ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.01 No.01 (2023)

Real-Time Survival Risk Prediction with Streaming Big Health Data: A Scalable Architecture

Adaeze Ojinika Ezeogu
Affiliation: Sheffield Hallam University, United Kingdom.
Department: Big Data Analytics
Email: Adaezeojinika@gmail.com
ORCID Number: https://orcid.org/ 0009-0002-7075-4345

Abstract
The healthcare sector faces significant changes as massive data generation grows, which offers new opportunities to improve patient care and resource management. Real-time survival risk prediction models utilize streaming big health data capabilities to enable personalized treatment strategies and proactive interventions that significantly improve patient outcomes. Dynamic assessment and updating patient risk scores when new information emerges proves vital for time-sensitive clinical environments, including emergency departments, intensive care units, and remote patient monitoring programs. Modern healthcare data streams that involve electronic health record updates, laboratory results, vital signs, and sensor data present challenges that traditional batch processing methods cannot adequately address due to their limitations in handling high velocity and volume, as well as diverse data types. Healthcare data streaming challenges require scalable and robust architectures that can manage large volumes of data and provide real-time survival risk predictions. Machine learning applications in healthcare depend on reliable data collection methods, essential for clinicians who require fast and precise information to deliver top-notch patient care. Healthcare settings require careful evaluation of tools, infrastructure, and regulatory standards for machine learning model deployment because behavioral and temporal data shifts add complexity to the process. With the widespread adoption of Electronic Health Records, healthcare providers now routinely collect data needed to develop clinical tools, which require adaptable prediction methods that handle EHR data constraints while updating predictions dynamically and focusing on individual patient clinical contexts.
Keywords: Survival Risk Prediction, Big Health Data, Scalability Challenges

Background
The widespread adoption of Electronic Health Records now supports machine learning applications that improve clinical analysis and sepsis prediction through physiological time-series analysis to advance clinical care standards (Chung et al., 2018). Machine learning enables the processing of extensive data within electronic health records to derive clinically important insights, according to Ghassemi et al. (2018). EHR data tracks nearly every aspect of patient care over time, thus containing an integrated, comprehensive clinical history that can be leveraged to construct risk models that help predict disease progression and reveal disease evolution, enabling accelerated clinical research and population-based predictive analysis (Ye et al., 2018). Electronic health record data analysis through machine learning methods has gained popularity for predicting disease onset and progression patterns (Dworzyński et al., 2020). The utility of machine learning algorithms extends to analyzing multiple data types, which they then use to form predictions about disease risk, diagnosis, and treatment plans (Ngiam & Khor, 2019). Predicting future patient health status enables proactive interventions and improved health outcomes, specifically for high-risk individuals facing chronic diseases or adverse events. Healthcare research frequently employs survival

analysis from statistics to model the periods before specific events, including death and disease progression, occur. While conventional methods require one cohort to derive and validate models, machine learning algorithms need customization per hospital, allowing for rule updates through a general method to be verified with future data from the hospital (Morgan et al., 2019). Machine learning proves its worth in healthcare through its capacity to analyze large data sets and apply generalized models to new cases, which enables the detection of patterns among similar patients that could be overlooked by time-constrained physicians (Liu et al., 2018).

Although traditional survival analysis techniques, including the Cox proportional hazards model and Kaplan-Meier estimator, have gained widespread acceptance, they do not effectively manage the complex nature of streaming big health data (Preetha et al., 2020)

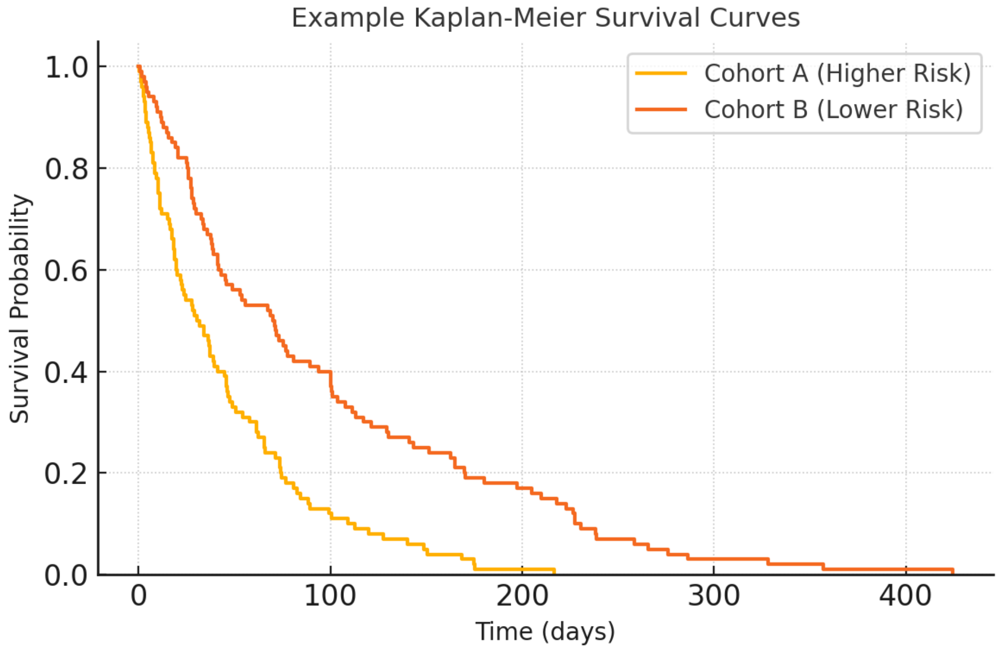


Chart 1 provides an example of ***Kaplan-Meier survival curves*** *for two patient cohorts with different risk profiles.*

The computational expense of these methods stems from their reliance on batch processing, which requires analyzing complete datasets at once, preventing them from delivering real-time predictions. Traditional survival models face difficulties when including time-varying covariates because these changing variables are crucial in determining an individual's survival chances. The development of scalable survival models capable of processing streaming data for dynamic risk prediction has become possible thanks to recent progress in distributed computing and machine learning. Modern machine learning approaches have proven their accuracy for predictive tasks and are becoming more common tools for disease diagnosis and health condition prediction (Xu et al., 2020). The ability of deep learning to derive core features from different data forms has established its position as a leading approach for predicting patient survival outcomes due to its strong representational power (Hagag et al., 2023). Researchers adopted machine learning methodologies for medical research because of their exceptional ability to predict outcomes and their flexibility with complex datasets, providing greater forecasting benefits than traditional statistical methods.

