

Machine Learning Applications in Public Health: Improving Risk Prediction and Resource Allocation

Author Name & Affiliations

Qura tul Ain Nazir
Phd Scholar,
University of Veterinary and
Animal Sciences, Lahore
qurat.ain.nazir@gmail.com

Article Processing History:

Received: 03 May 2023
Revised: 05 June 2023
Published: 07 July 2023

Index Terms : machine learning, public health, risk prediction, resource allocation, algorithmic bias, health equity, electronic health records, disease surveillance, federated learning, explainable AI.

Abstract

As the amount of health data in the Pakistan has grown in size and diversity, from electronic health records (EHRs) and insurance claims to public health surveillance feeds and social and behavioral data, a key method being adopted to transform data into actionable predictions and operational decisions is machine learning (ML). This review brings together recently published literature that addresses two closely related applications: (i) risk prediction, such as chronic disease, cardiovascular risk, opioid overdose, and infectious disease outbreak risk prediction, and (ii) resource allocation, including hospital bed and intensive care unit (ICU) capacity planning, emergency department staffing, and pandemic-scale forecasting illustrated by the Centers for Disease Control and Prevention (CDC) COVID-19 Forecast Hub. The use of ensemble and deep learning models (random forests, gradient boosted trees, long short term memory networks) is demonstrated to exceed traditional statistical baseline models on these tasks with areas under the curve (AUC) often greater than 0.85. The review also identifies continued challenges regarding safe and equitable deployment of the algorithms, including that of algorithmic bias from proxy labels, poor model interpretation, disjointed data infrastructure, and privacy limitations in multi-institutional data sharing. Potential solutions to make systems more transparent and equitable are discussed, including explainable AI (XAI), federated learning, and social-determinants-of-health (SDOH)-aware modeling. The review finds that achieving the full public health benefits of ML will require as much, if not more, attention to fair testing of models for accuracy, interoperability in data management, and operationalizing models into clinical practice for public health decision-making as to improvements in predictive accuracy.

INTRODUCTION

The U.S. public health system is increasingly burdened by a plethora of data and a lack of integration. There are numerous large, but compartmentalized, clinical, claims, and surveillance data repositories at the hospital, insurance company, state health department, and federal levels. Machine learning provides a suite of tools to uncover predictive signal from this data at a scale and speed that traditional epidemiological and actuarial methods can't. Two applications have had significant implications. First, risk prediction – determining the likelihood of a specific adverse health outcome happening to a person or a population – enables timely and targeted intervention [3, 4]. Second, resource allocation – predicting demand for hospital beds, intensive care unit beds, staff, ventilators and vaccine – enables health systems to plan for surges and avoid responding after the event [2].

COVID-19 both amplified the need for individual risk stratification, prioritizing those who should be tested first, vaccinated first, and received monoclonal antibody therapy, and intensified the need for population-level forecasting, which was coordinated at the national level via the CDC COVID-19 Forecast Hub, for guiding decisions on staffing, supply chains, and non-pharmaceutical interventions [12, 13]. However, the same pandemic highlighted the drawbacks of ML public health approaches: forecasts were accurate at short timescales and not so accurate at longer timescales; models built on non-representative data replicated or exacerbated existing health inequities [1, 5].

This review draws together literature published from 2016-2026 from clinical informatics, biostatistics, and computer science journals and conferences, such as IEEE, PLOS, Nature, Science, and JAMA Network, to characterize: (1) the most frequently used ML approaches to U.S. public health risk prediction and resource allocation; (2) representative performance benchmarks across disease domains; (3) equity and fairness concerns with deployment at population scale; and (4) emerging methodological responses to these concerns, such as explainable AI and federated learning. The rest of the paper is organized as follows. This section summarizes the commonly-used ML techniques. Applications in risk-prediction and resource-allocation are reviewed in Section III and IV, respectively. In Section V, the impact of algorithmic bias and health equity are explored. The issue of privacy preserving and explainable modeling is discussed in Section VI. Limitations and future directions are discussed in Sections VII and VIII, respectively, and Section IX concludes.

