CSE 6220: PROJECT PROPOSAL

Enhancing Patient Care: A Comprehensive Real-Time Healthcare Monitoring, Alerting, And Predictive System

Anshit Verma, Manvi Goel, Navdha Agarwal, Padma Priya Mukkamala,
Reshma Ramachandra, Sindhu Raghuram Panyam

GOAL AND SCOPE

Our project's overarching goal is to revolutionize healthcare monitoring by introducing a holistic and predictive system for real-time medical vitals monitoring. We aim to enhance patient care by proactively identifying critical conditions before they escalate and creating a comprehensive alerting mechanism for healthcare professionals.
Our project distinguishes itself by combining real-time monitoring, customizable alerts, predictive modeling (ARIMA, Prophet), and survival analysis into a single, unified system. We aspire to achieve the following primary objectives:

- Creating Personalized Alerts: Leverage real-time big data analysis to tailor alerts to individual patient profiles. This approach takes into account variations in vital sign baselines and medical history, providing healthcare professionals with highly specific alerts.
- Predicting Critical Events: Develop predictive models, including time series and survival analysis, to forecast deteriorating health conditions based on large datasets. These models will identify trends and issue early warnings about potential health crises, enabling timely interventions and reducing the risk of adverse events.
- Comparison and Analysis: Conduct a rigorous comparison and analysis of the behavior predicted by the models with respect to real-time alerts generated by the system. This evaluation will provide insights into the effectiveness of our predictive analytics in improving patient outcomes.

MOTIVATION & PROBLEM STATEMENT

The current landscape of healthcare monitoring is characterized by several pressing challenges that necessitate solutions:

1. With the advancement of healthcare technology, there's a surge in the volume of data generated. Managing and processing this large volume of data is a significant challenge. Delays in data transmission or processing can be detrimental to patient outcomes.
2. Traditional healthcare monitoring systems rely on reactive approaches, triggering alerts only when vital sign values breach predefined thresholds. In our project, we aim to produce alerts that are proactively generated using ML algorithms to improve healthcare scenarios.
3. Lack of personalization: Healthcare monitoring systems typically lack the personalization required to provide alerts tailored to individual patient profiles. Existing systems employ static thresholds for generating alerts, treating all patients uniformly. Our system will treat every patient uniquely and trigger alerts based on individual thresholds set for them.

RELATED WORK

Existing real-time monitoring solutions have significantly contributed to healthcare, emphasizing immediate awareness of vital sign fluctuations (Bernardos et al., Cooper et al., Komashie et al.). These systems aim to tailor alerts to individual patient profiles, accounting for variations in vital sign baselines and medical history. Our project builds upon this foundation, utilizing real-time big data analysis to further enhance the specificity and effectiveness of personalized alerts by concentrating on the early identification of critical conditions before escalation. This aligns with recent research highlighting the limitations of reactive monitoring and the potential for substantial improvements in patient care through early intervention.

The integration of predictive analytics into healthcare monitoring has gained traction in recent times. In medical survival analysis, traditional ML methods like SVM and Random Forests are commonly used but often limited by predefined features and difficulties capturing complex patterns (Singh & Mukhopadhyay, Ishwaran et al., Chen et al.). DL methods like DeepSurv and DeepHit offer promise in automatically extracting features from unstructured data but are seen as black boxes, hindering interpretability (Katzman et al., Lee et al.). Thus, we shifted our focus to Time series and survival analysis models (Bloch et al., Alghatani et al.) for forecasting deteriorating health conditions. These models analyze large datasets, identify trends, and provide early warnings about potential health crises. Our project focuses on the development of advanced predictive models to facilitate early intervention and improved patient outcomes. Further, our commitment to validating model predictions against real-time alerts sets our work apart, ensuring the accuracy and effectiveness of our approach.

DATASET:

The data has been obtained from Kaggle and is centered around Surgery Mortality Prediction. It provides patient data collected from a healthcare setting, primarily focusing on various health metrics and conditions such as blood pressure, heart rate, and temperature.

