

Predictive Healthcare Models Using Advanced Data Analytics

Tawanda Mushiri

Associate Professor, University of Zimbabwe, Zimbabwe

Abstract

Predictive healthcare models supported by advanced data analytics are increasingly important for improving disease-risk assessment, patient monitoring, clinical decision-making, and healthcare resource management. The growing availability of electronic health records, medical imaging, laboratory information, genomic datasets, wearable-device measurements, and longitudinal patient information has created opportunities for developing analytical models capable of identifying clinically relevant patterns. This study examines predictive healthcare models using advanced data analytics, with emphasis on machine learning, statistical modeling, deep learning, predictive classification, risk estimation, and healthcare decision support. A qualitative descriptive and conceptual research design was adopted using secondary academic literature.

The analysis indicates that predictive healthcare models can support early disease identification, patient-risk stratification, readmission prediction, clinical outcome estimation, personalized healthcare, and resource planning. Advanced data analytics can process large and heterogeneous datasets and identify relationships that may not be easily detected through conventional analytical methods. However, successful implementation is influenced by data quality, missing information, algorithmic bias, interpretability, privacy, cybersecurity, interoperability, model generalizability, and clinical validation.

The study concludes that predictive healthcare models can strengthen data-driven healthcare when developed using representative datasets, appropriate analytical techniques, rigorous validation, and clinically meaningful interpretation. Future research should emphasize multimodal data integration, explainable predictive models, real-time analytics, privacy-preserving learning, personalized prediction, and prospective clinical evaluation.

Keywords: Predictive Healthcare, Data Analytics, Machine Learning, Healthcare Prediction, Clinical Decision Support, Predictive Modeling, Electronic Health Records, Healthcare Informatics, Risk Assessment, Artificial Intelligence.

1. Introduction

Healthcare systems generate increasingly large quantities of data through electronic health records, medical imaging, laboratory investigations, genomic analysis, wearable devices, clinical monitoring, and administrative systems. These datasets contain valuable information concerning patient characteristics, disease progression, treatment patterns, and healthcare outcomes.

Advanced data analytics provides computational approaches for transforming such information into meaningful patterns that can support prediction, risk assessment, and healthcare decision-making.

Traditional healthcare analysis often focuses on diagnosing existing conditions or examining historical outcomes. Predictive healthcare models extend this approach by estimating the likelihood of future events or identifying patients who may face increased health risks. Predictive models can support early identification of disease, hospitalization risk assessment, readmission prediction, treatment-response estimation, and patient deterioration monitoring. These capabilities can contribute to more proactive healthcare management.

Statistical modeling represents an important foundation of predictive healthcare analytics. Logistic regression, survival analysis, regression models, and other statistical techniques can identify relationships between patient characteristics and healthcare outcomes. These methods can provide relatively interpretable results and remain valuable for clinical research. However, healthcare datasets may contain complex nonlinear relationships that require more advanced computational approaches.

Machine learning has expanded the scope of predictive healthcare modeling by enabling algorithms to learn patterns from historical data. Classification models can predict disease categories or clinical outcomes, while regression approaches can estimate continuous measures. Ensemble techniques can combine multiple predictive models, potentially improving performance. These methods can analyze large datasets and identify relationships among numerous clinical variables.

Deep learning provides additional capabilities for highly complex healthcare datasets. Neural networks can process medical images, clinical text, physiological signals, and other high-dimensional information. These systems can automatically learn representations from data and support prediction without relying entirely on manually selected variables. However, deep-learning models generally require substantial datasets and can present challenges related to interpretability and computational requirements.

Electronic health records represent a major source of predictive healthcare information. EHRs may include patient demographics, diagnoses, medications, laboratory results, procedures, vital signs, and longitudinal clinical observations. Predictive models can analyze these data to estimate risks such as hospitalization, readmission, disease progression, or clinical deterioration. Nevertheless, EHR information can contain missing values, inconsistent documentation, and differences in coding practices.

Predictive models can contribute to early disease detection. By analyzing combinations of risk factors and clinical indicators, models can estimate the probability of developing or experiencing particular conditions. Early risk identification may enable healthcare professionals to undertake additional assessment or preventive measures. However, predictive outputs should be interpreted carefully because incorrect predictions can create unnecessary interventions or missed opportunities for timely care.