Proposed Architecture for Streaming Survival Analysis

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.01 No.01 (2023)

Our scalable architecture solution for real-time survival risk prediction using streaming big health data utilizes Apache Kafka and Spark Streaming with distributed survival models.
The architecture manages modern health data's velocity, volume, and variety, which allows patient risk scores to be dynamically updated whenever new data emerges. The architecture consists of three main components: data ingestion, processing, and model deployment.

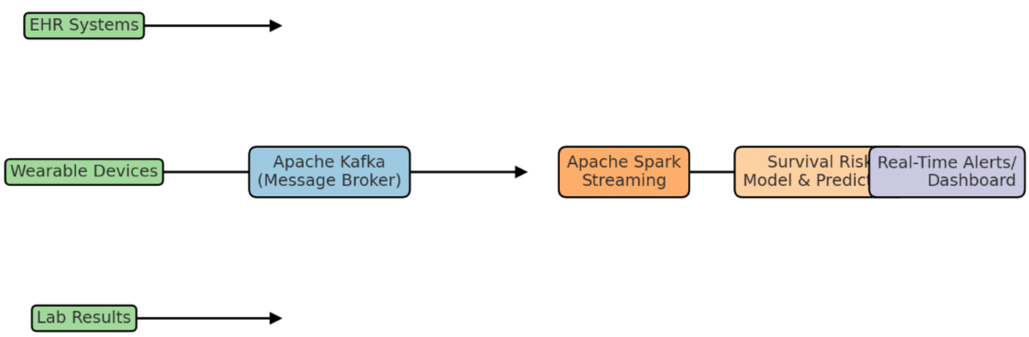


Figure 1 shows a high-level ***streaming architecture*** *for real-time survival risk prediction.*

The data ingestion component gathers and streams data from multiple sources, including EHR systems, laboratory information systems, and wearable devices. Real-time big data analytics depend on data ingestion, which requires acquiring raw data from multiple sources, followed by extracting and loading this data into a processing system.
Apache Kafka is a distributed streaming platform for managing high data volume and velocity. Data transportation between architecture components happens reliably through Kafka's fault-tolerant, scalable, high-throughput messaging system. Through its distributed architecture, Kafka maintains continuous data availability via horizontal scaling for handling increased loads and fault tolerance achieved through replication across multiple brokers.
The component that processes data transforms and analyzes streaming data to uncover relevant features and refine survival models.

Literature Review

New studies highlight the expanding use of machine learning techniques to process electronic health record data, which helps predict when diseases will develop and their progression patterns. Machine learning algorithms can handle different data formats to generate predictive models for disease risk assessment and treatment recommendations (Lee, 2017). Machine learning techniques have gained popularity because they demonstrate enhanced predictive capabilities compared to conventional statistical methods (Artetxe et al., 2018). Predicting a patient's future health status remains vital for launching preventive measures and boosting health results, particularly for those who face a high possibility of chronic illnesses or adverse health events. Survival analysis is a statistical discipline focusing on time-to-event data analysis, which healthcare researchers commonly use to predict the duration before events like death, disease progression, or hospital readmission happen.
The healthcare industry increasingly adopts machine learning because it efficiently processes various data types while improving predictive outcomes. Healthcare practitioners are witnessing a transformation in disease diagnosis, treatment planning, and patient monitoring due to implementing machine learning algorithms. Healthcare uses machine learning because

it can evaluate multiple data forms, such as electronic health records' structured data, clinical notes' unstructured data, and radiology exam imaging data. Machine learning algorithms enable medical professionals to gain valuable insights by identifying patterns within complex datasets, which enhances patient care. Machine learning methods are proving to be increasingly important for achieving various goals, such as risk assessment, according to Perveen et al. (2020).

The application of machine learning in healthcare faces multiple challenges, mainly concerning the quality and availability of data sources (Buabbas et al., 2023). Machine learning algorithms succeed through access to extensive datasets that maintain high quality while accurately representing the targeted patient population. Healthcare settings that deploy machine learning tools must carefully evaluate their impact to ensure these tools improve patient care delivery (Cabitza et al., 2017).

The practical and ethical application of machine learning technologies in healthcare requires collaboration between data scientists, clinicians, and healthcare administrators through a multidisciplinary approach (Woodman & Mangoni, 2023).

Scalability Challenges

Big health data scalability problems require innovative processing and deployment methods for data models.

The processing of raw medical data faces significant challenges due to data sparsity, missing entries, noise, and imbalanced outputs, which demand advanced algorithmic solutions. Traditional computer science concerns like scalability and convergence rate do not usually become significant problems for healthcare applications (Razzaghi et al., 2015). Healthcare workflows must incorporate machine learning technologies to enhance healthcare delivery without complicating the process.

Different data sources and formats call for strong data integration and preprocessing techniques. As data volumes increase, traditional healthcare systems become overloaded, forcing the need for scalable architectures and distributed computing solutions.

Low-latency processing of massive streaming data volumes remains essential for real-time survival risk prediction. Survival models exhibit high computational complexity, which presents difficulties when scaling to real-time processing requirements.