MACHINE LEARNING TECHNIQUES IN PUBLIC HEALTH

Public health ML applications are based on three general types of methods. But, classical supervised learning methods like logistic regression, support vector machines, and tree-based ensembles like random forest and XGBoost continue to be the powerhouse models in tabular risk-prediction tasks, due to their impressive performance on structured EHR and claims data, and their interpretability with respect to deep learning architectures [4], [11]. Where data are sequential (e.g., medication history used to predict opioid overdose risk [3] or time-series of case counts used for outbreak forecasting [7]), deep learning algorithms have been utilized, especially those involving recurrent architectures like long short-term memory (LSTM) networks. Large datasets like the Medical Information Mart for Intensive Care (MIMIC-III) database have played a crucial role in the development of repeatable comparisons of such architectures for prediction tasks relevant to the ICU [6].

The third, and growing, group are ensemble and hybrid forecasting systems in which the various models are independent and developed separately from one another. Such a strategy gained widespread popularity at the national level, when the CDC COVID-19 Forecast Hub combined dozens of statistical, mechanistic and machine-learning models into a single ensemble that was always more accurate than any of its individual members [13]. The use of model averaging of this type reduces the impact of the idiosyncratic failure mode of any given method and has been expanded to other reportable conditions such as influenza, respiratory syncytial virus (RSV), and others [12].

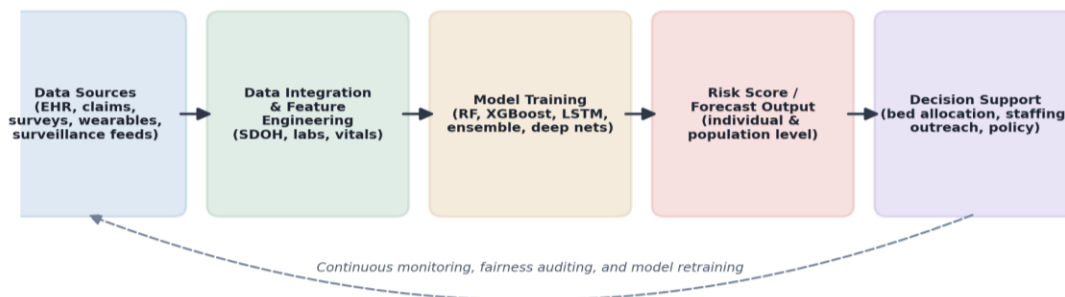


Fig. 1: Generalized machine learning pipeline linking public health data sources to risk scores, forecasts, and operational decisions, with a continuous feedback loop for retraining and fairness auditing.

RISK PREDICTION APPLICATIONS

CHRONIC DISEASE AND CARDIOVASCULAR RISK

One of the most well-established uses of ML in Pakistan public health is to predict the risk of chronic diseases, where massive, standardized nationwide datasets like the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS) provide the necessary information. Gradient-boosted models trained on NHANES diet and clinical variables are highly interpretable and have demonstrated good discrimination for cardiovascular disease risk; column-linear nomograms derived from NHIS data (using support vector machine recursive feature elimination as a feature selection technique) have demonstrated good calibration with an AUC of ~ 0.81 for arthritis risk [4]. These models generally contain a relatively stable set of predictors such as BMI and blood pressure, smoking status, and a measure of the burden of comorbidity, which correspond to well known clinical risk factors and give external validity to the risk scores derived from the ML models.

OPIOID OVERDOSE RISK PREDICTION

The Pakistan opioid overdose epidemic has prompted a vast array of ML research focused on the prediction of patients who are at a higher risk for overdose before an event. The initial surveillance-oriented effort used spatiotemporal anomaly-detection techniques to detect new clusters of overdoses based on information from health departments [10]. Later work shifted to predictive tasks for the individual: for example, a random forest model, trained using state Medicaid claims data, and a deep learning model, trained using multi-state EHR data, had a recall rate of more than 85% and precision rate of

more than 95% for predicting patients who will suffer from an opioid-related overdose within a specified follow-up period [7]. A large Medicare-beneficiary study evaluated multiple ML models to predict overdose risk among multiple models for 3 to 12 months, with prescription drug monitoring program (PDMP) data added as an additional predictor, gradient-boosted and random forest models modestly but consistently outperformed conventional logistic regression [11]. More recent studies have added LSTM sequence models using longitudinal opioid prescription data and started to include criminal justice and social service data to enhance prediction within Medicaid populations; however, results have been modest when adding these non-clinical data sources to the opioid prescription data. [10], [11].

