**Developing a Healthcare Monitoring System with a Comprehensive Dashboard**

|  |
| --- |
| **Paritosh Gandre**  **Farook Ahmed Ali Shaik**  **Deep Patel**  **Yash Kheni** |

**1)Author(s):** Balaram Yadav Kasula

**Date:** December 2023

**Title:** Machine Learning Applications in Diabetic Healthcare: A Comprehensive Analysis and Predictive Modeling.

**Journal: International Numeric Journal of Machine Learning and Robots.**

**Institution:** University of The Cumberlands, Williamsburg, KY, USA.

**URL:** <https://injmr.com/index.php/fewfewf/article/view/19/1>

**Summary:** Using data from 10,000 diabetic patients, the study offers a thorough analysis of the use of machine learning in diabetes treatment. The study compares many machine learning algorithms for predicting diabetes complications, and Random Forest outperformed SVM (81.2%) and Neural Networks (79.5%) with an accuracy of 85.6%. With a sensitivity of 88% in identifying diabetic nephropathy, the Decision Tree model indicates the study's effective risk stratification and early detection capabilities.

**Merits:**

1. Highly relevant to our project's predictive analytics goals, particularly in:
   * Implementation of multiple ML models for health risk prediction
   * Detailed analysis of model performance metrics
   * Feature importance analysis for healthcare data
   * Risk stratification approaches
2. Strong quantitative results that validate the effectiveness of ML in healthcare:
   * High prediction accuracy (85.6% with Random Forest)
   * Robust AUC scores (0.92 for Gradient Boosting)
   * Clear identification of key health indicators (HbA1c, BMI, diabetes duration)
3. Comprehensive methodology section that includes:
   * Data preprocessing
   * Feature selection
   * Model selection and evaluation
   * Cross-validation techniques

**Limitations:**

1. Focuses specifically on diabetes, while our project needs a broader healthcare scope
2. Limited discussion on:
   * Data security and privacy measures
   * Integration with existing EHR systems
   * Patient education components
   * Real-time monitoring capabilities
3. Does not address:
   * User interface design for healthcare dashboards
   * Patient engagement strategies
   * System scalability considerations
   * Interoperability standards

2**) Author(s):** Jiang, X., & Wang, B.   
  
**Year:** 2024  
  
**Title:** Enhancing Clinical Decision Making by Predicting Readmission Risk in Patients With Heart Failure Using Machine Learning: Predictive Model Development Study.   
  
**Website name:** JMIR Medical Informatics  
  
**URL:** <https://medinform.jmir.org/2024/1/e58812>

**Summary:** This study created a machine-learning model to forecast the 6-month readmission risks for 1,948 Chinese patients with heart failure. Out of 140 variables including demographics, lab findings, and comorbidities, 29 predictive characteristics were found using three variable selection techniques (LASSO regression, random forest, and χ² test). The following six models were put to the test: multilayer perceptrons (MLP), logistic regression, support vector machines (SVM), gradient boosting machines (GBM), Extreme Gradient Boosting (XGBoost), and graph convolutional networks (GCN). With the best results (AUC: 0.831, accuracy: 75%, specificity: 90.25%), the GCN model showed a high degree of predictive power for clinical decision assistance.

**Merits:**

1. Relevance to Predictive Analytics: Provides a model construction framework that directly supports using machine learning to anticipate health risks.
2. Sturdy Methodology: To fill up the gaps in localized predicting methods, a sizable, region-specific dataset (Sichuan Province, China) was used.

highlighted the importance of sophisticated deep learning for clinical outcomes by contrasting many algorithms (such as GCN vs. conventional models like logistic regression).

1. Data Preprocessing Rigor: KNN imputation was used for missing data and ethical compliance (IRB permission, anonymization).
2. Strategies for Feature Selection: Thorough identification of predictors provided by three different variable screening techniques.

**Failings:**

1. Limited Generalizability: Trained on a single Chinese regional dataset, raising questions about applicability to diverse populations or integration with EHRs in other regions.