Patient risk stratification is another important application of predictive analytics. Healthcare organizations can classify patients according to their estimated level of risk and prioritize appropriate monitoring or interventions. Risk-based approaches can support chronic disease management and resource allocation. The reliability of such models depends on appropriate outcome definitions, representative data, and continuous evaluation.

Predictive analytics can also support hospital resource management. Models can estimate patient admissions, length of stay, emergency department demand, and other operational outcomes. These predictions can assist healthcare organizations in planning staffing, bed capacity, and other resources. However, operational prediction should remain aligned with clinical priorities and should not compromise patient-centered care.

2. Objectives of the Study

Primary Objective

The primary objective of this study is to examine predictive healthcare models using advanced data analytics by analyzing statistical modeling, machine learning, deep learning, predictive classification, risk estimation, and multimodal healthcare data integration.

The study seeks to understand how predictive analytics can support disease-risk assessment, patient monitoring, clinical decision-making, healthcare planning, and personalized care while maintaining accuracy, transparency, fairness, privacy, security, and clinical validity.

2.1 Examination of Predictive Healthcare Models

The first objective is to examine major predictive healthcare models, including regression analysis, classification algorithms, ensemble learning, deep learning, time-series modeling, and predictive risk models.

The study evaluates their ability to process electronic health records, medical imaging, laboratory data, genomic information, physiological measurements, and other healthcare datasets for identifying patterns and estimating clinically relevant outcomes.

2.2 Assessment of Healthcare Applications

The second objective is to assess applications of advanced predictive analytics in disease-risk assessment, early detection, readmission prediction, patient deterioration monitoring, treatment-response estimation, personalized healthcare, and hospital resource management.

The study examines how predictive outputs can provide healthcare professionals with additional information for proactive patient management and evidence-based healthcare planning.

2.3 Identification of Implementation Requirements

The third objective is to identify important implementation requirements involving data quality, algorithmic fairness, model interpretability, privacy, cybersecurity, interoperability, generalizability, clinical validation, and continuous monitoring.

The study evaluates how these factors influence predictive reliability and emphasizes the need for secure, representative, transparent, clinically validated, and human-centered predictive healthcare systems.

3. Review of Literature

3.1 Predictive Analytics and Machine Learning in Healthcare

Deo (2015) discussed the growing application of machine learning in medicine and emphasized its potential for prediction and classification. Machine-learning models can process numerous clinical variables and identify relationships associated with health outcomes. Such methods provide opportunities for risk prediction, disease classification, and decision support, although their effectiveness depends on data quality and appropriate validation.

Rajkomar et al. (2019) highlighted the potential of machine learning for medical prediction using large healthcare datasets. Electronic health records can provide longitudinal information that supports predictive modeling. However, predictive performance may vary across institutions because of differences in patient populations, clinical practices, and data structures. External validation is therefore important for assessing real-world performance.

3.2 Deep Learning and Advanced Healthcare Data Analysis

Miotto et al. (2018) described deep learning as an important approach for analyzing complex healthcare information. Deep-learning models can process high-dimensional data and learn representations from clinical information. These capabilities can support predictive applications involving medical images, electronic health records, and other biomedical datasets.

Esteva et al. (2019) emphasized that deep learning has broad potential in healthcare applications. Deep-learning approaches can automatically extract features and support predictive classification. Nevertheless, substantial computational resources, large datasets, and appropriate validation may be required. Model interpretability also remains important when predictions are used in clinical environments.

3.3 Clinical Validation and Responsible Predictive Modeling

Kelly et al. (2019) identified important challenges associated with translating machine-learning systems into clinical practice. Predictive models must demonstrate clinical relevance rather than simply achieving strong performance on research datasets. Validation across diverse clinical environments is necessary to establish reliability and practical usefulness.

Wiens et al. (2019) emphasized responsible machine learning in healthcare, highlighting the importance of safety, fairness, data quality, and monitoring. Predictive healthcare systems can influence clinical decisions and therefore require appropriate governance. Transparent development, representative datasets, and continuous evaluation can support responsible implementation.

34. Research Methodology

4.1 Research Design

This study adopted a qualitative descriptive and conceptual research design to examine predictive healthcare models using advanced data analytics. Secondary academic literature concerning predictive analytics, machine learning, deep learning, healthcare informatics, clinical decision support, electronic health records, medical imaging, and risk prediction was reviewed.