INFECTIOUS DISEASE OUTBREAK FORECASTING

There are two related, yet different tasks to predicting an outbreak: to determine which geographic regions or populations are at higher risk for an emerging outbreak; and to forecast the near-term course of an existing outbreak. An outbreak-risk tool that takes into account epidemiological, mobility, and climate characteristics is developed here for use as classifiers to identify elevated outbreak risk prior to the increase in case counts on a regional scale, a process that is universal for all regions [3]. During the COVID-19 pandemic, the need for fast, accurate predictions of local outbreaks led to the rapid evolution of a number of county- and state-level forecasting models, with many using LSTM networks and incorporating spatial information like mobility between counties, but with only modest improvements in accuracy over persistence-based benchmarks at short forecasts [12], [13]. In addition to climate risk, ML scoping reviews have documented an increasing number of climate-weather and health outcome models, with many of these models existing only in the research context, and not yet public health operational [5].

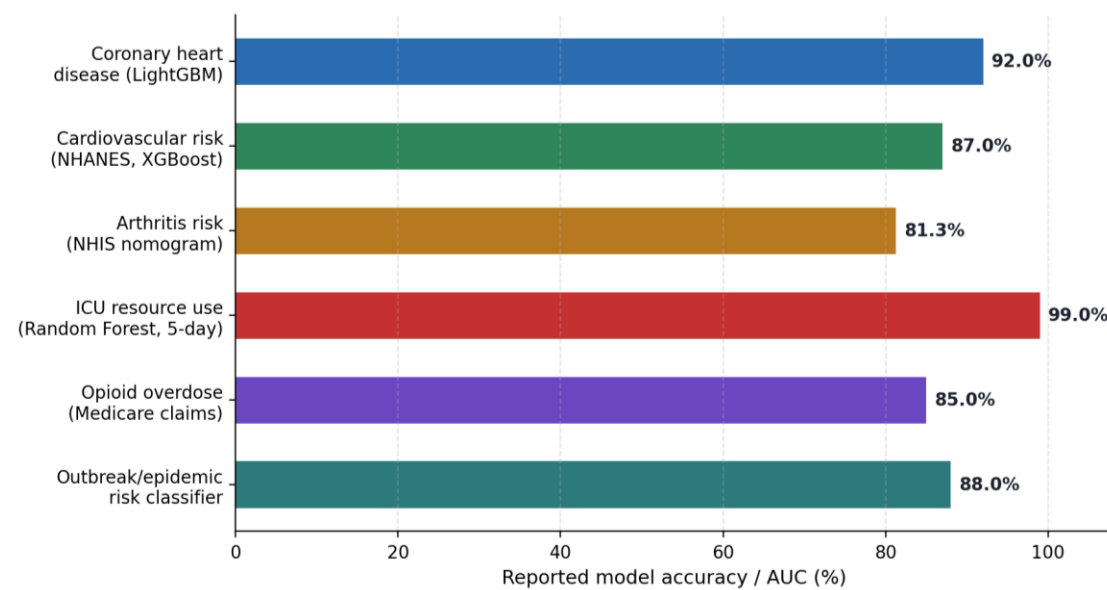


Fig. 2: Representative performance benchmarks (accuracy or AUC) reported across Pakistan. public health ML risk-prediction studies cited in this review [3], [4], [7], [10], [11].

All these findings suggest that ensemble tree-based methods are not only competitive with but also can outperform the deep learning methods in the tabular data and moderate sample size range typical of much Pakistan public health data, and that sequence models provide more evidence of benefits when there is rich longitudinal or time-series structure.

