International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V1I2P105
Eureka Vision Publication | Volume 1, Issue 2, 37-46, 2020

Original Article

ML-Based Risk Stratification of Patients Using Real-Time Clinical Streams on Cloud

Amit Taneja
Lead Data Engineer at Mitchell Martin, USA.

Abstract - Current health care systems are struggling with many issues to provide temporal and customized services to sick patients. The inability to stratify real-time and precise patient-level risk is one of the burning issues regarding the rapid expansion of the volume of clinical information, which includes the use of wearable devices, bedside monitors, and Electronic Health Records (EHRs). Cloud computing solutions can be readily integrated into existing Machine Learning (ML) algorithms to provide promising risk stratification solutions, enabling scalable and real-time analytics. In this paper, an entire framework is provided to support real-time risk stratification to fulfil the original quest of using ML on real-time clinical data streams to stratify yet unidentified patients using a cloud framework. The suggested method can absorb heterogeneous health data, preprocess it, select meaningful features, and then utilise predictive models to assess risks in real-time. By the use of Apache Kafka, Spark Streaming, and ML libraries, like TensorFlow and Scikit-learn, the system is scalable and has a low-latency processing rate. In the methodology section, the exact procedure of data collection, data preprocessing, feature engineering, and model training in ICU conditions is described. The models, such as Random Forests, Gradient Boosting Machines (GBM) and Long Short-Term Memory (LSTM) networks, were trained on publicly available datasets, namely MIMIC-III, and validated based on accuracy, precision, recall, and F1-score. In comparison with others, LSTM models are the most accurate because they are temporally sensitive to a patient's vital signs. The findings underscore a significant increase in the early identification of patient deterioration, providing healthcare workers with a real-time decision-making system. Future implications are described by model explainability, expanding patient privacy, and connecting with the hospital information system.

Keywords - Machine Learning, Risk Stratification, Real-Time Clinical Streams, Predictive Modeling, MIMIC-III, LSTM, EHR.

1. Introduction

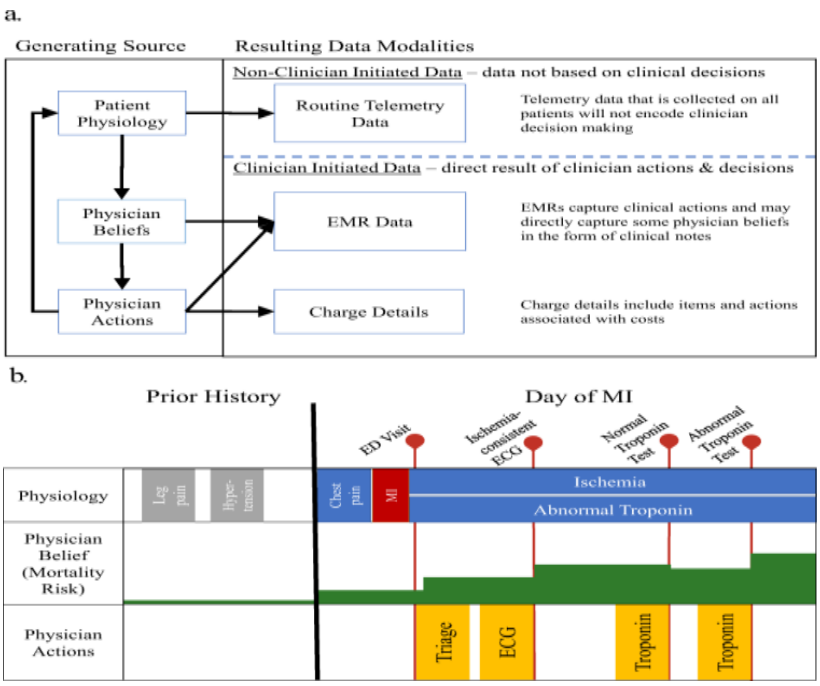

Figure 1. Clinician vs. Non-Clinician Initiated Data Modalities and Temporal Alignment During Myocardial Infarction (MI)

The number and types of clinical data have increased dramatically because of the digitalization of healthcare. Real-time health data are now monitored and produced in large volumes 24 hours a day, as Electronic Health Records (EHRs), wearable health technologies, bedside monitors, and smart medical devices provide an ongoing real-time stream of both structured and unstructured data. [1-3] This information explosion is an opportunity of paramount importance to improve patient care (especially in high-acuity settings like Intensive Care Units (ICUs), since timely decisions may save lives. Constant monitoring systems record vital signs, laboratory results, medication administration, and other key data, resulting in a rich yet complex flow of information. Nevertheless, it is a significant issue to create an efficient analysis of this data to assist in identifying patient worsening scores. More traditional rule-based monitoring methods, based on preselected heights or heuristics based on clinical knowledge, tend to fail to be sensitive and flexible enough to capture nuance, nonlinear trends, and be mindful of a critical event. These systems are typically reactive rather than predictive, often giving false alarms or overlooking early warning signs. Consequently, more and more people are becoming interested in using complex Machine Learning (ML) approaches, which can autonomously learn as clinicians input data and adjust to the dynamic clinical environments, and to detect patterns that might not be obvious to the human eye. The incorporation of these smart systems into clinical practices has the potential to perfect risk stratification, maximizing patient outcomes, and resource optimisation within the critical care units. Nevertheless, in order to make this potential a reality, there is a need to develop architectures capable not only of processing and analyze such a large amount of data in real time but also enabling scalability of such architectures, interpretability and adherence to healthcare standards.