The conceptual framework considered data collection, preprocessing, feature extraction, predictive modeling, outcome estimation, validation, interpretation, and healthcare decision support as interconnected stages. These components were examined according to predictive performance, clinical relevance, interpretability, fairness, privacy, security, generalizability, and practical usability.

4.2 Data Collection

Secondary information was collected from peer-reviewed academic articles, scholarly books, and established healthcare research publications. Literature concerning electronic health records, medical imaging, laboratory data, genomic information, wearable-device data, physiological monitoring, machine learning, deep learning, and predictive healthcare was considered.

The literature was organized according to analytical methods and healthcare applications. Analytical categories included regression, classification, ensemble learning, deep learning, time-series analysis, and multimodal modeling. Application categories included disease prediction, risk stratification, readmission prediction, patient deterioration, treatment-response prediction, personalized care, and healthcare resource management.

4.3 Data Analysis

The collected literature was analyzed through thematic analysis. Major themes included predictive modeling, data integration, clinical risk assessment, early disease identification, patient monitoring, explainability, fairness, privacy, cybersecurity, interoperability, generalizability, and clinical validation. A conceptual synthesis was developed to describe how advanced data analytics transforms healthcare information into predictive insights.

Data are collected and prepared, relevant characteristics are identified, predictive models estimate outcomes, and results are evaluated before being interpreted within healthcare contexts. Professional judgment remains essential when translating predictive information into clinical decisions.

5. Analysis

5.1 Predictive Models Used in Healthcare

Advanced analytical approaches provide different predictive capabilities.

Table 1: Predictive Healthcare Models and Applications

Predictive Model	Main Analytical Capability	Healthcare Application
Logistic Regression	Probability estimation	Disease-risk prediction
Linear Regression	Continuous prediction	Outcome estimation
Decision Tree	Rule-based prediction	Clinical classification
Random Forest	Ensemble prediction	Risk stratification
Support Vector Machine	Pattern classification	Disease prediction

Predictive Model	Main Analytical Capability	Healthcare Application
Gradient Boosting	Nonlinear prediction	Clinical risk analysis
Neural Networks	Complex pattern learning	Healthcare prediction
Deep Learning	High-dimensional analysis	Multimodal prediction

The analysis demonstrates that predictive healthcare models provide complementary capabilities. Statistical models can offer relatively interpretable relationships, while machine-learning and deep-learning approaches can capture complex nonlinear patterns. Model selection should consider the type of healthcare data, clinical objective, prediction horizon, interpretability requirements, computational resources, and consequences associated with incorrect predictions.

5.2 Healthcare Applications of Predictive Analytics

Table 2: Healthcare Applications of Predictive Models

Application	Main Data Sources	Predictive Contribution
Disease-Risk Prediction	EHR and Laboratory Data	Risk estimation
Readmission Prediction	Clinical Records	High-risk patient identification
Patient Deterioration	Vital Signs and EHR	Early warning
Treatment Response	Clinical and Molecular Data	Response estimation
Disease Progression	Longitudinal Data	Progression prediction
Hospital Demand	Administrative Data	Resource planning
Personalized Care	Multimodal Patient Data	Individual risk assessment
Remote Monitoring	Wearable Data	Emerging-risk identification

The analysis indicates that predictive analytics can contribute to both clinical and operational healthcare activities. Disease-risk prediction and patient deterioration monitoring can support proactive clinical management, while hospital-demand prediction can support organizational planning. However, predictive outputs require appropriate validation and interpretation to ensure that they provide meaningful information within specific healthcare contexts.

6. Discussion

Advanced data analytics provides healthcare organizations with opportunities to move from primarily reactive healthcare toward more proactive prediction and risk assessment. Predictive models can process large volumes of clinical information and identify patterns associated with future outcomes. These capabilities may support earlier intervention, improved patient monitoring, and more informed healthcare planning. Statistical models remain valuable because they can provide interpretable relationships between patient characteristics and outcomes. Logistic regression, for example, can estimate the probability of an event based on multiple variables. Such approaches can be useful when clinicians need understandable information. However, complex healthcare datasets may contain nonlinear relationships that require more advanced machine-learning methods.

