial jobs in terms of risk of low back disorders. *Int J Ind Ergon* 2010;**40**:629–35. doi:10.1016/j.ergon.2010.04.007

22 Ayyagari R, Vekeman F, Lefebvre P, *et al.* Pulse pressure and stroke risk: development and validation of a new stroke risk model. *Curr Med Res Opin* 2014;**30**:2453–60. doi:10.1185/03007995.2014.971357

23 Azeez A, Obaromi D, Odeyemi A, *et al.* Seasonality and Trend Forecasting of Tuberculosis Prevalence Data in Eastern Cape, South Africa, Using a Hybrid Model. *Int J Environ Res Public Health* 2016;**13**. doi:10.3390/ijerph13080757

24 Bakar AA, Kefli Z, Abdullah S, *et al.* Predictive models for dengue outbreak using multiple rulebase classifiers. 2011;:5 pp. doi:10.1109/ICEEI.2011.6021830

25 Balaraman S, Schafer JJ, Tseng AM, *et al.* Plasma miRNA Profiles in Pregnant Women Predict Infant Outcomes following Prenatal Alcohol Exposure. *PLoS One* 2016;**11**. doi:http://dx.doi.org/10.1371/journal.pone.0165081

26 Bandyopadhyay S, Wolfson J, Vock DM, *et al.* Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Min Knowl Discov* 2015;**29**:1033–69. doi:10.1007/s10618-014-0386-6

27 Bashir S, Qamar U, Khan FH. A Multicriteria Weighted Vote-Based Classifier Ensemble for Heart Disease Prediction. *Comput Intell* 2016;**32**:615–45. doi:http://dx.doi.org/10.1111/coin.12070

28 Bath PA, Pendleton N, Morgan K, *et al.* New approach to risk determination: development of risk profile for new falls among community-dwelling older people by use of a Genetic Algorithm Neural Network (GANN). *J Gerontol A Biol Sci Med Sci* 2000;**55**:M17-21.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=10719768

29 Bayati M, Bhaskar S, Montanari A. Statistical analysis of a low cost method for multiple disease prediction. *Stat Methods Med Res* 2018;**27**:2312–28. doi:http://dx.doi.org/10.1177/0962280216680242

30 Ben-Ari A, Hammond K. Text Mining the EMR for Modeling and Predicting Suicidal Behavior among US Veterans of the 1991 Persian Gulf War. 2015;:3168–75. doi:10.1109/HICSS.2015.382

31 Berchialla P, Scarinzi C, Snidero S, *et al.* Comparing models for quantitative risk assessment: an application to the European Registry of foreign body injuries in children. *Stat Methods Med Res* 2016;**25**:1244–59. doi:10.1177/0962280213476167

32 Bertsimas D, Bjarnadottir M V, Kane MA, *et al.* Algorithmic Prediction of Health-Care Costs. *Oper Res* 2008;**56**:1382–92. doi:10.1287/opre.1080.0619

33 Bhatt S, Cameron E, Flaxman SR, *et al.* Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *J R Soc Interface* 2017;**14**:20170520. doi:http://dx.doi.org/10.1098/rsif.2017.0520

34 Bhatt S, Gething PW, Brady OJ, *et al.* The global distribution and burden of dengue. *Nature* 2013;**496**:504–7.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1350966952?accountid=12347

35 Bi J, Sun J, Wu Y, *et al.* A machine learning approach to college drinking prediction and risk factor identification. *ACM Trans Intell Syst Technol* 2013;**4**:1–24. doi:10.1145/2508037.2508053

36 Bibi H, Nutman A, Shoseyov D, *et al.* Prediction of emergency department visits for respiratory symptoms using an artificial neural network. *Chest* 2002;**122**:1627–32.http://myaccess.library.utoronto.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=106806480&site=ehost-live NS -

37 Blankenberg S, Zeller T, Saarela O, *et al.* Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: the MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. *Circulation* 2010;**121**:2388–97. doi:https://dx.doi.org/10.1161/CIRCULATIONAHA.109.901413

38 Braithwaite SR, Giraud-Carrier C, West J, *et al.* Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. *JMIR Ment Heal* 2016;**3**:e21. doi:https://dx.doi.org/10.2196/mental.4822

39 Buczak AL, Baugher B, Guven E, *et al.* Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC Med Inform Decis Mak* 2015;**15**:47. doi:https://dx.doi.org/10.1186/s12911-015-0170-6

40 Buczak AL, Koshute PT, Babin SM, *et al.* A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. *BMC Med Inform Decis Mak* 2012;**12**:124. doi:https://dx.doi.org/10.1186/1472-6947-12-124

41 Caillet P, Klemm S, Ducher M, *et al.* Hip Fracture in the Elderly: A Re-Analysis of the EPIDOS Study with

Causal Bayesian Networks. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0120125
42 Cao S, Wang F, Tam W, *et al.* A hybrid seasonal prediction model for tuberculosis incidence in China. *BMC Med Inform Decis Mak* 2013;**13**:56. doi:https://dx.doi.org/10.1186/1472-6947-13-56
43 Caon F, Meneghel G, Zaghi P, *et al.* Applicability of neural networks to suicidological research: a pilot study. *Arch suicide Res* 2002;**6**:285–9.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/38408407?accountid=12347
44 Chae S, Kwon S, Lee D. Predicting Infectious Disease Using Deep Learning and Big Data. *Int J Env Res Public Heal* 2018;**15**. doi:10.3390/ijerph15081596
45 Chang JS, Yeh RF, Wiencke JK, *et al.* Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. *Cancer Epidemiol Biomarkers Prev* 2008;**17**:1368–73. doi:10.1158/1055-9965.EPI-07-2830
46 Chen Y, Chu CW, Chen MIC, *et al.* The utility of LASSO-based models for real time forecasts of endemic infectious diseases: a cross country comparison. *J Biomed Inform* Published Online First: 2018. doi:https://dx.doi.org/10.1016/j.jbi.2018.02.014
47 Chen Y-T, Miller PD, Barrett-Connor E, *et al.* An approach for identifying postmenopausal women age 50-64 years at increased short-term risk for osteoporotic fracture. *Osteoporos Int* 2007;**18**:1287–96.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17464525
48 Chenar SS, Deng Z. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res* 2018;**128**:20–37. doi:10.1016/j.watres.2017.10.032
49 Chenar SS, Deng Z. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ Int* 2018;**111**:212–23. doi:https://dx.doi.org/10.1016/j.envint.2017.11.032
50 Cheng Y, Jiang T, Zhu M, *et al.* Risk assessment models for genetic risk predictors of lung cancer using two-stage replication for Asian and European populations. *Oncotarget* 2017;**8**:53959–67. doi:https://dx.doi.org/10.18632/oncotarget.10403
51 Choi SB, Lee W, Yoon J-H, *et al.* Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. *J Affect Disord* 2018;**231**:8–14. doi:https://dx.doi.org/10.1016/j.jad.2018.01.019
52 Conrad D, Wilker S, Pfeiffer A, *et al.* Does trauma event type matter in the assessment of traumatic load? *Eur J Psychotraumatol* 2017;**8**. doi:10.1080/20008198.2017.1344079
53 Courtney KL, Stewart S, Popescu M, *et al.* Predictors of preterm birth in birth certificate data. *Stud Health Technol Inform* 2008;**136**:555–60.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=18487789
54 Crawford AG, Fuhr Jr JP, Clarke J, *et al.* Comparative effectiveness of total population versus disease-specific neural network models in predicting medical costs. *Dis Manag* 2005;**8**:277–87. doi:10.1089/dis.2005.8.277
55 Dartois L, Gauthier E, Heitzmann J, *et al.* A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort. *Breast Cancer Res Treat* 2015;**150**:415–26. doi:https://dx.doi.org/10.1007/s10549-015-3321-7
56 Darvishi E, Khotanlou H, Khoubi J, *et al.* Prediction Effects of Personal, Psychosocial, and Occupational Risk Factors on Low Back Pain Severity Using Artificial Neural Networks Approach in Industrial Workers. *J Manipulative Physiol Ther* 2017;**40**:486–93. doi:https://dx.doi.org/10.1016/j.jmpt.2017.03.012
57 Davidson MW, Haim DA, Radin JM. Using networks to combine 'big data' and traditional surveillance to improve influenza predictions. *Sci Rep* 2015;**5**:8154. doi:http://dx.doi.org/10.1038/srep08154
58 De Choudhury M, Gamon M, Counts S, *et al.* Predicting depression via social media. 2013;:128–37.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84900027892&partnerID=40&md5=37755eb6e5e5e0639df4438f14bf1c8f NS -
59 DeClaris N, Hammad T, Wahab AF, *et al.* Neural network models for studying and for managing in real-time schistosomiasis control programs. 1994;**vol.2**:1362–3. doi:10.1109/IEMBS.1994.415473
60 Delgado-Gomez D, Blasco-Fontecilla H, Sukno F, *et al.* Suicide attempters classification: Toward predictive models of suicidal behavior. *Neurocomputing An Int J* 2012;**92**:3–8. doi:http://dx.doi.org/10.1016/j.neucom.2011.08.033
61 Dicken RA, Fazle Rubby SAM, Naz S, *et al.* Risk assessment of the top five malignancies among males and females with respect to occupation, educational status and smoking habits. 2016;:6 pp. doi:10.1109/ICIS.2016.7550756
62 Dierker LC, Avenevoli S, Goldberg A, *et al.* Defining subgroups of adolescents at risk for experimental and

regular smoking. *Prev Sci* 2004;**5**:169–83. doi:10.1023/B:PREV.0000037640.66607.6b

63 Dinh TQ, Le H V, Cao TH, *et al.* Forecasting the magnitude of dengue in southern vietnam. 2016;**pt.I**:554–63. doi:10.1007/978-3-662-49381-6_53

64 Dogan M V, Grumbach IM, Michaelson JJ, *et al.* Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLoS One* 2018;**13**:e0190549. doi:http://dx.doi.org/10.1371/journal.pone.0190549

65 Dugan TM, Mukhopadhyay S, Carroll A, *et al.* Machine Learning Techniques for Prediction of Early Childhood Obesity. *Appl Clin Inform* 2015;**6**:506–20. doi:http://dx.doi.org/10.4338/ACI-2015-03-RA-0036

66 Duh MS, Walker AM, Pagano M, *et al.* Prediction and cross-validation of neural networks versus logistic regression: Using hepatic disorders as an example. *Am J Epidemiol* 1998;**147**:407–13.NS -

67 Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput Appl* 2018;**29**:685–93. doi:10.1007/s00521-016-2604-1

68 Dykxhoorn J, Hatcher S, Roy-Gagnon M-H, *et al.* Early life predictors of adolescent suicidal thoughts and adverse outcomes in two population-based cohort studies. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0183182

69 Félix Caballero F, Soulis G, Engchuan W, *et al.* Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: The ATHLOS project. *Sci Rep* 2017;**7**. doi:10.1038/srep43955

70 Gan R, Chen N, Huang D. Comparisons of forecasting for hepatitis in Guangxi Province, China by using three neural networks models. *PeerJ* 2016;**2016**:2684. doi:http://dx.doi.org/10.7717/peerj.2684

71 Gan X, Xu Y, Liu L, *et al.* Predicting the incidence risk of ischemic stroke in a hospital population of southern China: A classification tree analysis. *J Neurol Sci* 2011;**306**:108–14. doi:http://dx.doi.org/10.1016/j.jns.2011.03.032

72 Ganga GMD, Esposto KF, Braatz D. Application of discriminant analysis-based model for prediction of risk of low back disorders due to workplace design in industrial jobs. *Work J Prev Assess Rehabil* 2012;**41**:2370–6. doi:10.3233/WOR-2012-0467-2370

73 Gerardi DO, Monteiro LHA. System identification and prediction of dengue fever incidence in Rio de Janeiro. *Math Probl Eng* 2011;:720304 (13 pp.). doi:10.1155/2011/720304

74 Girela JL, Gil D, Johnsson M, *et al.* Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence methods. *Biol Reprod* 2013;**88**:99. doi:https://dx.doi.org/10.1095/biolreprod.112.104653

75 Gismondi RC, Almeida RMVR, Infantosi AFC. Artificial neural networks for infant mortality modelling. *Comput Methods Programs Biomed* 2002;**69**:237–47.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12204451

76 Gueli N, Piccirillo G, Troisi G, *et al.* The influence of lifestyle on cardiovascular risk factors: Analysis using a neural network. *Arch Gerontol Geriatr* 2005;**40**:157–72. doi:10.1016/j.archger.2004.08.002

77 Guo P, Liu T, Zhang Q, *et al.* Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis* 2017;**11**:e0005973. doi:https://dx.doi.org/10.1371/journal.pntd.0005973

78 Guo S, Lucas RM, Ponsonby A-L, *et al.* A novel approach for prediction of vitamin d status using support vector regression. *PLoS One* 2013;**8**:e79970. doi:https://dx.doi.org/10.1371/journal.pone.0079970

79 Harrigan RJ, Thomassen HA, Buermann W, *et al.* Economic Conditions Predict Prevalence of West Nile Virus. *PLoS One* 2010;**5**. doi:http://dx.doi.org/10.1371/journal.pone.0015437

80 Holmes JH, Durbin DR, Winston FK. Discovery of predictive models in an injury surveillance database: an application of data mining in clinical research. *Proceedings AMIA Symp* 2000;:359–63.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=11079905

81 Hu X, Quirchmayr G, Winiwarter W, *et al.* Influenza early warning model based on Yunqi theory. *Chin J Integr Med* 2012;**18**:192–6. doi:10.1007/s11655-012-1003-4

82 Hu ZK, Jin B, Shin AY, *et al.* Real-Time Web-Based Assessment of Total Population Risk of Future Emergency Department Utilization: Statewide Prospective Active Case Finding Study. *Interact J Med Res* 2015;**4**:39–51. doi:10.2196/ijmr.4022

83 Husam IS, Abuhamad AAB, Zainudin S, *et al.* Feature selection algorithms for Malaysian dengue outbreak detection model. *Sains Malaysiana* 2017;**46**:255–65. doi:10.17576/jsm-2017-4602-10

84 Husin NA, Mustapha N, Sulaiman MN, *et al.* Performance of hybrid GANN in comparison with other standalone models on dengue outbreak prediction. *J Comput Sci* 2016;**12**:300–6. doi:10.3844/jcssp.2016.300.306

85 Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity

prediction. *Accid Anal Prev* 2017;**108**:27–36. doi:10.1016/j.aap.2017.08.008
86 Jafari SA, Jahandideh S, Jahandideh M, *et al.* Prediction of road traffic death rate using neural networks optimised by genetic algorithm. *Int J Inj Contr Saf Promot* 2015;**22**:153–7. doi:https://dx.doi.org/10.1080/17457300.2013.857695
87 Jiang D, Hao M, Ding F, *et al.* Mapping the transmission risk of Zika virus using machine learning models. *Acta Trop* 2018;**185**:391–9. doi:https://dx.doi.org/10.1016/j.actatropica.2018.06.021
88 Joseph A, Ramamurthy B. Suicidal behavior prediction using data mining techniques. *Int J Mech Eng Technol* 2018;**9**:293–301.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045951449&partnerID=40&md5=74f2c78c8b63037785fb7ed22ef9b8a1 NS -
89 Juan YC, Chen CM, Chen SH. A classifier fusion approach to osteoporosis prediction for women in Taiwan. *J Ind Prod Eng* 2015;**32**:360–8. doi:10.1080/21681015.2015.1064484
90 Karstoft K-I, Statnikov A, Andersen SB, *et al.* Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers. *J Affect Disord* 2015;**184**:170–5. doi:https://dx.doi.org/10.1016/j.jad.2015.05.057
91 Kennedy EH, Wiitala WL, Hayward RA, *et al.* Improved Cardiovascular Risk Prediction Using Nonparametric Regression and Electronic Health Record Data. *Med Care* 2013;**51**:251–8. doi:10.1097/MLR.0b013e31827da594
92 Kesorn K, Ongruk P, Chompoosri J, *et al.* Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0125049
93 Kessler RC, Hwang I, Hoffmire CA, *et al.* Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res* 2017;**26**:e1575. doi:10.1002/mpr.1575
94 Kessler RC, Rose S, Koenen KC, *et al.* How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. *World Psychiatry* 2014;**13**:265–74. doi:http://dx.doi.org/10.1002/wps.20150
95 Kim BJ, Kim SH. Prediction of inherited genomic susceptibility to 20 common cancer types by a supervised machine-learning method. *Proc Natl Acad Sci U S A* 2018;**115**:1322–7. doi:10.1073/pnas.1717960115
96 Kim H, Chun H-W, Kim S, *et al.* Longitudinal Study-Based Dementia Prediction for Public Health. *Int J Environ Res Public Health* 2017;**14**. doi:https://dx.doi.org/10.3390/ijerph14090983
97 Kim JK, Kang S. Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis. *J Healthc Eng* Published Online First: 2017. doi:10.1155/2017/2780501
98 Kim MH, Banerjee S, Park SM, *et al.* Improving risk prediction for depression via Elastic Net regression - Results from Korea National Health Insurance Services Data. *AMIA . Annu Symp proceedings AMIA Symp* 2016;**2016**:1860–9.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexa&NEWS=N&AN=618140270 NS -
99 Kooperberg C, LeBlanc M, Obenchain V. Risk prediction using genome-wide association studies. *Genet Epidemiol* 2010;**34**:643–52. doi:10.1002/gepi.20509
100 Kwak J, Kim S, Kim G, *et al.* Scrub Typhus Incidence Modeling with Meteorological Factors in South Korea. *Int J Environ Res Public Health* 2015;**12**:7254–73. doi:10.3390/ijerph120707254
101 Lakshmi BN, Indumathi TS, Ravi N. A comparative study of classification algorithms for risk prediction in pregnancy. 2015;:6 pp. doi:10.1109/TENCON.2015.7373161
102 Lebrón-Aldea D, Dhurandhar EJ, Pérez-Rodríguez P, *et al.* Integrated genomic and BMI analysis for type 2 diabetes risk assessment. *Front Genet* 2015;**5**. doi:10.3389/fgene.2015.00075
103 Leclerc BS, Begin C, Cadieux E, *et al.* A classification and regression tree for predicting recurrent falling among community-dwelling seniors using home-care services. *Can J Public Health* 2009;**100**:263–7.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=19722338
104 Ledien J, Sorn S, Hem S, *et al.* Assessing the performance of remotely-sensed flooding indicators and their potential contribution to early warning for leptospirosis in Cambodia. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0181044
105 Lemke KW, Gudzune KA, Kharrazi H, *et al.* Assessing markers from ambulatory laboratory tests for predicting high-risk patients. *Am J Manag Care* 2018;**24**:e190–5.https://ajmc.s3.amazonaws.com/_media/_pdf/AJMC_06_2018_Lemke final.pdfhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexb&NEWS=N&AN=622514947 NS -

106 Leonenko VN, Bochenina KO, Kesarev SA, *et al.* Influenza peaks forecasting in Russia: Assessing the applicability of statistical methods. 2017;**108**:2363–7. doi:10.1016/j.procs.2017.05.196

107 Li H, Luo M, Zheng J, *et al.* An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. *Medicine (Baltimore)* 2017;**96**:e6090. doi:https://dx.doi.org/10.1097/MD.0000000000006090

108 Li J, Cao N, Li H, *et al.* ANN approach for modeling and prediction of water quality in Sichuan Kaschin-Beck disease districts. 2010;**vol.3**:1129–32. doi:10.1109/BMEI.2010.5639611

109 Li Q, Zhao L, Xue Y, *et al.* Exploring the impact of co-experiencing stressor events for teens stress forecasting. 2017;**10570 LNCS**:313–28. doi:10.1007/978-3-319-68786-5_25

110 Li ZB, Liu P, Wang W, *et al.* Using support vector machine models for crash injury severity analysis. *Accid Anal Prev* 2012;**45**:478–86. doi:10.1016/j.aap.2011.08.016

111 Liang F, Guan P, Wu W, *et al.* Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015. *PeerJ* 2018;**6**:e5134. doi:https://dx.doi.org/10.7717/peerj.5134

112 Liao YL, Wang JF, Wu JL, *et al.* PSO/ACO Algorithm-based Risk Assessment of Human Neural Tube Defects in Heshun County, China. *Biomed Environ Sci* 2012;**25**:569–76. doi:10.3967/0895-3988.2012.05.011

113 Liliana C, Oancea B, Nedelcu M, *et al.* Predicting cardiovascular diseases prevalence using neural networks. *Econ Comput Econ Cybern Stud Res* 2015;**49**:69–80.NS -

114 Liu J, Tang Z-H, Zeng F, *et al.* Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. *BMC Med Informatics Decis Mak* 2013;**13**:80. doi:10.1186/1472-6947-13-80

115 Liu K, Wang T, Yang Z, *et al.* Using Baidu Search Index to Predict Dengue Outbreak in China. *Sci Rep* 2016;**6**:38040. doi:https://dx.doi.org/10.1038/srep38040

116 Lopez E de M, Chanok SJ, Picornell AC, *et al.* Whole genome prediction of bladder cancer risk with the Bayesian LASSO. *Genet Epidemiol* 2014;**38**:467–76. doi:https://dx.doi.org/10.1002/gepi.21809

117 Luo W, Nguyen T, Nichols M, *et al.* Is demography destiny? Application of machine learning techniques to accurately predict population health outcomes from a minimal demographic dataset. *PLoS One* 2015;**10**:e0125602. doi:https://dx.doi.org/10.1371/journal.pone.0125602

118 Luo Y, Li Z, Guo H, *et al.* Predicting congenital heart defects: A comparison of three data mining methods. *PLoS One* 2017;**12**:e0177811. doi:https://dx.doi.org/10.1371/journal.pone.0177811