1.1. Importance of Risk Stratification

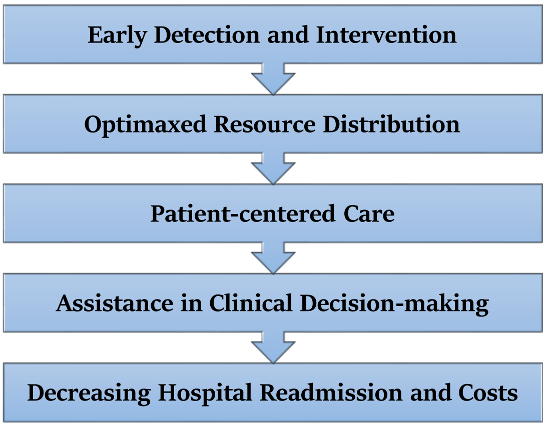

Figure 2. Importance of Risk Stratification

Stratification of risk is a fundamental process in modern healthcare, particularly in critical care settings, where a prompt response can mean the difference between life and death for a patient. It is a process of stratifying patients according to their predisposition to unfavourable outcomes, which facilitates both selective and prioritised actions in terms of clinical responses. The significance of the proper risk stratification is even multidimensional in patient care and hospital operations.

- **Early Detection and Intervention:** Among the most valuable advantages of risk stratification, one can distinguish the fact that it helps to prepare for early warning of patient deterioration. Connected to continuous sampling of vital signs, lab results, and clinical histories, predictive models will pick up on otherwise minor changes that can be predictors of future complications. With early warnings, clinicians can implement early interventions by tailoring medications, ordering tests, or intensifying care before a patient develops a potentially life-threatening condition. The proactive strategy manages to not only increase the survival rates but also decrease the intensity and length of hospitalizations.
- **Optimaxed Resource Distribution:** Risk stratification in resource-intensive settings helps allocate clinical resources more effectively, even with limited facilities in a given setting, such as an ICU. More frequent assessments and interventions may be provided to high-risk patients, whereas lower-risk individuals may not require as many assessments. This variant of targeted usage can be applied to optimize the workload of healthcare providers, focus on the areas where particular attention is required, and make the hospital more efficient.
- **Patient-centred Care:** Risk stratification also promotes the shift to personal medicine. Machine learning models will be able to customize risk assessment and treatment recommendations based on a specific physiological and clinical profile of a given patient. This patient-specific knowledge can be used to plan more personalized care with better results and patient satisfaction.
- **Assistance in Clinical Decision-making:** Physicians have to make decisions based on incomplete or quickly shifting information in high-paced clinical environments. An effective risk stratification system is a valuable aid to decision-making, providing evidence-based risk scores and predicted risks that enhance clinical judgment and informed decision-making. These tools, when incorporated into real-time dashboards, assist clinicians in making on-the-fly but informed judgments.
- **Decreasing Hospital Readmissions and Costs: Proper prediction of risk can also be important in curbing the cost of unnecessary readmissions,** as well as average healthcare costs. Through the identification of patients at risk

of relapse or complications before discharge or in the hospital setting, care teams can introduce specific interventions (e.g. follow-up plans or specialized care pathways) and minimize the risk of relapse or the occurrence of complications. To conclude, risk stratification is a principle of smart, data-driven healthcare. It also enables clinicians to provide timely, personalized, and resource-efficient care, ultimately helping make patient outcomes better and assisting in achieving broader goals of the modern health system.

1.2. Role of AI in Healthcare

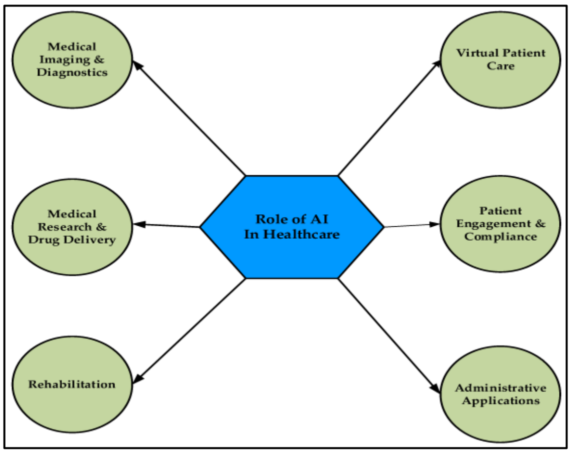

Figure 3. Role of AI in Healthcare

Machine Learning (ML) can be recognized as a game changer in the field of healthcare, offering the possibility to process huge amounts of big and complicated data and derive key insights that cannot be achieved with conventional statistical tools. [4,5] ML offers computationally efficient methods to identify previously unknown trends, make clinically based predictions, and enable data-driven decision-making, particularly with the exponential increase in Electronic Health Records (EHRs), imaging systems, genomics, and real-time patient monitoring. The primary strength of ML is its ability to learn without being explicitly programmed to handle each specific situation, which enables algorithms to adapt to the complex and, in many cases, nonlinear patterns of relationships that occur in clinical data. Numerous ML algorithms are successfully used in different areas of healthcare. Decision Trees are rule-based and have an intuitive pattern of classification, often used as a triage or diagnostic decision support procedure. Support Vector Machines (SVM) are useful with regard to high-dimensional data and have been used in cancer detection, image analysis and disease classifications. Neural networks, specifically deep learning models, have shown outstanding results in tasks such as radiological image interpretation, ECG signal interpretation, and natural language processing of clinical notes. Ensemble Learning techniques, including Random Forests and Gradient Boosting Machines, enhance the predictive performance of a set of models by reducing variance and bias. This has proved extremely useful in critical care units where patients require constant monitoring as they are continuously fed with physiological data that is analyzed in real-time to predict imminent complications. Predictive models can also anticipate the complexity of sepsis, cardiac arrest, and respiratory arrest, allowing them to be addressed effectively beforehand. In addition, ML also helps in the prediction of an outcome, where the clinician may predict the survival of the patient, prolonged in-hospital stay, or the likelihood of readmission. With the paradigm of healthcare changing toward personalized and precision medicine, ML is needed to individualize the ways of treatment according to the profile of the patient. Although issues like data quality, interpretability, and compliance are still present, the increased use of ML in healthcare results in increased effectiveness, precision, and patient outcomes.