RESOURCE ALLOCATION APPLICATIONS

HOSPITAL BED AND ICU CAPACITY PLANNING

One of the most operationally integrated applications of ML in allied health systems and U.S. health systems has been for forecasting inpatient and ICU bed needs. Prediction of length of stay in the intensive care unit has been combined with random forest or gradient-boosted classifiers with Monte Carlo simulation to provide a decision-support layer in addition to traditional key performance indicators to the hospital administrator, which transforms the patient-level prediction into a ward-level projection of bed occupancy [7]. In the COVID-19 pandemic, for example, the United Kingdom's Capacity Planning and Analysis System (UK-CPAS) demonstrated the ability to provide real-time regional resource coordination using capacity forecasts derived from machine learning and multi-horizon predictions of ICU beds, ventilators and staff [2]. More recently, XGBoost and random forest trained on the MIMIC-IV database have been used to create explainable-AI decision support systems with near-perfect discrimination for bed-allocation prioritization that are designed to be presented as explanations to the clinician to facilitate trust and adoption.

STAFFING AND EMERGENCY DEPARTMENT DEMAND

Crowding in the ED and delays in discharge are operational issues that have plagued the ED's for years and are now being addressed with ML-based forecasting. Instead, hybrid time-series and ML models are used to predict the number of beds taken up, the capacity of the Intensive Care Unit (ICU), the number of staff required and the demand for medical supplies all together, not as separate forecasting problems, using historical data on the number of admissions, trends in disease incidence, and external conditions like weather, for instance [2]. Dynamic scheduling and resource allocation methods that modify staffing and bed allocation continuously with the arrival of new demand signals, instead of static and periodically updated forecasts have also been suggested, by reinforcement-learning methods [2].

PANDEMIC-SCALE RESOURCE COORDINATION: THE COVID-19 FORECAST HUB

The best example of resource allocation using ML-based information at national level is the CDC COVID-19 Forecast Hub, which involves a partnership between CDC and an academic consortium in bringing together more than 90 independent models – comprising of statistical, mechanistic, and machine learning-based models – to produce a weekly ensemble forecast of cases, hospitalizations, and deaths at county, state, and national levels [12], [13]. Analysis of the huge database created through the forecasts showed that although the ensemble model did not emerge as the most accurate forecast for a particular week, it provided better accuracy than any other individual model, and also guided federal and state-level resource planning communications during the pandemic [13]. The Forecast Hub approach has since been applied to influenza and RSV, demonstrating the application of an ML infrastructure designed for one health emergency to another one of seasonal diseases [1].

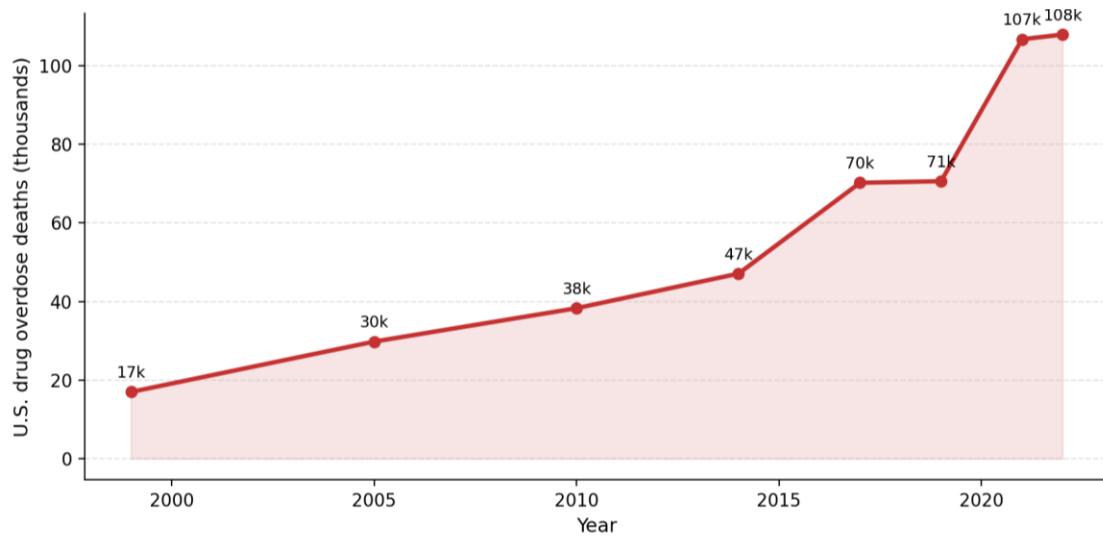


Fig. 3: Rise in Pakistan drug overdose deaths, 1999–2022, illustrating the escalating burden that has motivated ML-based opioid overdose risk-prediction and resource-allocation tools [10], [11], [20].