119 Makar M, Ghassemi M, Cutler DM, *et al.* Short-term Mortality Prediction for Elderly Patients Using Medicare Claims Data. *Int J Mach Learn Comput* 2015;**5**:192–7. doi:https://dx.doi.org/10.7763/IJMLC.2015.V5.506

120 Malovini A, Nuzzo A, Ferrazzi F, *et al.* Phenotype forecasting with SNPs data through gene-based Bayesian networks. *BMC Bioinformatics* 2009;**10**. doi:10.1186/1471-2105-10-S2-S7

121 Marcus MW, Raji OY, Duffy SW, *et al.* Incorporating epistasis interaction of genetic susceptibility single nucleotide polymorphisms in a lung cancer risk prediction model. *Int J Oncol* 2016;**49**:361–70. doi:https://dx.doi.org/10.3892/ijo.2016.3499

122 Mardones-Restat F, Jones G, Mardones-Santander F, *et al.* Growth failure prediction in chile. *Int J Epidemiol* 1989;**18**:S44–9. doi:10.1093/ije/18.Supplement_2.S44

123 Mathulamuthu SS, Asirvadam VS, Dass SC, *et al.* Predicting dengue cases by aggregation of climate variable using manifold learning. In: *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 12-14 Sept. 2017*. Piscataway, NJ, USA: : IEEE 2017. 535–40. doi:10.1109/ICSIPA.2017.8120670

124 McCoy Jr. TH, Castro VM, Roberson AM, *et al.* Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. *JAMA psychiatry* 2016;**73**:1064–71. doi:https://dx.doi.org/10.1001/jamapsychiatry.2016.2172

125 McKenzie DP, Toumbourou JW, Forbes AB, *et al.* Predicting future depression in adolescents using the Short Mood and Feelings Questionnaire: A two-nation study. *J Affect Disord* 2011;**134**:151–9. doi:10.1016/j.jad.2011.05.022

126 Mendelek F, Caby I, Pelayo P, *et al.* The Application of a Classification-Tree Model for Predicting Low Back Pain Prevalence Among Hospital Staff. *Arch Environ Occup Heal* 2013;**68**:135–44. doi:10.1080/19338244.2012.663010

127 Miller PD, Barlas S, Brenneman SK, *et al.* An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med* 2004;**164**:1113–

20.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15159269
128 Miotto R, Li L, Dudley JT. Deep Learning to Predict Patient Future Diseases from the Electronic Health Records. In: *Advances in Information Retrieval. 38th European Conference on IR Research, ECIR 2016, 20-23 March 2016*. Cham, Switzerland: : Springer International Publishing 2016. 768–74. doi:10.1007/978-3-319-30671-1_66
129 Mittag F, Römer M, Zell A. Influence of Feature Encoding and Choice of Classifier on Disease Risk Prediction in Genome-Wide Association Studies. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0135832
130 Mo H, Liu L, Li J, *et al.* Risk Factors Selection for SGA Prediction. 2016;**vol.2**:627–32. doi:10.1109/COMPSAC.2016.180
131 Modu B, Polovina N, Lan Y, *et al.* Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System. *Appl Sci-Basel* 2017;**7**. doi:10.3390/app7080836
132 Nalini C, Meera D. Breast cancer prediction system using Data mining methods. *Int J Pure Appl Math* 2018;**119**:10901–11.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85048797875&partnerID=40&md5=2f62acf2b7bc59f17af18d08fe43c210 NS -
133 Nieto P, Malone JB, Bavia ME. Ecological niche modeling for visceral leishmaniasis in the state of Bahia, Brazil, using genetic algorithm for rule-set prediction and growing degree day-water budget analysis. *Geospat Health* 2006;**1**:115–26.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=18686237
134 Nuutinen M, Leskela RL, Suojalehto E, *et al.* Development and validation of classifiers and variable subsets for predicting nursing home admission. *BMC Med Inform Decis Mak* 2017;**17**. doi:10.1186/s12911-017-0442-4
135 Oswaldo Santos B, Reis Santana LM, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS One* 2018;**13**. doi:http://dx.doi.org/10.1371/journal.pone.0195065
136 Özcan T, Tüysüz F. Healthcare expenditure prediction in Turkey by using genetic algorithm based grey forecasting models. Int. Ser. Oper. Res.Manag. Sci. 2018;**262**:159–90. doi:10.1007/978-3-319-65455-3_7
137 Panaretos D, Koloverou E, Dimopoulos AC, *et al.* A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002-2012): the ATTICA study. *Br J Nutr* 2018;**120**:1–9. doi:https://dx.doi.org/10.1017/S0007114518001150
138 Park J, Edington DW. A sequential neural network model for diabetes prediction. *Artif Intell Med* 2001;**23**:277–93.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=11704441
139 Parsaeian M, Mohammad K, Mahmoudi M, *et al.* Comparison of Logistic Regression and Artificial Neural Network in Low Back Pain Prediction: Second National Health Survey. *Iran J Public Heal* 2012;**41**:86–92.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3469002/pdf/ijph-41-86.pdf NS -
140 Pei Z, Liu J, Liu M, *et al.* Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine. *Interdiscip Sci* 2018;**10**:126–30. doi:https://dx.doi.org/10.1007/s12539-017-0271-2
141 Pekkala T, Hall A, Lotjonen J, *et al.* Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study. *J Alzheimers Dis* 2017;**55**:1055–67.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=27802228
142 Peterson AT, Martinez-Campos C, Nakazawa Y, *et al.* Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. *Trans R Soc Trop Med Hyg* 2005;**99**:647–55.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15979656
143 Poole S, Grannis S, Shah NH. Predicting Emergency Department Visits. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci* 2016;**2016**:438–45.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=27570684
144 Popescu M, Khalilia M. Improving Disease Prediction Using ICD-9 Ontological Features. In: *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), 27-30 June 2011*. Piscataway, NJ, USA: : IEEE 2011. 1805–9.NS -
145 Potash E, Brew J, Loewi A, *et al.* Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2015. doi:10.1145/2783258.2788629
146 Puddu PE, Menotti A. Artificial neural networks versus proportional hazards Cox models to predict 45-year all-cause mortality in the Italian Rural Areas of the Seven Countries Study. *BMC Med Res Methodol*

2012;**12**. doi:10.1186/1471-2288-12-100
147 Puddu PE, Menotti A. Artificial neural network versus multiple logistic function to predict 25-year coronary heart disease mortality in the Seven Countries Study. *Eur J Cardiovasc Prev Rehabil* 2009;**16**:583–91. doi:https://dx.doi.org/10.1097/HJR.0b013e32832d49e1
148 Qiang X, Kou Z. Prediction of interspecies transmission for avian influenza A virus based on a back-propagation neural network. *Math Comput Model* 2010;**52**:2060–5. doi:10.1016/j.mcm.2010.06.008
149 Rajliwall NS, Chetty G, Davey R. Chronic disease risk monitoring based on an innovative predictive modelling framework. Institute of Electrical and Electronics Engineers Inc. 2018. 1–8. doi:10.1109/SSCI.2017.8285257
150 Ram S, Zhang W, Williams M, *et al.* Predicting asthma-related emergency department visits using big data. *IEEE J Biomed Heal informatics* 2015;**19**:1216–23. doi:https://dx.doi.org/10.1109/JBHI.2015.2404829
151 Ramos R, Silva C, Moreira MWL, *et al.* Using predictive classifiers to prevent infant mortality in the Brazilian northeast. Institute of Electrical and Electronics Engineers Inc. 2017. 1–6. doi:10.1109/HealthCom.2017.8210811
152 Rasmy L, Zheng WJ, Xu H, *et al.* A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J Biomed Inform* 2018;**84**:11–6. doi:http://dx.doi.org/10.1016/j.jbi.2018.06.011
153 Ray EL, Sakrejda K, Lauer SA, *et al.* Infectious disease prediction with kernel conditional density estimation. *Stat Med* 2017;**36**:4908–29. doi:http://dx.doi.org/10.1002/sim.7488
154 Razavian N, Blecker S, Schmidt AM, *et al.* Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. *Big data* 2015;**3**:277–87. doi:https://dx.doi.org/10.1089/big.2015.0020
155 Ren H, Li J, Yuan Z-A, *et al.* The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China. *BMC Infect Dis* 2013;**13**:421. doi:https://dx.doi.org/10.1186/1471-2334-13-421
156 Robson JO, Verstraete SG, Shiboski S, *et al.* A Risk Score for Childhood Obesity in an Urban Latino Cohort. *J Pediatr* 2016;**172**:29-34.e1. doi:10.1016/j.jpeds.2016.01.055
157 Rodriguez EA, Estrada FE, Torres WC, *et al.* Early prediction of severe maternal morbidity using machine learning techniques. Lect. Notes Comput. Sci. 2016;**10022 LNAI**:259–70. doi:10.1007/978-3-319-47955-2_22
158 Rose S. Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol* 2013;**177**:443–52. doi:https://dx.doi.org/10.1093/aje/kws241
159 Saberian F, Zamani A, Shoorehdeli MA, *et al.* Prediction of seasonal influenza epidemics in Tehran using artificial neural networks. In: *2014 22nd Iranian Conference on Electrical Engineering (ICEE), 20-22 May 2014*. Piscataway, NJ, USA: : IEEE 2014. 1921–3. doi:10.1109/IranianCEE.2014.6999855
160 Sadilek A, Kautz H, Silenzio V. Modeling spread of disease from social interactions. 2012. 322–9.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871981836&partnerID=40&md5=ddc1c242ae203453a8846fa59de09240 NS -
161 Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. 2012. 136–42.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84868268429&partnerID=40&md5=adeafe49ba93bc3c38d55525b18c1243 NS -
162 Safi N, Adimi F, Soebiyanto RP, *et al.* Toward malaria risk prediction in Afghanistan using remote sensing. International Society for Photogrammetry and Remote Sensing 2010. 339–42.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84924065446&partnerID=40&md5=db144d9b2dbec33ad5dc068d11cfed8d NS -
163 Sanchez AS, Iglesias-Rodriguez FJ, Fernandez PR, *et al.* Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *Int J Ind Ergon* 2016;**52**:92–9. doi:10.1016/j.ergon.2015.09.012
164 Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. *Theor Biol Med Model* 2014;**11 Suppl 1**:S6. doi:https://dx.doi.org/10.1186/1742-4682-11-S1-S6
165 Schalkwijk D, Graaf A, Tsivtsivadze E, *et al.* Lipoprotein metabolism indicators improve cardiovascular risk prediction. PLoS One. 2014;**9**:e92840. doi:10.1371/journal.pone.0092840
166 Shaoyan Z, Tjortjis C, Xiaojun Z, *et al.* Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf Syst Front* 2009;**11**:449–60. doi:10.1007/s10796-009-9157-0
167 Shearer FM, Longbottom J, Browne AJ, *et al.* Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Heal* 2018;**6**:e270–8. doi:https://dx.doi.org/10.1016/S2214-109X(18)30024-X

168 Sheets L, Petroski GF, Zhuang Y, *et al.* Combining Contrast Mining with Logistic Regression To Predict Healthcare Utilization in a Managed Care Population. *Appl Clin Inform* 2017;**8**:430–46. doi:http://dx.doi.org/10.4338/ACI-2016-05-RA-0078

169 Shen F, Yuan J, Sun Z, *et al.* Risk Identification and Prediction of Coal Workers' Pneumoconiosis in Kailuan Colliery Group in China: A Historical Cohort Study. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0082181

170 Shi Y, Liu X, Kok S-Y, *et al.* Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect* 2016;**124**:1369–75. doi:https://dx.doi.org/10.1289/ehp.1509981

171 Simon GE, Johnson E, Lawrence JM, *et al.* Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records. *Am J Psychiatry* 2018;:appiajp201817101167. doi:https://dx.doi.org/10.1176/appi.ajp.2018.17101167

172 Soebiyanto RP, Kiang R. Meteorological parameters as predictors for seasonal influenza. *Geocarto Int* 2014;**29**:39–47. doi:10.1080/10106049.2013.799717

173 Song J. SVR model for prediction of incidence influenza based on automated method. Adv. Mater. Res. 2014;**926–930**:1159–63. doi:10.4028/www.scientific.net/AMR.926-930.1159

174 Song X, Mitnitski A, MacKnight C, *et al.* Assessment of individual risk of death using self-report data: An artificial neural network compared with a frailty index. *J Am Geriatr Soc* 2004;**52**:1180–4. doi:http://dx.doi.org/10.1111/j.1532-5415.2004.52319.x

175 Song X, Mitnitski A, Cox J, *et al.* Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform* 2004;**107**:736–40.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15360910

176 Spencer KL, Olson LM, Schnetz-Boutaud N, *et al.* Using Genetic Variation and Environmental Risk Factor Data to Identify Individuals at High Risk for Age-Related Macular Degeneration. *PLoS One* 2011;**6**. doi:http://dx.doi.org/10.1371/journal.pone.0017784

177 Sung Kean K, Tae Keun Y, Ein O, *et al.* Osteoporosis risk prediction using machine learning and conventional methods. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3-7 July 2013*. Piscataway, NJ, USA: : IEEE 2013. 188–91. doi:10.1109/EMBC.2013.6609469

178 Sushmita S, Newman S, Marquardt J, *et al.* Population Cost Prediction on Public Healthcare Datasets. Proc. 5th Int. Conf. Digit. Heal. 2015. 2015. doi:10.1145/2750511.2750521

179 Tafaro L, Cicconetti P, Piccirillo G, *et al.* Is it possible to predict one-year survival in centenarians? A neural network study. *Gerontology* 2005;**51**:199–205.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15832048

180 Tamaki Y, Nomura Y, Katsumura S, *et al.* Construction of a dental caries prediction model by data mining. *J Oral Sci* 2009;**51**:61–8. doi:http://dx.doi.org/10.2334/josnusd.51.61

181 Toschke AM, Beyerlein A, von Kries R. Children at high risk for overweight: a classification and regression trees analysis approach. *Obes Res* 2005;**13**:1270–4.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16076998

182 Tran T, Luo W, Phung D, *et al.* Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry* 2014;**14**. doi:10.1186/1471-244X-14-76

183 van den Kommer TN, Comijs HC, Dik MG, *et al.* Development of classification models for early identification of persons at risk for persistent cognitive decline. *J Neurol* 2008;**255**:1486–94. doi:10.1007/s00415-008-0942-3

184 Van Voorhees BW, Paunesku D, Gollan J, *et al.* Predicting future risk of depressive episode in adolescents: the Chicago Adolescent Depression Risk Assessment (CADRA). *Ann Fam Med* 2008;**6**:503–11. doi:https://dx.doi.org/10.1370/afm.887

185 Vestergaard P, Kruse C, Goemaere S, *et al.* Predicting mortality and incident immobility in older Belgian men by characteristics related to sarcopenia and frailty. *Osteoporos Int* 2018;**29**:1437–45. doi:10.1007/s00198-018-4467-z

186 Volkova S, Ayton E, Porterfield K, *et al.* Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLoS One* 2017;**12**. doi:10.1371/journal.pone.0188941

187 Voss R, Cullen P, Schulte H, *et al.* Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. *Int J Epidemiol* 2002;**31**:1253–64.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12540731

188 Vovsha I, Rajan A, Salleb-Aouissi A, *et al.* Predicting preterm birth is not elusive: Machine learning paves the way to individual wellness. AI Access Foundation 2014. 82–9.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84904912879&partnerID=40&md5=02cbb14baeb811a357606bcc5e9d11ca NS -

189 Walton NA, Poynton MR, Gesteland PH, *et al.* Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *BMC Med Informatics Decis Mak* 2010;**10**:68. doi:10.1186/1472-6947-10-68

190 Walz Y, Wegmann M, Leutner B, *et al.* Use of an ecologically relevant modelling approach to improve remote sensing-based schistosomiasis risk profiling. *Geospat Health* 2015;**10**:271–9. doi:10.4081/gh.2015.398

191 Wang AG, An N, Chen GL, *et al.* Predicting hypertension without measurement: A non-invasive, questionnaire-based approach. *Expert Syst Appl* 2015;**42**:7601–9. doi:10.1016/j.eswa.2015.06.012

192 Wang CJ, Li YQ, Wang L, *et al.* Development and Evaluation of a Simple and Effective Prediction Approach for Identifying Those at High Risk of Dyslipidemia in Rural Adult Residents. *PLoS One* 2012;**7**:e43834. doi:http://dx.doi.org/10.1371/journal.pone.0043834

193 Wang JF, Liu X, Liao YL, *et al.* Prediction of Neural Tube Defect Using Support Vector Machine. *Biomed Environ Sci* 2010;**23**:167–72. doi:10.1016/S0895-3988(10)60048-7

194 Wang J, Jia P, Cuadros DF, *et al.* A Remote Sensing Data Based Artificial Neural Network Approach for Predicting Climate-Sensitive Infectious Disease Outbreaks: A Case Study of Human Brucellosis. *Remote Sens* 2017;**9**. doi:10.3390/rs9101018

195 Wang J, Zhou Y, Kou Z. One class support vector machine for predicting avian-to-human transmission of avian influenza a virus. 2012. 184–8. doi:10.1109/CSAE.2012.6272935

196 Wang J, Deng Z. Modeling and Prediction of Oyster Norovirus Outbreaks along Gulf of Mexico Coast. *Environ Heal Perspect* 2016;**124**:627.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1806104356?accountid=12347

197 Wang KW, Deng C, Li JP, *et al.* Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol Infect* 2017;**145**:1118–29. doi:https://dx.doi.org/10.1017/S0950268816003216

198 Wang Y, Gu J. Comparative study among three different artificial neural networks to infectious diarrhea forecasting. Institute of Electrical and Electronics Engineers Inc. 2014. 40–6. doi:10.1109/BIBM.2014.6999373

199 Wang Y, Li J, Gu J, *et al.* Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Appl Soft Comput J* 2015;**35**:280–90. doi:10.1016/j.asoc.2015.05.047

200 Warren H, Casas J-P, Hingorani A, *et al.* Genetic prediction of quantitative lipid traits: comparing shrinkage models to gene scores. *Genet Epidemiol* 2014;**38**:72–83. doi:https://dx.doi.org/10.1002/gepi.21777

201 Wei WD, Jiang JJ, Gao L, *et al.* A New Hybrid Model Using an Autoregressive Integrated Moving Average and a Generalized Regression Neural Network for the Incidence of Tuberculosis in Heng County, China. *Am J Trop Med Hyg* 2017;**97**:799–805. doi:10.4269/ajtmh.16-0648

202 Wei W, Visweswaran S, Cooper GF, *et al.* The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *J Am Med Informatics Assoc* 2011;**18**:370–5. doi:10.1136/amiajnl-2011-000101

203 Wei W, Jiang J, Liang H, *et al.* Application of a Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China. *PLoS One* 2016;**11**:e0156768. doi:https://dx.doi.org/10.1371/journal.pone.0156768

204 Wen L, Li-Zhong Z, Dong-Wang M, *et al.* Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. *Medicine (Baltimore)* 2018;**97**:1–7. doi:10.1097/MD.0000000000010529

205 Woolery LK, Grzymala-Busse J. Machine learning for an expert system to predict preterm birth risk. *J Am Med Inform Assoc* 1994;**1**:439–46.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=7850569

206 Wu C, Kao SC, Shih CH, *et al.* Open data mining for Taiwan's dengue epidemic. *Acta Trop* 2018;**183**:1–7. doi:http://dx.doi.org/10.1016/j.actatropica.2018.03.017

207 Wu H, Cai Y, Wu Y, *et al.* Time series analysis of weekly influenza-like illness rate using a one-year period of factors in random forest regression. *Biosci Trends* 2017;**11**:292–6. doi:https://dx.doi.org/10.5582/bst.2017.01035

208 Wu JL, Roy J, Stewart WF. Prediction Modeling Using EHR Data Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Med Care* 2010;**48**:S106–13. doi:10.1097/MLR.0b013e3181de9e17

209 Wu W, Guo J, An S, *et al.* Comparison of Two Hybrid Models for Forecasting the Incidence of Hemorrhagic Fever with Renal Syndrome in Jiangsu Province, China. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0135492

210 Wu Y, Yang Y, Nishiura H, *et al.* Deep Learning for Epidemiological Predictions. 41st Int. ACM SIGIR Conf. Res. & Dev. Inf. Retr. 2018. doi:10.1145/3209978.3210077

211 Yan W, Xu Y, Yang X, *et al.* A hybrid model for short-term bacillary dysentery prediction in Yichang city, China. *Jpn J Infect Dis* 2010;**63**:264–70.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=20657066

212 Yang C, Delcher C, Shenkman E, *et al.* Machine Learning Approaches for Predicting High Utilizers in Health Care. 2017;**pt.II**:382–95. doi:10.1007/978-3-319-56154-7_35

213 Ye C, Fu T, Hao S, *et al.* Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *J Med Internet Res* 2018;**20**:e22. doi:https://dx.doi.org/10.2196/jmir.9268

214 Yiheng Z, Jingyao Z, Jiebo L. Predicting Multiple Risky Behaviors via Multimedia Content. In: *Social Informatics. 9th International Conference, SocInfo 2017, 13-15 Sept. 2017*. Cham, Switzerland: : Springer International Publishing 2017. 65–73. doi:10.1007/978-3-319-67256-4_7

215 Yoo TK, Kim SK, Kim DW, *et al.* Osteoporosis Risk Prediction for Bone Mineral Density Assessment of Postmenopausal Women Using Machine Learning. *Yonsei Med J* 2013;**54**:1321–30. doi:10.3349/ymj.2013.54.6.1321

216 Yoon Y, Song J, Hong SH, *et al.* Analysis of multiple single nucleotide polymorphisms of candidate genes related to coronary heart disease susceptibility by using support vector machines. *Clin Chem Lab Med* 2003;**41**:529–34. doi:http://dx.doi.org/10.1515/CCLM.2003.080

217 Young SD, Torrone EA, Urata J, *et al.* Using Search Engine Data as a Tool to Predict Syphilis. *Epidemiology* 2018;**29**:574–8. doi:10.1097/EDE.0000000000000836