No external validation cohort, reducing confidence in model robustness.

1. Clinical Interpretability: The high-performing GCN model’s "black box" nature limits clinician trust, a critical gap for our project’s goal to enhance doctor decision-making.
2. Sensitivity Concerns: Model sensitivity (52.12%) is relatively low, potentially missing nearly half of the at-risk patients, a critical flaw for early chronic condition detection.
3. Integration Challenges: Lacks discussion of real-world deployment (e.g., EHR interoperability), which is central to the scalability and integration objectives of our project.

3) **Author(s):** 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, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S. L., Chou, K., Pearson, M., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M. D., Cui, C., Corrado, G. S., & Dean, J.

**Year:** 2018

**Title:** Scalable and accurate deep learning with electronic health records

**Website name:** npj Digital Medicine

**URL:** <https://www.nature.com/articles/s41746-018-0029-1>

**Summary:** Using raw electronic health record (EHR) data from two university institutions in the United States, the study investigates the use of deep learning for clinical outcome prediction. The authors trained models to predict outcomes like in-hospital mortality (AUROC 0.93–0.94), 30-day unplanned readmissions (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.86), and discharge diagnoses (frequency-weighted AUROC 0.90) by presenting patient data in the FHIR (Fast Healthcare Interoperability Resources) format. By using unstructured data (such as clinical notes) and avoiding the need for human feature engineering, the models outperformed conventional clinical scoring systems and showed scalability across healthcare institutions.

**Merits:**

1. Comprehensive Data Utilization: To prevent time-consuming data harmonization, the method processes whole EHRs, including unstructured clinical notes. This supports the objective of our project, which is to centralize a variety of patient data.
2. High Predictive Accuracy: Proved the usefulness of predictive analytics in healthcare monitoring systems by outperforming conventional models (such as aEWS and HOSPITAL) across tasks.
3. Scalability: The approach has been validated in two hospitals and facilitates interoperability with current EHR systems, which is a primary goal of our project.
4. Interpretability: Case studies demonstrated how models found clinically significant characteristics (such as drugs or test results), assisting physicians in taking action.

**Failings:**

1. Limited Generalizability: Focused on academic medical centers; performance in community hospitals or outpatient settings remains unverified.
2. Computational Complexity: Training deep learning models on billions of data points requires significant infrastructure, which may challenge scalability for smaller institutions.
3. Ethical Considerations: No discussion of data privacy or bias mitigation, critical for our project’s security and compliance goals.
4. Clinical Integration: While predictive accuracy is high, real-world deployment challenges (e.g., clinician trust, and workflow integration) are unaddressed.

4) **Author(s):** Tianyi Liu, Andrew Krentz, Lei Lu, Vasa Curcin

**Year:** 2024

**Title:** Machine Learning based prediction models for Cardiovascular Disease Risk Using Electronic Health Records Data, Systematic Review and Meta-analysis

**Website name:** European Heart Journal - Digital Health

**URL:** <https://academic.oup.com/ehjdh/article/6/1/7/7845948#google_vignette>

**Summary:** This systematic review evaluates machine learning (ML) models for predicting cardiovascular disease (CVD) risk using electronic health records (EHRs) and compares their performance to conventional risk scores like QRISK3 and ASCVD. Analyzing 20 studies (2010–2024), the authors found ML models—particularly random forest (pooled AUC: 0.865) and deep learning (AUC: 0.847)—outperformed traditional statistical methods (AUC: 0.765). The study shows ML’s potential to improve risk stratification through EHR data integration but finds significant heterogeneity (I² > 99%) and methodological constraints across trials.

**Merits:**

1. Comprehensive Analysis: Provides a strong comparison of performance indicators (e.g., AUC, calibration) by combining evidence from 26 traditional models and 32 machine learning models.
2. Clinical Relevance: Shows how ML is better at using EHR data to predict long-term CVD risk, which is in line with the demands of individualized prevention in modern healthcare.
3. Methodological Rigor: Complies with PRISMA criteria, guaranteeing openness in the selection of studies and the evaluation of bias.
4. Practical Insights: Emphasizes how ML may manage patient comorbidities and heterogeneity, which traditional models frequently fail to take into account.