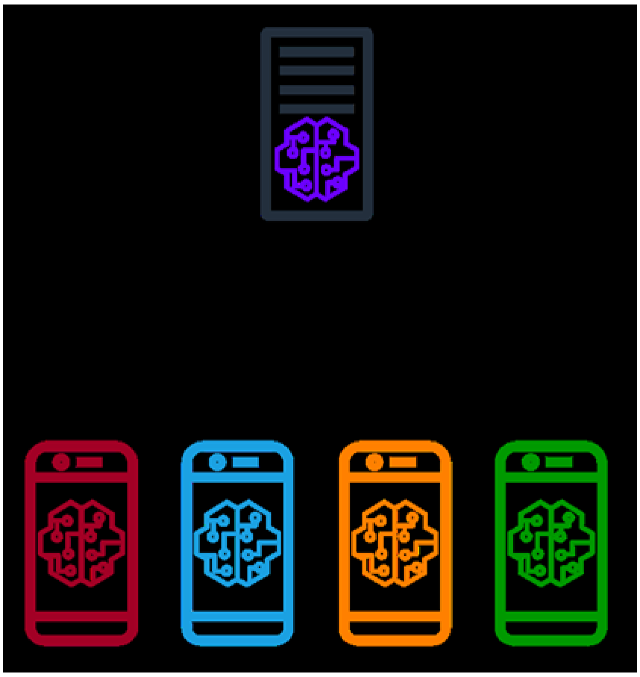
Data Privacy and Security

Data privacy and security protection are essential issues when streaming big health data.

Handling sensitive patient information requires organizations to comply with data privacy regulations, including HIPAA and GDPR (Weng, 2019).

To safeguard patient privacy during data analysis, researchers must utilize data anonymization and de-identification methods (Feldman et al., 2017). Data protection requires secure methods for transmitting and storing data to prevent unauthorized access and breaches. Implementing access controls and authentication mechanisms is necessary to permit data access only for authorized personnel. Model training using federated learning and differential privacy methods allows for machine learning development without violating patient privacy, as shown by Chou et al. in 2018.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.01 No.01 (2023)

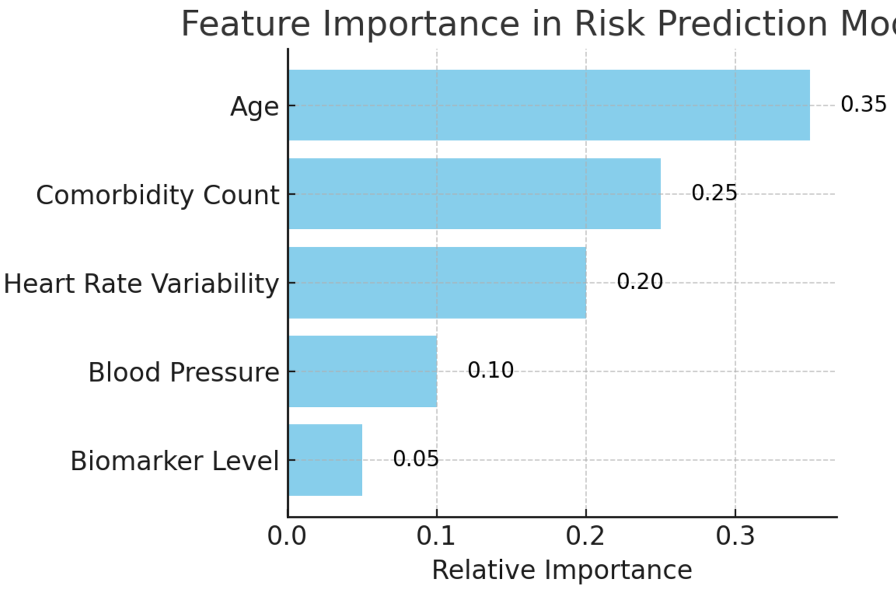


*This schematic illustrates a **federated learning approach** to model training*

Protecting privacy demands balancing data usefulness with privacy safeguards, which mandates a thorough evaluation of the unique application and legal obligations (Price & Cohen, 2018).

Model Interpretability and Explainability

Survival risk prediction models require interpretability and explainability to establish trust and confidence.



*Bar chart 2: Highlights the **relative importance of features** in the survival risk prediction model (for illustration purposes).*

To make informed patient care decisions, clinicians must understand the factors that drive model predictions.

Deep neural networks represent black-box models that present interpretation challenges and thus face limited clinical implementation.

Methods like SHAP values and LIME enable an understanding of feature significance in model predictions.
Interactive tools and visualizations enable clinicians to examine model predictions while gaining insights into their reasoning processes.
When choosing a survival model, you must balance accuracy and ease of understanding. According to Girardi et al. (2021), explainability techniques vary across different AI models and feature subsets.