218 Young SG, Tullis JA, Cothren J. A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus. *Appl Geogr* 2013;**45**:241–9. doi:10.1016/j.apgeog.2013.09.022

219 Young SD, Zhang × Qingpeng. Using search engine big data for predicting new HIV diagnoses. *PLoS One* 2018;**13**. doi:http://dx.doi.org/10.1371/journal.pone.0199527

220 Yu L, Zhou L, Tan L, *et al.* Application of a New Hybrid Model with Seasonal Auto-Regressive Integrated Moving Average (ARIMA) and Nonlinear Auto-Regressive Neural Network (NARNN) in Forecasting Incidence Cases of HFMD in Shenzhen, China. *PLoS One* 2014;**9**. doi:http://dx.doi.org/10.1371/journal.pone.0098241

221 Zeng Q, Huang HL, Pei X, *et al.* Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Anal Methods Accid Res* 2016;**10**:12–25. doi:10.1016/j.amar.2016.03.002

222 Zhang G, Huang S, Duan Q, *et al.* Application of a Hybrid Model for Predicting the Incidence of Tuberculosis in Hubei, China. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0080969

223 Zhang X, Liu Y, Yang M, *et al.* Comparative Study of Four Time Series Methods in Forecasting Typhoid Fever Incidence in China. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0063116

224 Zhang X, Zhang T, Young AA, *et al.* Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS One* 2014;**9**:e88075. doi:https://dx.doi.org/10.1371/journal.pone.0088075

225 Zhao D, Weng C, Zhao D, *et al.* Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *J Biomed Inform* 2011;**44**:859–68. doi:10.1016/j.jbi.2011.05.004

226 Zhao Y, Xiong P, McCullough LE, *et al.* Comparison of Breast Cancer Risk Predictive Models and Screening Strategies for Chinese Women. *J Womens Heal* 2017;**26**:294–302. doi:10.1089/jwh.2015.5692

227 Zhongwen G, Meng L. Design of H7N9 Avian Influenza Management and Forecasting System Based on GIS. In: *2015 IEEE 5th International Conference on Electronics, Information and Emergency Communication (ICEIEC), 14-16 May 2015*. Piscataway, NJ, USA: : IEEE 2015. 376–9. doi:10.1109/ICEIEC.2015.7284562

228 Zhou LL, Xia J, Yu LJ, *et al.* Using a Hybrid Model to Forecast the Prevalence of Schistosomiasis in Humans. *Int J Environ Res Public Heal* 2016;**13**. doi:10.3390/ijerph13040355

229 Zhou L, Yu L, Wang Y, *et al.* A hybrid model for predicting the prevalence of schistosomiasis in humans of

Qianjiang City, China. *PLoS One* 2014;**9**:e104875. doi:https://dx.doi.org/10.1371/journal.pone.0104875

230 Zhou Y, Glenn C, Luo J. Understanding and predicting multiple risky behaviors from social media. AI Access Foundation 2017. 600–5.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029534550&partnerID=40&md5=ae29c604e757e03f7561a191d4b296e1 NS -

231 Zhuo Z, Jiang L, Chee Keong K, *et al.* Learning in Glaucoma Genetic Risk Assessment. In: *2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010), 31 Aug.-4 Sept. 2010*. Piscataway, NJ, USA: : IEEE 2010. 6182–5. doi:10.1109/IEMBS.2010.5627757

Appendix D: Narrative Synthesis of Aspects of Discussion

Rationale for applying machine learning approaches mainly centered around it being “state of the art” or better suited to modeling complex data than regression. Machine learning was thought to be “state of the art” due to improved accuracy and deeper insights. Discussions of complex modeling focused on capturing non-linear relationships, interactions, and high-dimensionality.

When authors discussed model limitations, frequent concerns were an inadequate sample size, too few features, questionable generalizability, and a lack of interpretability. Aspects of the data other than sample size and feature number, such as potential measurement error or selection bias, were infrequently mentioned.

When discussing model implementation, many articles stated that predictive accuracy would be improved; but they did not frequently discuss how this could be translated to specific health-related policies or actions. Additionally, they rarely mentioned organizations and knowledge users that would be best suited to leverage the model.

BMJ Open

Predicting population health with machine learning: a scoping review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-037860.R1
Article Type:	Original research
Date Submitted by the Author:	27-Jul-2020
Complete List of Authors:	Morgenstern, Jason; McMaster University, Health Research Methods, Evidence, and Impact Buajitti, Emmalin; University of Toronto; Institute for Clinical Evaluative Sciences, O'Neill, Meghan; University of Toronto, Piggott, Thomas; McMaster University, Health Research Methods, Evidence, and Impact Goel, Vivek; University of Toronto; Institute for Clinical Evaluative Sciences, Fridman, Daniel; Hospital for Sick Children Kornas, Kathy ; University of Toronto Rosella, Laura; University of Toronto; Institute for Clinical Evaluative Sciences,
Primary Subject Heading:	Public health
Secondary Subject Heading:	Epidemiology, Research methods
Keywords:	PUBLIC HEALTH, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

PREDICTING POPULATION HEALTH WITH MACHINE LEARNING: A SCOPING REVIEW

AUTHORS

Jason D. Morgenstern[1], MD
Emmalin Buajitti[2,3], MPH
Meghan O'Neill[2], MPH
Thomas Piggott[1], MD, MSc
Vivek Goel[2,3], MD, MSc, PhD
Daniel Fridman[4], MPH, MA
Kathy Kornas[2], MSc
Laura C. Rosella[2,3,5,6], MHSc, PhD

AFFILIATIONS

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
3. Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada
4. The Hospital for Sick Children, Toronto, Ontario, Canada
5. Public Health Ontario, Toronto, Ontario, Canada
6. Vector Institute, Toronto, Ontario, Canada

CORRESPONDENCE TO:

Dr. Laura Rosella
Dalla Lana School of Public Health
University of Toronto
27 King's College Circle
Toronto, M5S 1A1, Canada
laura.rosella@utoronto.ca
Tel: (416) 978 6064

ABSTRACT (300 / 300 words)

Objective. To determine how machine learning has been applied to prediction applications in population health contexts. Specifically, to describe which outcomes have been studied, the data sources most widely used, and whether reporting of machine learning predictive models aligns with established reporting guidelines.

Design. A scoping review.

Data Sources. MEDLINE, EMBASE, CINAHL, ProQuest, Scopus, Web of Science, Cochrane Library, INSPEC, and ACM Digital Library were searched on July 18th, 2018.

Eligibility criteria. We included English articles published between 1980 and 2018 that used machine learning to predict population health-related outcomes. We excluded studies that only used logistic regression or were restricted to a clinical context.

Data extraction and synthesis. We summarized findings extracted from published reports, which included general study characteristics, aspects of model development, reporting of results, and model discussion items.

Results. Of 22 618 articles found by our search, 231 were included in the review. The United States (n=71, 30.74%) and China (n=40, 17.75%) produced the most studies. Cardiovascular disease (n=22, 9.52%) was the most studied outcome. The median number of observations was 5414 (interquartile range (IQR)=16543.5) and the median number of features was 17 (IQR=31). Health records (n=126, 54.5%) and investigator-generated data (n=86, 37.2%) were the most common data sources. Many studies did not incorporate recommended guidelines on machine learning and predictive modeling. Predictive discrimination was commonly assessed using area under the receiver operator curve (n=98, 42.42%) and calibration was rarely assessed (n=22, 9.52%).

Conclusions. Machine learning applications in population health have concentrated on regions and diseases well-represented in traditional data sources, infrequently using big data. Important aspects of model development were under-reported. Greater use of big data and reporting guidelines for predictive modeling could improve machine learning applications in population health.

Registration. Registered on the Open Science Framework on July 17th, 2018 (available at: https://osf.io/rnqe6/).

Strengths and limitations of this study

- Our review is one of the first syntheses of machine learning applications in population and public health.
- We used a robust search strategy, including nine peer-reviewed databases, grey literature, and reference searching, to comprehensively describe the literature.
- We compared reported study characteristics to established predictive modeling reporting guidelines, which provide an objective measure of the quality of reporting.
- Since both machine learning and population health have broad definitions, there may be some relevant articles that were not included.
- Given our focus on prediction, we could not address many other important intersections of machine learning and population health, such as surveillance and health promotion.

Word Count: 3663

INTRODUCTION

Predictive models have a long history in clinical medicine. One well-known example is the Framingham risk score, which was first developed in 1967.[1] Such models have proliferated throughout clinical practice to inform management and interventions, including preventive approaches. More recently, researchers have developed prediction models beyond individual clinical applications, for population health uses.[2,3] While there is no universal definition of population health, it generally encompasses "the health outcomes of a group of individuals, including the distribution of such outcomes within the group."[4] Similarly to clinical medicine, population-level models can be used to identify high-risk groups, directing the implementation of preventive interventions. Additionally, population health prediction models can inform policymakers about future disease burden and help to assess the impact of public health actions. Thus far, most predictive modeling in both medicine and population health has used parametric statistical regression models. More recently, there has been increasing interest in the use of a broader range of machine learning methods for prediction tasks.[5–7]

Machine learning can be loosely defined as the study and development of algorithms that learn from data with little or no human assistance.[8] These approaches have been increasingly applied in the past two decades as a result of the enabling growth of big data reserves and computational power.[9] Recent machine learning applications to prediction in population health contexts include forecasting childhood lead poisoning,[10] yellow fever incidence,[11] and the onset of suicidal ideation.[12]

The distinction between machine learning algorithms and parametric regression models is debated.[13] Regression models tend to impose more structure on the data, requiring greater human input for the verification of distributional assumptions and incorporation of domain knowledge in choosing the input parameters.[14] Algorithms employed in machine learning often derive more structure directly from the data, making fewer distributional assumptions about the data or variables. The literature remains divided on the relative advantages of more traditional approaches compared to newer methods;[15] however, given the wide variation in applications and the data used in these examples, broad assessments of superiority are often not appropriate. Also, there are debates regarding the differences in developing and validating machine learning approaches for health applications.[15,16]

Population health applications of prediction models are relatively new compared to clinical applications; correspondingly, the role of machine learning in these applications has been far less studied and discussed in the health literature. The goals of our review are to determine how machine learning has been applied to prediction in population health, the nature of the models and data used, and how the models have been developed. We also sought to assess how well the published literature aligns with recommended guidelines for reporting of predictive models and machine learning, by extracting features related to model development and performance that are highlighted by two such guidelines.[16,17]

METHODS

We based our scoping review on the framework proposed by Arksey and O'Malley[18] and refined by the Joanna Briggs Institute.[19] We also followed the more recent Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews.[20] Our study protocol was registered on the Open Science Framework on July 17th, 2018 (available at: https://osf.io/rnqe6/).

Our initial goal was to scope out all machine learning applications in population health. However, the screening process identified a much larger number of publications than anticipated. Consequently, to describe the subject area comprehensively, we restricted our scope to articles predicting future outcomes.

Search Strategy

Our search strategy consisted of peer-reviewed literature databases, grey literature, and reference searches. First, we searched nine interdisciplinary, indexed databases (MEDLINE, EMBASE, CINAHL, ProQuest, Scopus, Web of Science, Cochrane Library, INSPEC, and ACM Digital Library) on July 18th, 2018 for papers published between 1980 and 2018. Our search was informed by consultation with a health science librarian, a machine learning textbook,[21] and a similar registered review.[15] Supplementary Table A includes the full MEDLINE search strategy and filters, and serves an example search query for all database searches.

Our grey literature search included Google Scholar and Google. We developed a Google Scholar search based on terms related to 'machine learning' and 'population health', which was refined based on the relevance of initial results. The first 200 results were included in screening. A similar approach was used for the general Google search, which we restricted to the first 30 results. We examined relevant websites for publications. Results were limited to articles published on or before the date of the peer-reviewed literature search. Finally, we searched the references of relevant reviews for additional articles. Most of these reviews were identified during screening.

Eligibility Criteria

We included articles if they used machine learning to develop a predictive model that could be applied in a population health context. Therefore, we excluded articles where the model was trained primarily on people with a pre-existing disease. We also excluded articles that were only indirectly related to population health; for example, traffic accident models that did not predict a health outcome. Studies predicting individual outcomes were included if the approach was determined to be scalable to a population level. Finally, articles using only logistic regression were excluded. See Appendix A for the full eligibility criteria.

In order to manage the scope, articles were excluded if their full text could not be retrieved with our institutional licenses and if they were not written in English. Finally, articles published prior to 1980 were excluded as earlier machine learning investigators lacked comparable amounts of digitized data, software, and computational resources.

Screening Process

Initially, individual reviewers screened titles for obvious irrelevance to the review topic (JDM and EB). An example of an obviously irrelevant topic would be a paper describing the *machine health* lifespan of a piece of industrial equipment; specific examples of articles removed at this stage are listed in Appendix B. Then, we imported remaining references into Covidence systematic review management software.[22] Two reviewers screened the abstracts of remaining articles (JDM, EB, MO, and DF). Prior to evaluating full texts using all eligibility criteria, we then screened out articles that did not focus on a prediction application (JDM, EB, MO). Finally, two reviewers screened the full text of remaining articles (JDM, EB, MO). Conflicts were resolved by discussion between at least two reviewers.

Data Extraction and Synthesis

Individual authors extracted article data (JDM, EB, MO, and DF). We based our extraction items on features identified in a recent biomedical guideline for reporting of machine learning predictive models [16] and on the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement.[17] Major extraction categories identified from these guidelines included general study characteristics (e.g. geographic location and sample size), model development (e.g. algorithms used and type of validation), results (e.g. discrimination and calibration measures), and model discussion (e.g. practical costs of errors and implementation). See Supplementary Table B for a description of each extraction item.

We computed descriptive statistics for all extraction items. For categorical extracted features (e.g. whether or not unstructured text was used, the method of validation used), we calculated the total number and percent of all studies in a particular category. For continuous extracted features (e.g. number of observations in the study sample), we calculated the median value and the interquartile range (range between quartile 1 and quartile 3 in the value distribution). We also completed a narrative synthesis of discussion elements based on the text of included manuscripts.

Patient and Public Involvement Statement

There was no patient or public involvement in this study.

RESULTS

We initially retrieved 16 172 articles, after removing duplicates (Figure 1). We excluded 6494 articles after title screening, 7860 after abstract screening, 1453 when screening out non-prediction articles, and 121 after full-text screening. This resulted in 231 articles being included in the final review (Appendix C).

General Study Characteristics

The number of articles published in the population health prediction area that used machine learning increased dramatically after 2007 (Supplementary Figure A). Studies were undertaken worldwide, with the largest representation from the United States (US) (n=71, 30.74%) and China (n=40, 17.75%) (Table 1). Relatively few articles came from Oceania (n=2, 0.87%), Africa (n=5, 2.16%), and the Americas outside of the US (n=13, 5.63%).

Characteristic*	Number of Articles#	Percent of Articles**
Region		
United States	71	30.74%
Asia Excluding China	41	17.75%
China	40	17.32%
Europe	36	15.58%
Americas Excluding United States	13	5.63%
Africa	5	2.16%
Oceania	2	0.87%
Multi-region	15	6.49%
Not Reported	8	3.46%
Year published		
before 1990	1	0.4%
1990-1999	3	1.3%
2000-2004	13	5.6%
2005-2009	18	7.8%
2010-2014	70	30.3%
2015-2018	126	54.5%
Outcome level†		
Individual Risk Prediction	139	60.17%
Population Risk prediction	92	39.83%
Number of observations	Median = 5414#	IQR = 16543.5**
Not reported	72	31.2%
Number of features	Median = 17#	IQR = 31**
Not reported	59	25.5%
Used any unstructured text		
Yes	24	10.4%

Characteristic*	Number of Articles#	Percent of Articles**
No	207	89.6%
Machine learning model was compared with other statistical methods	111	48.1%
Reported data pre-processing‡		
Yes	160	69.3%
No	71	30.7%
Reported method of feature selection		
Yes	164	71.0%
No	67	29.0%
Reported hyper-parameter search		
Yes	114	49.4%
No	117	50.6%
Method of Validation		
Holdout	112	48.5%
Cross-validation or bootstrap	84	36.4%
External	15	6.5%
Not reported	32	13.9%
Reported descriptive statistics§		
Yes	140	60.6%
No	91	39.4%
Discussed the practical costs of prediction errors¶		
Yes	36	15.6%
No	195	84.4%
Stated rationale for using machine learning		
Yes	179	77.5%
No	52	22.5%

Characteristic*	Number of Articles#	Percent of Articles**
Discussed model usability		
Yes	91	39.4%
No	140	60.6%
Stated model limitations		
Yes	161	69.7%
No	70	30.3%
Discussed model implementation		
Yes	184	79.7%
No	47	20.3%
Dataset Availability by Study‖		
Closed	149	64.5%
Public	42	18.2%
Closed and Public	38	16.5%
Unknown	1	0.4%

*Refer to Supplementary Table A for a description of each characteristic and rationales for including some elements.

†Individual risk prediction refers to studies that developed models to predict the health outcomes of individuals, while population risk prediction refers to studies that developed models to predict aggregated population-level health outcomes.

‡Whether any aspects of data cleaning or pre-processing were reported. Examples include how missing data was handled, whether log transformations were done, and if derived variables were generated.

§Included a broad array of descriptive statistics such as sample population demographics, feature distributions, and outcome distributions.

¶Whether the article discussed the relative risks of false negative and false positive results based on their predictive model in contexts where it might be used.

‖Closed refers to datasets that were not immediately available in the public domain or were not identifiable as such.

#In rows where the characteristic being measured is an integer count (e.g. number of features), this column refers to the median value.

**In rows where the characteristic being measured is an integer count (e.g. number of features), this column refers to the interquartile range (IQR; quartile 3 – quartile 1).

Table 1: **Summary statistics of included articles**

The median number of observations in each article was 5414 (interquartile range (IQR)=16 543.5) and the median number of features (i.e. independent variables) used was 17 (IQR=31) (Table 1). Seventy-two studies (31.2%) did not report the number of observations. These studies often used data from reportable disease databases, which do not necessarily have a firm sampling frame, making ascertainment of the number of observations difficult.

Algorithms

The most frequently used machine learning algorithms were neural networks (n=95, 41.13%), followed by support vector machines (n=59, 25.54%), single tree-based methods (n=52, 22.51%), and random forests (n=48, 20.78%) (Supplementary Table C). About half of the articles made a comparison with statistical methods (n=111, 48.1%), which were generally logistic regression or autoregressive integrated moving average models (Table 1).

Outcomes

Non-communicable disease outcomes were assessed by many articles (n=95, 41.13%), with communicable diseases (n=76, 32.90%) and non-disease outcomes (n=60, 25.97%) studied somewhat less often. The outcome most frequently predicted was cardiovascular disease (n=22, 9.52%) (Figure 2). Other commonly forecasted non-communicable disease outcomes were suicidality (n=13, 5.63%), cancer (n=12, 5.19%), and perinatal health (n=12, 5.19%). Influenza (n=15, 6.49%) and dengue fever (n = 14, 6.06%) were the most predicted communicable disease outcomes. Aside from non-communicable and communicable disease, mortality (n=13, 5.63%) and healthcare utilization (n=14, 6.06%) were also frequently predicted.

Data

Data sources were usually structured (n=207, 89.6%) and closed, i.e. not publicly available (n=189, 81.8%) (Table 1). In general, high-dimensional data with many observations, such as multi-linked electronic medical records (EMRs) or internet-based data, may offer the most value for machine learning applications. These data types were represented in some of the articles captured, for which the most frequently reported data sources were health records (n=126, 54.5%) and investigator-generated (e.g. cohort studies) (n=86, 37.2%) (Table 2). A large proportion of studies (n=42, 18.2%) used an environmental data source (e.g. satellite imagery), mostly for prediction of infectious disease. Government databases (n=32, 13.9%) and internet-based data (n=21, 9.1%) were less frequently used. Among studies from China and the US, 80.0% and 67.6% respectively used health records data, whereas 54.5% of studies overall used these data sources (Supplementary Figure B).

Sources of Data Used[*]	**Number**	**Percent**
Environmental	42	18.2%
Geographical Information Database	12	5.2%
Meteorological/Air Quality Datasets	32	13.9%
Satellite Imagery	21	9.1%
Health Records Database	126	54.5%
Clinical Record Database[†]	46	19.9%
Disease Registry	2	0.9%
Population Health Survey	15	6.5%
Reportable Disease Database	42	18.2%

Other Health Records Database	30	13.0%
Government Database	32	13.9%
Census	11	4.8%
Vital Statistics	13	5.6%
Other Government Database	14	6.1%
HealthMap	3	1.3%
Private Insurance Data	9	3.9%
Private Insurance Claims	9	3.9%
Private Insurance Questionnaire	3	1.3%
Internet-based	21	9.1%
Search engine	12	5.2%
Social Media	12	5.2%
Investigator-generated‡	86	37.2%
Public Repositories§	19	8.2%
Health Organization Reports¶	5	2.2%
Not Reported	6	2.6%

*Categories are not mutually exclusive.

†Any dataset produced primarily for the purpose of delivering clinical care, such as electronic medical records and administrative healthcare databases produced by hospitals.

‡Any datasets resulting from researcher-driven studies such as randomized controlled trials, cohort studies, and case-control studies.

§Any freely available datasets such as MIMIC or the UC Irvine Machine Learning Repository.

¶Health-related reports, typically including disease burden estimates, produced by non-governmental or governmental organizations such as the World Health Organization.