**Failings:**

1. Heterogeneity: Generalizability is constrained by the high degree of variation in study designs, datasets, and model architectures.
2. Publication Bias: Results may be skewed by the possible overrepresentation of high-performing models.
3. Clinical Applicability Gaps: Insufficient attention is paid to practical implementation issues, such as EHR interoperability and clinician interpretability of models.
4. Ethnic Diversity: The majority of research concentrated on US and UK populations, which limited its generalizability to cohorts worldwide.

5) **Authors**: Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, Kyung-Sup Kwak

**Year**: 2020

**Article Title**: A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Ensemble Deep Learning and Feature Fusion

**Journal**: *Information Fusion*

**Volume and Issue**: Volume 63, 2020, pp. 208–222

**URL**: <https://www.sciencedirect.com/science/article/pii/S1566253520303055>

**Summary:** This paper introduces a smart healthcare monitoring system for predicting heart disease using ensemble deep learning and feature fusion techniques. The system integrates data from wearable sensors and electronic medical records (EMRs) to create a comprehensive dataset. Feature selection is performed using information gain and conditional probability to reduce noise and enhance prediction accuracy. The deep learning model, combined with the LogitBoost algorithm, achieves an accuracy of 98.5%, outperforming traditional machine learning classifiers. Additionally, the system offers personalized health recommendations based on Semantic Web Rule Language (SWRL) rules. The proposed approach provides a highly accurate and robust prediction model, significantly improving heart disease diagnosis.

**Merits**:

* Achieves high accuracy (98.5%) using a combination of sensor and EMR data.
* Reduces noise and irrelevant features with feature fusion and information gain techniques.
* Provides personalized health recommendations using an ontology-based approach.

**Demerits**:

* Integrating multiple data sources can increase system complexity.
* Requires extensive preprocessing of sensor data to remove noise and artifacts.

6) **Authors:** Mohammad Ayoub Khan, Fahad Algarni

**Year**: 2020

**Article Title**: A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS

**Journal**: *IEEE Access*

**Volume and Issue**: Volume 8, 2020

**DOI/URL**: <https://ieeexplore.ieee.org/abstract/document/9131756>

**Summary:** This paper proposes a machine learning-based framework for heart disease diagnosis in the Internet of Medical Things (IoMT) environment. It combines Modified Salp Swarm Optimization (MSSO) with an Adaptive Neuro-Fuzzy Inference System (ANFIS) for improved prediction. The model optimizes ANFIS parameters using MSSO and Levy flight-based crow search to avoid local minima and achieve better accuracy. The system achieves an accuracy of 99.45%, which is higher than other comparable models. This IoMT-based framework is designed for real-time patient monitoring, providing accurate and scalable solutions for heart disease prediction in clinical and remote settings.

**Merits**:

* Utilizes MSSO to optimize ANFIS learning, achieving a prediction accuracy of 99.45%.
* Enhances search capability with Levy flight algorithm to prevent local minima.
* Suitable for real-time prediction in IoMT environments.

**Demerits**:

* High computational cost may limit scalability.
* Complex parameter tuning is required for the MSSO-ANFIS model.

7) **Authors**: Julia Hippisley-Cox, Carol Coupland, Peter Brindle

**Year**: 2017  
**Article Title**: Development and Validation of QRISK3 Risk Prediction Algorithms to Estimate Future Risk of Cardiovascular Disease: Prospective Cohort Study  
**Journal**: *BMJ (British Medical Journal)*  
**Volume and Issue**: BMJ 2017; 357:j2099  
**DOI/URL**: <https://www.bmj.com/content/357/bmj.j2099>

**Summary**: QRISK3 is an updated risk prediction algorithm that estimates the 10-year risk of cardiovascular disease by incorporating both traditional and newly identified clinical risk factors. The paper describes the development and validation of QRISK3 using data from 10 million patients in the UK. The algorithm includes new variables like systolic blood pressure variability, migraine, lupus, and severe mental illness, which improve prediction accuracy. Cox proportional hazards models are used to develop the algorithm, with multiple imputation for missing data. QRISK3 provides high calibration and discrimination (C-statistic of 0.88 for women and 0.86 for men), making it a reliable tool for clinical use in identifying individuals at high risk of heart disease.