2. Literature Survey

2.1. Traditional Risk Stratification Techniques

Standardised methods of risk stratification largely include the use of standardised scoring schemes within critical care, such as the Acute Physiology and Chronic Health Evaluation II (APACHE II), Sequential Organ Failure Assessment (SOFA), and Simplified Acute Physiologic Score (SAPS). [6-9] Those models are founded on the static samples of the health condition of the patient by means of vital signs, lab results, and doctor conclusions taken at a particular time. Although these tools help provide prognostic data, they have a weakness in that they are static and thus not responsive to the real-time changes in the patient's condition. Furthermore, they are rule-driven and not flexible enough to accommodate specific patient progression or changing clinical scenarios. They thus are not well-suited to a dynamic, high-stakes setting such as the Intensive Care Unit (ICU).

2.2. Historical Applications of Machine Learning

The early application of Machine Learning (ML) in healthcare primarily involved traditional approaches to ML, including decision trees and logistic regression. As another illustration, a recent study conducted by Johnson et al. (2016) used logistic regression over ICU datasets to predict mortality and demonstrated its moderate success relative to properly structured, tabular data. However, such a method did not account for the fact that clinical data streams can have a temporal relationship, e.g., the development of vital signs over time or the changing treatment interventions. Being aware of this limitation, examined the possible application of time-series models on Electronic Health Records (EHRs) and proposed the possibility of Recurrent Neural Networks (RNNs) to identify any longitudinal trends and patterns in patients' data. It was a significant milestone on the way toward more advanced and time-sensitive predictive analytics in healthcare.

2.3. Real-Time Health Monitoring Frameworks

Various projects have been pursued to come up with a continuous health monitoring system that is capable of monitoring and comparing the data of a patient in real-time. Striking examples are PhysioNet and OpenICE, which provide open-source systems for gathering and communicating physiological signals. The frameworks, however, are especially concentrated on data collection and signal processing, not flowing easily with machine learning-based analytics, particularly those that can operate at scale and in real-time. Further experiments, including those by Zhang et al. (2019), have introduced distributed stream processing platforms to detect anomalies in data streams arising in the ICU, such as Apache Spark. These systems were more responsive and scalable, yet they continued to deal with the problem of integration with cloud infrastructure and advanced ML models in order to implement thorough risk evaluation.

2.4. Importance of Cloud in Processing of Health Data

The advent of cloud computing has consequently changed the outlook of data storage and processing in the world of health in a profound way. Cloud-based systems are paramount when dealing with large quantities of sensitive health data, where adherence to regulatory data privacy compliance requirements, including the Health Insurance Portability and Accountability Act (HIPAA), is important. Within recent years, cloud vendors like Google Cloud Platform (GCP) and Microsoft Azure have added machine learning practices to their cloud computing portfolios, meaning scalable and real-time analysis via frameworks such as TensorFlow and Azure Machine Learning. The platforms have the computing resources and infrastructure in place that can facilitate the use of complex predictive models while ensuring security and regulatory compliance, and thus are becoming quite appealing in healthcare.

2.5. Vacuum within the Previous Research

Even with improvements in risk stratification, machine learning, real-time monitoring, and cloud computing, a major gap has been identified in the failure to integrate all these technologies holistically. There is a relative scarcity of research demonstrating the efficacy of streamed data analysis in combination with large-scale, cloud-based processing and training of deep learning algorithms, such as the Long Short-term Memory (LSTM), to generate an orchestrated system of patient risk stratification that can be deployed. This poor integration has prevented effective formulation of highly usable adaptive tools that would give a dogged exactitude of personalised risk evaluations in a clinical set-up. Such a gap will be filled through a multidisciplinary approach, where knowledge from the fields of healthcare, data science, software engineering, and cloud architecture can be unified.

3. Methodology

3.1. System Architecture

The architecture of the proposed real-time system of patient stratification by risks is divided into four main levels, each of which implements a particular set of functions [10-12], which represents an efficient answer to the scales of data processing, its visualization and processing.