Machine learning can identify complex relationships within healthcare datasets. Classification models can estimate whether patients are likely to experience specific outcomes, while regression models can estimate continuous outcomes. Ensemble methods can combine multiple predictive approaches and may provide improved robustness. However, model performance must be evaluated using clinically meaningful measures. Deep learning can provide additional capabilities for complex and high-dimensional healthcare data. Medical images, physiological signals, clinical text, and other information can be analyzed using neural networks. Deep-learning models can automatically extract relevant representations from data. Their complexity, however, can make interpretation difficult and may require substantial computational resources.

Electronic health records provide extensive opportunities for predictive healthcare analytics. Longitudinal EHR information can support predictions concerning hospitalization, readmission, disease progression, and clinical deterioration. However, EHR datasets frequently contain missing information and documentation inconsistencies. Models should therefore be developed using appropriate preprocessing and validation strategies. Disease-risk prediction is an important application of predictive healthcare models. By combining demographic, clinical, laboratory, and historical variables, predictive systems can estimate the likelihood of disease or adverse outcomes. Such predictions may help healthcare professionals identify patients requiring additional assessment. However, risk predictions should not be treated as definitive diagnoses.

Readmission prediction can support post-discharge management. Patients with elevated predicted risk may receive additional follow-up, education, or monitoring. The effectiveness of such systems depends not only on prediction accuracy but also on whether healthcare organizations can provide appropriate interventions after risk identification. Patient deterioration prediction is another important application. Continuous analysis of vital signs, laboratory measurements, and clinical observations can identify changes associated with worsening health conditions. Early warning systems may support timely clinical attention. However, excessive false-positive alerts can create alert fatigue, making threshold selection and workflow integration important.

Predictive analytics can also support personalized healthcare. Individuals with similar diagnoses may have substantially different risks and treatment responses. By analyzing patient-specific information, predictive models can estimate individualized outcomes. Personalized prediction can therefore provide more targeted information, although it requires sufficient patient-level data and appropriate validation.

Multimodal analytics can integrate different types of healthcare information. Combining EHR data with imaging, laboratory measurements, genomic information, and wearable-device data may improve the representation of patient health. However, multimodal integration increases analytical complexity and requires appropriate data alignment and preprocessing.

7. Challenges

7.1 Data Quality and Heterogeneity

Predictive healthcare models depend on reliable data, yet healthcare datasets frequently contain missing values, measurement errors, inconsistent documentation, duplicate records, and incomplete histories.

These problems can reduce predictive reliability and create misleading relationships between variables. Data preprocessing and quality-control procedures are therefore necessary before predictive modeling. Healthcare data also vary across hospitals, laboratories, medical devices, and geographic environments.

Differences in patient populations, measurement techniques, documentation practices, and data formats can influence model performance. Standardization and external validation are essential for determining whether predictive models can maintain reliable performance across different healthcare environments.

7.2 Bias and Generalizability

Predictive models can produce biased outcomes when training datasets do not adequately represent relevant patient populations. Unequal representation can create differences in prediction performance across demographic or clinical groups. Such disparities may reduce fairness and reliability. Researchers should therefore examine subgroup performance and ensure that development datasets include appropriate population diversity.

Generalizability remains difficult because predictive models may learn patterns specific to their development environment. A model trained using data from one healthcare organization may not perform similarly elsewhere. Differences in clinical workflows, patient characteristics, disease prevalence, and data collection can affect predictions. Multi-institutional datasets and external validation can improve confidence in generalizability.

7.3 Interpretability and Clinical Adoption

Complex predictive models may generate accurate risk estimates without providing explanations that healthcare professionals can easily understand. Limited interpretability can make it difficult to evaluate whether a prediction is clinically reasonable. Explainable AI methods can provide information concerning influential variables, although explanations should be interpreted carefully and validated for clinical relevance.

Clinical adoption also depends on workflow compatibility. Predictive systems that generate excessive alerts or require complicated procedures may increase workload. Healthcare professionals need clear information concerning prediction meaning, uncertainty, limitations, and appropriate responses. Human-centered design, professional training, and continuous feedback can improve the practical integration of predictive analytics.