**Merits**:

* Incorporates a comprehensive set of risk factors, improving prediction accuracy for specific populations.
* Validated on a large real-world dataset of over 10 million patients.
* Widely used in clinical settings for long-term cardiovascular risk prediction.

**Demerits**:

* Not suitable for real-time monitoring or short-term prediction.
* Relies on historical data, which may not reflect real-time patient conditions.

8) **Authors:** Min Chen, Yixue Hao, Kai Hwang, Lu Wang, Lin Wang

**Year:** 2017 **Article Title:** Disease Prediction by Machine Learning Over Big Data from Healthcare Communities  
**Journal**: *IEEE Access* **Volume and Issue:** Volume 5, 2017, pp. 8869–8870  
**DOI/URL:** <https://ieeexplore.ieee.org/document/7912315>

**Summary**: This paper focuses on predicting chronic diseases by combining structured and unstructured data using a convolutional neural network (CNN)-based multimodal prediction model. The study utilizes real-life hospital data from central China, including electronic health records (EHRs), medical image data, and gene data. A latent factor model is used to reconstruct missing data, while CNNs automatically extract features from unstructured text data. The proposed model achieves an accuracy of 94.8%, outperforming other conventional machine learning algorithms like Naive Bayes, K-Nearest Neighbor (KNN), and Decision Trees. The integration of structured and unstructured data offers a holistic approach to disease prediction, addressing key challenges in healthcare big data analysis.

**Merits:**

* Combines structured and unstructured data for a more holistic prediction model.
* CNN-based approach for unstructured data enhances prediction accuracy (94.8%).
* Addresses missing data with a latent factor model.

**Demerits:**

* Limited generalizability due to reliance on regional data.
* Complex CNN-based models require significant computational resources.

9) **Authors:** Chen, A., & Chen, D. O.

**Year:** 2022

**Article Title:** Simulation of a machine learning-enabled learning health system for risk prediction using synthetic patient data

**Journal:** Scientific Reports

**Volume and Issue:** 12(17917)

**DOI/URL:** <https://www.nature.com/articles/s41598-022-23011-4>

**Summary of the Study**

Chen and Chen (2022) propose a simulated ML-enabled LHS designed for risk prediction in healthcare. The study employs synthetic patient data generated by Synthea, a tool that mimics real-world patient data while avoiding privacy concerns. The authors develop an XGBoost-based risk prediction model for lung cancer, progressively improving its accuracy by incrementally adding more synthetic patient data. Their approach is then verified by applying the same methodology to stroke risk prediction, demonstrating the adaptability of their ML-enabled LHS.

**Key Contributions**

The study introduces a data-centric approach to LHS, which emphasizes continuous data updates to refine predictive models. Unlike traditional algorithm-centric ML models that focus on tuning algorithms, this study underscores the importance of improving data quality and quantity. Another major contribution is the public availability of both the ML code and synthetic patient data, facilitating further research. Furthermore, the study successfully showcases how synthetic data can be used to simulate real-world ML applications, overcoming privacy constraints associated with electronic health records (EHRs).

**Comparison with Existing Work**

While previous studies have explored ML applications in healthcare, most rely on real patient data, making replication and further advancements challenging due to privacy restrictions. For example, Wang et al. (2019) developed an ML-based lung cancer risk prediction model using real EHR data, achieving an AUC of 0.88. In contrast, Chen and Chen's (2022) synthetic data approach achieved an AUC of 0.96, highlighting its potential for model training without privacy risks. Other studies, such as Yeh et al. (2021), focused on deep learning models for lung cancer prediction, yet lacked the continuous learning aspect that this study integrates into its LHS framework.