Table 2: **Data sources**

Features

The median number of features used in a machine learning algorithm was 17 (IQR = 31; Table 1). The frequency of specific feature categories used are shown in Supplementary Figure C and

Supplementary Table D. Biomedical and sociodemographic features were frequently used (Supplementary Figure C). Of these, the most commonly used were disease history (43.3%), age (48.5%), and sex/gender (41.1%). Among lifestyle features, smoking was the most frequently used (25.1%) and of environmental features, meteorology was common (17.3%). Social media posts (5.2%) and web search queries (5.2%) were not often used. In general, most studies focused on features typical of clinical prediction models, such as subject demographics, behaviours, and medical histories. We observed limited use of other data, such as unstructured text or image-based features, which are difficult to parse using traditional statistical approaches and could benefit more from machine learning applications

Model Development and Validation

The majority of articles reported how data pre-processing (n=160, 69.3%) and feature selection (n=164, 71%) were done (Table 1). Fewer authors reported how hyperparameters were selected (n=114, 49.4%). Most studies used a holdout method of validation (n=112, 48.5%), fifteen (6.5%) externally validated their models, and thirty-two (13.9%) did not report how models were validated.

Performance Metrics

Most articles reported a prediction discrimination metric (n=172, 74.46%), which quantifies a model's ability to correctly rank-order individuals (Table 3).[23] Discrimination is a useful performance metric in cases where classification is the primary goal, including many machine learning-relevant tasks such as image recognition. The most common discrimination metrics employed were area under the receiver operator curve (n=98, 42.42%), accuracy (n=76, 32.90%), and recall (n=68, 29.44%).

In clinical and public health settings, accurate prediction of outcome probabilities is important to the practical utility of a tool, so assessing model calibration is very important.

Few articles in our study reported a measure of calibration (n=21, 9.09%), which describes how well a model predicts the absolute probability of outcomes(table 3).[23] Calibration was mostly assessed with graphing methods (n=9, 3.90%) and Hosmer-Lemeshow statistics (n=8, 3.46%).

Some articles also reported a measure of overall model fit (n=77, 33.33%). Overall performance was usually measured with a form of mean error, such as root mean squared error (n=35, 15.15%).

Prediction Performance Metrics Used	**Number**	**Percent**
Any overall performance metric	77	33.33%
RMSE	35	15.15%
MSE	26	11.26%
MAE	24	10.39%
MAPE	23	9.96%
R2*	19	8.23%
Correlation	8	3.46%
AIC or BIC	8	3.46%
Other performance metric†	21	9.09%
Any discrimination metric	172	74.46%
Area under the curve‡	98	42.42%
Accuracy§	76	32.90%
Recall¶	68	29.44%
Precision‖	39	16.88%

F statistics	10	4.33%
Likelihood Ratio**	4	1.73%
Youden Index	3	1.30%
Manual or visual comparison	3	1.30%
Other discrimination metric††	4	1.73%
Any calibration metric	21	9.09%
Manual or visual comparison‡‡	9	3.90%
Hosmer-Lemeshow	8	3.46%
Observed/Expected	5	2.16%
Other calibration metric§§	3	1.30%
Any reclassification metric	6	2.60%
Net Reclassification Index	5	2.16%
Integrated Discrimination Improvement	3	1.30%

RMSE = Root Mean Squared Error; MSE = Mean Squared Error; MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion.

*Includes R2 and pseudo-R2 metrics.

†Includes penalty error, Total Sum of Squares, proportional reduction in error, overall prediction error, specific prediction error, Nash-Sutcliffe, Root Mean Squared Percentage Error (2), mean relative absolute error, Analysis of Variance F-stat, 2LogLikelihood, relative efficiency, deviance, Ljung-Box test, mean absolute deviation, standard error, Mean Percentage Error, Brier score, and log score.

‡Includes c-statistic, s-index, area under ROC / AUC.

§Includes accuracy, misclassification, and error rate.

¶Includes sensitivity, specificity, true/false positive, and true/false negative.

‖Includes positive predictive value, negative predictive value, and precision.

**Includes positive/negative LR.

††Includes G-means (2), k-statistic, Matthews correlation coefficient.

‡‡Includes calibration plots.

§§Includes mean bias (from Bland-Altman plot), calibration factoring, and Calibration statistic.

Table 3: **Prediction Performance Metrics**

Study Discussion and Narrative Synthesis

Most articles included some discussion of their rationale for using machine learning (n=179, 77.5%), although some articles did not mention or explain their rationale (n = 52, 22.5%) (Table 1). Rationale for applying machine learning approaches mainly focused on being "state of the art" or better suited to modeling complex data than regression.

Most articles also had some discussion of the limitations of their study (n=161, 69.7%), and how the model might be implemented (n=184, 79.7%) (Table 1). Frequent concerns were an inadequate sample size, too few features, questionable generalizability, and a lack of interpretability. When discussing model implementation, many articles stated that predictive accuracy would be improved; but they did not frequently discuss how this could be translated to specific health-related policies or actions.

Less than half of the articles discussed model usability (n=91, 39.4%); that is, whether and how the model could practically be used in a relevant context. This is an important reporting component of the TRIPOD statement ("Discuss the potential clinical use of the model and implications for future research") and is relevant for understanding real-word applications of prediction models.[17] Also, only a small number discussed the costs of prediction errors in real-world contexts (n=36, 15.6%).

See Appendix D for further narrative synthesis of discussion reporting items.

DISCUSSION

Our results show that machine learning is increasingly being applied to make predictions related to population health. However, applications of machine learning to population health prediction tasks have not capitalized fully on the opportunities presented by emerging big data resources and efficient machine learning algorithms. Furthermore, reporting of these models often does not align with established guidelines for reporting of prediction models, which limits their ability to be critically appraised, compared with existing statistical models, or implemented in clinical or public health practice.

Applications of Machine Learning Prediction Models

Nearly half of the included studies were conducted in the US or China. Both countries produce the greatest number of scientific publications in general;[24] however, they also likely benefited from robust health data infrastructures. The US has rapidly digitized much of its healthcare system, resulting in large EMRs linked with government data through public-private partnerships, including processes to make these data available to researchers.[25,26] Both the US and China made greater use of health records and less use of investigator-generated data relative to other regions, which may have made machine learning projects more tractable. They also used more internet-based data, which typically includes many observations and is high-dimensional, making it amenable to machine learning methods. We noted that studies from Oceania, Africa, and the Americas (outside of the US) were limited. This may be partly due to less availability of traditional sources of structured health data. However, given that machine learning methods can incorporate non-traditional data sources, there is the potential to expand use of these methods even when structured health data is unavailable.

We found that a wide range of population health outcomes have been the focus of machine learning prediction models. However, relative to morbidity and mortality, multiple outcome categories like cancer, human immunodeficiency virus, dementia, gastroenteritis, pneumococcal disease, perinatal health, tuberculosis, and malaria appear understudied.[27] Many of these conditions are most prevalent in regions with decreased access to traditional health data, perhaps stymieing research. If machine learning methods are used to leverage novel data sources for research in these regions, it could enable greater study of neglected diseases.

Most investigators did not analyze a large number of observations and features. We observed a high reliance on electronic health records and investigator-generated data, including the use of relatively small study cohorts. Small study sample sizes or narrow data collection associated with these data sources can make it difficult to achieve high sample sizes or high dimensional data, which may impact machine learning algorithm performance. Specifically, the use of smaller investigator-generated datasets may affect the performance of studied models, as machine learning algorithms generally require a high number of observations relative to features.[28] Additionally, most studies focused on features typical of clinical prediction models, such as biomedical factors and limited aspects of broader socioeconomic or environmental determinants of health. We also observed infrequent use of unstructured data and wearable data for prediction purposes. A reliance on small datasets and traditional numbers and types of features is unlikely to fully leverage any benefits of machine learning. This may be contributing to the small differences frequently seen between parametric regression and machine learning model performance. Greater use of linked population-level databases, large EMRs, internet data, and unstructured features would likely improve these approaches.

Reporting of Machine Learning Prediction Models

Based on the elements of model development that we studied, adherence to existing machine learning[16] and prediction model[17] guidelines appears limited. Most articles did not report their method of hyper-parameter selection, discuss practical costs of prediction errors, or consider model usability, which are needed for transparency and model assessment. Many studies did not report the number of features included, method of validation, method of feature selection, or any performance metric. Given these issues, it would be difficult or impossible to compare many of these machine learning models with existing approaches. However, we acknowledge that existing guidelines were not available when many included studies were published. Future work should apply existing guidance,[16] including from TRIPOD,[17] and anticipate the forthcoming TRIPOD-ML statement.[29]

Lastly, we noted that included studies rarely assessed predictive performance in terms of calibration, which refers to a model's ability to accurately predict the absolute probability of outcomes.[23] In contrast, discrimination measures of predictive performance quantify a model's ability to correctly rank-order individuals. Many traditional machine learning tasks, such as image recognition, often have a high signal to noise ratio. In these cases, discrimination may be a suitable lone performance metric, as the algorithm can achieve near perfect performance. Conversely, health outcomes tend to be more stochastic. As a result, accurate prediction of probabilities is more important.[23] Models can have good predictive discrimination, but poor calibration, making them less useful in practice, particularly for population health applications. A further issue is that many measures of discrimination, such as accuracy and recall, artificially impose a threshold for calling events. Thresholds should ideally be ascertained by decision-makers based on their cost-utility curves.[23] Overall, applications of machine learning in population health would benefit from greater use of calibration performance metrics.

Strengths and Limitations of this Review

A strength of our study is that we addressed an understudied area, the intersection of machine learning and population health. Additionally, prediction is an application with untapped potential in population health, and where machine learning has the potential to make significant improvements. Our study also employed a comprehensive search strategy, including numerous multidisciplinary peer-reviewed databases, alongside a grey literature search. Furthermore, we applied insights from the field of clinical prediction modeling to population health and machine learning. Finally, given the focus on prediction, we were able to take a comprehensive approach to data extraction and synthesis.

In terms of limitations, concentrating on prediction prevented us from exploring applications of machine learning to other important aspects of population health, such as disease surveillance. These should be the focus of future research. Our review was also limited by including only English articles and articles with available full text, which may have introduced selection bias. Because of the broad scope of this review, and inconsistent reporting of model development and validation in reviewed articles, we were unable to carry out a critical appraisal of the literature and are unable to comment significantly on the overall performance of published machine learning population health prediction tools. This would be of great value for understanding the clinical and population health relevance of machine learning prediction tools. Lastly, the two main concepts underlying our review, machine learning and population health, are not universally defined. As a result, we may have excluded articles that may be relevant to these fields.

Research Recommendations and Conclusion

This was the first scoping review specifically focused on machine learning prediction in population health applications. Predictive modeling in population health can help to inform preventive interventions, anticipate future disease burden, and assess the impact of health policies and programs. Advances in machine learning offer opportunities to improve these models, particularly when incorporating big data. Countries with substantial EMR-use and government database linkage such as Finland, Singapore, and Denmark[30] likely have untapped potential for machine learning research. This is still a nascent field, but based on our findings more research in Oceania, Africa, and South America would also be particularly beneficial. Diseases with a high global burden of disease that were underrepresented in our findings include malaria, tuberculosis, and dementia, which may be opportune for further study.[31] Additionally, future machine learning projects could incorporate larger datasets and more non-traditional features. Greater use of resources such as HealthMap, social media, web search patterns, remote sensing, and WHO reports would enable more work in regions without formal data sources and enrich research in others. Another largely untapped prospect is using machine learning and high-dimensional data to incorporate richer representations of the social determinants of health. Opportunities should continue to grow as governments increasingly digitize their health service records and link databases to both health and non-health data. Overall, as applications of machine learning in population health develop, adherence to existing guidance[16,17,29] will improve our ability to assess and advance machine learning applications. We hope that our results will help to inform future research in this area, including the development of guidelines for machine learning applications in population health. Finally, it will be important to evaluate the impact of prediction models on decisions made in population health and the practice of public health.

CONTRIBUTORS

JDM contributed to the design of the study and led the literature search, article screening, data extraction, analysis, and writing of the manuscript. EB contributed to the design of the study, the literature search, article screening, data extraction, and analysis. MO contributed to article screening, data extraction, analysis, and writing of the manuscript. TP and VG contributed to the design of the study and supervised work. DF contributed to article screening and data extraction. KK contributed to the design of the study. LCR led the design of the study. All authors interpreted study results and contributed to drafting of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Catherine Bornbaum for her assistance with the initial design of the study.

FUNDING

This work was funded by the Canadian Institutes of Health Research (FRN: 72054363). LCR is funded by a Canada Research Chair in Population Health Analytics. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

COMPETING INTERESTS

None declared.

PATIENT CONSENT FOR PUBLICATION

Not required.

ETHICS APPROVAL

Not required as only prior published research was included in the review.

DATA AVAILABILITY

The full data extraction table used for this review will be made publicly available after publication with no end date on Mendeley Data (DOI: 10.17632/7rrz9xrp2j.1).

REFERENCES

1 Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chronic Dis* 1967;**20**:511–24. doi:10.1016/0021-9681(67)90082-3

2 Nunes MB, McPherson M, Kommers P, *et al.* Proceedings of the International Association for Development of the Information Society (IADIS) International Conference on E-Learning (Lisbon, Portugal, July 20-22, 2017). 2017;:251.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/2013525439?accountid=12347

3 Manuel DG, Tuna M, Bennett C, *et al.* Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). *CMAJ* 2018;**190**:E871–82. doi:10.1503/cmaj.170914

4 Kindig D, Stoddart G. What is population health? *Am J Public Health* 2003;**93**:380–3.http://www.ncbi.nlm.nih.gov/pubmed/12604476 (accessed 24 Jun 2018).

5 Panch T, Pearson-Stuttard J, Greaves F, *et al.* Artificial intelligence: opportunities and risks for public health. *Lancet Digit Heal* 2019;**1**:e13–4. doi:10.1016/s2589-7500(19)30002-0

6 Aldridge RW. Research and training recommendations for public health data science. *Lancet Public Heal* 2019;**4**:e373. doi:10.1016/S2468-2667(19)30112-4

7 Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annu Rev Public Health* 2018;**39**:95–112. doi:https://dx.doi.org/10.1146/annurev-publhealth-

040617-014208

8 Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev* 1959;**3**:210–29. doi:10.1147/rd.33.0210

9 Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. 2nd ed. Upper Saddle River, New Jersey: : Prentice Hall 2003.

10 Potash E, Brew J, Loewi A, *et al.* Predictive modeling for public health: Preventing childhood lead poisoning. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2015;**2015-Augus**:2039–47. doi:10.1145/2783258.2788629

11 Shearer FM, Longbottom J, Browne AJ, *et al.* Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Heal* 2018;**6**:e270–8. doi:10.1016/S2214-109X(18)30024-X

12 De Choudhury M, Kiciman E, Dredze M, *et al.* Discovering shifts to suicidal ideation from mental health content in social media. In: *Conference on Human Factors in Computing Systems - Proceedings*. Association for Computing Machinery 2016. 2098–110. doi:10.1145/2858036.2858207

13 K.G.M. M, J.A.H. DG, W. B, *et al.* Checklist for data extraction and critical appraisal for systematic reviews of prediction modelling studies: The charms checklist. *Eur J Epidemiol* 2015;**30**:904. doi:http://dx.doi.org/10.1007/s10654-015-0072-z

14 Breiman L. Statistical modeling: The two cultures. *Stat Sci* 2001;**16**:199–215. doi:10.1214/ss/1009213726

15 Christodoulou E, Ma J, Collins GS, *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;**110**:12–22. doi:10.1016/j.jclinepi.2019.02.004

16 Luo W, Phung D, Tran T, *et al.* Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res* 2016;**18**:e323. doi:10.2196/jmir.5870

17 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;**13**:1. doi:https://dx.doi.org/10.1186/s12916-014-0241-z

18 Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;**8**:19–32. doi:10.1080/1364557032000119616

19 Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual 2015 - Methodology for JBI Scoping Reviews. Adelaide: 2015. doi:10.1017/CBO9781107415324.004

20 Tricco AC, Lillie E, Zarin W, *et al.* PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* Published Online First: 4 September 2018. doi:10.7326/M18-0850

21 Hastie T, Tibshirani R, Witten D, *et al. An Introduction to Statistical Learning: With Applications in R*. New York, NY: : Springer 2013.

22 Veritas Health Innovation. Covidence systematic review software. www.covidence.org

23 Steyerberg EW. *Clinical prediction models*. New York: : Springer 2009.

24 Country outputs | Nature Index. https://www.natureindex.com/country-outputs/generate/All/global/All/score (accessed 15 Nov 2019).

25 Hecht J. The future of electronic health records. *Nature* 2019;**573**:S114–6. doi:10.1038/d41586-019-02876-y

26 Gliklich RE, Dreyer NA, Leavy MB. Public-Private Partnerships. 2014.

27 Naghavi M, Wang H, Lozano R, *et al.* Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;**385**:117–71. doi:10.1016/S0140-6736(14)61682-2

28 van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;**14**:137. doi:10.1186/1471-2288-14-137

29 Collins GS, M Moons KG. Reporting of artificial intelligence prediction models. Published Online

First: 2019. doi:10.1016/S0140-6736(19)30235-1

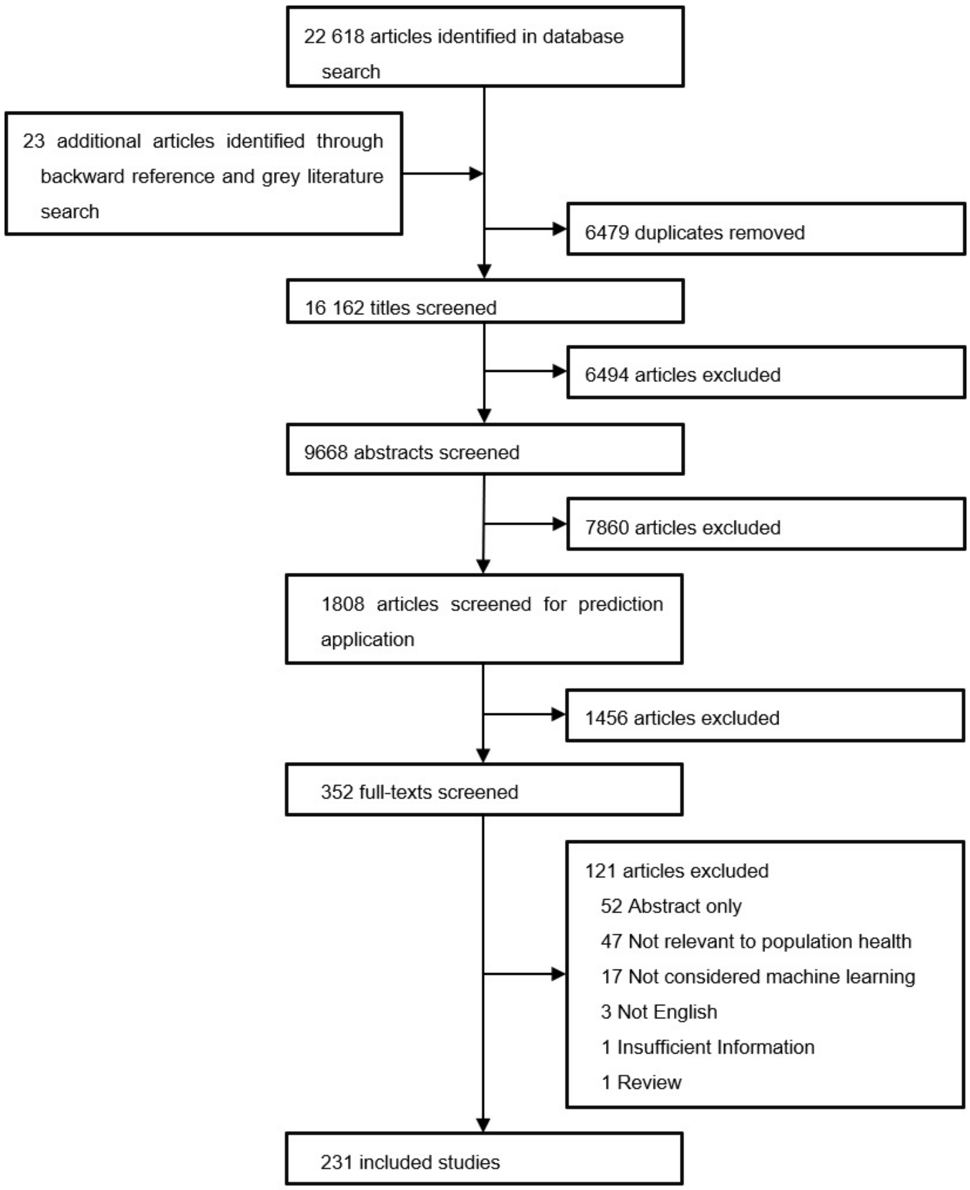
30 OECD. Health Data Governance: Privacy, Monitoring and Research. Paris: 2015. http://dx.doi.org/10.1787/9789264244566-en

31 Kyu HH, Abate D, Abate KH, *et al.* Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018;**392**:1859–922. doi:10.1016/S0140-6736(18)32335-3

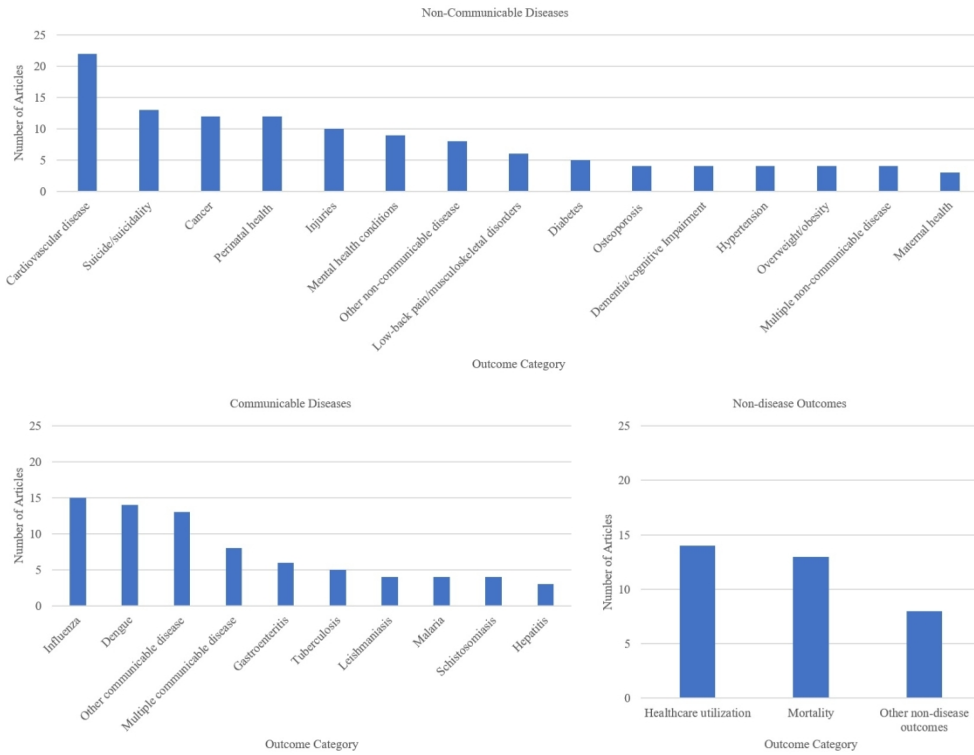
FIGURES LEGENDS

Figure 1: PRISMA flowchart of article screening process.