Architecture for Streaming Survival Analysis

Real-time risk prediction in healthcare depends on a scalable streaming survival analysis architecture.
Multiple essential components make up this architectural design.
The data ingestion layer begins by gathering streaming data from multiple sources, including EHR systems, wearable devices, and lab results.
The message queue system Apache Kafka acts as a buffer for incoming data and manages its dependable delivery to downstream processing elements.
A stream processing engine like Apache Spark Streaming or Apache Flink conducts real-time data transformation alongside feature extraction.
The distributed survival model, including options like the Cox proportional hazards model or random survival forest, computes patient risk scores using extracted features.
The model serving layer enables automatic access to the survival model, which conducts risk prediction in real-time.
The monitoring and alerting system tracks the model's performance and sends alerts for detected anomalies and critical events.
The system architecture requires fault tolerance and high availability features to guarantee uninterrupted service in essential healthcare environments.
Selecting appropriate technologies and components relies on analyzing the healthcare application requirements, which include data volume, velocity, and latency constraints.
In healthcare systems, risk prediction models analyze patient characteristics to forecast the likelihood of a particular outcome within a specific period (Shipe et al., 2019). Survival analysis represents a specialized statistical field often used to model time-to-event data, including the time until death or disease progression. Azuaje (2019) and Katuwal & Chen (2016) show that machine learning applications in survival analysis have grown recently to enhance predictive accuracy and tailor treatment options. Clinical application of these models encounters significant difficulties during real-time deployment, which becomes more intense with the continuous flow of large-scale health data.
Recent computational model developments that allow for extensive patient data capture are driving a swift transformation of healthcare services (Battineni et al., 2020). Analyzing health data in real time requires a scalable system that can effectively manage incoming data's high volume and speed (Bradley, 2013). Prospective studies form the basis for traditional risk scoring models. However, they incur significant expenses and time while potentially introducing bias from differential loss to follow-up and expensive screening methods (Perveen et al., 2020). The static characteristics of these models fail to capture the dynamic alterations in patient health conditions (Nazer et al., 2023). The system design should enable handling of time-varying covariates so the model can continuously adjust according to patient condition fluctuations.
Integrating medical knowledge into clinical machine learning models is a crucial area of study within research.

Methodology

Developing real-time survival risk prediction models from streaming big health data requires multiple essential steps.

Data Preprocessing and Feature Engineering

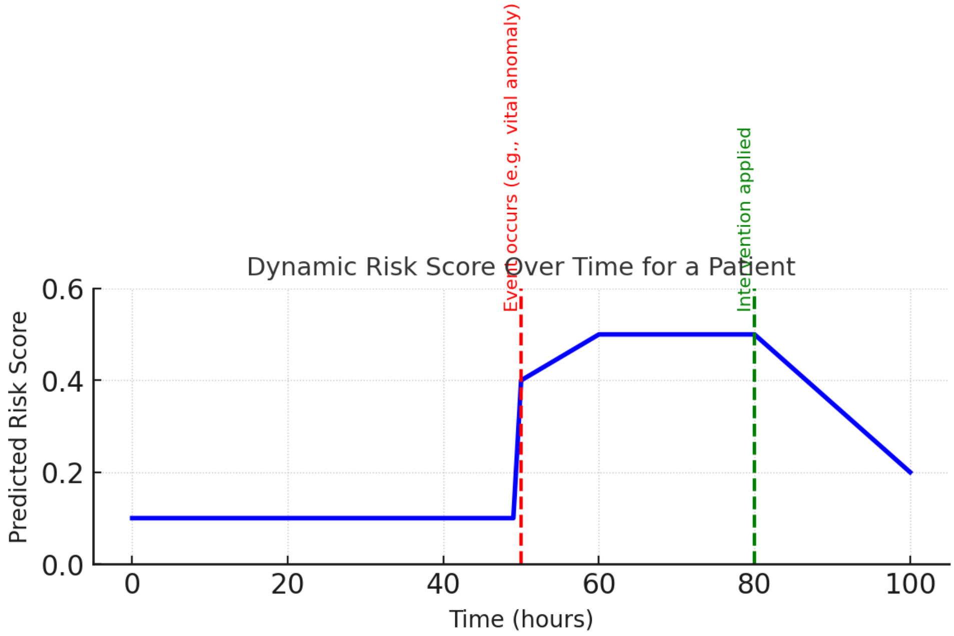
The first stage requires processing the data and constructing relevant features.

Streaming data usually presents challenges because it contains noisy information, incomplete values, and inconsistent data points.

Outlier removal and imputation represent data cleaning techniques used to enhance data quality.

Feature engineering converts raw data into useful features that survival models utilize.

The process includes generating time-varying covariates, which can be rolling averages of vital signs or adjustments in medication dosages.



*The graph demonstrates a **simulated patient's risk score over time**, updating in real-time as new data streams in*

Feature selection methods enable identifying essential features for survival prediction while decreasing model complexity and boosting performance.

Developing efficient techniques for identifying relevant features from large datasets is a fundamental element in achieving accurate and reliable predictive outcomes. The effectiveness of the model and its ability to pinpoint individuals at increased risk depend heavily on which features are selected. The feature engineering and selection process requires statistical analyses and domain expertise to identify meaningful variables from raw data, enhancing the prediction models' precision and robustness.

Model Training and Evaluation

The following stage requires the training and evaluation of the model.

The distributed survival model receives training from historical data through a Cox proportional hazards model or a random survival forest.

Suitable evaluation metrics like the C-index or Brier score measure the model's prediction accuracy.

Cross-validation methods help verify that the model performs effectively on data it has never seen before.

The model undergoes continuous retraining with newly acquired data to maintain its prediction accuracy while adjusting to the evolving patient population.

Model Deployment and Monitoring

The last procedure requires deploying the model and monitoring its operation.

A trained survival model becomes part of a model serving layer that enables real-time model access for risk prediction.

The model undergoes regular performance checks to identify any signs of degradation and data drift.

Notifications activate when the system finds either anomalies or critical events.

Clinical Validation and Refinement

Clinical validation is required in real-world healthcare settings to assess the model's performance. Assessing failure modes and clinical mimics from model predictions enables clinicians to gain insights into the dataset or labels (Lu et al., 2020). The iterative validation and refinement process establishes the model's clinical relevance and reliability (Elias et al., 2022).

Data Streams and Real-Time Processing

The emergence of real-time data streams requires specialized processing methods to handle the data effectively (Sanchez- Pinto et al., 2018).

Continuous information streams deliver vital signs, lab results, and medication updates regularly.

Real-time processing methods, such as stream and complex event processing, allow for immediate analysis of incoming data to support prompt risk predictions.

Before prediction models can use the data, it needs preprocessing, which encompasses cleaning steps and transformation tasks, followed by integration procedures.

Infrastructure for Scalable Processing

Streaming health data requires a scalable infrastructure to manage its high volume and rapid speed.