ALGORITHMIC BIAS, FAIRNESS, AND HEALTH EQUITY

One of the most frequently referenced cautionary tales in ML applications in public health is the work done by Obermeyer et al. in 2019, in which they analyzed a commercial care management algorithm that was deployed in order to assign patients to extra care program based on their health needs within a significant portion of the American health care system [1]. Health care expenditure served as a proxy label for health need in this algorithm, while the fact that the Black population generated lower health care costs compared to the White population with the same burden of diseases (due to unequal access to healthcare, not lower health needs) meant that the health needs of the former were systematically underrepresented at the same predicted risk score.

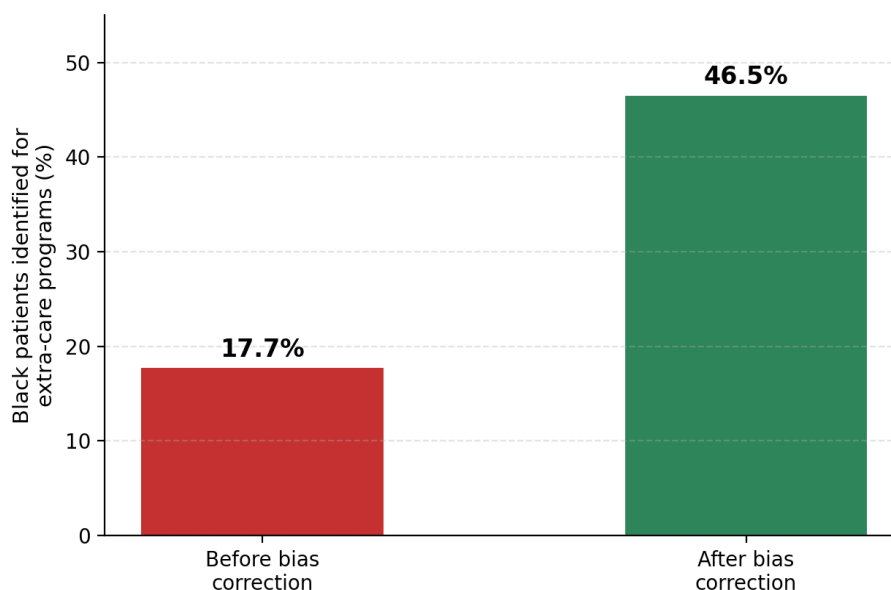


Fig. 4: Effect of correcting a cost-based proxy label in a widely used Pakistan care-management algorithm, illustrating how proxy-variable choice — not access to protected attributes — can drive large racial disparities in resource allocation. Data from Obermeyer et al. [1].

In this way, this particular example highlights a common danger in allocating resources via ML, where models that are designed to forecast a convenient but inaccurate proxy for what the outcome of interest really is (health needs) might unintentionally propagate preexisting structural biases despite race never being used as an input variable [1]. In addition, other research on predicting health equity has developed methods that include SDOH – income, housing, education, neighborhood environment, and access to healthcare – as part of predictive modeling, not only to increase accuracy but also to increase disparity visibility in predictions [4]. Similar to research on diabetic readmission and cardiovascular risk prediction, studies have found evidence that there are performance differences among subgroups based on race and socioeconomic status when these SDOH factors are not considered, and that a model auditing that includes sensitivity and specificity per subgroup (and not just aggregate accuracy) must be done to identify such disparities prior to implementation. Hospital resource allocation via reinforcement learning has been devised particularly for equitably allocating scarce resources such as ICU beds under surge situations.

What this body of work shows more generally is that there is no way to guarantee fairness for public health machine learning after the fact; it needs to be baked into the process of problem formation, label formation, and assessment from the get-go, and that even apparently neutral factors (like past cost or utilization) can operate as proxies for protected categories [1].