Figure 2: Number of articles by outcome.

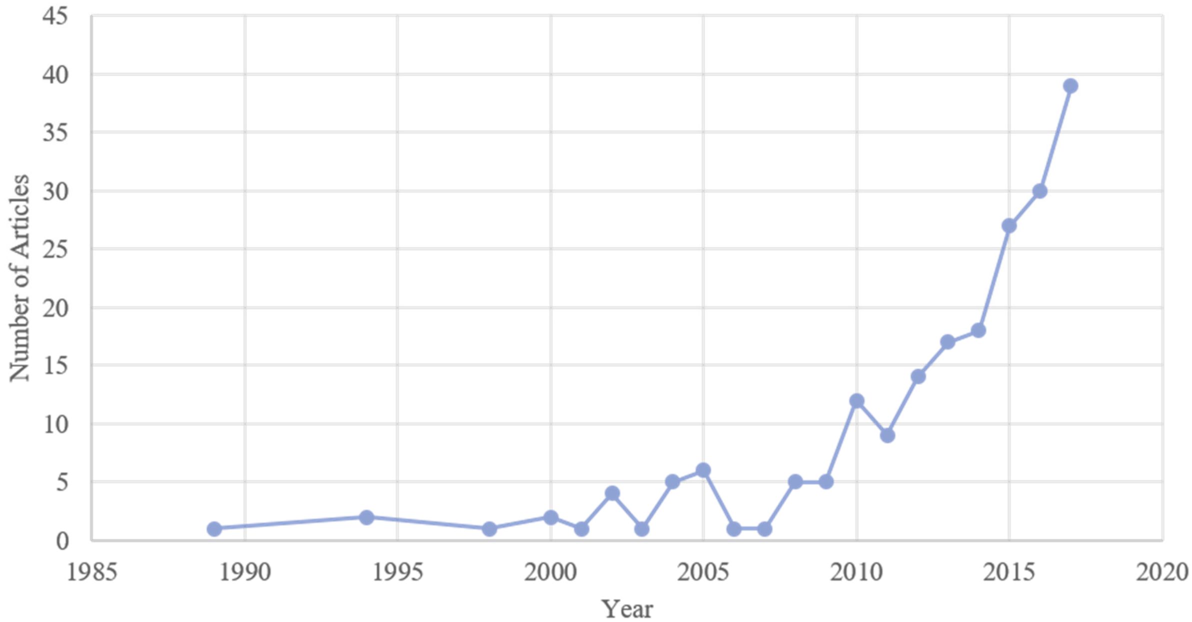


144x180mm (144 x 144 DPI)

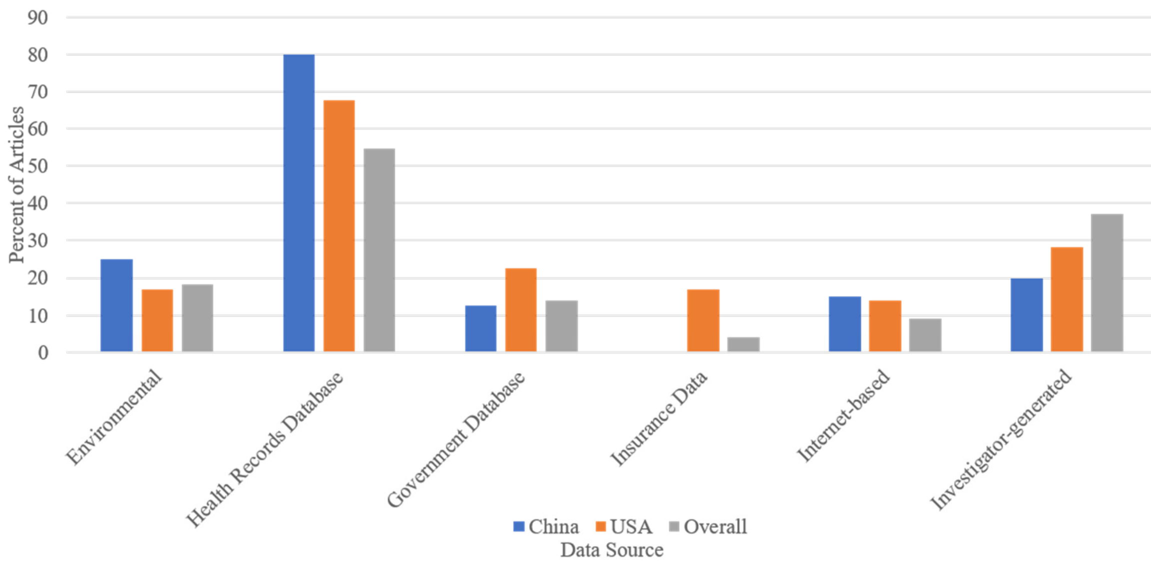


298x230mm (144 x 144 DPI)

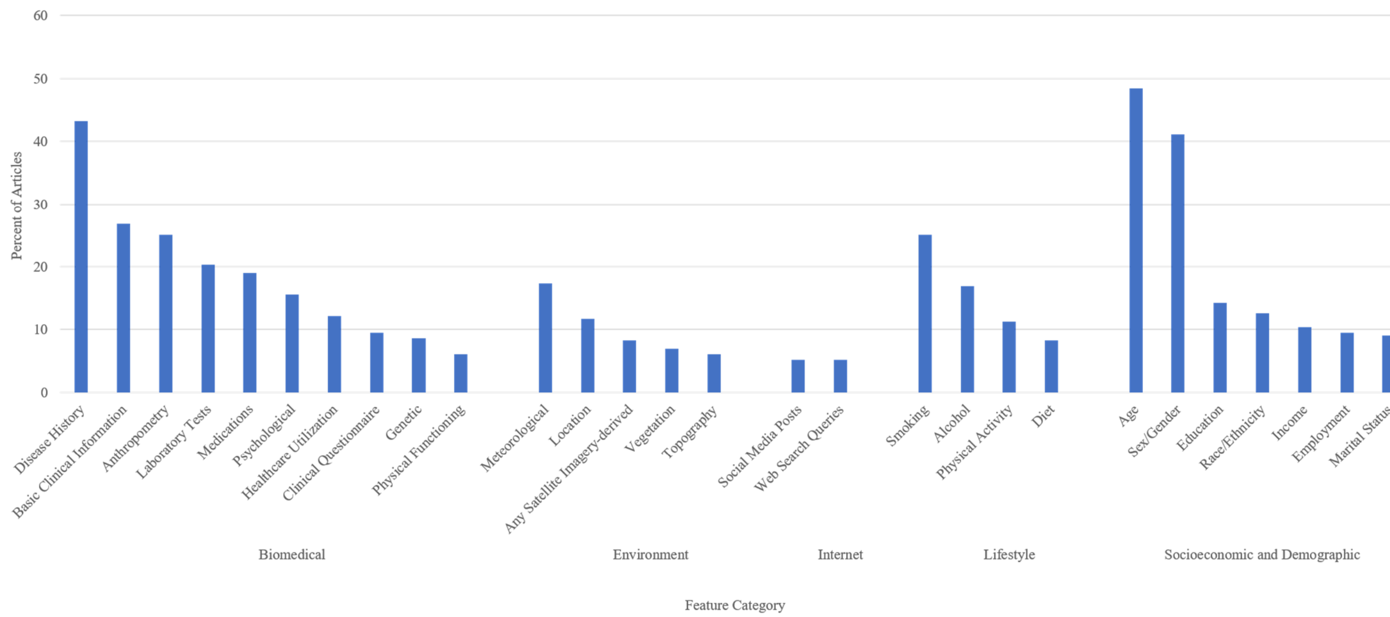
Supplementary Figure A: Number of articles by publication year. Articles from 2018 are not plotted in this figure because the review does not include all studies published that year.



Supplementary Figure B: Data Source by selected region.



Supplementary Figure C: Most commonly used feature categories.



Supplementary Table A: MEDLINE search query[1]

Machine Learning Terms	Population Health Terms
1. Exp Artificial Intelligence/	24. Exp Population Health/
2. Exp “neural networks (computer)”/	25. Exp Population Surveillance/
3. Support vector machine*.kf,tw	26. Exp Health Equity/
4. Neural net*.kf,tw	27. Health status/
5. Perceptron*.kf,tw	28. Health status disparities/
6. Deep learning.kf,tw	29. Public health systems research/
7. Random forest*.kf,tw	30. “Social determinants of health”/
8. Lasso*.kf,tw	31. Health surveys/
9. Gaussian mixture*.kf,tw	32. Health status indicators/
10. Bayesian network*.kf,tw	33. “global burden of disease”/
11. Classification tree*.kf,tw	34. Global health/
12. Regression tree*.kf,tw	35. Environmental health/
13. Relevance vector machine*.kf,tw	36. Harm reduction/
14. Nearest neighbo*.kf,tw	37. Public health informatics/
15. Probability estimation tree*.kf,tw	38. Community medicine/
16. Elastic net*.kf,tw	39. Public health/
18. Naive bayes.kf,tw	40. Epidemiology/
19. Genetic algorithm*.kf,tw	41. Preventive medicine/
20. Artificial intelligence.kf,tw	42. Occupational medicine/
21. Machine learning.kf,tw	43. Environmental medicine/
22. Statistical learning.kf,tw	44. Public health practice/
23. /or 1-22	45. Preventive health services/
	46. Health promotion/
	47. public health.kf,tw
	48. population health.kf,tw
	49. health promot*.kf,tw
	50. population surveillance.kf,tw
	51. health surveillance.kf,tw
	52. health equity.kf,tw
	53. preventive medicine.kf,tw
	54. health protection.kf,tw
	55. disease prevention.kf,tw
	56. social determinant* of health.kf,tw
	57. health determinant*.kf,tw
	58. determinant* of health.kf,tw
	59. occupational medicine.kf,tw
	60. community medicine.kf,tw
	61. epidemiolog*.kf,tw
	62. health status*.kf,tw
	63. global health.kf,tw

	64. environmental health.kf,tw
	65. harm reduction.kf,tw
	66. environmental medicine.kf,tw
	67. /or 24-66
	68. 23 and 67

[1]Limited to articles published in 1980 or after.

Supplementary Table B: Data Extraction Field Descriptions

Data Extraction Field	**Description**
Title	The article titles.
First Author	The last name and first initial of the first listed author of each article
Year of Publication	The year of publication noted for each article.
Outcome level	One of two categories: 1. *Population risk prediction:* the aggregated outcome of a whole population was predicted 2. *Individual risk prediction*: outcomes of individual participants were predicted
Outcome	Selected from the following, which are not mutually exclusive, as some articles predicted multiple outcomes: **Non-communicable Disease** 1. *Cardiovascular disease*: any disease characterized by atherosclerosis and resulting ischemia, including myocardial infarction and stroke 2. *Suicide/suicidality* 3. *Cancer* 4. *Perinatal health*: including pre-term birth, fetal alcohol spectrum disorder, congenital heart disease, growth failure, and neural tube defects 5. *Mental health conditions* 6. *Osteoporosis* 7. *Low-back pain and other musculoskeletal disorders* 8. *Diabetes* 9. *Dementia and cognitive Impairment* 10. *Hypertension* 11. *Injuries*: including fractures, falls, traffic injury, and foreign body injuries 12. *Overweight and obesity* 13. *Maternal health*: including fertility, pregnancy risk, and severe maternal morbidity 14. *Multiple non-communicable disease* 15. *Other non-communicable disease*: including liver disorders, Crohn's disease, glaucoma, dental caries, and lead poisoning **Communicable Disease** 16. *Influenza* 17. *Dengue* 18. *Gastroenteritis* 19. *Tuberculosis* 20. *Leishmaniasis* 21. *Malaria* 22. *Schistosomiasis* 23. *Hepatitis*: of viral origin 24. *Multiple communicable disease* 25. *Other communicable disease*: including zika, hand food and mouth disease, leptospirosis, yellow fever, West Nile, and typhoid fever **Non-disease Outcomes** 26. *Mortality* 27. *Healthcare utilization* 28. *Other non-disease outcomes*: including health behaviours, vitamin d status, and wellness score
Region	Categorized based on Organisation for Economic Cooperation and Development (OECD) region except for the United States and China, which were given their own categories due to the high number of publications. One of the following: *1. Africa* *2. Americas except for the United States* *3. Asia except for China* *4. China* *5. Europe* *6. Oceania* *7. United States* *8. Multi-region* *9. Other/Unknown*
Study Setting	One of two categories:

	1. *Clinical*: when data was collected in any type of clinical setting 2. *Community*: when data was collected in a community setting
Data Source Categories	Selected from the following categories, which were not mutually exclusive, and often more than one was used: 1. *Geographical Information Database*: any dataset containing basic map-based spatial information such as distances and topography 2. *Meteorological/Air Quality Datasets* 3. *Satellite Imagery*: examples include the moderate resolution imaging spectroradiometer (MODIS) and the Shuttle Radar Topography Mission (SRTM) 4. *Clinical Record Database*: any dataset produced primarily for the purpose of delivering clinical care, such as electronic medical records and administrative healthcare databases produced by hospitals 5. *Disease Registry*: a dataset maintained to monitor and/or provide care for a specific disease 6. *Population Health Survey*: a regular epidemiological survey administered periodically to assess the health of populations 7. *Reportable Disease Database*: a dataset containing reports of diseases for which it is mandatory for healthcare providers to report 8. *Other Health Records Database*: any other health records dataset not encompassed in other categories, including various surveillance systems 9. *Census* 10. *Vital Statistics*: information regularly collected by governments regarding births and deaths 11. *Other Government Database*: other governmental datasets including socioeconomic and demographic information 12. *HealthMap*: a public health surveillance system using natural language processing to analyze informal data sources such as online news, individual reports, expert-curated discussions 13. *Private Insurance Claims:* including medical, hospital, and prescription drug claims 14. *Private Insurance Questionnaires* 15. *Internet Search*: including the number of searches of certain key terms and meta data such as the location of the searches 16. *Social Media*: both posts and metadata 17. *Investigator-generated*: any datasets resulting from researcher-driven studies such as randomized controlled trials, cohort studies, and case-control studies 18. *Public Repositories*: any freely available datasets such as MIMIC 19. *Health Organization Reports*: health-related reports, typically including disease burden estimates, produced by non-governmental or governmental organizations such as the World Health Organization 20. *Not Reported*
Feature Categories	Selected from the following categories, which were not mutually exclusive, as often more than one category was used (if more than one instance of a feature category was found in an article it was only counted once): **Biomedical** 1. *Anthropometry*: measurements of the human body such as height and weight 2. *Basic Clinical Information*: information typically collected during a brief physician encounter such as a focused medical history and physical examination, including blood pressure 3. *Basic Medical Tests*: any test requiring somewhat specialized equipment such as an electrocardiogram 4. *Clinical Questionnaire*: a standardized questionnaire administered in a clinical context such as the Montreal Cognitive Assessment or Patient Health Questionnaire-9 5. *Disease History*: information regarding present and/or past diagnoses of an individual 6. *Genetic* 7. *Healthcare Utilization* 8. *Instrumental Activities of Daily Living*: features relating to an individual's daily functioning in areas such as cooking and shopping 9. *Laboratory Tests*: any features derived from human specimens requiring specialized equipment for analysis, such as hematological and microbiological results 10. *Medical Imaging* 11. *Medications* 12. *Physical Functioning*: features including the presence of any physical disabilities or the status of activities of daily living 13. *Prenatal*: relevant aspects of the period before birth such as the use of prenatal vitamins or the results of routine lab results

14. *Psychological*: features including mood or anxiety symptoms
15. *Self-Reported Health Status*

Internet-based

16. *Social Media Images*
17. *Social Media Location*: either aggregated or individual
18. *Social Media Metadata*: any information other than the content of social media posts, such as the frequency of general posts and time of posting
19. *Social Media Posts*: social media post content
20. *Social Network*: the interconnections among individuals in a social media platform
21. *Web Search Metadata*: any aspects of web searches other than their content
22. *Web Search Queries*: the content of web search queries either individual or aggregated

Lifestyle

23. *Alcohol*
24. *Diet*
25. *Physical Activity*
26. *Sleep*
27. *Smoking*
28. *Unspecified*
29. *Other Substance-use*
30. *Other Lifestyle*

Environment

31. *Air Quality*
32. *Any Satellite Imagery-derived*
33. *Biodiversity and Domestic Animals*
34. *Satellite-based Built Environment*
35. *Other Built Environment*
36. *Connectivity*: the ease of access to large urban centers and/or general services
37. *Electrical Lighting (satellite-based)*
38. *General Environmental Exposures (not included in other categories)*
39. *Hazard*: characteristics of an external hazard such as the presence of lighting on a roadway
40. *Satellite-based Land-use*
41. *Other Land-use*
42. *Location*
43. *Meteorological*
44. *Surface Water Distribution/Flooding (satellite-based)*
45. *Satellite-based Topography*
46. *Other Topography*
47. *Vector/Reservoir Characteristics*: including mosquito surveillance numbers and the population of non-human primates in the case of yellow fever
48. *Vegetation (satellite-based)*: such as the normalized difference vegetation index (NDVI)
49. *Water Composition*
50. *Other Satellite Imagery-derived*
51. ***Population Disease or Healthcare Statistics***

Socioeconomic and Demographic

52. *Adverse Adult Experiences/Trauma*
53. *Adverse Childhood Experiences*
54. *Age*
55. *Antisocial Behaviour*
56. *Economy Makeup*: such as the number of individuals working in various types of occupations
57. *Education*
58. *Electricity*
59. *Employment*
60. *Garbage Collection*
61. *Healthcare System*: such as the availability of universal, public healthcare
62. *Household Characteristics*: the number of individuals in the household and their ages
63. *Housing Structure*: aspects of the physical structure of housing such as the number of units and age of the building
64. *Human Development Index*
65. *Immigration Status*
66. *Income*
67. *Income Inequality*
68. *Language*
69. *Legal System*
70. *Literacy*

