IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 8, 2022, accepted May 1, 2022, date of publication May 16, 2022, date of current version May 20, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3175304*

# Precision Clinical Medicine Through Machine Learning: Using High and Low Quantile Ranges of Vital Signs for Risk Stratification of ICU Patients

**KHALID ALGHATANI[1], (Member, IEEE), NARIMAN AMMAR[2], (Member, IEEE), ABDELMOUNAAM REZGUI[3], AND ARASH SHABAN-NEJAD[2]**

[1]King Fahad Medical City, Riyadh 11525, Saudi Arabia
[2]UTHSC-ORNL Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, The University of Tennessee Health Science Center, Memphis, TN 38103, USA
[3]School of Information Technology, Illinois State University, Normal, IL 61790, USA

Corresponding authors: Khalid Alghatani (kalghatani@kfmc.med.sa) and Arash Shaban-Nejad (ashabann@uthsc.edu)

The work of Abdelmounaam Rezgui was supported by the College of Applied Science and Technology (CAST), Illinois State University, Normal, IL, USA.

**ABSTRACT** Remote monitoring of patients in the intensive care unit (ICU) is a crucial observation and assessment task that is necessary for precision medicine. We have recently built a cloud-based intelligent remote patient monitoring (IRPM) framework in which we follow the state-of-the-art in risk stratification through machine learning-based prediction, but with minimal features that rely on vital signs, the most commonly used physiological variables obtained inside and outside hospitals. In this work, we significantly improve the functionality of the initial IRPM framework by building three machine learning models for readmission, abnormality, and next-day vital sign measurements. We provide a formal representation of a feature engineering algorithm and report the development and validation of three reproducible machine learning prediction models: ICU patient readmission, abnormality, and next-day vital sign measurements. For the readmission model, we proposed two solutions for data with imbalanced classes and applied five binary classification algorithms to each solution. For the abnormality model, we applied the same five algorithms to predict whether a patient will show abnormal health conditions. Our findings indicate that we can still achieve a reasonable performance using these machine learning models by focusing on low and high quantile ranges of vital signs. The best accuracy achieved in the readmission model was around 67.53%, with an area under the receiver operating characteristic (AUROC) of 0.7376. The highest accuracy achieved in the abnormality model was around 67.40%, with an AUROC of 0.7379. For the next-day vital sign measurements model, we provide three approaches for selecting model predictors and apply the eXtreme Gradient Boosting (XGB) and Random Forest Regression (RFR) algorithms to each solution. We found that, in general, the use of the most recent vital sign measurements achieves the least prediction error. Considering the large investment from the medical industry in patient monitoring devices, the developed models will be incorporated into an Intelligent ICU Patient Monitoring (IICUPM) module that can potentially facilitate the delivery of high quality care by implementing cost-efficient policies for handling the patients who utilize ICU resources the most.

**INDEX TERMS** Precision medicine, artificial intelligence, ICU patient monitoring, prediction models, biomedical informatics.

## I. INTRODUCTION

Precision health and medicine [1] is a relatively new initiative that aims to improve equitable care and overall individual and population health through targeted observation and assessment, early detection, prevention, and treatment, as well as precision health promotion and engagement. Intensive care units (ICUs) often collect and use a high volume of patients' data to enable physicians and ICU nurses to make timely decisions in delivering high-quality critical care. The use

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.


This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

of artificial intelligence (AI) and machine learning (ML) methods to improve the care and treatment of critically ill patients has grown substantially in recent years [2], [3]. ML approaches for analysis of clinical datasets offer great promise for the delivery of personalized medicine and targeted treatment, but these approaches must be customized and optimized [4], [5] for each specific application.

We have recently developed a cloud-based intelligent remote patient monitoring (IRPM) framework [6]. In this work, we significantly improve the functionality of the initial IRPM system by building three machine learning models for three clinical outcomes: readmission, abnormality, and next-day vital sign measurements. Our motivation for building those three machine learning models is two-fold: 1) to achieve the goal of studying the three clinical outcomes (readmission, abnormality, and next-day vital sign); 2) to test if more powerful machine learning models can be deployed in the IRPM framework.