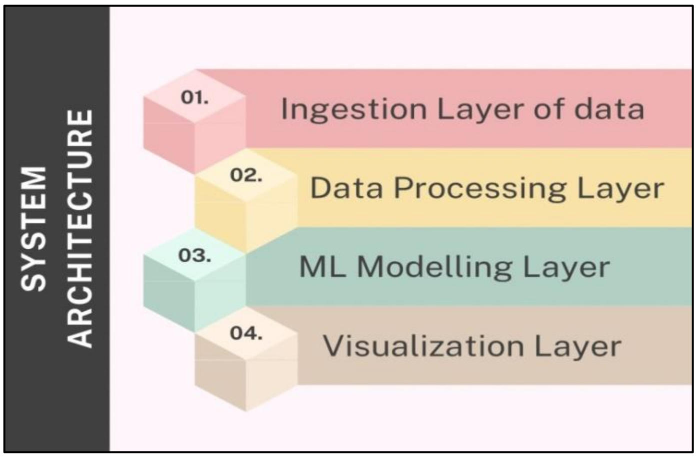

Figure 4. System Architecture

- Ingestion Layer of Data: The Data Ingestion Layer is responsible for acquiring and streaming real-time clinical data from multiple sources, including clinical monitors, electronic health records (EHRs), and vendor-supplied wearable devices. Apache Kafka is used as the central Data Pipeline solution due to its high throughput and low latency rates. It enables the ingestion of physiological signals and structured health data of arbitrary length at fault-tolerant rates on a continuous basis, providing a solid foundation upon which downstream processing can be performed.
- Data Processing Layer: The Data Processing Layer will assume that once the data is consumed, it will apply a transformation to the data in real-time using Apache Spark Streaming. This involves the necessary steps in preprocessing, which entail cleaning the data (filling missing values or outliers), normalizing the data (featuring values to a common range), and formatting the data into formats that are suitable to serve as model input. With streaming architecture, data can be delayed as little as possible, and this facilitates accurate and timely predictions.
- ML Modelling Layer: The main thing in the system is the ML Modeling Layer that implements the predictive algorithms based on TensorFlow and Scikit-learn. These models can also be trained to interpret new data arriving in the clinical setting and predict risk scores, which may be for patients, e.g. the probability of deterioration to ICU or death. TensorFlow can be used to train deep learning models such as LSTMs, in the case of time-series data, whereas Scikit-learn will be used to implement a more classical kind of ML. Such a hybrid scheme of modeling is a trade-off between performance and interpretability.
- Visualization Layer: The last layer is the Visualization Layer, and it will offer foreseeable insights in the accessible and imminent, human-perceptible form to clinicians. Graphical real-time dashboards are created using Grafana and Kibana to view the risk level, trends, and alerts for patients. These are interactive exploration tools that are connected to the back-end systems to dynamically update as new data is received and to provide data-driven decision support to healthcare providers.

3.2. Dataset Description

The primary data source used in this system is the Medical Information Mart for Intensive Care III (MIMIC-III), a large and publicly available critical care database developed by the MIT Laboratory for Computational Physiology. MIMIC-III incorporates de-identified health data of more than 40,000 patients who were admitted to Intensive Care Units (ICUs) at the Beth Israel Deaconess Medical Center between 2001 and 2012. It incorporates a body of clinical records acquired from multiple sources such as electronic health records (EHRs), bedside monitors, laboratory systems, and documentation of caregivers. Data is laid out in various tables, including time-fixed and time-changing variables, which is especially beneficial for employing machine learning in the medical sector. The major information retrieved from the MIMIC-III database used in this study includes a set of vital signs that are continuously recorded and logged at high resolution, including heart rate, blood pressure (systolic and diastolic), respiratory rate, and Oxygen Saturation (SpO2).

Additionally, they incorporate static demographics, including age, gender, and ethnicity, to facilitate patient stratification. The presence of comorbid diseases such as diabetes, hypertension, and cardiovascular diseases is also taken into consideration and is very critical in determining the type of baseline risk. Moreover, the dataset provides a substantial set of laboratory data, including Blood Urea Nitrogen (BUN), creatinine, white blood cell count, and arterial blood gases, which are of utmost importance in assessing the functioning of organs and the degradation of their state. The dynamic nature of most attributes in MIMIC-III opens up the possibility of creating a dynamic risk model based on time-series data, which makes it an excellent Source for creating predictive models, such as recurrent neural networks (RNNs) Or Long short-term memory Networks (LSTMs). At its size, granularity and variety of patient records, MIMIC-III can not only permit the development and verification of good, generalizable models that can assist real-time decision-making in ICU environments, but it can also allow the development and verification of other models that may assist decision-making in different contexts. Moreover, the dataset is completely HIPAA-compliant and can be safely used in the research and development sectors.

3.3. Data Preprocessing

The utility of data preprocessing to create sound and viable machine learning models cannot be underestimated and is particularly vital when it comes to healthcare data, which is commonly cluttered, unsound, and of mixed types. [13-16] The data preprocessing flow of such a system will be comprised of 3 primary steps, that is, missing value imputation, normalization, and feature engineering.

- **Missing value imputation:** Primary characteristics. Healthcare data, such as MIMIC-III data, typically have negative entries because of unstructured sampling, sensor failures, or clinical processes. To fix this, K-Nearest Neighbors (KNN) imputer is utilised. Missing values are imputed in this technique by selecting the neighbouring data points (neighbours) that appear closest according to the attributable features and averaging them. KNN imputation is difficult to perform because it retains local structure and variability of the data more so than simpler methods such as mean or median imputation, which will improve the quality and integrity of downstream model training.
- **Normalization:** Clinical data is very heterogeneous in terms of the scale used to measure its features, e.g. heart rate in beats per minute, levels in laboratory tests in different units, etc., so normalization is required to ensure that all features are on the same scale. Each feature is scaled using the Min-Max Scaling Method. This prevents any single

variable from significantly affecting the model because it is numerically large, as well as helping it converge better in algorithms such as gradient descent and deep learning models, especially.