	71. *Marital Status* 72. *Occupational Risk*: including risk factors for low-back pain such as prolonged sitting or injury from repetitive movements 73. *Parental*: including disciplinary styles and the amount of time spent at home and number of parent-child activities 74. *Peer Group*: behaviours of peer group 75. *Political Stability* 76. *Population and Population Density* 77. *Population Growth* 78. *Race/Ethnicity* 79. *Religion* 80. *Sanitation*: availability of sewage systems 81. *Sex/Gender* 82. *Social Support* 83. *Unspecified* 84. *Vehicle Ownership*: at population level 85. *Water Supply Quality* 86. *Wealth* 87. *Other Socioeconomic and Demographic* 88. ***Other Features*** 89. ***Not Reported***
Number of Datasets Used	The number of distinct datasets used regardless of the number of sources.
Dataset Availability	Selected from the following categories: 1. *Public*: all the datasets used by article authors were publicly available 2. *Closed*: all the datasets were not publicly available or appeared not to be available 3. *Closed and Public*: the datasets used were a mix of available and not available
Any Unstructured Text Used	Natural human language was included in the model as a feature with no initial ordinal/nominal structure imposed.
Number of Observations	The number of individuals or other units of observations (such as countries) included in the predictive model. If multiple subsets of the data and/or distinct datasets were used for different models, the largest number was used.
Machine Learning Algorithm Type	The algorithm type used to build the predictive model, with multiple types often used in the same article. Algorithms were only counted once when used in each article, even if used to build multiple different models in the same article. Selected from the following categories: 1. *Neural Networks*: includes deep learning/deep neural networks as well as other simpler neural networks 2. *Support Vector Machine* 3. *Single Tree-based Methods*: includes classification trees, regression trees, and decision trees 4. *Random Forest* 5. *Least Absolute Shrinkage and Selection Operator (LASSO)* 6. *Bayesian Networks*: includes naïve bayes 7. *Feature Selection Methods*: includes k-means clustering and genetic algorithms; these were often used as a pre-processing step and in a few cases this was the only use of machine learning (i.e. a machine learning model was not used to build the predictive model itself) 8. *Boosted Tree-based Methods*: includes gradient boosting and boosted trees 9. *K-Nearest Neighbour* 10. *Elastic Net* 11. *Ridge Regression* 12. *Other*: includes association rule learning, single task learning, multitask learning, rough set classifier, associative classification, bagging, partial least squares discriminant analysis, Just-Add-Data Bio Tool, super learner, particle swarm optimization, ant colony optimization, Isomap, PCA, Disease State Index, Stacking, kernel conditional density estimation, stepwise deletion, conditional random fields, contrast mining, grammatical evolution, Learning from Examples Using ROugh Sets, AUtoregression with exogenous outputs, and natural language processing
Compared with Other Statistical Methods	Whether the machine learning method's predictive performance was compared with a traditional parametric statistical regression model such as logistic regression (yes/no).
Reported Data Pre-processing	Whether any aspects of data cleaning or pre-processing were reported (yes/no). Examples include how missing data was handled, whether log transformations were done, and if derived variables were generated. Missing data and all model development processes have been identified as important to report by TRIPOD.[1]
Reported Method of Feature Selection	Whether the method of feature selection was reported (yes/no). When there is a high number of features initially, this is usually done using algorithmic, domain knowledge-

	informed, or mixed approaches. Feature selection is an important element of reporting as identified by TRIPOD.[1]
Number of Features	The number of features included in the final prediction model after feature selection. If multiple models were used in one article, the largest number of features was chosen.
Reported Hyper-parameter Search	Whether the process for determining the hyper-parameters of the machine learning model, such as the number of features used to build each tree in a random forest, was reported (yes/no). This is an important aspect of model development[2], and thus considered an important element to report by the TRIPOD statement.[1]
Method of Validation	How the authors validated the predictive performance of their model, selected from one of the following categories: 1. *Holdout*: the dataset was divided into two parts; one part was used to train the model and the other was used to test the model 2. *Cross-validation and bootstrap*: the dataset was either divided into more than two parts and repeatedly trained and tested on different parts of the dataset or random sampling with replacement was used to train the model 3. *External*: the model was tested on a completely separate dataset
Reported Descriptive Statistics	Whether the article reported any descriptive statistics regarding their sample (yes/no). We considered a broad array of descriptive statistics including sample population demographics, feature distributions, and outcome distributions. These are all important reporting elements according to TRIPOD.[1]
Calibration Metrics	The types of calibration predictive performance metrics used to evaluate models, which could be more than one. Calibration refers to a model's ability to accurately predict absolute probabilities of the outcome occurring.[3] One or more of the following categories was selected if a calibration metric was used: 1. *Manual or visual comparison*: includes calibration plots 2. *Hosmer-Lemeshow* 3. *Observed/Expected*: is a ratio or comparison of observed and predicted/expected probabilities 4. *Other calibration metric*: includes mean bias (from Bland-Altman plot), calibration factoring, calibration statistic
Discrimination Metrics	The types of discrimination predictive performance metrics used to evaluate models, which could be more than one. Discrimination refers to a model's ability to correctly rank-order individuals according to their likelihood of developing the outcome.[3] One or more of the following categories was selected if a discrimination metric was used: 1. *Area under the curve*: meaning receiver operator curve; also includes c-statistic and s-index 2. *Accuracy*: includes accuracy, misclassification, and error rate 3. *Recall*: includes sensitivity, specificity, true/false positive, and true/false negative 4. *Precision*: includes positive predictive value, negative predictive value, and precision 5. *F statistics* 6. *Likelihood Ratio*: includes both positive and negative likelihood ratios 7. *Youden Index* 8. *Manual or visual comparison* 9. *Other discrimination metric*: includes G-means, k-statistic, and Matthews correlation coefficient
Overall Goodness of Fit Metrics	The types of overall goodness of fit performance metrics used to evaluate models, which could be more than one. Overall goodness of fit refers to a model's predictions' concordance with observed outcomes. One or more of the following categories was selected if an overall performance metric was used: 1. *Root mean squared error* 2. *Mean squared error* 3. *Mean absolute error* 4. *Mean absolute percentage error* 5. R^2: includes pseudo-R2s 6. *Correlation* 7. *Akaike Information Criterion or Bayesian Information Criterion* 8. *Other performance metric*: includes penalty error, total sum of squares, proportional reduction in error, overall prediction error, specific prediction error, Nash-Sutcliffe, root mean squared percentage error, mean relative absolute error, analysis of variance F-stat, -2LogLikelihood, relative efficiency, deviance, Ljung-Box test, mean absolute deviation, standard error, Brier score, log score, and mean percentage error
Did Machine Learning Models Outperform Traditional Methods?	Whether the machine learning-based predictive models outperformed the statistical parametric regression models based on the performance metrics supplied by the authors (yes/no). However, this should not be taken to mean that the difference in model performance was reliable or valid. Often, important performance metrics and essential aspects of model development were not reported, making accurate comparisons difficult.

Discussed the Practical Costs of Prediction Errors	Whether the article discussed the relative risks of false negative and false positive results based on their predictive model in contexts where it might be used (yes/no). These costs are important for determining the usefulness and application of predictive models.[3]
Stated Rationale for Using Machine Learning	Whether the article stated any reasons for using a machine learning approach instead of a statistical parametric regression approach (yes/no).
Rationale for Using Machine Learning - Free Text	Reviewers included article quotations and summaries in this section to capture different rationales for using machine learning. Reviewers attempted to only extract free text regarding each specific type of rationale once
Discussed Model Usability	Whether the article discussed any aspect of how the model could be practically used in a relevant context (yes/no).
Stated Model Limitations	Whether the article discussed any potential limitations of the research (yes/no).
Model limitations - Free Text	Reviewers included article quotations and summaries in this section to capture different reported limitations. Reviewers attempted to only extract free text regarding each specific type of limitation once.
Discussed Model Implementation	Whether the article included discussion of any consequences of model implementation such as potential clinical, population-health, and policy-level impacts (yes/no).
Model Implementation - Free Text	Reviewers included article quotations and summaries in this section to capture different reported consequences of model implementation. Reviewers attempted to only extract free text regarding each specific type of implementation impact once.

Supplementary Table C: Types of machine learning algorithms used.

Types of Algorithms	Number	Percent
Neural Networks*	95	41.13%
Support Vector Machine	59	25.54%
Single tree-based methods†	52	22.51%
Random Forest	48	20.78%
LASSO	25	10.82%
Bayesian Networks‡	23	9.96%
Feature selection methods§	20	8.66%
Boosted tree-based methods¶	19	8.23%
K-Nearest Neighbour	19	8.23%
Elastic Net	9	3.90%
Ridge regression	5	2.16%
Other‖	22	9.52%

*Includes deep neural networks.
†Includes CART, decision trees.
‡Includes naive bayes.
§Includes cluster methods (e.g. k-means clustering) and genetic algorithms.
¶Includes gradient boosting and boosted trees.
‖Including (all algorithms used once unless otherwise specified) association rule learning (n=3), single task learning, multitask learning, rough set classifier, associative classification, bagging, partial least squares discriminant analysis, Just-Add-Data Bio Tool, super learner, particle swarm optimization, ant colony optimization, isomap, principal components analysis, disease state Index, stacking, kernel conditional density estimation, stepwise deletion, conditional random fields, contrast mining, grammatical evolution, Learning from Examples Using ROugh Sets, AUtoregression with exogenous outputs, and natural language processing (n=2).

Supplementary Table D: Detailed feature categories included in studies.

Feature Category	Number of Articles	Percent
Biomedical	141	61.04
Anthropometry	58	25.11
Basic Clinical Information	62	26.84
Basic Medical Tests	10	4.33
Clinical Questionnaire	22	9.52
Disease History	100	43.29
Genetic	20	8.66
Healthcare Utilization	28	12.12
Instrumental Activities of Daily Living	6	2.60
Laboratory Tests	47	20.35
Medical Imaging	10	4.33
Medications	44	19.05
Physical Functioning	14	6.06
Prenatal	10	4.33
Psychological	36	15.58
Self-Reported Health Status	7	3.03
Internet-based	21	9.09
Social Media Images	1	0.43
Social Media Location	5	2.16
Social Media Metadata	4	1.73
Social Media Posts	12	5.19
Social Network	3	1.30
Web Search Metadata	1	0.43
Web Search Queries	12	5.19
Lifestyle	81	35.06
Alcohol	39	16.88
Diet	19	8.23
Physical Activity	26	11.26
Sleep	11	4.76
Smoking	58	25.11
Unspecified	4	1.73
Other Substance-use	13	5.63
Other Lifestyle	13	5.63
Environment	82	35.50
Air Quality	5	2.16
Any Satellite Imagery-derived	19	8.23
Biodiversity and Domestic Animals	2	0.87
Built Environment	8	3.46
Satellite	4	1.73
Other	4	1.73

Connectivity	4	1.73
Electrical Lighting[1]	1	0.43
General Environmental Exposures (not included in other categories)	9	3.90
Hazard	10	4.33
Land-use	2	0.87
Satellite	1	0.43
Other	1	0.43
Location	27	11.69
Meteorological	40	17.32
Surface Water Distribution/Flooding[1]	6	2.60
Topography	14	6.06
Satellite	12	5.19
Other	2	0.87
Vector/Reservoir Characteristics	9	3.90
Vegetation[1]	16	6.93
Water Composition	1	0.43
Other Satellite Imagery-derived	7	3.03
Population-level Disease or Healthcare Statistics	38	16.45
Socioeconomic and Demographic Factors	150	64.94
Adverse Adult Experiences/Trauma	5	2.16
Adverse Childhood Experiences	4	1.73
Age	112	48.48
Antisocial Behaviour	2	0.87
Economy Makeup	1	0.43
Education	33	14.29
Electricity	2	0.87
Employment	22	9.52
Garbage Collection	1	0.43
Healthcare System	5	2.16
Household Characteristics	10	4.33
Housing Structure	4	1.73
Human Development Index	1	0.43
Immigration Status	5	2.16
Income	24	10.39
Income Inequality	3	1.30
Language	2	0.87
Legal System	1	0.43
Literacy	2	0.87
Marital Status	21	9.09
Occupational Risk	10	4.33
Parental	3	1.30
Peer Group	1	0.43
Political Stability	1	0.43

Population and Population Density	11	4.76
Population Growth	2	0.87
Race/Ethnicity	29	12.55
Religion	3	1.30
Sanitation	5	2.16
Sex/Gender	95	41.13
Social Support	10	4.33
Unspecified	6	2.60
Vehicle Ownership	2	0.87
Water Supply Quality	5	2.16
Wealth	2	0.87
Other Socioeconomic and Demographic	29	12.55
Other Features	17	7.36
Not Reported	1	0.43

[1]See supplementary table B for greater detail regarding feature categories
[2]Satellite-derived

Appendix A: Eligibility Criteria

The following types of articles were excluded:

- Reviews;
- Focused on a methodological development;
- Only included an abstract;
- Only used linear regression, logistic regression, generalized additive models, or other approaches not considered machine learning for the purpose of this review;
- Only applied models to diagnosis, treatment decisions, or prognosis of individuals who already had a disease;
- Only related to logistics, human resources, finance, or management involved in provision of public health services;
- Focused on occupational health, traffic accidents, or environmental monitoring, with no direct link to population health outcomes;
- Used smart home or home monitoring systems;
- Used advanced imaging or other expensive predictors that would be difficult or unsafe to scale to a population level;
- Focused on clinical decision support systems;
- Predicted adverse drug effects, except vaccines.

Appendix B: Examples of article titles removed during title screening

1. Improved classification of mangroves health status using hyperspectral remote sensing data
2. Diesel engine and propulsion diagnostics of a mini-cruise ship by using artificial neural networks
3. Relationship between benthic macroinvertebrate bio-indices and physicochemical parameters of water: A tool for water resources managers
4. Adaptive one-switch row-column scanning
5. Development of a distributed bearing health monitoring and assessing system
6. Neural networks based sensor validation and recovery methodology for advanced aircraft engines
7. Mining images in publicly-available cameras for homeland security
8. The human pulvinar and attentional processing of visual distractors
9. Text classification techniques in oil industry applications
10. Research on acoustic mechanical fault diagnosis method of high voltage circuit breaker based on improved EEMD

Appendix C: All studies included in the review.

1 Achrekar H, Gandhe A, Lazarus R, *et al. Predicting Flu Trends using Twitter data*. 2011. doi:10.1109/INFCOMW.2011.5928903

2 Adamou M, Antoniou G, Greasidou E, *et al.* Mining Free-Text Medical Notes for Suicide Risk Assessment. Proc. 10th Hell. Conf. Artif. Intell. 2018. doi:10.1145/3200947.3201020

3 Adams LJ, Bello G, Dumancas GG. Development and Application of a Genetic Algorithm for Variable Optimization and Predictive Modeling of Five-Year Mortality Using Questionnaire Data. *Bioinform Biol Insights* 2015;**9**:31–41. doi:https://dx.doi.org/10.4137/BBI.S29469

4 Agarwal A, Baechle C, Behara RS, *et al.* Multi-method approach to wellness predictive modeling. *J Big Data* 2016;**3**:1–23. doi:http://dx.doi.org/10.1186/s40537-016-0049-0

5 Agarwal V, Zhang L, Zhu J, *et al.* Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis. *J Med Internet Res* 2016;**18**:e251.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medp&NEWS=N&AN=27655225

6 Agopian AJ, Lupo PJ, Tinker SC, *et al.* Working towards a risk prediction model for neural tube defects. *Birth Defects Res A Clin Mol Teratol* 2012;**94**:141–6. doi:https://dx.doi.org/10.1002/bdra.22883

7 Ahn C, Hwang Y, Park SK. Predictors of all-cause mortality among 514,866 participants from the Korean National Health Screening Cohort. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0185458

8 Aichele S, Rabbitt P, Ghisletta P. Illness and intelligence are comparatively strong predictors of individual differences in depressive symptoms following middle age. *Aging Ment Health* 2017;:1–10. doi:https://dx.doi.org/10.1080/13607863.2017.1394440

9 Akbulut A, Ertugrul E, Topcu V. Fetal health status prediction based on maternal clinical history using machine learning techniques. *Comput Methods Programs Biomed* 2018;**163**:87–100. doi:http://dx.doi.org/10.1016/j.cmpb.2018.06.010

10 Akhavan P, Karimi M, Pahlavani P, *et al.* Risk mapping of Cutaneous Leishmaniasis via a fuzzy C Means-based Neuro-Fuzzy inference system. 2014;**40**:19–23. doi:10.5194/isprsarchives-XL-2-W3-19-2014

11 Alby S, Shivakumar BL. A prediction model for type 2 diabetes risk among Indian women. *ARPN J Eng Appl Sci* 2016;**11**:2037–43.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959387072&partnerID=40&md5=0fde9764a6290488b1c3472e2bbb5f7c NS -

12 Allen T, Murray KA, Zambrana-Torrelio C, *et al.* Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun* 2017;**8**:1124. doi:https://dx.doi.org/10.1038/s41467-017-00923-8

13 Allore H, Tinetti ME, Araujo KLB, *et al.* A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and Social and Productive Activities scores. *J Clin Epidemiol* 2005;**58**:154–61.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15680749

14 Al-Mallah MH, Elshawi R, Ahmed AM, *et al.* Using Machine Learning to Define the Association between Cardiorespiratory Fitness and All-Cause Mortality (from the Henry Ford Exercise Testing Project). *Am J Cardiol* 2017;**120**:2078–84. doi:10.1016/j.amjcard.2017.08.029

15 Almeida AS, Werneck GL. Prediction of high-risk areas for visceral leishmaniasis using socioeconomic indicators and remote sensing data. *Int J Health Geogr* 2014;**13**. doi:10.1186/1476-072X-13-13

16 Alves EB, Costa CHN, de Carvalho FAA, *et al.* Risk Profiles for Leishmania infantum Infection in Brazil. *Am J Trop Med Hyg* 2016;**94**:1276–81. doi:10.4269/ajtmh.15-0513

17 Amini P, Ahmadinia H, Poorolajal J, *et al.* Evaluating the High Risk Groups for Suicide: A Comparison of Logistic Regression, Support Vector Machine, Decision Tree and Artificial Neural Network. *Iran J Public Health* 2016;**45**:1179–87.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5149472/pdf/IJPH-45-1179.pdf NS -

18 Amini P, Maroufizadeh S, Samani RO, *et al.* Factors Associated with Macrosomia among Singleton Live-births: A Comparison between Logistic Regression, Random Forest and Artificial Neural Network Methods. *Epidemiol Biostat Public Heal* 2016;**13**. doi:10.2427/11985

19 Anand A, Shakti D. Prediction of diabetes based on personal lifestyle indicators. 2015;:673–6. doi:10.1109/NGCT.2015.7375206

20 Anderson RT, Balkrishnan R, Camacho F. Risk classification of Medicare HMO enrollee cost levels using a decision-tree approach. *Am J Manag Care* 2004;**10**:89–98.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15011809

21 Asensio-Cuesta S, Diego-Mas JA, Alcaide-Marzal J. Applying generalised feedforward neural networks to

classifying industrial jobs in terms of risk of low back disorders. *Int J Ind Ergon* 2010;**40**:629–35. doi:10.1016/j.ergon.2010.04.007

22 Ayyagari R, Vekeman F, Lefebvre P, *et al.* Pulse pressure and stroke risk: development and validation of a new stroke risk model. *Curr Med Res Opin* 2014;**30**:2453–60. doi:10.1185/03007995.2014.971357

23 Azeez A, Obaromi D, Odeyemi A, *et al.* Seasonality and Trend Forecasting of Tuberculosis Prevalence Data in Eastern Cape, South Africa, Using a Hybrid Model. *Int J Environ Res Public Health* 2016;**13**. doi:10.3390/ijerph13080757

24 Bakar AA, Kefli Z, Abdullah S, *et al.* Predictive models for dengue outbreak using multiple rulebase classifiers. 2011;:5 pp. doi:10.1109/ICEEI.2011.6021830

25 Balaraman S, Schafer JJ, Tseng AM, *et al.* Plasma miRNA Profiles in Pregnant Women Predict Infant Outcomes following Prenatal Alcohol Exposure. *PLoS One* 2016;**11**. doi:http://dx.doi.org/10.1371/journal.pone.0165081

26 Bandyopadhyay S, Wolfson J, Vock DM, *et al.* Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data. *Data Min Knowl Discov* 2015;**29**:1033–69. doi:10.1007/s10618-014-0386-6

27 Bashir S, Qamar U, Khan FH. A Multicriteria Weighted Vote-Based Classifier Ensemble for Heart Disease Prediction. *Comput Intell* 2016;**32**:615–45. doi:http://dx.doi.org/10.1111/coin.12070

28 Bath PA, Pendleton N, Morgan K, *et al.* New approach to risk determination: development of risk profile for new falls among community-dwelling older people by use of a Genetic Algorithm Neural Network (GANN). *J Gerontol A Biol Sci Med Sci* 2000;**55**:M17-21.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=10719768

29 Bayati M, Bhaskar S, Montanari A. Statistical analysis of a low cost method for multiple disease prediction. *Stat Methods Med Res* 2018;**27**:2312–28. doi:http://dx.doi.org/10.1177/0962280216680242

30 Ben-Ari A, Hammond K. Text Mining the EMR for Modeling and Predicting Suicidal Behavior among US Veterans of the 1991 Persian Gulf War. 2015;:3168–75. doi:10.1109/HICSS.2015.382

31 Berchialla P, Scarinzi C, Snidero S, *et al.* Comparing models for quantitative risk assessment: an application to the European Registry of foreign body injuries in children. *Stat Methods Med Res* 2016;**25**:1244–59. doi:10.1177/0962280213476167

32 Bertsimas D, Bjarnadottir M V, Kane MA, *et al.* Algorithmic Prediction of Health-Care Costs. *Oper Res* 2008;**56**:1382–92. doi:10.1287/opre.1080.0619

33 Bhatt S, Cameron E, Flaxman SR, *et al.* Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *J R Soc Interface* 2017;**14**:20170520. doi:http://dx.doi.org/10.1098/rsif.2017.0520

34 Bhatt S, Gething PW, Brady OJ, *et al.* The global distribution and burden of dengue. *Nature* 2013;**496**:504–7.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1350966952?accountid=12347

35 Bi J, Sun J, Wu Y, *et al.* A machine learning approach to college drinking prediction and risk factor identification. *ACM Trans Intell Syst Technol* 2013;**4**:1–24. doi:10.1145/2508037.2508053

36 Bibi H, Nutman A, Shoseyov D, *et al.* Prediction of emergency department visits for respiratory symptoms using an artificial neural network. *Chest* 2002;**122**:1627–32.http://myaccess.library.utoronto.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=106806480&site=ehost-live NS -

37 Blankenberg S, Zeller T, Saarela O, *et al.* Contribution of 30 biomarkers to 10-year cardiovascular risk estimation in 2 population cohorts: the MONICA, risk, genetics, archiving, and monograph (MORGAM) biomarker project. *Circulation* 2010;**121**:2388–97. doi:https://dx.doi.org/10.1161/CIRCULATIONAHA.109.901413

38 Braithwaite SR, Giraud-Carrier C, West J, *et al.* Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. *JMIR Ment Heal* 2016;**3**:e21. doi:https://dx.doi.org/10.2196/mental.4822

39 Buczak AL, Baugher B, Guven E, *et al.* Fuzzy association rule mining and classification for the prediction of malaria in South Korea. *BMC Med Inform Decis Mak* 2015;**15**:47. doi:https://dx.doi.org/10.1186/s12911-015-0170-6

40 Buczak AL, Koshute PT, Babin SM, *et al.* A data-driven epidemiological prediction method for dengue outbreaks using local and remote sensing data. *BMC Med Inform Decis Mak* 2012;**12**:124. doi:https://dx.doi.org/10.1186/1472-6947-12-124