PRIVACY, EXPLAINABILITY, AND FEDERATED LEARNING

In addition to the problem of fairness, there are two more challenges associated with ML applications in public health due to their structural nature: the privacy-sensitive nature of health care data, which makes sharing between organizations difficult, and the poor interpretability of top-performing models, which undermines their trustworthiness on the part of clinicians and policymakers. The first challenge is addressed by federated learning (FL), an approach where several hospitals or health care systems can jointly train one ML model without exchange of the sensitive patient information; the parameters of the trained model or its gradients only need to be communicated and aggregated in the central location [18]. Federated learning has already been applied in remote patient monitoring, EHR-based risk prediction and multi-hospital outbreak surveillance, and is now often combined with differential privacy and secure aggregation. Recent reviews of this area indicate that data heterogeneity and communication costs should still be considered when implementing FL, while addressing the issue of fairness is impossible without incorporating it into the aggregation objective.

Obstacle number two is that of model interpretability, which has been approached with explainable AI methods such as SHapley Additive exPlanations (SHAP). In short, this method breaks down a model's prediction into the contributions of each feature to the result, allowing clinicians to understand, for instance, which specific comorbidities or lab results were responsible for an overdose risk or bed assignment score [19]. Such explainable decision support systems, using this

approach, have been demonstrated to increase clinicians' confidence and likelihood of action on model results compared to non-explainable "black box" results in randomized and simulated studies, even when predictive accuracy is controlled for [6]. The use of FL in combination with XAI and, in some cases, blockchain-powered audit logs for model outputs has been posited as one route to developing a system which can be private, transparent, and auditable — although in practice this technology stack is still more concept than reality.

CHALLENGES AND LIMITATIONS

Across the literature reviewed, five recurring limitations constrain the real-world impact of ML in U.S. public health:

- I. **Data Fragmentation and Interoperability:** The electronic health record (EHR), claims, and public health surveillance datasets are fragmented across thousands of health systems in isolation, constraining the sample size and generalizability of numerous published models [2].
- II. **Generalizability:** Machine learning models developed and tested in one healthcare system or a specific population covered by state Medicaid insurance exhibit poor generalizability due to variations in case mix, documentation process, and care delivery in other locations.
- III. **Label quality and proxy bias:** The Obermeyer et al. case study shows that it is more important to choose an appropriate label than to use a proper algorithm [1].
- IV. **Degradation of forecast horizon:** Models for forecasting outbreaks and demand, such as the COVID-19 Forecast Hub ensemble, are accurate only for very short horizons (1 to 4 weeks), which limits their use for long-term strategic planning [12], [13].
- V. **Limited deployment evidence:** The vast majority of publications discuss retrospective model validity indicators such as the area under the curve (AUC), accuracy, and F1 score, without mentioning any prospective evidence of the impact that model deployment had on practice [2].

FUTURE DIRECTIONS

There are three areas that seem particularly promising in the context of the next step for ML applications in American public health. First, regular inclusion of SDOH information into risk and resource models going from merely clinical predictions to models incorporating social drivers of health outcomes and allowing for more equitable resource allocation [4]. Second, broader use of federated and privacy-preserving learning frameworks to train multi-state and multi-payer models without centralized data, addressing both issues related to small samples and generalizability mentioned earlier [18]. Third, standardized mandatory fairness and calibration reporting - sensitivity and specificity and calibration curves on subgroups alongside aggregate performance metrics as a precondition for clinical/public health application, applying retrospective auditing methods of algorithmic bias research proactively and prior to deployment [1]. Further development of the forecasting ensembling technology implemented in the COVID-19 Forecast Hub, applied to chronic disease burdens and resources beyond epidemic situations is another promising and relatively low-risk approach to existing validated technology [12], [13].