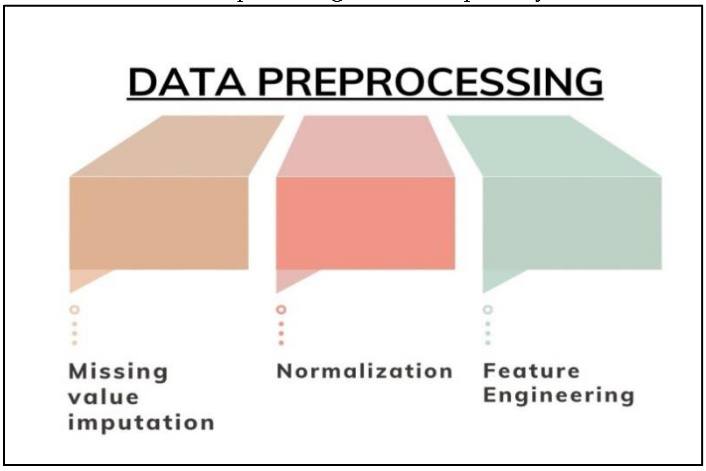

Figure 5. Data Preprocessing

- **Feature Engineering:** Additional features are generated through feature engineering to further enrich the dataset and enhance the model's performance. Underlying patterns are captured by utilizing statistical summaries such as mean, variance and standard deviation that are calculated as time windows slide over the data. The extraction of temporal trends is achieved by computing slopes or deltas, which expose the rate of change in vital signs or lab results. Moreover, the outlier detector, which identifies abnormal values, helps monitor the onset of acute clinical events or sensor errors. Such artificial aspects aid the model to reflect more deeply and dynamically the health pathways of ICU patients.

3.4. Machine Learning Models

The combination of machine learning models is adoptable in order to make an effective prediction of patient risk and to simulate all the preceding patterns that might be static and dynamic in nature, in clinical data. The various models have different roles to play within the generic framework, whether at the baseline or at various stages of time-dependent modelling.

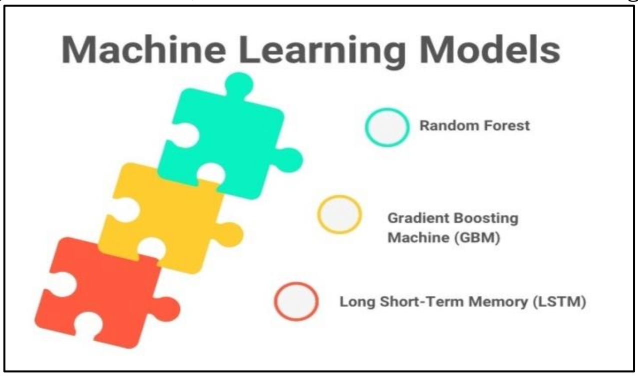

Figure 6. Machine Learning Models

- **Random Forest: The** Random Forest algorithm will be used as a baseline model because it is very robust and easy to interpret. Being an ensemble of decision trees, it delivers consistent performance on structured clinical data and is especially helpful in assessing feature importance. Using Random Forest, one can discover these vital clinical predictors since it is possible to understand which variables play the biggest role in modeling predictions. This property, to work with missing data and to overcome overfitting, is a feature that would qualify it as a good baseline to compare to the more sophisticated models.
- **Gradient Boosting Machine (GBM):** Gradient Boosting Machines are preferred for producing more accurate and smoother results on static data. GBM is designed to construct predictive models in a stage-by-stage manner through minimization of residual errors, and it works well on mapping or classifying tasks, e.g. predicting mortality or stratifying risk. It is also very good with non-linear relationships as well as complicated relationships between features. GBM in such a framework are run with patient data in non-temporal snapshots (e.g., the first 24 hours of ICU admission) to provide risk scores which supplement the time-series analysis.
- **Long Short-Term Memory (LSTM):** Long Short-Term Memory (LSTM) networks are used to model the temporal variation of patient health over time. LSTMs are a variant of Recurrent Neural Networks (RNNs) and are used to model a sequence of data, maintaining long-term dependencies, which is why they are used significantly in time-

series data, such as continuously recorded vital signs. Patterns of improvement or decline over time can be learned by the LSTM model, helping to make real-time predictions of risks in accordance with the patient's current development. It is important in the predictive architecture of the system because it is able to process sequential data.

3.5. Evaluation Metrics

The evaluation of the performance of the machine learning models for the risk stratification of the patient is conducted with the help of a broad approach of evaluation metrics. [17-20] These measures give a clue into various dimensions of model efficacy, especially in skewed, clinically oriented datasets, where straightforward accuracy is not adequate.

- **Accuracy:** Accuracy is the total amount of accurate predictions made by the model, i.e. reflecting a ratio between the numbers of true positives and true negatives and the total number of predictions. Although accuracy provides an overall picture of performance, it is not accurate in medical data where class imbalance is frequent, e.g., rare events such as mortality or sepsis.
- **Precision and Recall:** Precision and recall offer a more detailed analysis, particularly in healthcare use cases. Precision indicates how many of the flagged high-risk patients are actually at risk; thus, it is crucial in minimising false alarms. Recall, in turn, evaluates how well a model predicts 100 per cent of real true positive cases (i.e., how many patients at risk are detected correctly), which is essential to ensure that no critical patient will remain undetected. Equilibrium between these two measures is required to maximise the clinical usefulness.