### A. INTELLIGENT REMOTE PATIENT MONITORING (IRPM) FRAMEWORK

Our IRPM framework consists of the following modules (Figure 1):

1) Intelligent ICU patient monitoring (IICUPM) module: through the IICUPM module's interface, the hospital system can load clinical and demographic characteristics for either an individual patient or a patient population. The module provides five predictive ML models to process the data and return risk scoring results.
2) Out-of-hospital module: unlike the IICUPM module, this module targets individual patients. Through the module's interface, a patient's readings from wearable devices (e.g., heart rate and SPO2) are uploaded into an abnormality ML prediction model, which returns an overall assessment or risk scoring result.
3) IRPM core system and database: the IICUPM and out of hospital interfaces interact with the core IRPM system, which sends the data to the ML model development modules. After processing the data, the ML models send the results back to the IRPM framework to record them in the database.

### B. CLINICAL CONSIDERATIONS

The use of prognostic models is a common practice in precision clinical medicine to formally combine predictors from which risks of a specific endpoint can be calculated for individual patients. The clinical goal is to ensure that patients are placed on the appropriate care pathway, including proper ICU type and level of care. The proposed framework provides not only healthcare provider-focused services but also patient-facing digital health services through the out-of-hospital module. This article sheds light on a portion of the ongoing development of the IRPM framework. We developed a prototype for a dashboard that can be used for triaging patients and navigating care to support clinical

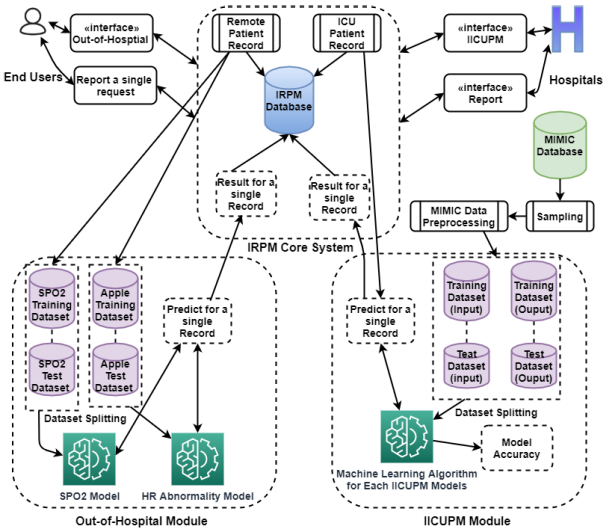



**FIGURE 1. The intelligent remote patient monitoring (IRPM) framework is comprised of intelligent ICU patient monitoring (IICUPM) and out-of-hospital modules. MIMIC stands for the medical information mart for intensive care.**

decision-making by using predicted risk to preemptively triage patients across different levels of the healthcare system. The results of applying the feature engineering algorithm and prognostic ML models will help determine the proper level of care based on the predicted risk level. Therefore, high-risk patients can be managed at higher level care facilities while lower-risk patients can be managed at lower levels of care.

### C. SUMMARY OF CONTRIBUTIONS

The contributions of this article are three-fold:

- Building on our recent development of the IRPM framework, we provide three more ML models in the IICUPM module for risk stratification of ICU patients, including readmission, abnormality, and next-day vital sign measurements.
- In addition to providing three transparent and reproducible ML models, we also present details of a feature engineering algorithm that can be deployed in different critical care settings.
- We provide two different solutions for data with imbalanced classes and three different variations for predicting next day vital sign measures to show that our solutions can predict health outcomes with reasonable performance. We evaluated all proposed solutions and applied them to ICU patient data from a publicly available database to determine the best use on our IICUPM module.

## II. RELATED WORK

We discuss research efforts on building ML models to predict ICU-related outcomes. Some researchers have studied patient readmission to the ICU as an outcome. For example, Rajkomar *et al.* [7] developed a deep learning (DL) model to predict 30-day unplanned hospital readmission. In our

study, we predict readmission to the ICU during the same hospitalization rather than readmission after discharge from the hospital. Some research has involved the development and deployment of ML models to predict health status using the Medical Information Mart for Intensive Care (MIMIC) database [8]–[10]. However, most of these investigators have used an exhaustive list of features to achieve higher accuracy in their models. We discuss some of them below.