**Merits and Strengths**

* **Use of Synthetic Data -** The study demonstrates how synthetic patient data can effectively train ML models while maintaining privacy.
* **Data-Centric Approach -** The iterative methodology ensures continuous improvement in model accuracy with incremental data updates.
* **Open Access -** The availability of ML code and synthetic datasets allows other researchers to build upon the findings.

**Limitations and Challenges**

* **Synthetic vs. Real Data** - Although synthetic data is useful, it does not fully replicate the complexity of real-world patient data.
* **Generalizability -** The findings may not directly translate to actual clinical settings without additional real-world validation.
* **Algorithm Scope -** The study primarily focuses on XGBoost and does not extensively explore deep learning techniques.

10) **Authors:** Shinde, S. A., & Rajeswari, P. R.

**Year:** 2018

**Article Title:** Intelligent health risk prediction systems using machine learning: A review

**Journal:** International Journal of Engineering & Technology

**Volume and Issue:** 7(3), 1019-1023

**DOI/URL:** <https://www.researchgate.net/publication/326253594_Intelligent_health_risk_prediction_systems_using_machine_learning_A_review>

**Summary of the Study**

Shinde and Rajeswari (2018) provide a comprehensive review of ML applications in health risk prediction. Their study categorizes ML methods into supervised, unsupervised, semi-supervised, reinforcement, and deep learning techniques. The authors explore various ML models, such as classification, clustering, regression, and optimization, applied to EHRs for disease prediction. The study also examines binary and multi-class classification techniques used in health informatics, emphasizing the role of ML in developing intelligent decision-support systems.

**Key Contributions**

* **Categorization of ML Methods -** The paper systematically classifies ML techniques based on their learning paradigms and problem-solving approaches.
* **Survey of Health Risk Prediction Models -** A broad analysis of existing ML-based disease prediction systems is provided.
* **Challenges in ML Applications -** The study highlights difficulties in handling incomplete and noisy biomedical datasets.
* **Future Research Directions -** The authors suggest improvements in data quality, feature selection, and ML model optimization.

**Comparison with Existing Work**

Compared to other studies, Shinde and Rajeswari (2018) focus more on categorizing ML approaches rather than implementing a specific model. Previous research, such as Wu et al. (2017), has applied ML to specific diseases like breast cancer, achieving high prediction accuracy. In contrast, this study provides a broader perspective, covering multiple ML techniques and their applications. Similarly, studies like Yeh et al. (2021) delve deeper into deep learning models for health prediction, whereas this review emphasizes a wide range of ML methods, including traditional approaches.

**Merits and Strengths**

* **Comprehensive ML Categorization -** The study effectively organizes ML techniques, making it easier for new researchers to understand the field.
* **Broad Application Scope -** Covers multiple diseases and ML approaches rather than focusing on a single health condition.
* **Identification of Challenges -** Discusses real-world challenges in applying ML to healthcare, such as data quality and computational constraints.

**Limitations and Challenges**

* **Lack of Experimental Validation -** The study reviews existing methods but does not implement or test any specific ML model.
* **Generalization Issues -** Covers a wide range of techniques without providing in-depth analysis of their individual effectiveness.
* **Limited Focus on Deep Learning -** While deep learning is mentioned, the study does not explore its potential in detail.
* **Challenges in Data Processing -** Highlights data-related issues but does not propose concrete solutions for overcoming them.

11) **Authors:** Abhadiomhen, S. E., Nzeakor, E. O., & Oyibo, K.