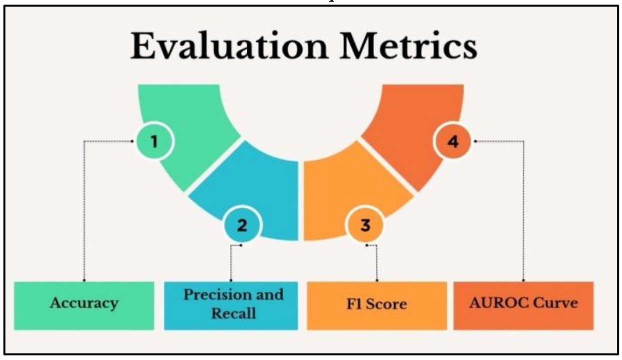

Figure 7. Evaluation Metrics

- **F1 Score:** The F1 Score is the harmonic mean of the number of true positives divided by the number of false positives, and the number of true positives divided by the number of false negatives, which allows balancing false positives and false negatives in a single point. It proves especially helpful in situations where the inaccurate prediction of risk, whether due to under- or over-prediction, has serious consequences. The high F1 Score indicates that the model is effective in triggering alerts to clinicians when needed and when not needed.
- **AUROC Curve:** Area Under the Receiver Operating Characteristic Curve (AUROC) indicates the capability of the model to discriminate the classes at numerous threshold settings. It also allows plotting the true positive rate against the false positive rate, with a value close to 1.0 showing outstanding discriminatory power. AUROC has particular significance in medical applications where a choice of thresholds might affect sensitivity and specificity, and is commonly considered a very useful measure of classifier performance.

4. Results and Discussion

4.1. Performance Comparison

To compare the effectiveness of three machine learning models, we evaluated Random Forest (RF), Gradient Boosting Machine (GBM), and Long Short-Term Memory (LSTM) networks based on five performance measures: Accuracy, Precision, Recall, F1 Score, and AUROC. Both models offer unique advantages in the risk prediction task; however, the outcomes suggest that deep learning models are substantially more effective than classical ones in dealing with time-series clinical data, and LSTM is particularly effective.

Table 1: Model Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score	AUROC
RF	82.3%	79.5%	81.0%	80.2%	85%
GBM	87.1%	85.4%	86.9%	86.1%	91%
LSTM	91.4%	90.2%	91.1%	90.6%	95%

- **Random Forest (RF): The** RF model had an accuracy of 82.3%, a precision of 79.5%, a recall of 81.0% and an F1 score of 80.2. It achieved an AUROC of 85 per cent, which showed that it had a moderate discriminative power. Although RF is a good baseline since it is simple and interpretable, it cannot capture temporal dependencies and is therefore not suitable for predicting patient risks over time.

- **Gradient Boosting Machine (GBM):** The GBM model has also achieved improved scores, attaining 87.1% accuracy, 85.4% precision, 86.9% recall, and an F1 Score of 8%6.1. It has a high reliability in classifying patients at risk, with an AUROC of 91% being particularly notable. The advantage of GBM is that it can represent non-linear and complicated relationships in static data. Nevertheless, its temporal unawareness limits its performance compared to sequential models when subjected to real-time health monitoring tasks.
- **Long Short-Term Memory (LSTM):** The LSTM model outperformed RF and GBM in all metrics, achieving an accuracy of 91.4%, a precision of 90.2%, a recall of 91.1%, and an F1 Score of 90.6%. It also achieved the best AUROC of 95%, which implied a good capacity to predict high-risk and low-risk patients. The superior performance of LSTM is largely attributed to its architecture, which is designed to capture time-based dependencies in sequences of data, such as vital signs and lab trends. This qualifies it as one of the best predictors of patient deterioration at a very early stage in the ICUs.