Organizations use distributed computing frameworks like Apache Kafka and Spark Streaming to enable parallel data processing across multiple nodes.

Despite system failures, these systems deliver fault tolerance capabilities and maintain reliable data processing.

Digital devices have resulted in exponential data growth alongside expanded storage capabilities and increased computational power, which enables complex databases to be processed at unprecedented speeds (Papachristou et al., 2023).

Integration with Clinical Workflows

The model's predictions require integration with clinical workflows to remain actionable and benefit patient care.

Clinicians require the model's output to be delivered in an understandable and direct format, highlighting patient risk factors and possible treatment options.

During the integration process, developers need to account for ethical concerns when implementing predictive healthcare models, including maintaining fairness while ensuring transparency and accountability.

Results

According to Moazemi et al. (2023, the primary input modalities for clinical analysis included time series data and electronic health records. Typically, researchers evaluate real-time survival risk prediction models using streaming big health data based on multiple key performance indicators.

Predictive Accuracy: The most important metric evaluates how precisely the model predicts the risk of patient survival. The C-index evaluates how well the model can distinguish patients with different survival durations, while the Brier score determines how precise the survival probability predictions are (Deznabi & Fiterau, 2023).

Calibration: Calibration describes how closely predicted survival probabilities match observed survival probabilities. When a model is calibrated correctly, its predictions match the actual results.

Timeliness: The predictions must be generated promptly, which is crucial for real-time applications. The model must produce predictions at a speed suitable for clinical decision-making applications.

Scalability: Scalability describes how the model manages large datasets and fast data streams.

Interpretability: Interpretability measures how well the predictions made by a model can be understood and explained.

Ethical and Fairness Considerations

According to Calster et al. (2019), medical predictions must accurately distinguish between patients with disease conditions and those who do not. Accurate risk predictions are a requirement, but algorithm development often faces the problem of overfitting, leading to reduced discrimination and calibration when tested on new data (Calster et al., 2019).

Simulation Setup

Researchers developed a simulated healthcare dataset to test their proposed method's practicality and effectiveness.

The dataset features constantly changing EHR updates, lab results, and vital signs measurements representing real-world clinical data dynamics.

Our simulation contains a patient population with a range of demographic characteristics, diverse as stories, and disease conditions.

Discussion

Predictive models in healthcare serve to optimize the expected benefit they offer according to Pfohl et al. (2022. Real-time survival risk prediction models that analyze streaming big health data can deliver multiple advantages to healthcare delivery systems, according to Jilanee et al. (2021.

Early Intervention: Real-time risk prediction through continuous patient data monitoring enables clinicians to detect high-risk patients and take timely actions to prevent or lessen adverse outcomes (Houfani et al., 2021).

Improved Resource Allocation: These models enable efficient healthcare resource distribution by focusing medical interventions on patients who benefit most.

Personalized Treatment: Healthcare professionals can customize treatment approaches using models to match interventions with each patient's risk profile.

Data heterogeneity, time-varying covariates, censoring, and non-linear interactions impact the accuracy level of these models.

While AI implementation in healthcare presents several public health benefits, it also raises ethical and legal issues, including data privacy and surveillance, safety concerns, transparency, fairness, algorithmic biases, and the philosophical question about human judgment, which need to be addressed when incorporating this technology into healthcare systems (Olawade et al., 2023).

Healthcare is adopting machine learning technology because it works faster and makes more precise predictions.

Training AI algorithms with extensive data sets enables them to find patterns and create predictions, which allows healthcare professionals to analyze patient information to predict health issues and develop customized care strategies.

Implementing AI technology in healthcare encounters significant challenges related to data protection and algorithmic fairness while building trust in AI-generated recommendations.

Challenges and Future Directions

Healthcare AI adoption encounters multiple obstacles, including poor data quality, opaque model decision-making, and complex regulatory requirements.

Data Quality: AI models' reliability and accuracy rely on the quality of the training data.

Interpretability: Deep learning models' opaque nature means users cannot easily understand their decision-making processesusers cannot easily understand their decision-making processes.

Regulatory Hurdles: Healthcare AI applications must comply with various regulations governing data security and privacy protection.

The next phase of research needs to enhance data quality, create more transparent AI models, and overcome regulatory barriers to help AI become mainstream in healthcare.

The broader adoption of healthcare AI systems depends on addressing ethical concerns, data limitations, and model trustworthiness. Future research needs to expand data diversity and build standardized ethical guidelines and transparent models.

AI implementation in healthcare faces ethical dilemmas and data protection issues, necessitating healthcare professionals to apply medical ethics principles of autonomy, beneficence, nonmaleficence, and justice when merging AI technologies with healthcare systems (Farhud & Zokaei, 2021). Healthcare AI development and application must address ethical and regulatory challenges for practical and responsible implementation. The development of AI technologies in healthcare demands strong regulations to maintain their safety and effectiveness while ensuring ethical operations.

Conclusion

The emergence of real-time survival risk prediction models utilizing streaming big health data marks a significant breakthrough in healthcare analytics.

The architecture uses Apache Kafka and Apache Spark Streaming alongside distributed survival models to update patient risk scores dynamically when new data becomes available.

AI's potential to overcome healthcare challenges positions it as a central decision-maker for future healthcare system operations (Malik et al., 2020).

This study demonstrates how the methodology discussed could enhance intelligent alerting systems, as well as ICU triage and remote patient monitoring solutions.