41 Caillet P, Klemm S, Ducher M, *et al.* Hip Fracture in the Elderly: A Re-Analysis of the EPIDOS Study with

Causal Bayesian Networks. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0120125
42 Cao S, Wang F, Tam W, *et al.* A hybrid seasonal prediction model for tuberculosis incidence in China. *BMC Med Inform Decis Mak* 2013;**13**:56. doi:https://dx.doi.org/10.1186/1472-6947-13-56
43 Caon F, Meneghel G, Zaghi P, *et al.* Applicability of neural networks to suicidological research: a pilot study. *Arch suicide Res* 2002;**6**:285–9.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/38408407?accountid=12347
44 Chae S, Kwon S, Lee D. Predicting Infectious Disease Using Deep Learning and Big Data. *Int J Env Res Public Heal* 2018;**15**. doi:10.3390/ijerph15081596
45 Chang JS, Yeh RF, Wiencke JK, *et al.* Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. *Cancer Epidemiol Biomarkers Prev* 2008;**17**:1368–73. doi:10.1158/1055-9965.EPI-07-2830
46 Chen Y, Chu CW, Chen MIC, *et al.* The utility of LASSO-based models for real time forecasts of endemic infectious diseases: a cross country comparison. *J Biomed Inform* Published Online First: 2018. doi:https://dx.doi.org/10.1016/j.jbi.2018.02.014
47 Chen Y-T, Miller PD, Barrett-Connor E, *et al.* An approach for identifying postmenopausal women age 50-64 years at increased short-term risk for osteoporotic fracture. *Osteoporos Int* 2007;**18**:1287–96.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17464525
48 Chenar SS, Deng Z. Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res* 2018;**128**:20–37. doi:10.1016/j.watres.2017.10.032
49 Chenar SS, Deng Z. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ Int* 2018;**111**:212–23. doi:https://dx.doi.org/10.1016/j.envint.2017.11.032
50 Cheng Y, Jiang T, Zhu M, *et al.* Risk assessment models for genetic risk predictors of lung cancer using two-stage replication for Asian and European populations. *Oncotarget* 2017;**8**:53959–67. doi:https://dx.doi.org/10.18632/oncotarget.10403
51 Choi SB, Lee W, Yoon J-H, *et al.* Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea. *J Affect Disord* 2018;**231**:8–14. doi:https://dx.doi.org/10.1016/j.jad.2018.01.019
52 Conrad D, Wilker S, Pfeiffer A, *et al.* Does trauma event type matter in the assessment of traumatic load? *Eur J Psychotraumatol* 2017;**8**. doi:10.1080/20008198.2017.1344079
53 Courtney KL, Stewart S, Popescu M, *et al.* Predictors of preterm birth in birth certificate data. *Stud Health Technol Inform* 2008;**136**:555–60.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=18487789
54 Crawford AG, Fuhr Jr JP, Clarke J, *et al.* Comparative effectiveness of total population versus disease-specific neural network models in predicting medical costs. *Dis Manag* 2005;**8**:277–87. doi:10.1089/dis.2005.8.277
55 Dartois L, Gauthier E, Heitzmann J, *et al.* A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort. *Breast Cancer Res Treat* 2015;**150**:415–26. doi:https://dx.doi.org/10.1007/s10549-015-3321-7
56 Darvishi E, Khotanlou H, Khoubi J, *et al.* Prediction Effects of Personal, Psychosocial, and Occupational Risk Factors on Low Back Pain Severity Using Artificial Neural Networks Approach in Industrial Workers. *J Manipulative Physiol Ther* 2017;**40**:486–93. doi:https://dx.doi.org/10.1016/j.jmpt.2017.03.012
57 Davidson MW, Haim DA, Radin JM. Using networks to combine 'big data' and traditional surveillance to improve influenza predictions. *Sci Rep* 2015;**5**:8154. doi:http://dx.doi.org/10.1038/srep08154
58 De Choudhury M, Gamon M, Counts S, *et al.* Predicting depression via social media. 2013;:128–37.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84900027892&partnerID=40&md5=37755eb6e5e5e0639df4438f14bf1c8f NS -
59 DeClaris N, Hammad T, Wahab AF, *et al.* Neural network models for studying and for managing in real-time schistosomiasis control programs. 1994;**vol.2**:1362–3. doi:10.1109/IEMBS.1994.415473
60 Delgado-Gomez D, Blasco-Fontecilla H, Sukno F, *et al.* Suicide attempters classification: Toward predictive models of suicidal behavior. *Neurocomputing An Int J* 2012;**92**:3–8. doi:http://dx.doi.org/10.1016/j.neucom.2011.08.033
61 Dicken RA, Fazle Rubby SAM, Naz S, *et al.* Risk assessment of the top five malignancies among males and females with respect to occupation, educational status and smoking habits. 2016;:6 pp. doi:10.1109/ICIS.2016.7550756
62 Dierker LC, Avenevoli S, Goldberg A, *et al.* Defining subgroups of adolescents at risk for experimental and

regular smoking. *Prev Sci* 2004;**5**:169–83. doi:10.1023/B:PREV.0000037640.66607.6b
63 Dinh TQ, Le H V, Cao TH, *et al.* Forecasting the magnitude of dengue in southern vietnam. 2016;**pt.I**:554–63. doi:10.1007/978-3-662-49381-6_53
64 Dogan M V, Grumbach IM, Michaelson JJ, *et al.* Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. *PLoS One* 2018;**13**:e0190549. doi:http://dx.doi.org/10.1371/journal.pone.0190549
65 Dugan TM, Mukhopadhyay S, Carroll A, *et al.* Machine Learning Techniques for Prediction of Early Childhood Obesity. *Appl Clin Inform* 2015;**6**:506–20. doi:http://dx.doi.org/10.4338/ACI-2015-03-RA-0036
66 Duh MS, Walker AM, Pagano M, *et al.* Prediction and cross-validation of neural networks versus logistic regression: Using hepatic disorders as an example. *Am J Epidemiol* 1998;**147**:407–13.NS -
67 Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput Appl* 2018;**29**:685–93. doi:10.1007/s00521-016-2604-1
68 Dykxhoorn J, Hatcher S, Roy-Gagnon M-H, *et al.* Early life predictors of adolescent suicidal thoughts and adverse outcomes in two population-based cohort studies. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0183182
69 Félix Caballero F, Soulis G, Engchuan W, *et al.* Advanced analytical methodologies for measuring healthy ageing and its determinants, using factor analysis and machine learning techniques: The ATHLOS project. *Sci Rep* 2017;**7**. doi:10.1038/srep43955
70 Gan R, Chen N, Huang D. Comparisons of forecasting for hepatitis in Guangxi Province, China by using three neural networks models. *PeerJ* 2016;**2016**:2684. doi:http://dx.doi.org/10.7717/peerj.2684
71 Gan X, Xu Y, Liu L, *et al.* Predicting the incidence risk of ischemic stroke in a hospital population of southern China: A classification tree analysis. *J Neurol Sci* 2011;**306**:108–14. doi:http://dx.doi.org/10.1016/j.jns.2011.03.032
72 Ganga GMD, Esposto KF, Braatz D. Application of discriminant analysis-based model for prediction of risk of low back disorders due to workplace design in industrial jobs. *Work J Prev Assess Rehabil* 2012;**41**:2370–6. doi:10.3233/WOR-2012-0467-2370
73 Gerardi DO, Monteiro LHA. System identification and prediction of dengue fever incidence in Rio de Janeiro. *Math Probl Eng* 2011;:720304 (13 pp.). doi:10.1155/2011/720304
74 Girela JL, Gil D, Johnsson M, *et al.* Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence methods. *Biol Reprod* 2013;**88**:99. doi:https://dx.doi.org/10.1095/biolreprod.112.104653
75 Gismondi RC, Almeida RMVR, Infantosi AFC. Artificial neural networks for infant mortality modelling. *Comput Methods Programs Biomed* 2002;**69**:237–47.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12204451
76 Gueli N, Piccirillo G, Troisi G, *et al.* The influence of lifestyle on cardiovascular risk factors: Analysis using a neural network. *Arch Gerontol Geriatr* 2005;**40**:157–72. doi:10.1016/j.archger.2004.08.002
77 Guo P, Liu T, Zhang Q, *et al.* Developing a dengue forecast model using machine learning: A case study in China. *PLoS Negl Trop Dis* 2017;**11**:e0005973. doi:https://dx.doi.org/10.1371/journal.pntd.0005973
78 Guo S, Lucas RM, Ponsonby A-L, *et al.* A novel approach for prediction of vitamin d status using support vector regression. *PLoS One* 2013;**8**:e79970. doi:https://dx.doi.org/10.1371/journal.pone.0079970
79 Harrigan RJ, Thomassen HA, Buermann W, *et al.* Economic Conditions Predict Prevalence of West Nile Virus. *PLoS One* 2010;**5**. doi:http://dx.doi.org/10.1371/journal.pone.0015437
80 Holmes JH, Durbin DR, Winston FK. Discovery of predictive models in an injury surveillance database: an application of data mining in clinical research. *Proceedings AMIA Symp* 2000;:359–63.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=11079905
81 Hu X, Quirchmayr G, Winiwarter W, *et al.* Influenza early warning model based on Yunqi theory. *Chin J Integr Med* 2012;**18**:192–6. doi:10.1007/s11655-012-1003-4
82 Hu ZK, Jin B, Shin AY, *et al.* Real-Time Web-Based Assessment of Total Population Risk of Future Emergency Department Utilization: Statewide Prospective Active Case Finding Study. *Interact J Med Res* 2015;**4**:39–51. doi:10.2196/ijmr.4022
83 Husam IS, Abuhamad AAB, Zainudin S, *et al.* Feature selection algorithms for Malaysian dengue outbreak detection model. *Sains Malaysiana* 2017;**46**:255–65. doi:10.17576/jsm-2017-4602-10
84 Husin NA, Mustapha N, Sulaiman MN, *et al.* Performance of hybrid GANN in comparison with other standalone models on dengue outbreak prediction. *J Comput Sci* 2016;**12**:300–6. doi:10.3844/jcssp.2016.300.306
85 Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity

prediction. *Accid Anal Prev* 2017;**108**:27–36. doi:10.1016/j.aap.2017.08.008

86 Jafari SA, Jahandideh S, Jahandideh M, *et al.* Prediction of road traffic death rate using neural networks optimised by genetic algorithm. *Int J Inj Contr Saf Promot* 2015;**22**:153–7. doi:https://dx.doi.org/10.1080/17457300.2013.857695

87 Jiang D, Hao M, Ding F, *et al.* Mapping the transmission risk of Zika virus using machine learning models. *Acta Trop* 2018;**185**:391–9. doi:https://dx.doi.org/10.1016/j.actatropica.2018.06.021

88 Joseph A, Ramamurthy B. Suicidal behavior prediction using data mining techniques. *Int J Mech Eng Technol* 2018;**9**:293–301.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045951449&partnerID=40&md5=74f2c78c8b63037785fb7ed22ef9b8a1 NS -

89 Juan YC, Chen CM, Chen SH. A classifier fusion approach to osteoporosis prediction for women in Taiwan. *J Ind Prod Eng* 2015;**32**:360–8. doi:10.1080/21681015.2015.1064484

90 Karstoft K-I, Statnikov A, Andersen SB, *et al.* Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers. *J Affect Disord* 2015;**184**:170–5. doi:https://dx.doi.org/10.1016/j.jad.2015.05.057

91 Kennedy EH, Wiitala WL, Hayward RA, *et al.* Improved Cardiovascular Risk Prediction Using Nonparametric Regression and Electronic Health Record Data. *Med Care* 2013;**51**:251–8. doi:10.1097/MLR.0b013e31827da594

92 Kesorn K, Ongruk P, Chompoosri J, *et al.* Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the Aedes aegypti Infection Rate in Similar Climates and Geographical Areas. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0125049

93 Kessler RC, Hwang I, Hoffmire CA, *et al.* Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res* 2017;**26**:e1575. doi:10.1002/mpr.1575

94 Kessler RC, Rose S, Koenen KC, *et al.* How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. *World Psychiatry* 2014;**13**:265–74. doi:http://dx.doi.org/10.1002/wps.20150

95 Kim BJ, Kim SH. Prediction of inherited genomic susceptibility to 20 common cancer types by a supervised machine-learning method. *Proc Natl Acad Sci U S A* 2018;**115**:1322–7. doi:10.1073/pnas.1717960115

96 Kim H, Chun H-W, Kim S, *et al.* Longitudinal Study-Based Dementia Prediction for Public Health. *Int J Environ Res Public Health* 2017;**14**. doi:https://dx.doi.org/10.3390/ijerph14090983

97 Kim JK, Kang S. Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis. *J Healthc Eng* Published Online First: 2017. doi:10.1155/2017/2780501

98 Kim MH, Banerjee S, Park SM, *et al.* Improving risk prediction for depression via Elastic Net regression - Results from Korea National Health Insurance Services Data. *AMIA . Annu Symp proceedings AMIA Symp* 2016;**2016**:1860–9.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexa&NEWS=N&AN=618140270 NS -

99 Kooperberg C, LeBlanc M, Obenchain V. Risk prediction using genome-wide association studies. *Genet Epidemiol* 2010;**34**:643–52. doi:10.1002/gepi.20509

100 Kwak J, Kim S, Kim G, *et al.* Scrub Typhus Incidence Modeling with Meteorological Factors in South Korea. *Int J Environ Res Public Health* 2015;**12**:7254–73. doi:10.3390/ijerph120707254

101 Lakshmi BN, Indumathi TS, Ravi N. A comparative study of classification algorithms for risk prediction in pregnancy. 2015;:6 pp. doi:10.1109/TENCON.2015.7373161

102 Lebrón-Aldea D, Dhurandhar EJ, Pérez-Rodríguez P, *et al.* Integrated genomic and BMI analysis for type 2 diabetes risk assessment. *Front Genet* 2015;**5**. doi:10.3389/fgene.2015.00075

103 Leclerc BS, Begin C, Cadieux E, *et al.* A classification and regression tree for predicting recurrent falling among community-dwelling seniors using home-care services. *Can J Public Health* 2009;**100**:263–7.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=19722338

104 Ledien J, Sorn S, Hem S, *et al.* Assessing the performance of remotely-sensed flooding indicators and their potential contribution to early warning for leptospirosis in Cambodia. *PLoS One* 2017;**12**. doi:http://dx.doi.org/10.1371/journal.pone.0181044

105 Lemke KW, Gudzune KA, Kharrazi H, *et al.* Assessing markers from ambulatory laboratory tests for predicting high-risk patients. *Am J Manag Care* 2018;**24**:e190–5.https://ajmc.s3.amazonaws.com/_media/_pdf/AJMC_06_2018_Lemke final.pdfhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emexb&NEWS=N&AN=622514947 NS -

106 Leonenko VN, Bochenina KO, Kesarev SA, *et al.* Influenza peaks forecasting in Russia: Assessing the applicability of statistical methods. 2017;**108**:2363–7. doi:10.1016/j.procs.2017.05.196
107 Li H, Luo M, Zheng J, *et al.* An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. *Medicine (Baltimore)* 2017;**96**:e6090. doi:https://dx.doi.org/10.1097/MD.0000000000006090
108 Li J, Cao N, Li H, *et al.* ANN approach for modeling and prediction of water quality in Sichuan Kaschin-Beck disease districts. 2010;**vol.3**:1129–32. doi:10.1109/BMEI.2010.5639611
109 Li Q, Zhao L, Xue Y, *et al.* Exploring the impact of co-experiencing stressor events for teens stress forecasting. 2017;**10570 LNCS**:313–28. doi:10.1007/978-3-319-68786-5_25
110 Li ZB, Liu P, Wang W, *et al.* Using support vector machine models for crash injury severity analysis. *Accid Anal Prev* 2012;**45**:478–86. doi:10.1016/j.aap.2011.08.016
111 Liang F, Guan P, Wu W, *et al.* Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015. *PeerJ* 2018;**6**:e5134. doi:https://dx.doi.org/10.7717/peerj.5134
112 Liao YL, Wang JF, Wu JL, *et al.* PSO/ACO Algorithm-based Risk Assessment of Human Neural Tube Defects in Heshun County, China. *Biomed Environ Sci* 2012;**25**:569–76. doi:10.3967/0895-3988.2012.05.011
113 Liliana C, Oancea B, Nedelcu M, *et al.* Predicting cardiovascular diseases prevalence using neural networks. *Econ Comput Econ Cybern Stud Res* 2015;**49**:69–80.NS -
114 Liu J, Tang Z-H, Zeng F, *et al.* Artificial neural network models for prediction of cardiovascular autonomic dysfunction in general Chinese population. *BMC Med Informatics Decis Mak* 2013;**13**:80. doi:10.1186/1472-6947-13-80
115 Liu K, Wang T, Yang Z, *et al.* Using Baidu Search Index to Predict Dengue Outbreak in China. *Sci Rep* 2016;**6**:38040. doi:https://dx.doi.org/10.1038/srep38040
116 Lopez E de M, Chanok SJ, Picornell AC, *et al.* Whole genome prediction of bladder cancer risk with the Bayesian LASSO. *Genet Epidemiol* 2014;**38**:467–76. doi:https://dx.doi.org/10.1002/gepi.21809
117 Luo W, Nguyen T, Nichols M, *et al.* Is demography destiny? Application of machine learning techniques to accurately predict population health outcomes from a minimal demographic dataset. *PLoS One* 2015;**10**:e0125602. doi:https://dx.doi.org/10.1371/journal.pone.0125602
118 Luo Y, Li Z, Guo H, *et al.* Predicting congenital heart defects: A comparison of three data mining methods. *PLoS One* 2017;**12**:e0177811. doi:https://dx.doi.org/10.1371/journal.pone.0177811
119 Makar M, Ghassemi M, Cutler DM, *et al.* Short-term Mortality Prediction for Elderly Patients Using Medicare Claims Data. *Int J Mach Learn Comput* 2015;**5**:192–7. doi:https://dx.doi.org/10.7763/IJMLC.2015.V5.506
120 Malovini A, Nuzzo A, Ferrazzi F, *et al.* Phenotype forecasting with SNPs data through gene-based Bayesian networks. *BMC Bioinformatics* 2009;**10**. doi:10.1186/1471-2105-10-S2-S7
121 Marcus MW, Raji OY, Duffy SW, *et al.* Incorporating epistasis interaction of genetic susceptibility single nucleotide polymorphisms in a lung cancer risk prediction model. *Int J Oncol* 2016;**49**:361–70. doi:https://dx.doi.org/10.3892/ijo.2016.3499
122 Mardones-Restat F, Jones G, Mardones-Santander F, *et al.* Growth failure prediction in chile. *Int J Epidemiol* 1989;**18**:S44–9. doi:10.1093/ije/18.Supplement_2.S44
123 Mathulamuthu SS, Asirvadam VS, Dass SC, *et al.* Predicting dengue cases by aggregation of climate variable using manifold learning. In: *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 12-14 Sept. 2017*. Piscataway, NJ, USA: : IEEE 2017. 535–40. doi:10.1109/ICSIPA.2017.8120670
124 McCoy Jr. TH, Castro VM, Roberson AM, *et al.* Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing. *JAMA psychiatry* 2016;**73**:1064–71. doi:https://dx.doi.org/10.1001/jamapsychiatry.2016.2172
125 McKenzie DP, Toumbourou JW, Forbes AB, *et al.* Predicting future depression in adolescents using the Short Mood and Feelings Questionnaire: A two-nation study. *J Affect Disord* 2011;**134**:151–9. doi:10.1016/j.jad.2011.05.022
126 Mendelek F, Caby I, Pelayo P, *et al.* The Application of a Classification-Tree Model for Predicting Low Back Pain Prevalence Among Hospital Staff. *Arch Environ Occup Heal* 2013;**68**:135–44. doi:10.1080/19338244.2012.663010
127 Miller PD, Barlas S, Brenneman SK, *et al.* An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med* 2004;**164**:1113–

20.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15159269
128 Miotto R, Li L, Dudley JT. Deep Learning to Predict Patient Future Diseases from the Electronic Health Records. In: *Advances in Information Retrieval. 38th European Conference on IR Research, ECIR 2016, 20-23 March 2016*. Cham, Switzerland: : Springer International Publishing 2016. 768–74. doi:10.1007/978-3-319-30671-1_66
129 Mittag F, Römer M, Zell A. Influence of Feature Encoding and Choice of Classifier on Disease Risk Prediction in Genome-Wide Association Studies. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0135832
130 Mo H, Liu L, Li J, *et al.* Risk Factors Selection for SGA Prediction. 2016;**vol.2**:627–32. doi:10.1109/COMPSAC.2016.180
131 Modu B, Polovina N, Lan Y, *et al.* Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System. *Appl Sci-Basel* 2017;**7**. doi:10.3390/app7080836
132 Nalini C, Meera D. Breast cancer prediction system using Data mining methods. *Int J Pure Appl Math* 2018;**119**:10901–11.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85048797875&partnerID=40&md5=2f62acf2b7bc59f17af18d08fe43c210 NS -
133 Nieto P, Malone JB, Bavia ME. Ecological niche modeling for visceral leishmaniasis in the state of Bahia, Brazil, using genetic algorithm for rule-set prediction and growing degree day-water budget analysis. *Geospat Health* 2006;**1**:115–26.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=18686237
134 Nuutinen M, Leskela RL, Suojalehto E, *et al.* Development and validation of classifiers and variable subsets for predicting nursing home admission. *BMC Med Inform Decis Mak* 2017;**17**. doi:10.1186/s12911-017-0442-4
135 Oswaldo Santos B, Reis Santana LM, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS One* 2018;**13**. doi:http://dx.doi.org/10.1371/journal.pone.0195065
136 Özcan T, Tüysüz F. Healthcare expenditure prediction in Turkey by using genetic algorithm based grey forecasting models. Int. Ser. Oper. Res.Manag. Sci. 2018;**262**:159–90. doi:10.1007/978-3-319-65455-3_7
137 Panaretos D, Koloverou E, Dimopoulos AC, *et al.* A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002-2012): the ATTICA study. *Br J Nutr* 2018;**120**:1–9. doi:https://dx.doi.org/10.1017/S0007114518001150
138 Park J, Edington DW. A sequential neural network model for diabetes prediction. *Artif Intell Med* 2001;**23**:277–93.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=11704441
139 Parsaeian M, Mohammad K, Mahmoudi M, *et al.* Comparison of Logistic Regression and Artificial Neural Network in Low Back Pain Prediction: Second National Health Survey. *Iran J Public Heal* 2012;**41**:86–92.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3469002/pdf/ijph-41-86.pdf NS -
140 Pei Z, Liu J, Liu M, *et al.* Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine. *Interdiscip Sci* 2018;**10**:126–30. doi:https://dx.doi.org/10.1007/s12539-017-0271-2
141 Pekkala T, Hall A, Lotjonen J, *et al.* Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study. *J Alzheimers Dis* 2017;**55**:1055–67.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=27802228
142 Peterson AT, Martinez-Campos C, Nakazawa Y, *et al.* Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. *Trans R Soc Trop Med Hyg* 2005;**99**:647–55.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15979656
143 Poole S, Grannis S, Shah NH. Predicting Emergency Department Visits. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci* 2016;**2016**:438–45.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=prem&NEWS=N&AN=27570684
144 Popescu M, Khalilia M. Improving Disease Prediction Using ICD-9 Ontological Features. In: *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), 27-30 June 2011*. Piscataway, NJ, USA: : IEEE 2011. 1805–9.NS -
145 Potash E, Brew J, Loewi A, *et al.* Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning. Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2015. doi:10.1145/2783258.2788629
146 Puddu PE, Menotti A. Artificial neural networks versus proportional hazards Cox models to predict 45-year all-cause mortality in the Italian Rural Areas of the Seven Countries Study. *BMC Med Res Methodol*