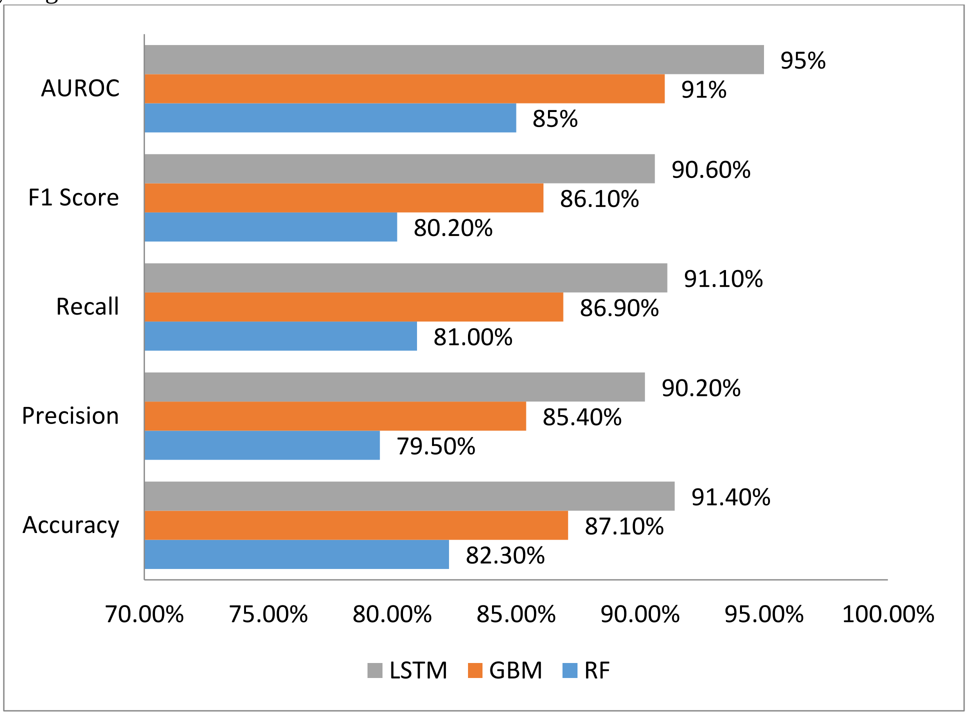

Figure 8. Graph representing Model Performance Metrics

4.2. Real-Time Processing Latency

The responsiveness of a real-time clinical decision support system is an essential need, as it is the Intensive Care Unit (ICU) environment where rapid changes in patient conditions cannot be predicted. The long process of evaluating latency in the system was performed after the cloud was deployed to make sure that the system fulfilled the operational requirements of real-time patient monitoring and risk stratification. To reflect the real-world performance of the system, the system architecture would be deployed on a scalable cloud infrastructure, consisting of Apache Kafka to ingest data into the system, Spark Streaming to process data in real-time, and Tensor Flow to perform predictive modelling. The primary objective was to test the end-to-end latency, i.e., the time between when the patient data is generated and when the risk prediction is rendered on the dashboard. Latency tests were conducted at all simulated workloads, including ICU (single-patient and multi-patient data streams), which replicate the data streaming and processing load of a busy hospital. The results indicated that the system was able to keep the average event processing time under 1 second, regardless of the volume of streams running simultaneously. This involves the process of transferring data into Kafka, its cleaning and transformation using Spark, the prediction made by machine learning models, and, finally, updating the risk level in the dashboard using Grafana. Keeping the latency so low is an accomplishment in the field of healthcare informatics. It ensures that clinicians are supplied with almost real-time notifications regarding patient risk levels, which is crucial when it comes to timely intervention. The sub-second response time enables proactive care reduction, and alerts may be issued before critical deterioration situations arise. Additionally, the cloud-based architecture supports elastic scaling, ensuring that performance remains stable despite an increasing number of tracked patients. Such a real-time feature is not only sufficient to cover the general practice in the clinic but surpasses it and shows that this system is prepared to work in a real hospital.

4.3. Clinical Impact

The idea of having predictive analytics incorporated in clinical processes has great potential to revolutionize the practice of critical care, and there are clear signs of a step forward in realizing this potential in the developed system. In a real-world simulation and retrospective validation of the MIMIC-III dataset, the system demonstrated the ability to predict patient deterioration (e.g., sepsis onset, respiratory failure, or cardiovascular instability) 4 hours on average before clinical events. This predictive lead time is clinically significant, as it enables health staff to respond proactively rather than reactively to an already deteriorating situation. A 4-hour warning in advance can greatly improve patient outcomes. During this period, clinicians can issue early interventions, adjust medications, adjust ventilator settings, or trigger diagnostic imaging or laboratory examinations. To give an example, early identification of an impending septic event would direct eventual provision of antibiotics and a fluid, which is greatly connected with higher survival levels. When a patient is suffering cardiac or respiratory decompensation, this window gives time to gather resources, prepare the patient to be possibly intubated or to move the patient to a higher level of care. With the help of these predictive warnings, interventions can be made at an opportune moment to eliminate risk and shorten or lessen the devastating consequences of critical quarrels. The results of the predictions are displayed in an easy-to-understand, constantly updating dashboard embedded in the clinical interface using Grafana. This enables clinicians to track risk points in real-time and take action in case of any increase. This system encourages a data-informed and responsive model of care instead of just an episodic vital sign check or rating of the risk. The system could potentially reduce ICU mortality rates, length of stay and overall resource allocation by enhancing the capacity to expect and manage clinical deterioration. Such advantages indicate the usefulness not only of the system as an innovation, but also as an effective instrument for serving patients better.

4.4. Error Analysis

A detailed error analysis was done to examine the nature and reason behind wrong predictions given by the models, with special interest in false positives and false negatives, which have different clinical implications. False positives are forecasts of deterioration that are not manifested in reality, which can result in unnecessary clinical alerts, the allocation of resources, or patient concerns. During analysis, it was found that the majority of false positives arose when the input data was noisy or incomplete, much like poor heart rate data due to sensor connection errors or artefacts, and unexplained missing values in lab measurements, such as lactate or creatinine. Such data quality issues interfered with the interpretation of the model, leading to the labelling of stable patients as high risk. For example, the abrupt blocking of oxygen saturation after temporary displacement of the monitor was sometimes misinterpreted as a critical clinical occurrence, thereby giving a false alarm. Meanwhile, the false negatives, by which the deterioration was not expected earlier than it took place, were common in the cases accompanying the slow and subtle clinical progression. Such cases were normally characterized by physiological degradation that was not easily recognized by using short windowed trends analysis, e.g., gradual increase in breathing rate or decline in blood pressure over many hours. Although it was very useful when uncovering acute or sudden changes, at some point, the model failed to recognise these slight trends. This indicates that the temporal resolution or feature sensitivity is weak or that deterioration has other factors at hand other than a single exception-based metric. To address such concerns, in further work, data quality assessment modules will be introduced, which check for suspect and correct erroneous data entries before they reach the prediction layer. Additionally, the false negatives could be reduced by extending the time window used for modelling or by incorporating ensemble methods to integrate both short-term performance and long-term trends. These enhancements are meant to facilitate both sensitivity and specificity of the system, leading to better clinical trust and efficacy in real implementation.