Contributions

The primary contributions of this article include:

A scalable architecture for streaming survival analysis: The architecture processes the high speed and diverse nature of contemporary health information to facilitate instantaneous risk predictions.
A simulated healthcare dataset: Researchers developed the dataset using updates from EHRs, lab results, and dynamic vital signs to simulate a realistic environment for model testing.
This is a demonstration of the feasibility of real-time survival risk prediction. The results show how this model could improve healthcare decisions.
Broader adoption requires addressing ethical concerns and data limitations alongside model trustworthiness. At the same time, future work should aim to improve data diversity, create standardized ethical guidelines, and enhance model transparency for equitable and effective healthcare applications, according to Kuwaiti et al. (2023. Through vast data repositories to train AI algorithms for pattern identification and prediction creation healthcare providers can now evaluate extensive patient datasets to anticipate health issues and develop personalized treatment plans (Lisacek-Kiosoglous et al., 2023). Predictive modeling through AI enables individualized patient treatment with the creation of specialized treatment strategies.
AI systems generate specific therapy suggestions based on genetic information alongside demographic and lifestyle factors, proving most beneficial in oncology practices. Through AI-based technologies, healthcare delivery can be revolutionized as they enhance diagnostics while personalizing treatments and optimizing administrative tasks, leading to better patient outcomes and increased healthcare efficiency (Serag et al., 2019). AI adoption in healthcare faces significant obstacles such as maintaining data privacy standards while addressing algorithmic bias and building trust in AI-generated recommendations (Kuwaiti et al., 2023). Organizational inertia must be overcome, and trust established to achieve successful and safe AI integration.
Implementing AI in healthcare systems enables a transformation to fairer cancer treatment by supporting healthcare delivery in under-resourced areas. Through improved detection methods and treatment strategies, AI technologies can improve cancer care outcomes while minimizing access disparities (Patel et al., 2020). The successful implementation of AI technologies in clinical settings requires additional clinical trials and structured regulatory systems, creating a path for innovative AI applications in healthcare.
Integrating AI into clinical workflows requires responsible practices that address data quality and interpretability challenges along with regulatory compliance needs (Kelly et al., 2019). AI promises beneficial outcomes but requires resolution of several hurdles, including algorithmic bias and inadequate model robustness, before it can be applied effectively in clinical environments (Markus et al., 2020). AI functions as a supportive tool that amplifies human skills rather than substituting them and requires careful implementation to assist healthcare professionals while advancing patient care. AI will play a significant role in shaping healthcare's future, yet demands collaborative efforts from developers, clinicians, and regulators to ensure its implementation remains practical and ethical.

References

Abroshan, M., Giles, O., Greenbury, S. F., Roberts, J. C., Schaar, M. van der, Steyn, J. S., Wilson, A., & Yong, M. (2023). Safe AI for health and beyond -- Monitoring to transform a health service. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2303.01513

Artetxe, A., Beristain, A., & Graña, M. (2018). Predictive models for hospital readmission risk: A systematic review of methods [Review of Predictive models for hospital readmission risk: A systematic review of methods]. Computer Methods and Programs in Biomedicine, 164, 49. Elsevier BV. https://doi.org/10.1016/j.cmpb.2018.06.006

Asri, H., Mousannif, H., Moatassime, H. A., & Zahir, J. (2020). Big Data and Reality Mining in Healthcare: Promise and Potential. In Lecture Notes in Computer Science (p. 122). Springer Science+Business Media. https://doi.org/10.1007/978-3-030-51935-3_13

Azuaje, F. (2019). Artificial intelligence for precision oncology: beyond patient stratification [Review of Artificial intelligence for precision oncology: beyond patient stratification]. Npj Precision Oncology, 3(1). Nature Portfolio. https://doi.org/10.1038/s41698-019-0078-1

Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis [Review of Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis]. Journal of Personalized Medicine, 10(2), 21. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/jpm10020021

Bradley, P. S. (2013). Implications of Big Data Analytics on Population Health Management. Big Data, 1(3), 152. https://doi.org/10.1089/big.2013.0019

Buabbas, A. J., Miskin, B., Alnaqi, A. A., Ayed, A. K., Shehab, A. A., Syed-Abdul, S., & Uddin, M. (2023). Investigating Students' Perceptions towards Artificial Intelligence in Medical Education. Healthcare, 11(9), 1298. https://doi.org/10.3390/healthcare11091298

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended Consequences of Machine Learning in Medicine. JAMA, 318(6), 517. https://doi.org/10.1001/jama.2017.7797

Calster, B. V., Wynants, L., Timmerman, D., Steyerberg, E. W., & Collins, G. S. (2019). Predictive analytics in health care: How can we know it works? Journal of the American Medical Informatics Association, 26(12), 1651. https://doi.org/10.1093/jamia/ocz130

Calvaresi, D., & Calbimonte, J.-P. (2020). Real-Time Compliant Stream Processing Agents for Physical Rehabilitation. Sensors, 20(3), 746. https://doi.org/10.3390/s20030746

Chou, E., Nguyen, T., Beal, J., Haque, A., & Fei-Fei, L. (2018). A Fully Private Pipeline for Deep Learning on Electronic Health Records. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1811.09951

Chung, I., Kim, S., Lee, J., Kim, K. J., Hwang, S. J., & Yang, E. (2018). Deep Mixed Effect Model using Gaussian Processes: A Personalized and Reliable Prediction for Healthcare. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1806.01551

Deznabi, I., & Fiterau, M. (2023). MultiWave: Multiresolution Deep Architectures through Wavelet Decomposition for Multivariate Time Series Prediction. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2306.10164

Dworzyński, P., Aasbrenn, M., Rostgaard, K., Melbye, M., Gerds, T. A., Hjalgrim, H., & Pers, T. H. (2020). Nationwide prediction of type 2 diabetes comorbidities. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-58601-7

Elias, P., Poterucha, T. J., Rajaram, V., Moller, L. M., Rodriguez, V., Bhave, S., Hahn, R. T., Tison, G. H., Abreau, S., Barrios, J. P., Torres, J. N., Hughes, J. W., Pérez, M., Finer, J., Kodali, S., Khalique, O., Hamid, N., Schwartz, A., Homma, S., … Perotte, A. (2022). Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease. Journal of the American College of Cardiology, 80(6), 613. https://doi.org/10.1016/j.jacc.2022.05.029