2012;**12**. doi:10.1186/1471-2288-12-100
147 Puddu PE, Menotti A. Artificial neural network versus multiple logistic function to predict 25-year coronary heart disease mortality in the Seven Countries Study. *Eur J Cardiovasc Prev Rehabil* 2009;**16**:583–91. doi:https://dx.doi.org/10.1097/HJR.0b013e32832d49e1
148 Qiang X, Kou Z. Prediction of interspecies transmission for avian influenza A virus based on a back-propagation neural network. *Math Comput Model* 2010;**52**:2060–5. doi:10.1016/j.mcm.2010.06.008
149 Rajliwall NS, Chetty G, Davey R. Chronic disease risk monitoring based on an innovative predictive modelling framework. Institute of Electrical and Electronics Engineers Inc. 2018. 1–8. doi:10.1109/SSCI.2017.8285257
150 Ram S, Zhang W, Williams M, *et al.* Predicting asthma-related emergency department visits using big data. *IEEE J Biomed Heal informatics* 2015;**19**:1216–23. doi:https://dx.doi.org/10.1109/JBHI.2015.2404829
151 Ramos R, Silva C, Moreira MWL, *et al.* Using predictive classifiers to prevent infant mortality in the Brazilian northeast. Institute of Electrical and Electronics Engineers Inc. 2017. 1–6. doi:10.1109/HealthCom.2017.8210811
152 Rasmy L, Zheng WJ, Xu H, *et al.* A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J Biomed Inform* 2018;**84**:11–6. doi:http://dx.doi.org/10.1016/j.jbi.2018.06.011
153 Ray EL, Sakrejda K, Lauer SA, *et al.* Infectious disease prediction with kernel conditional density estimation. *Stat Med* 2017;**36**:4908–29. doi:http://dx.doi.org/10.1002/sim.7488
154 Razavian N, Blecker S, Schmidt AM, *et al.* Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. *Big data* 2015;**3**:277–87. doi:https://dx.doi.org/10.1089/big.2015.0020
155 Ren H, Li J, Yuan Z-A, *et al.* The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China. *BMC Infect Dis* 2013;**13**:421. doi:https://dx.doi.org/10.1186/1471-2334-13-421
156 Robson JO, Verstraete SG, Shiboski S, *et al.* A Risk Score for Childhood Obesity in an Urban Latino Cohort. *J Pediatr* 2016;**172**:29-34.e1. doi:10.1016/j.jpeds.2016.01.055
157 Rodriguez EA, Estrada FE, Torres WC, *et al.* Early prediction of severe maternal morbidity using machine learning techniques. Lect. Notes Comput. Sci. 2016;**10022 LNAI**:259–70. doi:10.1007/978-3-319-47955-2_22
158 Rose S. Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol* 2013;**177**:443–52. doi:https://dx.doi.org/10.1093/aje/kws241
159 Saberian F, Zamani A, Shoorehdeli MA, *et al.* Prediction of seasonal influenza epidemics in Tehran using artificial neural networks. In: *2014 22nd Iranian Conference on Electrical Engineering (ICEE), 20-22 May 2014*. Piscataway, NJ, USA: : IEEE 2014. 1921–3. doi:10.1109/IranianCEE.2014.6999855
160 Sadilek A, Kautz H, Silenzio V. Modeling spread of disease from social interactions. 2012. 322–9.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871981836&partnerID=40&md5=ddc1c242ae203453a8846fa59de09240 NS -
161 Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. 2012. 136–42.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84868268429&partnerID=40&md5=adeafe49ba93bc3c38d55525b18c1243 NS -
162 Safi N, Adimi F, Soebiyanto RP, *et al.* Toward malaria risk prediction in Afghanistan using remote sensing. International Society for Photogrammetry and Remote Sensing 2010. 339–42.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84924065446&partnerID=40&md5=db144d9b2dbec33ad5dc068d11cfed8d NS -
163 Sanchez AS, Iglesias-Rodriguez FJ, Fernandez PR, *et al.* Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *Int J Ind Ergon* 2016;**52**:92–9. doi:10.1016/j.ergon.2015.09.012
164 Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. *Theor Biol Med Model* 2014;**11 Suppl 1**:S6. doi:https://dx.doi.org/10.1186/1742-4682-11-S1-S6
165 Schalkwijk D, Graaf A, Tsivtsivadze E, *et al.* Lipoprotein metabolism indicators improve cardiovascular risk prediction. PLoS One. 2014;**9**:e92840. doi:10.1371/journal.pone.0092840
166 Shaoyan Z, Tjortjis C, Xiaojun Z, *et al.* Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf Syst Front* 2009;**11**:449–60. doi:10.1007/s10796-009-9157-0
167 Shearer FM, Longbottom J, Browne AJ, *et al.* Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. *Lancet Glob Heal* 2018;**6**:e270–8. doi:https://dx.doi.org/10.1016/S2214-109X(18)30024-X

168 Sheets L, Petroski GF, Zhuang Y, *et al.* Combining Contrast Mining with Logistic Regression To Predict Healthcare Utilization in a Managed Care Population. *Appl Clin Inform* 2017;**8**:430–46. doi:http://dx.doi.org/10.4338/ACI-2016-05-RA-0078

169 Shen F, Yuan J, Sun Z, *et al.* Risk Identification and Prediction of Coal Workers' Pneumoconiosis in Kailuan Colliery Group in China: A Historical Cohort Study. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0082181

170 Shi Y, Liu X, Kok S-Y, *et al.* Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore. *Environ Health Perspect* 2016;**124**:1369–75. doi:https://dx.doi.org/10.1289/ehp.1509981

171 Simon GE, Johnson E, Lawrence JM, *et al.* Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records. *Am J Psychiatry* 2018;:appiajp201817101167. doi:https://dx.doi.org/10.1176/appi.ajp.2018.17101167

172 Soebiyanto RP, Kiang R. Meteorological parameters as predictors for seasonal influenza. *Geocarto Int* 2014;**29**:39–47. doi:10.1080/10106049.2013.799717

173 Song J. SVR model for prediction of incidence influenza based on automated method. Adv. Mater. Res. 2014;**926–930**:1159–63. doi:10.4028/www.scientific.net/AMR.926-930.1159

174 Song X, Mitnitski A, MacKnight C, *et al.* Assessment of individual risk of death using self-report data: An artificial neural network compared with a frailty index. *J Am Geriatr Soc* 2004;**52**:1180–4. doi:http://dx.doi.org/10.1111/j.1532-5415.2004.52319.x

175 Song X, Mitnitski A, Cox J, *et al.* Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform* 2004;**107**:736–40.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15360910

176 Spencer KL, Olson LM, Schnetz-Boutaud N, *et al.* Using Genetic Variation and Environmental Risk Factor Data to Identify Individuals at High Risk for Age-Related Macular Degeneration. *PLoS One* 2011;**6**. doi:http://dx.doi.org/10.1371/journal.pone.0017784

177 Sung Kean K, Tae Keun Y, Ein O, *et al.* Osteoporosis risk prediction using machine learning and conventional methods. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 3-7 July 2013*. Piscataway, NJ, USA: : IEEE 2013. 188–91. doi:10.1109/EMBC.2013.6609469

178 Sushmita S, Newman S, Marquardt J, *et al.* Population Cost Prediction on Public Healthcare Datasets. Proc. 5th Int. Conf. Digit. Heal. 2015. 2015. doi:10.1145/2750511.2750521

179 Tafaro L, Cicconetti P, Piccirillo G, *et al.* Is it possible to predict one-year survival in centenarians? A neural network study. *Gerontology* 2005;**51**:199–205.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=15832048

180 Tamaki Y, Nomura Y, Katsumura S, *et al.* Construction of a dental caries prediction model by data mining. *J Oral Sci* 2009;**51**:61–8. doi:http://dx.doi.org/10.2334/josnusd.51.61

181 Toschke AM, Beyerlein A, von Kries R. Children at high risk for overweight: a classification and regression trees analysis approach. *Obes Res* 2005;**13**:1270–4.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16076998

182 Tran T, Luo W, Phung D, *et al.* Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. *BMC Psychiatry* 2014;**14**. doi:10.1186/1471-244X-14-76

183 van den Kommer TN, Comijs HC, Dik MG, *et al.* Development of classification models for early identification of persons at risk for persistent cognitive decline. *J Neurol* 2008;**255**:1486–94. doi:10.1007/s00415-008-0942-3

184 Van Voorhees BW, Paunesku D, Gollan J, *et al.* Predicting future risk of depressive episode in adolescents: the Chicago Adolescent Depression Risk Assessment (CADRA). *Ann Fam Med* 2008;**6**:503–11. doi:https://dx.doi.org/10.1370/afm.887

185 Vestergaard P, Kruse C, Goemaere S, *et al.* Predicting mortality and incident immobility in older Belgian men by characteristics related to sarcopenia and frailty. *Osteoporos Int* 2018;**29**:1437–45. doi:10.1007/s00198-018-4467-z

186 Volkova S, Ayton E, Porterfield K, *et al.* Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLoS One* 2017;**12**. doi:10.1371/journal.pone.0188941

187 Voss R, Cullen P, Schulte H, *et al.* Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. *Int J Epidemiol* 2002;**31**:1253–64.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12540731

188 Vovsha I, Rajan A, Salleb-Aouissi A, *et al.* Predicting preterm birth is not elusive: Machine learning paves the way to individual wellness. AI Access Foundation 2014. 82–9.https://www.scopus.com/inward/record.uri?eid=2-s2.0-84904912879&partnerID=40&md5=02cbb14baeb811a357606bcc5e9d11ca NS -

189 Walton NA, Poynton MR, Gesteland PH, *et al.* Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *BMC Med Informatics Decis Mak* 2010;**10**:68. doi:10.1186/1472-6947-10-68

190 Walz Y, Wegmann M, Leutner B, *et al.* Use of an ecologically relevant modelling approach to improve remote sensing-based schistosomiasis risk profiling. *Geospat Health* 2015;**10**:271–9. doi:10.4081/gh.2015.398

191 Wang AG, An N, Chen GL, *et al.* Predicting hypertension without measurement: A non-invasive, questionnaire-based approach. *Expert Syst Appl* 2015;**42**:7601–9. doi:10.1016/j.eswa.2015.06.012

192 Wang CJ, Li YQ, Wang L, *et al.* Development and Evaluation of a Simple and Effective Prediction Approach for Identifying Those at High Risk of Dyslipidemia in Rural Adult Residents. *PLoS One* 2012;**7**:e43834. doi:http://dx.doi.org/10.1371/journal.pone.0043834

193 Wang JF, Liu X, Liao YL, *et al.* Prediction of Neural Tube Defect Using Support Vector Machine. *Biomed Environ Sci* 2010;**23**:167–72. doi:10.1016/S0895-3988(10)60048-7

194 Wang J, Jia P, Cuadros DF, *et al.* A Remote Sensing Data Based Artificial Neural Network Approach for Predicting Climate-Sensitive Infectious Disease Outbreaks: A Case Study of Human Brucellosis. *Remote Sens* 2017;**9**. doi:10.3390/rs9101018

195 Wang J, Zhou Y, Kou Z. One class support vector machine for predicting avian-to-human transmission of avian influenza a virus. 2012. 184–8. doi:10.1109/CSAE.2012.6272935

196 Wang J, Deng Z. Modeling and Prediction of Oyster Norovirus Outbreaks along Gulf of Mexico Coast. *Environ Heal Perspect* 2016;**124**:627.http://libaccess.mcmaster.ca/login?url=https://search.proquest.com/docview/1806104356?accountid=12347

197 Wang KW, Deng C, Li JP, *et al.* Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol Infect* 2017;**145**:1118–29. doi:https://dx.doi.org/10.1017/S0950268816003216

198 Wang Y, Gu J. Comparative study among three different artificial neural networks to infectious diarrhea forecasting. Institute of Electrical and Electronics Engineers Inc. 2014. 40–6. doi:10.1109/BIBM.2014.6999373

199 Wang Y, Li J, Gu J, *et al.* Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Appl Soft Comput J* 2015;**35**:280–90. doi:10.1016/j.asoc.2015.05.047

200 Warren H, Casas J-P, Hingorani A, *et al.* Genetic prediction of quantitative lipid traits: comparing shrinkage models to gene scores. *Genet Epidemiol* 2014;**38**:72–83. doi:https://dx.doi.org/10.1002/gepi.21777

201 Wei WD, Jiang JJ, Gao L, *et al.* A New Hybrid Model Using an Autoregressive Integrated Moving Average and a Generalized Regression Neural Network for the Incidence of Tuberculosis in Heng County, China. *Am J Trop Med Hyg* 2017;**97**:799–805. doi:10.4269/ajtmh.16-0648

202 Wei W, Visweswaran S, Cooper GF, *et al.* The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *J Am Med Informatics Assoc* 2011;**18**:370–5. doi:10.1136/amiajnl-2011-000101

203 Wei W, Jiang J, Liang H, *et al.* Application of a Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China. *PLoS One* 2016;**11**:e0156768. doi:https://dx.doi.org/10.1371/journal.pone.0156768

204 Wen L, Li-Zhong Z, Dong-Wang M, *et al.* Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. *Medicine (Baltimore)* 2018;**97**:1–7. doi:10.1097/MD.0000000000010529

205 Woolery LK, Grzymala-Busse J. Machine learning for an expert system to predict preterm birth risk. *J Am Med Inform Assoc* 1994;**1**:439–46.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=7850569

206 Wu C, Kao SC, Shih CH, *et al.* Open data mining for Taiwan's dengue epidemic. *Acta Trop* 2018;**183**:1–7. doi:http://dx.doi.org/10.1016/j.actatropica.2018.03.017

207 Wu H, Cai Y, Wu Y, *et al.* Time series analysis of weekly influenza-like illness rate using a one-year period of factors in random forest regression. *Biosci Trends* 2017;**11**:292–6. doi:https://dx.doi.org/10.5582/bst.2017.01035

208 Wu JL, Roy J, Stewart WF. Prediction Modeling Using EHR Data Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Med Care* 2010;**48**:S106–13. doi:10.1097/MLR.0b013e3181de9e17
209 Wu W, Guo J, An S, *et al.* Comparison of Two Hybrid Models for Forecasting the Incidence of Hemorrhagic Fever with Renal Syndrome in Jiangsu Province, China. *PLoS One* 2015;**10**. doi:http://dx.doi.org/10.1371/journal.pone.0135492
210 Wu Y, Yang Y, Nishiura H, *et al.* Deep Learning for Epidemiological Predictions. 41st Int. ACM SIGIR Conf. Res. & Dev. Inf. Retr. 2018. doi:10.1145/3209978.3210077
211 Yan W, Xu Y, Yang X, *et al.* A hybrid model for short-term bacillary dysentery prediction in Yichang city, China. *Jpn J Infect Dis* 2010;**63**:264–70.http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med6&NEWS=N&AN=20657066
212 Yang C, Delcher C, Shenkman E, *et al.* Machine Learning Approaches for Predicting High Utilizers in Health Care. 2017;**pt.II**:382–95. doi:10.1007/978-3-319-56154-7_35
213 Ye C, Fu T, Hao S, *et al.* Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *J Med Internet Res* 2018;**20**:e22. doi:https://dx.doi.org/10.2196/jmir.9268
214 Yiheng Z, Jingyao Z, Jiebo L. Predicting Multiple Risky Behaviors via Multimedia Content. In: *Social Informatics. 9th International Conference, SocInfo 2017, 13-15 Sept. 2017*. Cham, Switzerland: : Springer International Publishing 2017. 65–73. doi:10.1007/978-3-319-67256-4_7
215 Yoo TK, Kim SK, Kim DW, *et al.* Osteoporosis Risk Prediction for Bone Mineral Density Assessment of Postmenopausal Women Using Machine Learning. *Yonsei Med J* 2013;**54**:1321–30. doi:10.3349/ymj.2013.54.6.1321
216 Yoon Y, Song J, Hong SH, *et al.* Analysis of multiple single nucleotide polymorphisms of candidate genes related to coronary heart disease susceptibility by using support vector machines. *Clin Chem Lab Med* 2003;**41**:529–34. doi:http://dx.doi.org/10.1515/CCLM.2003.080
217 Young SD, Torrone EA, Urata J, *et al.* Using Search Engine Data as a Tool to Predict Syphilis. *Epidemiology* 2018;**29**:574–8. doi:10.1097/EDE.0000000000000836
218 Young SG, Tullis JA, Cothren J. A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus. *Appl Geogr* 2013;**45**:241–9. doi:10.1016/j.apgeog.2013.09.022
219 Young SD, Zhang × Qingpeng. Using search engine big data for predicting new HIV diagnoses. *PLoS One* 2018;**13**. doi:http://dx.doi.org/10.1371/journal.pone.0199527
220 Yu L, Zhou L, Tan L, *et al.* Application of a New Hybrid Model with Seasonal Auto-Regressive Integrated Moving Average (ARIMA) and Nonlinear Auto-Regressive Neural Network (NARNN) in Forecasting Incidence Cases of HFMD in Shenzhen, China. *PLoS One* 2014;**9**. doi:http://dx.doi.org/10.1371/journal.pone.0098241
221 Zeng Q, Huang HL, Pei X, *et al.* Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. *Anal Methods Accid Res* 2016;**10**:12–25. doi:10.1016/j.amar.2016.03.002
222 Zhang G, Huang S, Duan Q, *et al.* Application of a Hybrid Model for Predicting the Incidence of Tuberculosis in Hubei, China. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0080969
223 Zhang X, Liu Y, Yang M, *et al.* Comparative Study of Four Time Series Methods in Forecasting Typhoid Fever Incidence in China. *PLoS One* 2013;**8**. doi:http://dx.doi.org/10.1371/journal.pone.0063116
224 Zhang X, Zhang T, Young AA, *et al.* Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS One* 2014;**9**:e88075. doi:https://dx.doi.org/10.1371/journal.pone.0088075
225 Zhao D, Weng C, Zhao D, *et al.* Combining PubMed knowledge and EHR data to develop a weighted bayesian network for pancreatic cancer prediction. *J Biomed Inform* 2011;**44**:859–68. doi:10.1016/j.jbi.2011.05.004
226 Zhao Y, Xiong P, McCullough LE, *et al.* Comparison of Breast Cancer Risk Predictive Models and Screening Strategies for Chinese Women. *J Womens Heal* 2017;**26**:294–302. doi:10.1089/jwh.2015.5692
227 Zhongwen G, Meng L. Design of H7N9 Avian Influenza Management and Forecasting System Based on GIS. In: *2015 IEEE 5th International Conference on Electronics, Information and Emergency Communication (ICEIEC), 14-16 May 2015*. Piscataway, NJ, USA: : IEEE 2015. 376–9. doi:10.1109/ICEIEC.2015.7284562
228 Zhou LL, Xia J, Yu LJ, *et al.* Using a Hybrid Model to Forecast the Prevalence of Schistosomiasis in Humans. *Int J Environ Res Public Heal* 2016;**13**. doi:10.3390/ijerph13040355
229 Zhou L, Yu L, Wang Y, *et al.* A hybrid model for predicting the prevalence of schistosomiasis in humans of

Qianjiang City, China. *PLoS One* 2014;**9**:e104875. doi:https://dx.doi.org/10.1371/journal.pone.0104875
230 Zhou Y, Glenn C, Luo J. Understanding and predicting multiple risky behaviors from social media. AI Access Foundation 2017. 600–5.https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029534550&partnerID=40&md5=ae29c604e757e03f7561a191d4b296e1 NS -
231 Zhuo Z, Jiang L, Chee Keong K, *et al.* Learning in Glaucoma Genetic Risk Assessment. In: *2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010), 31 Aug.-4 Sept. 2010*. Piscataway, NJ, USA: : IEEE 2010. 6182–5. doi:10.1109/IEMBS.2010.5627757

Appendix D: Narrative Synthesis of Aspects of Discussion

Rationale for applying machine learning approaches mainly centered around it being “state of the art” or better suited to modeling complex data than regression. Machine learning was thought to be “state of the art” due to improved accuracy and deeper insights. Discussions of complex modeling focused on capturing non-linear relationships, interactions, and high-dimensionality.

When authors discussed model limitations, frequent concerns were an inadequate sample size, too few features, questionable generalizability, and a lack of interpretability. Aspects of the data other than sample size and feature number, such as potential measurement error or selection bias, were infrequently mentioned.

When discussing model implementation, many articles stated that predictive accuracy would be improved; but they did not frequently discuss how this could be translated to specific health-related policies or actions. Additionally, they rarely mentioned organizations and knowledge users that would be best suited to leverage the model.