7.4 Privacy, Security, and Interoperability

Predictive healthcare systems process sensitive information such as clinical histories, laboratory results, genomic information, and physiological measurements. Unauthorized access or inappropriate data sharing can compromise confidentiality. Strong encryption, authentication, access controls, secure storage, and privacy-preserving methods are therefore important for protecting healthcare information throughout the predictive analytics lifecycle.

Interoperability presents another challenge because healthcare information is frequently distributed across different digital systems. EHRs, laboratories, imaging platforms, pharmacies, and wearable devices may use different formats. Limited interoperability can prevent predictive models from accessing complete information. Standardized data structures and communication mechanisms can improve integration and support more comprehensive predictive healthcare analysis.

8. Future Research

Future research should emphasize multimodal predictive healthcare models that integrate EHRs, medical imaging, laboratory results, genomic information, clinical text, and wearable-device data. Explainable and fairness-aware algorithms should be developed to improve transparency and equitable performance. Research should also investigate real-time prediction, personalized risk estimation, early warning systems, and uncertainty-aware healthcare analytics.

Further research should prioritize large-scale multi-institutional datasets and prospective clinical validation. Federated learning and privacy-preserving analytics could support collaborative predictive modeling while protecting sensitive information. Future studies should examine model drift, interoperability, cybersecurity, clinical workflow integration, and patient outcomes. These developments can strengthen the reliability and practical usefulness of advanced predictive healthcare systems.

9. Conclusion

Predictive healthcare models using advanced data analytics provide significant opportunities for strengthening modern healthcare decision-making. Statistical models, machine-learning algorithms, and deep-learning approaches can analyze complex clinical information and estimate risks or future outcomes. These capabilities can support disease prediction, patient monitoring, early intervention, clinical decision support, and healthcare resource planning.

Predictive analytics can be applied across multiple healthcare domains, including disease-risk assessment, readmission prediction, patient deterioration monitoring, treatment-response estimation, personalized healthcare, and operational planning. Electronic health records, laboratory information, medical imaging, genomic data, and wearable-device measurements can provide valuable predictive information. Multimodal integration may provide more comprehensive patient representations.

The reliability of predictive healthcare models depends strongly on data quality, representativeness, and validation. Missing information, inconsistent documentation, dataset heterogeneity, algorithmic bias, and differences between healthcare environments can influence predictive performance. External validation, subgroup evaluation, continuous monitoring, and appropriate model recalibration are therefore important for maintaining reliable predictions.

Interpretability, privacy, cybersecurity, and interoperability are also essential considerations. Healthcare professionals need meaningful information concerning predictions and their limitations, while sensitive patient information must be protected. Interoperable systems are necessary to integrate information from different healthcare platforms. Human oversight should remain central to ensure that predictive outputs are interpreted appropriately.

Overall, advanced predictive healthcare analytics can contribute to more proactive, personalized, and data-driven healthcare when responsibly implemented. Successful adoption requires high-quality data, appropriate analytical methods, clinical validation, transparent interpretation, secure infrastructure, fairness, and professional oversight. Future research should focus on multimodal prediction, explainable AI, privacy-preserving learning, real-time analytics, personalized risk assessment, and prospective clinical evaluation.