Farhud, D. D., & Zokaei, S. (2021). Ethical Issues of Artificial Intelligence in Medicine and Healthcare. In the Iranian Journal of Public Health. Knowledge E. https://doi.org/10.18502/ijph.v50i11.7600

Feldman, K., Faust, L., Wu, X., Huang, C., & Chawla, N. V. (2017). Beyond Volume: The Impact of Complex Healthcare Data on the Machine Learning Pipeline. In Lecture

Notes in Computer Science (p. 150). Springer Science+Business Media. https://doi.org/10.1007/978-3-319-69775-8_9

Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., & Ranganath, R. (2018). Opportunities in Machine Learning for Healthcare. arXiv (Cornell University). https://arxiv.org/pdf/1806.00388

Girardi, I., Vagenas, P., Arcos-Díaz, D., Bessaï, L., Büsser, A., Furlan, L., Furlan, R., Gatti, M., Giovannini, A., Hoeven, E., & Marchiori, C. (2021). On the explainability of hospitalization prediction on a large COVID-19 patient dataset. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2110.15002

Hagag, A., Gomaa, A. M., Kornek, D., Maier, A., Fietkau, R., Bert, C., Putz, F., & Huang, Y. (2023). Deep Learning for Cancer Prognosis Prediction Using Portrait Photos by StyleGAN Embedding. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2306.14596

Houfani, D., Slatnia, S., Kazar, O., Saouli, H., & Merizig, A. (2021). Artificial intelligence in healthcare: a review on predicting clinical needs [Review of Artificial intelligence in healthcare: a review on predicting clinical needs]. International Journal of Healthcare Management, 15(3), 267. Taylor & Francis. https://doi.org/10.1080/20479700.2021.1886478

Juba, O. O., Olumide, A. O., Ochieng, J. O., & Aburo, N. A. (2022). Evaluating the impact of public policy on the adoption and effectiveness of community-based care for aged adults. *International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence*, *13*(1), 65-102

Jilanee, D., Khan, S. A., Shah, S. M. H., Capriles, N. M. A., Capriles, C. A. A., Tahir, H., Gul, A., Ashraf, S. U., Tousif, S., & Jiwani, A. (2021). Comparison of the Performance of Various Scores in Predicting Mortality Among Patients Hospitalized With COVID-19. Cureus. https://doi.org/10.7759/cureus.20751

Juba, O. O., Lawal, O., David, J. I., & Olumide, B. F. (2023). Developing and assessing care strategies for dementia patients during unsupervised periods: Balancing safety with independence. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(04), 322-349.

Katuwal, G. J., & Chen, R. F. (2016). Machine Learning Model Interpretability for Precision Medicine. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1610.09045

Kelly, C., Karthikesalingam, A., Suleyman, M., Corrado, G. S., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17(1). https://doi.org/10.1186/s12916-019-1426-2

Kuwaiti, A. A., Nazer, K., Alreedy, A. H., AlShehri, S. D., Almuhanna, A., Subbarayalu, A. V., Muhanna, D. A., & Al- Muhanna, F. (2023). A Review of the Role of Artificial Intelligence in Healthcare [Review of A Review of the Role of Artificial Intelligence in Healthcare]. Journal of Personalized Medicine, 13(6), 951. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/jpm13060951

Lee, J. (2017). Patient-Specific Predictive Modeling Using Random Forests: An Observational Study for the Critically Ill. JMIR Medical Informatics, 5(1). https://doi.org/10.2196/medinform.6690

Lisacek-Kiosoglous, A., Powling, A. S., Fontalis, A., Gabr, A., Mazomenos, E. B., & Haddad, F. S. (2023). Artificial intelligence in orthopaedic surgery. Bone and Joint Research, 12(7), 447. https://doi.org/10.1302/2046-3758.127.bjr-2023-0111.r1

Liu, J., Zhang, Z., & Razavian, N. (2018). Deep EHR: Chronic Disease Prediction Using Medical Notes. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1808.04928

Lu, C., Strout, J., Gauriau, R., Wright, B. A., Marcruz, F. B. D. C., Buch, V., & Andriole, K. P. (2020). An Overview and Case Study of the Clinical AI Model Development Life Cycle for Healthcare Systems. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2003.07678

Malik, Y. S., Sircar, S., Bhat, S., Ansari, M. I., Pande, T., Kumar, P., Mathapati, B., Ganesh, B., Kaushik, R., Natesan, S., Ezzikouri, S., Zowalaty, M. E. E., & Dhama, K. (2020). How artificial intelligence may help the Covid‐ 19 pandemic: Pitfalls and lessons for the future [Review of How artificial intelligence may help the Covid‐ 19 pandemic: Pitfalls and lessons for the future]. Reviews in Medical Virology, 31(5), 1. Wiley. https://doi.org/10.1002/rmv.2205

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2020). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies [Review of The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies]. Journal of Biomedical Informatics, 113, 103655. Elsevier BV. https://doi.org/10.1016/j.jbi.2020.103655

Moazemi, S., Vahdati, S., Li, J., Kalkhoff, S., Castano, L. J. V., Dewitz, B., Bibo, R., Sabouniaghdam, P., Tootooni, M. S., Bundschuh, R. A., Lichtenberg, A., Aubin, H., & Schmid, F. (2023). Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review [Review of Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: A systematic review]. Frontiers in Medicine, 10. Frontiers Media. https://doi.org/10.3389/fmed.2023.1109411

Morgan, D. J., Bame, B., Zimand, P., Dooley, P. M., Thom, K. A., Harris, A. D., Bentzen, S. M., Ettinger, W. H., Garrett-Ray, S., Tracy, J. K., & Liang, Y. (2019). Assessment of Machine Learning vs Standard Prediction Rules for Predicting Hospital Readmissions. JAMA Network Open, 2(3). https://doi.org/10.1001/jamanetworkopen.2019.0348