Link to the dataset - https://www.kaggle.com/datasets/hitmanhulk/surgery-mortality-prediction
Size: 50240 records and 106 columns

SOLUTION APPROACH AND ARCHITECTURAL DESIGN:

A) REAL TIME BIG DATA SYSTEM:

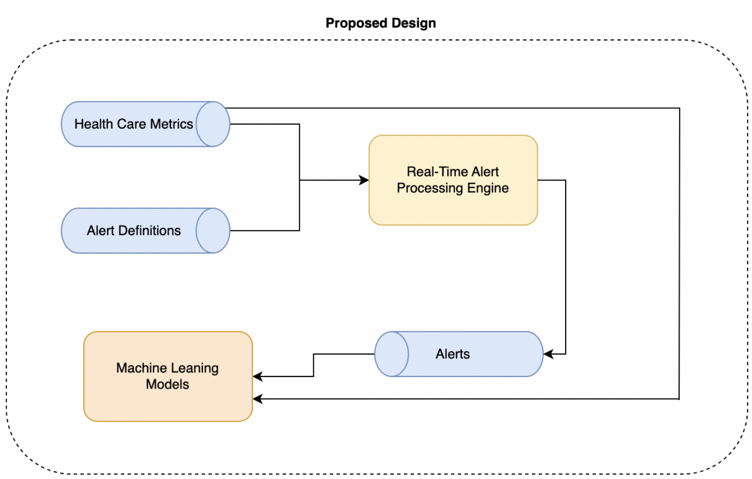
Our project aims to develop a robust real-time healthcare monitoring and alerting system that enables healthcare professionals to monitor patient's vitals and receive immediate alerts in case of anomalies. In the following section, we describe the solution approach and the architectural design of our alerting system ensuring its reliability, scalability, and performance.

- **Data Preprocessing and Ingestion:** We require time-series data encompassing vital signs such as heart rate and blood pressure, etc., simulating continuous generation from monitoring devices for each patient, with updates recorded every minute. Once we get this real-time data, it is continuously ingested from a queuing service like Apache Kafka queue and passed to the data processing component.

- **Real-Time Data Processing:** We will use Apache Flink as our real-time data processing framework, which will be responsible for handling incoming data streams. To ensure isolation between patients, each patient's data will be independently processed in parallel threads.

- **Alert Definition Stream:** We would require a separate data stream for healthcare professionals to provide new alert definitions. This stream would enable doctors to customize alert criteria for individual patients. These alert definitions are consumed from a separate Kafka queue and seamlessly integrated into the data processing component.

- **Anomaly Detection and Alerting Mechanism:** We will integrate a real-time anomaly detection algorithm to trigger alerts when significant deviations occur in patient data with respect to the alert definitions. These alerts would include information about the detected deviation and will be promptly communicated to the doctor.



B) MACHINE LEARNING (ML) PIPELINE

This section focuses on using historical patient data to predict future health risks. This system complements the real-time monitoring, allowing for proactive patient care. In our ML tasks, we will employ Time Series models for predictive analysis, and Survival Analysis Models for predicting the time remaining before actionable events.

Time Series Model:

Time Series Analysis (TSA) focuses on studying how a variable changes over time. In our use case, we harness real-time data streams to train a time series model, working alongside our alerting system. This predicts future metric values, allowing healthcare professionals to take preventive actions before thresholds are exceeded. For our experiments, we employ two time-series models and compare their predictions to real-time predictions.

ARIMA signifies the model's utilization of the dependent relationship between an observation and lagged observations from previous time steps. This empowers ARIMA to forecast future trends and deviations within the healthcare vitals, enhancing the predictive capabilities of our monitoring system.

PROPHET is an open-source library for univariate time series that implements an additive time series forecasting model. It works best with time series that have strong seasonality and several seasons of historical data. Since we aim to identify trends from historical data, the prophet seems to be fitting appropriately for our use case.

***Survival Model*:**

Incorporating survival analysis into our healthcare monitoring system represents a proactive approach to time-to-event analysis. By utilizing historical patient data, analyzing the alerts generated from our real-time system, and survival analysis techniques, we develop predictive models that continuously assess the likelihood of critical healthcare events occurring. In tandem with the real-time monitoring system, survival analysis can not only trigger immediate alerts when metrics fall outside the healthy range but also forecast future periods of potential instability. This dual approach ensures immediate intervention when needed, while also allowing healthcare professionals to proactively manage and monitor patients deemed at risk in the future.