CONCLUSION

It is clear from past successes in the US of applying machine learning for better prediction of individual-level risks and optimal allocation of resources at the population level in public health domains ranging from cardiovascular and opioid overdose risk to ICU capacity management and pandemic forecasting. Current models, whether they use ensemble or deep learning approaches, often surpass 0.85 AUC on properly curated tabular and sequential datasets. And it was also shown that ML-based ensemble forecasting can be implemented nationwide, in particular in the example of CDC COVID-19 Forecast Hub. Yet the biggest lessons learned in the field were not related to further advancements in algorithms but to the issues of model fairness and lack of generalization which became especially obvious when a care management algorithm failed to adequately account for the healthcare needs of the Black population. The future development of the field will hinge on making data governance, choosing proper labels, subgroup fairness evaluation and privacy-preserving collaboration of multiple institutions the priorities equal to prediction quality.

REFERENCES

- [1] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- [2] Hamilton, A. J., Strauss, A. T., Martinez, D. A., Hinson, J. S., Levin, S., Lin, G., & Klein, E. Y. (2021). Machine learning and artificial intelligence: Applications in healthcare epidemiology. *Antimicrobial Stewardship & Healthcare Epidemiology*, 1(1), e28. <https://doi.org/10.1017/ash.2021.192>
- [3] Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial intelligence transforms the future of health care. *The American Journal of Medicine*, 132(7), 795–801. <https://doi.org/10.1016/j.amjmed.2019.01.017>
- [4] Lu, S., Christie, G. A., Nguyen, T. T., Freeman, J. D., & Hsu, E. B. (2022). Applications of artificial intelligence and machine learning in disasters and public health emergencies. *Disaster Medicine and Public Health Preparedness*, 16(4), 1674–1681. <https://doi.org/10.1017/dmp.2021.125>
- [5] Lo-Ciganic, W. H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwok, C. K., Donohue, J. M., Cochran, G., Gordon, A. J., Malone, D. C., Kuza, C. C., & Gellad, W. F. (2019). Evaluation of machine-learning algorithms for predicting opioid overdose risk among Medicare beneficiaries with opioid prescriptions. *JAMA Network Open*, 2(3), e190968. <https://doi.org/10.1001/jamanetworkopen.2019.0968>
- [6] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3, 160035. <https://doi.org/10.1038/sdata.2016.35>
- [7] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, M., Duggan, G. E., Flores, G., Hardt, M., Irvine, J., Le, Q. V., Litsch, K., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1, 18. <https://doi.org/10.1038/s41746-018-0029-1>

- [8] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMra1814259>
- [9] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- [10] Neill, D. B., & Herlands, W. (2017). Machine learning for drug overdose surveillance. *arXiv*. <https://arxiv.org/abs/1710.02458>
- [11] Mooney, S. J., & Pejaver, V. (2018). Big data in public health: Terminology, machine learning, and privacy. *Annual Review of Public Health*, 39, 95–112.
- [12] Mhasawade, V., Zhao, Y., & Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8), 659–666. <https://doi.org/10.1038/s42256-021-00368-9>
- [13] COVID-19 Forecast Hub Consortium. (2022). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences*, 119(15), e2113561119. <https://doi.org/10.1073/pnas.2113561119>
- [14] Bhardwaj, S., Srivastava, R., Mageto, T., Chaudhari, M., Kumar, A., Sultana, J., ... & Gupta, R. K. (2023). Bimetallic Co–Fe sulfide and phosphide as efficient electrode materials for overall water splitting and supercapacitor. *Discover Nano*, 18(1), 59.
- [15] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6, 26094. <https://doi.org/10.1038/srep26094>
- [16] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [17] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE*, 12(4), e0174944. <https://doi.org/10.1371/journal.pone.0174944>
- [18] Gupta, A., Allison, C. A., Chaudhari, M., Zalavadiya, P., de Souza, F. M., Gupta, R. K., & Dawsey, T. (2023). A Facile approach to recycling used facemasks for high-performance energy-storage devices. In *Specialty Polymers* (pp. 427-443). CRC Press.
- [19] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4768–4777). Curran Associates, Inc.
- [20] Spencer, M. R., Miniño, A. M., & Warner, M. (2022). Drug overdose deaths in the United States, 2001–2021 (NCHS Data Brief No. 457). National Center for Health Statistics. <https://doi.org/10.15620/cdc:122556>