Nazer, L., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation [Review of Bias in artificial intelligence algorithms and recommendations for mitigation]. PLOS Digital Health, 2(6). Public Library of Science. https://doi.org/10.1371/journal.pdig.0000278

Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for healthcare delivery [Review of Big data and machine learning algorithms for healthcare delivery]. The Lancet Oncology, 20(5). Elsevier BV. https://doi.org/10.1016/s1470-2045(19)30149-4

Olawade, D. B., Wada, O. J., David-Olawade, A. C., Kunonga, E., Abaire, O. J., & Ling, J. (2023). Using artificial intelligence to improve public health: a narrative review [Review of Using artificial intelligence to improve public health: a narrative review]. Frontiers in Public Health, 11. Frontiers Media. https://doi.org/10.3389/fpubh.2023.1196397

Papachristou, N., Kotronoulas, G., Δικαίος, N., Allison, S. J., Eleftherochorinou, H., Rai, T., Kunz, H. R., Barnaghi, P., Miaskowski, C., & Bamidis, P. D. (2023). Digital Transformation of Cancer Care in the Era of Big Data, Artificial Intelligence and Data-Driven Interventions: Navigating the Field. Seminars in Oncology Nursing, 39(3), 151433. https://doi.org/10.1016/j.soncn.2023.151433

Patel, S. K., George, B., & Rai, V. (2020). Artificial Intelligence to Decode Cancer Mechanism: Beyond Patient Stratification for Precision Oncology [Review of Artificial Intelligence to Decode Cancer Mechanism: Beyond Patient Stratification for Precision Oncology]. Frontiers in Pharmacology, 11. Frontiers Media. https://doi.org/10.3389/fphar.2020.01177

Perveen, S., Shahbaz, M., Ansari, M. S., Keshavjee, K., & Guergachi, A. (2020). A Hybrid Approach for Modeling Type 2 Diabetes Mellitus Progression. Frontiers in Genetics, 10. https://doi.org/10.3389/fgene.2019.01076

Pfohl, S., Xu, Y., Foryciarz, A., Ignatiadis, N., Genkins, J. Z., & Shah, N. H. (2022). Net benefit, calibration, threshold selection, and training objectives for algorithmic fairness in healthcare. 2022 ACM Conference on Fairness, Accountability, and Transparency, 1039. https://doi.org/10.1145/3531146.3533166

Preetha, S., Manohar, A., M, A. A. H., & Tarikere, N. Y. (2020). A REVIEW OF MACHINE LEARNING ALGORITHMS IN HEALTHCARE [Review of A REVIEW OF MACHINE LEARNING ALGORITHMS IN HEALTHCARE]. International Journal of Recent Trends in Engineering and Research, 6(5), 44. https://doi.org/10.23883/ijrter.2020.6030.nwzoa

Price, W. N., & Cohen, I. G. (2018). Privacy in the age of medical big data [Review of Privacy in the age of medical big data]. Nature Medicine, 25(1), 37. Nature Portfolio. https://doi.org/10.1038/s41591-018-0272-7

Razzaghi, T., Roderick, O., Safro, I., & Marko, N. (2015). Fast Imbalanced Classification of Healthcare Data with Missing Values. arXiv (Cornell University). https://arxiv.org/pdf/1503.06250

Sanchez- Pinto, L. N., Venable, L. R., Fahrenbach, J., & Churpek, M. M. (2018). Comparison of variable selection methods for clinical predictive modeling. International Journal of Medical Informatics, 116, 10. https://doi.org/10.1016/j.ijmedinf.2018.05.006

Serag, A., Ion- Mărgineanu, A., Qureshi, H., McMillan, R. B., Martin, M.-J. S., Diamond, J., O'Reilly, P. G., & Hamilton, P. (2019). Translational AI and Deep Learning in Diagnostic Pathology [Review of Translational AI and Deep Learning in Diagnostic Pathology]. Frontiers in Medicine, 6. Frontiers Media. https://doi.org/10.3389/fmed.2019.00185

Shipe, M. E., Deppen, S. A., Farjah, F., & Grogan, E. L. (2019). Developing prediction models using logistic regression: an overview [Review of Developing prediction models for clinical use using an overview]. Journal of Thoracic Disease, 11. AME Publishing Company. https://doi.org/10.21037/jtd.2019.01.25

Supriya, M., & Deepa, A. J. (2017). A Survey on Prediction Using Big Data Analytics. International Journal of Big Data and Analytics in Healthcare, 2(1), 1. https://doi.org/10.4018/ijbdah.2017010101

Tidke, B., Mehta, R. G., & Dhanani, J. (2018). Real-Time Bigdata Analytics: A Stream Data Mining Approach. In Advances in intelligent systems and computing (p. 345). Springer Nature. https://doi.org/10.1007/978-981-10-8636-6_36

Weng, W. (2019). Machine Learning for Clinical Predictive Analytics. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1909.09246

Woodman, R., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future [Review of A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future]. Aging Clinical and Experimental Research,

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

35(11), 2363. Springer Science+Business Media. https://doi.org/10.1007/s40520-023-02552-2

Xu, Y., Ju, L.-S., Tong, J., Zhou, C., & Yang, J. (2020). Machine Learning Algorithms for Predicting the Recurrence of Stage IV Colorectal Cancer After Tumor Resection. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-59115-y

Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., Xia, M., Liu, M., Zhou, X., Wu, Q., Guo, Y., Zhu, C., Li, Y., Culver, D. S., Alfreds, S. T., Stearns, F., Sylvester, K. G., Widen, E., McElhinney, D. B., & Ling, X. B. (2018). Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. Journal of Medical Internet Research, 20(1). https://doi.org/10.2196/jmir.9268