Survival analysis can be used to predict both the time until an event happens and the likelihood of an event occurring within a specified time frame. To do this, the data is first organized to show when each patient's metric goes out of range or if it hasn't yet. Using this data, a model is fit, often the Cox Proportional Hazards model, which considers factors like age or gender that might influence the metric's stability.

EVALUATION METRICS USED AND EXPECTED RESULTS:

For our predictive model, we are using real-time alerts as our true labels for training. We will use metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) for evaluation. We expect our predictive model to predict the alerts at the same time as our real-time monitoring system.

For our survival analysis model, we will employ cross-validation techniques and a concordance index to indicate predictive accuracy. We expect to produce a survival curve that depicts the likelihood of a health metric remaining within its acceptable range over time. Steeper declines in these curves indicate periods of heightened risk. We also aim to identify risk factors such as age, and gender, that influence the time it takes for a metric to deviate from its range. Lastly, for any given patient or set of conditions, the model can provide an estimated time frame within which a metric is likely to go out of its acceptable range.

TIMELINE

Dates	Activities/Tasks
25-29 Sept	Project Ideas Exploration, Finalization, Proposal Writing
2-8 Oct	Start technical implementation with Data Preprocessing of Health data
9-16 Oct	Real-Time Data Processing Setup utilizing Apache Flink for data processing and Kafka as an incoming ingestion source to our application. Build predictive model - ARIMA
17-26 Oct	Alert Definition Kafka Stream Implementation Gathering results from the ARIMA model
27-5 Nov	Anomaly Detection Algorithm Development using Apache Flink Build predictive model - Prophet
6-12 Nov	Alert Mechanism Integration by building an outgoing Kafka Queue for alerts Build Survival Analysis Model
12-18 Nov	Gather predictive results and perform evaluation and comparisons
19-26 Nov	Test the integration of the entire pipeline and deploy it to the Test Environment
27-1 Dec	Project Wrap-up and Handover

BIBLIOGRAPHY:

Lehman, Li-Wei, Mohammad Ghassemi, Jasper Snoek, and Shamim Nemati. "Patient prognosis from vital sign time series: Combining convolutional neural networks with dynamical systems approach." In 2015 Computing in Cardiology Conference (CinC), pp. 1069-1072. IEEE, 2015.

Bloch, Eli, Tammy Rotem, Jonathan Cohen, Pierre Singer, and Yehudit Aperstein. "Machine learning models for the analysis of vital signs dynamics: a case for sepsis onset prediction." Journal of Healthcare Engineering 2019 (2019).

Alghatani, Khalid, Nariman Ammar, Abdelmounaam Rezgui, and Arash Shaban-Nejad. "Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation." JMIR medical informatics 9, no. 5 (2021): e21347.

Katzman, Jared L., Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network." BMC Medical Research Methodology 18, no. 1 (2018): 1-12.

Singh, Ritesh, and Keshab Mukhopadhyay. "Survival analysis in clinical trials: Basics and must know areas." Perspectives in clinical research 2, no. 4 (2011): 145.

Chen, Zhanbo, and Qiufeng Wei. "Developing an Improved Survival Prediction Model for Disease Prognosis." Biomolecules 12, no. 12 (2022): 1751.

Ana M. Bernardos, Marcos Pires, Daniel Ollé, and José R. Casar. 2019. Digital Phenotyping as a Tool for Personalized Mental Healthcare. In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'19). Association for Computing Machinery, New York, NY, USA, 403–408.

Clare J. Hooper, Mark Bernstein, and Mark Weal. 2020. Personalization and Community 2020: User Modelling and Social Connections in Web Science, Healthcare and Education: Chairs' Workshop Summary. In Companion Publication of the 12th ACM Conference on Web Science

Alexander Komashie, Ali Mousavi, and Mustafa Ozbayrak. 2005. A real-time knowledge-based decision support system for health care quality improvement using discrete event simulation techniques. In Proceedings of the 37th conference on Winter simulation (WSC '05).