**Year:** 2024

**Article Title:** Health risk assessment using machine learning: Systematic review

**Journal:** Electronics

**Volume and Issue:** 13(4405)

**DOI/URL:** <https://www.mdpi.com/2079-9292/13/22/4405>

**Summary of the Study**

Abhadiomhen et al. (2024) conduct a systematic review of ML applications in HRA. Using a structured three-phase approach, they analyze 26 peer-reviewed studies from five databases. The study highlights that 42% of these studies focus on general health risks, while others target specific conditions. Secondary data sources dominate, with random forest emerging as the most frequently used algorithm. Additionally, the study identifies a gap in diverse sample representation and emphasizes the need for improving model interpretability to enhance trustworthiness in healthcare applications.

**Key Contributions**

* **Comprehensive Review -** It systematically categorizes ML applications in HRA, differentiating between general and condition-specific assessments.
* **Analysis of Data Sources -** Highlights the reliance on secondary datasets and the underrepresentation of diverse populations.
* **Popular ML Algorithms -** Identifies random forest as the most utilized algorithm, followed by ensemble methods and deep learning techniques.

**Comparison with Existing Work**

Compared to previous studies, Abhadiomhen et al. (2024) provide a broader perspective on ML applications in HRA. Prior works, such as Mishra et al. (2024), focus exclusively on pancreatic cancer risk prediction using electronic health records, evaluating the effectiveness of various ML techniques for this specific disease. In contrast, this review covers a wider range of conditions, making it more comprehensive. Similarly, Singh et al. (2024) investigate AI applications in cardiovascular disease risk assessment, emphasizing a personalized approach. However, their study does not generalize well to other diseases. Unlike previous research that focuses on specific health conditions, this study synthesizes findings across various diseases, offering a more holistic understanding of ML applications in HRA.

**Merits and Strengths**

* **Comprehensive Scope -** Covers a wide range of ML techniques and their applications in different health risk assessments.
* **Identification of Research Gaps -** Highlights the need for diverse datasets and improved interpretability.
* **Systematic Methodology -** Follows a structured approach adhering to PRISMA guidelines, ensuring the reliability of findings.
* **Emphasis on Model Interpretability -** Raises awareness of the need for transparent ML models in healthcare.

**Limitations and Challenges**

* **Lack of Experimental Validation -** The study reviews existing literature but does not implement ML models for performance comparison.
* **Over-Reliance on Secondary Data -** Many included studies use publicly available datasets, limiting real-world applicability.
* **Generalization Issues -** The study acknowledges that existing ML models lack validation on diverse populations.

12) **Authors:** Liu, Y., Qin, S., Yepes, A. J., Shao, W., Zhang, Z., & Salim, F. D.

**Year:** 2022

**Article Title:** Integrated convolutional and recurrent neural networks for health risk prediction using patient journey data with many missing values

**Journal:** Journal of Medical Informatics

**Volume and Issue:** 45(3), 112-126

**DOI/URL:** [**https://arxiv.org/pdf/2211.06045**](https://arxiv.org/pdf/2211.06045)

**Summary of the Study**

Liu et al. (2022) propose a novel end-to-end deep learning model that integrates CNNs and RNNs for health risk prediction. Unlike conventional methods that rely on imputing missing EHR data, their model processes missing values without generating synthetic data, thereby preserving clinical accuracy. The CNN component captures short-term temporal patterns within patient records, while the Gated Recurrent Unit (GRU) models long-term dependencies. Experimental results on real-world datasets show superior performance compared to state-of-the-art imputation-based models.

**Key Contributions**

* **Integration of CNN and RNN -** Combines CNNs for short-term dependencies and RNNs for long-term relationships in patient data.
* **Handling Missing Data Without Imputation -** Avoids the bias introduced by imputed values, ensuring data integrity.
* **End-to-End Learning Model -** Provides a robust deep learning architecture that directly processes raw EHR data.
* **Comparison with Existing Models -** Demonstrates higher predictive accuracy than traditional imputation-based methods.