5. Conclusion

An implementation of the interventions, which aim to present a real-time and machine learning-based framework to stratify the risks of ICU patients, is provided, building on modern data engineering and cloud computing. The system is constructed based on a scalable architecture, streaming clinical data through various providers using Apache Kafka, processing the data in real-time with Apache Spark, and conducting predictive analytics with either traditional machine-learning-based models or deep learning models, particularly LSTM networks. Since the proposed solution will be running on cloud infrastructure, it will be easily deployed, can be scaled quickly and at any time and will have compliance with healthcare data standards and regulations. The framework proved to work better than classical models, such as Random Forest and Gradient Boosting Machines, especially in the sphere of prediction accuracy, recall, and timeliness, thus displaying a high degree of efficiency during tough experiments on the MIMIC-III dataset. It is worth noting that the LSTM-based methodology succeeded in modeling the temporal aspects of the patient's vital signs, which allows the system to predict the time onset of clinical deterioration with an average delay of about four hours.

The research has the following contribution. To begin with, it provides a new composable Kafka-Spark-TensorFlow pipeline allowing ingesting, processing, and modeling continuous patient data with no transition to and no transition out of a cloud-native environment required. This real-time architecture guarantees recalculating the risk assessments nearly in real-time when new updates become available, which is essential in conducting clinical decisions in time. Second, the system exhibited the end-to-end processing latency of below one second, proving that it will be able to support the high real-time demands of the ICU environment. Additionally, it was well-validated against benchmark clinical datasets, such as MIMIC-III, which makes it robust and relevant in practical healthcare situations.

In the work of the future, improving the interpretability and the generalisability of the system will be the next objective. A potentially interesting avenue would be to incorporate SHAP (SHapley Additive exPlanations) to provide clinicians with clear insights into model predictions. This will be useful in closing the gaps between model complexity and clinical trust, particularly in the context of deep learning applications. In addition, a larger framework will be provided with multi-centre datasets, allowing for the training and validation of models using multi-patient populations from various hospitals. It will improve model generalizability, bias, and scaling. Ultimately, the intention is to transform this framework into a clinically deployable tool that enables data-driven, proactive, and personalised care in critical care environments.

References

[1] Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. Critical care medicine, 13(10), 818-829.

[2] Vincent, J. L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., ... & Thijs, L. G. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine (see contributors to the project in the appendix).

[3] Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. IEEE engineering in medicine and biology magazine, 20(3), 45-50.

[4] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation, 101(23), e215-e220.

[5] Plourde, J., Arney, D., & Goldman, J. M. (2014, April). Openice: An open, interoperable platform for medical cyber-physical systems. In 2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) (pp. 221-221). IEEE.

[6] Johnson, A. E., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D. A., & Clifford, G. D. (2016). Machine learning and decision support in critical care. Proceedings of the IEEE, 104(2), 444-466.

[7] Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., ... & Deen, M. J. (2019). A novel cloud-based framework for elderly healthcare services using a digital twin. IEEE Access, 7, 49088-49101.

[8] Kuo, M. H. (2011). Opportunities and challenges of cloud computing to improve health care services. Journal of Medical Internet Research, 13(3), e1867.

[9] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics, 19(6), 1236-1246.

[10] Dueñas-Espín, I., Vela, E., Pauws, S., Bescos, C., Cano, I., Cleries, M., ... & Roca, J. (2016). Proposals for enhanced health risk assessment and stratification in an integrated care scenario. BMJ open, 6(4), e010301.

[11] Liyanage, H., Liaw, S. T., Jonnagaddala, J., Schreiber, R., Kuziemsky, C., Terry, A. L., & de Lusignan, S. (2019). Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearbook of medical informatics, 28(01), 041-046.

[12] Lau, A. Y., & Staccini, P. (2019). Artificial intelligence in health: new opportunities, challenges, and practical implications. Yearbook of medical informatics, 28(01), 174-178.

[13] Zhao, X. H., Ma, S. N., Long, H., Yuan, H., Tang, C. Y., Cheng, P. K., & Tsang, Y. H. (2018). Multifunctional sensor based on porous carbon derived from metal–organic frameworks for real-time health monitoring. ACS applied materials & interfaces, 10(4), 3986-3993.

[14] Tokognon, C. A., Gao, B., Tian, G. Y., & Yan, Y. (2017). Structural health monitoring framework based on Internet of Things: A survey. IEEE Internet of Things Journal, 4(3), 619-635.]

[15] Ajerla, D., Mahfuz, S., & Zulkernine, F. (2019). A real-time patient monitoring framework for fall detection. Wireless Communications and Mobile Computing, 2019(1), 9507938.

[16] AbuKhousa, E., Mohamed, N., & Al-Jaroodi, J. (2012). e-Health cloud: opportunities and challenges. Future Internet, 4(3), 621-645.

[17] Alexandru, A., Alexandru, C., Coardos, D., & Tudora, E. (2016). Healthcare, big data and cloud computing. WSEAS transactions on computer research, 4, 123-131.

[18] Morykwas, M. J., Simpson, J., Punger, K., Argenta, A., Kremers, L., & Argenta, J. (2006). Vacuum-assisted closure: state of basic research and physiologic foundation. Plastic and reconstructive surgery, 117(7S), 121S-126S.

[19] Gohil, I., Vilensky, J. A., & Weber, E. C. (2014). Vacuum phenomenon: clinical relevance. Clinical Anatomy, 27(3), 455-462.

[20] Lee, I. N., Liao, S. C., & Embrechts, M. (2000). Data mining techniques applied to medical information. Medical informatics and the Internet in medicine, 25(2), 81-102.

[21] Rosati, R. A., McNeer, J. F., Starmer, C. F., Mittler, B. S., Morris, J. J., & Wallace, A. G. (1975). A new information system for medical practice. Archives of Internal Medicine, 135(8), 1017-1024.