References

1. Chen, I. Y., Joshi, S., Ghassemi, M., & Ranganath, R. (2021). Probabilistic machine learning for healthcare. *Annual Review of Biomedical Data Science*, 4, 393–415. <https://doi.org/10.1146/annurev-biodatasci-092820-033938>
2. de Hond, A. A. H., Leeuwenberg, A. M., Hooft, L., Kant, I. M. J., Nijman, S. W. J., van Os, H. J. A., Aardoom, J. J., Debray, T. P. A., Schuit, E., van Smeden, M., Reitsma, J. B., Steyerberg, E. W., Chavannes, N. H., & Moons, K. G. M. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: A scoping review. *npj Digital Medicine*, 5, 2. <https://doi.org/10.1038/s41746-021-00549-7>
3. McAdams, R. M., Kaur, R., Sun, Y., Bindra, H., Cho, S. J., et al. (2022). Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: A systematic review. *Journal of Perinatology*, 42, 1561–1575.
4. Zhang, A., Xing, L., Zou, J., & Wu, J. C. (2022). Shifting machine learning for healthcare from development to deployment and from models to data. *Nature Biomedical Engineering*, 6, 1330–1345. <https://doi.org/10.1038/s41551-022-00898-y>
5. Steinberg, E., Jung, K., Fries, J. A., Corbin, C. K., Pfohl, S. R., & Shah, N. H. (2021). Language models are an effective representation learning technique for electronic health record data. *Journal of Biomedical Informatics*, 113, 103637. <https://doi.org/10.1016/j.jbi.2020.103637>
6. Waring, J., Lindvall, C., & Umeton, R. (2020). Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial Intelligence in Medicine*, 104, 101822. <https://doi.org/10.1016/j.artmed.2020.101822>
7. Pearson, T. A., Califf, R. M., Roper, R., Engelgau, M. M., Khoury, M. J., Alcantara, C., Blakely, C., et al. (2020). Precision health analytics with predictive analytics and implementation research: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 76(3). <https://doi.org/10.1016/j.jacc.2020.05.043>
8. Ayala Solares, J. R., Raimondi, F. E. D., Zhu, Y., Rahimian, F., Canoy, D., Tran, J., Gomes, A. C. P., Payberah, A. H., Zottoli, M., Nazarzadeh, M., Conrad, N., Rahimi, K., & Salimi-Khorshidi, G. (2020). Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics*, 101, 103337. <https://doi.org/10.1016/j.jbi.2019.103337>
9. Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *npj Digital Medicine*, 3, 136. <https://doi.org/10.1038/s41746-020-00341-z>

10. Landi, I., Glicksberg, B. S., Lee, H. C., Cherng, S., Landi, G., Danieleto, M., Dudley, J. T., Furlanello, C., & Miotto, R. (2020). Deep representation learning of electronic health records to unlock patient stratification at scale. *npj Digital Medicine*, 3, 96. <https://doi.org/10.1038/s41746-020-0301-z>
11. Hilton, C. B., Milinovich, A., Felix, C., Vakharia, N., Crone, T., Donovan, C., Proctor, A., et al. (2020). Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. *npj Digital Medicine*, 3, 51.
12. Tucker, A., Wang, Z., Rotalinti, Y., & Myles, P. (2020). Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *npj Digital Medicine*, 3, 147.
13. Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *Journal of Electrical Systems and Information Technology*, 10, 40. Note: this one is 2023, so exclude it if your cutoff is strictly before 2023.
14. Ashfaq, A., Sant'Anna, A., Lingman, M., & Nowaczyk, S. (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics*, 97, 103256. <https://doi.org/10.1016/j.jbi.2019.103256>
15. Chen, P., Dong, W., Lu, X., Kaymak, U., He, K., & Huang, Z. (2019). Deep representation learning for individualized treatment effect estimation using electronic health records. *Journal of Biomedical Informatics*, 100, 103303. <https://doi.org/10.1016/j.jbi.2019.103303>
16. Li, R., Chen, Y., Ritchie, M. D., & Moore, J. H. (2020). Electronic health records and polygenic risk scores for predicting disease risk. *Nature Reviews Genetics*, 21, 493–502. <https://doi.org/10.1038/s41576-020-0224-1>
17. Steele, A. J., Denaxas, S. C., Shah, A. D., Hemingway, H., & Luscombe, N. M. (2018). Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLOS ONE*, 13(8), e0202344. <https://doi.org/10.1371/journal.pone.0202344>
18. Wong, J., Horwitz, M. M., Zhou, L., & Toh, S. (2018). Using machine learning to identify health outcomes from electronic health record data. *Current Epidemiology Reports*, 5(4), 331–342. <https://doi.org/10.1007/s40471-018-0165-9>
19. Wong, A., Young, A. T., Liang, A. S., Gonzales, R., Douglas, V. C., & Hadley, D. (2018). Development and validation of an electronic health record-based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. *JAMA Network Open*, 1(4), e181018. <https://doi.org/10.1001/jamanetworkopen.2018.1018>
20. Brisimi, T. S., Xu, T., Wang, T., Dai, W., Adams, W. G., & Paschalidis, I. C. (2018). Predicting chronic disease hospitalizations from electronic health records: An interpretable classification approach. *Proceedings of the IEEE*, 106(4), 690–707. <https://doi.org/10.1109/JPROC.2017.2789319>