**Comparison with Existing Work**

While previous studies, such as Che et al. (2018), rely on imputation techniques to manage missing EHR data, Liu et al. (2022) eliminate the need for imputed data, reducing classification bias. Similarly, work by Tan et al. (2020) explores attention-based GRUs for health prediction but does not incorporate CNNs to capture local trends in patient journeys. In contrast, Liu et al.'s (2022) integrated CNN-RNN model provides a more comprehensive solution for analyzing patient data with missing values.

**Merits and Strengths**

* **Eliminates Data Imputation Bias -** Maintains clinical integrity by avoiding artificial data generation.
* **Enhanced Temporal Feature Extraction -** CNNs capture short-term patterns, while RNNs handle long-term dependencies.
* **Improved Prediction Accuracy -** Outperforms state-of-the-art models on multiple ICU datasets.

**Limitations and Challenges**

* **Computational Complexity -** Deep learning models require significant processing power and memory.
* **Limited Interpretability -** CNN-RNN architectures lack transparency in decision-making processes.
* **Dependency on High-Quality Data -** Although it avoids imputation, the model still relies on structured EHR inputs.
* **Generalization Concerns -** The approach may not generalize well to healthcare settings with highly heterogeneous data.

**References:**

1. Chen, A., & Chen, D. O. (2022). Simulation of a machine learning-enabled learning health system for risk prediction using synthetic patient data. Scientific Reports, 12(17917). <https://www.nature.com/articles/s41598-022-23011-4>
2. Wang, X., et al. (2019). Prediction of the 1-year risk of incident lung cancer: Prospective study using electronic health records from the State of Maine. Journal of Medical Internet Research, 21(5), e13260. <https://www.jmir.org/2019/5/e13260/>
3. Yeh, M. C. (2021). Artificial intelligence-based prediction of lung cancer risk using nonimaging electronic medical records: Deep learning approach. Journal of Medical Internet Research, 23(8), e26256. <https://www.jmir.org/2021/8/e26256>
4. Shinde, S. A., & Rajeswari, P. R. (2018). Intelligent health risk prediction systems using machine learning: A review. International Journal of Engineering & Technology, 7(3), 1019-1023. <https://www.researchgate.net/publication/326253594_Intelligent_health_risk_prediction_systems_using_machine_learning_A_review>
5. Wu, Y., Burnside, E. S., Fan, J., & Craven, M. (2017). Breast cancer risk prediction using electronic health records. 2017 IEEE International Conference on Healthcare Informatics (ICHI), 224-228. <https://ieeexplore.ieee.org/document/8031151>
6. Abhadiomhen, S. E., Nzeakor, E. O., & Oyibo, K. (2024). Health risk assessment using machine learning: Systematic review. Electronics, 13(4405). <https://www.mdpi.com/2079-9292/13/22/4405>
7. Mishra, A., et al. (2024). Machine learning models for pancreatic cancer risk prediction using electronic health record data: A systematic review and assessment. *The American Journal of Gastroenterology*. <https://pubmed.ncbi.nlm.nih.gov/38752654>
8. Singh, R., et al. (2024). Artificial intelligence for cardiovascular disease risk assessment in personalized framework: A scoping review. *PubMed Central*. <https://pubmed.ncbi.nlm.nih.gov/38846068>
9. Liu, Y., Qin, S., Yepes, A. J., Shao, W., Zhang, Z., & Salim, F. D. (2022). Integrated convolutional and recurrent neural networks for health risk prediction using patient journey data with many missing values. Journal of Medical Informatics, 45(3), 112-126. <https://arxiv.org/pdf/2211.06045>
10. Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. Scientific Reports, 8(1), 1-12. <https://www.nature.com/articles/s41598-018-24271-9>
11. Tan, Q., Ye, M., Yang, B., Liu, S., Ma, A. J., Yip, T. C.-F., Wong, G. L.-H., & Yuen, P. (2020). Data-GRU: Dual-attention time-aware gated recurrent unit for irregular multivariate time series. Proceedings of the AAAI Conference on Artificial Intelligence, 34(1), 930-937. <https://ojs.aaai.org/index.php/AAAI/article/view/5440>