According to the MIMIC website, only a few studies describe the development and deployment of prediction models for ICU readmission using the MIMIC database. To the best of our knowledge, predicting ICU patient readmission and abnormality based on balanced classification using only vital signs and demographic attributes from MIMIC has not been studied previously. Nor are we aware of any research performed on the prediction of ICU patients' next-day vital sign measurements. Lin *et al.* [11] developed models to predict ICU patient readmission within 30 days of discharge using recurrent neural network (RNN) with long short-term memory (LSTM). They used several features from the MIMIC III database, including 17 chart events, 4 demographic features, and the International Classification of Diseases, 9th Revision (ICD-9) code. In our study, we predict patient readmission without prior knowledge of their medical conditions or diagnoses.

Fialho *et al.* [12] developed a model for predicting ICU readmission using patient data before ICU discharge. Their dataset included data from 893 non-readmitted patients and 135 readmitted patients. They used several features from MIMIC-II, including 6 monitoring signals, the results of 15 laboratory tests, and urine output. In our research, we used data from only the first day of each patient's ICU stay and only 11 input features. We include imbalanced techniques to produce a balanced classification with the same total number of samples in both the readmitted and non-readmitted patient classes.

Shin *et al.* [13] built three prediction models using two ML algorithms to predict hospital readmission in pediatric asthma patients from a regional hospital in Memphis, TN. They also used 12 features, including demographic attributes, biomarkers, and socioeconomic factors. The goal was to predict readmission within one year of the initial hospitalization. They compared a model based solely on socioeconomic features derived from the patients' residential neighborhood to a model that is based on clinical features derived from the patient record. They found that the model based on socioeconomic factors achieved accuracy that is as good as that of the biomarker-based model.

Golmaei and Luo [14] examined a novel DL framework based on a deep patient representation. They assessed the framework on 30-day hospital readmission using patient data extracted from the MIMIC-III database. Their results show that the novel framework achieves a better predictive power than do the baseline models.

Pakbin *et al.* [15] developed risk-of-ICU-readmission models for predicting readmission to the ICU at different target times. They studied patient ICU readmission after a hospital discharge and during single hospital admission and extracted several features from the MIMIC III database.

Momenzadeh *et al.* [16] applied cluster centroids as the undersampling method aimed to reduce the majority class samples. The k-means clustering technique was used to divide the majority class and generate k clusters. Their strategy is to use different numbers of clusters. The number of clusters is equal to the minority class and two and three times greater than the minority class. Their strategy has an impact on accuracy and avoids model over-fitting. Akter *et al.* [17] applied several unsuperivsed approaches to cluster breast cancer patients, including K-Means and several Hierarchical Clustering techniques. The highest accuracy achieved in their case was 70.91% using Hierarchical Clustering with average linkage.

## III. THE QUANTILES APPROACH

In this section, we introduce the quantiles approach that not only focuses on patient's characteristics in the baseline, but also performs feature engineering steps to add richer features to the data.

### A. PATIENT BASELINE VITAL SIGN FEATURES

In general, a patient will have a set of vital sign readings in the baseline that are often normally distributed [18]. Vital signs include body temperature (BT), heart rate (HR), respiration rate (RR), arterial systolic blood pressure (SBP), arterial diastolic blood pressure (DBP), peripheral oxygen saturation (SpO2), and blood glucose level (GL). These vital signs are readily obtainable from the electronic health record (EHR) because they are measured frequently.

### B. PATIENT DETERIORATING CONDITIONS

Vital signs are recorded in a sequential manner and research studies often process these features using either time series analyses or by aggregating them for each patient using some measure (e.g., mean or median). Thus, when dealing with sequential vital sign readings, some researchers [12] use the mean value of the vital sign observations rather than using observations that may deviate far from the median. In our approach, we argue that a patient's deteriorating condition often happens at a high or low level of measurement. Thus, we believe that these observations are essential, as they capture dramatic changes in a patient's health status.

### C. FEATURE ENGINEERING ALGORITHM IN THE QUANTILES APPROACH

We have previously proposed the notion of the quantiles approach [6], in which we perform feature engineering by emphasizing high and low quantiles of a patient's vital sign observations. Our previous work showed that the quantiles approach provides a richer dataset by engineering a list of new features. Algorithm 1 shows the steps performed by the feature engineering algorithm. The algorithm has three inputs: a list of patient samples $P$, and two desired probabilities

**Algorithm 1** Feature Engineering Algorithm

**INPUT:** $P$, $PPF_H$, $PPF_L$
**OUTPUT:** $V_{Quantiles}$
**for** $p_i \in P$ **do**
  $S \leftarrow ICUStays(p_i)$
  **for** $s_j \in S$ **do**
    $V_{baseline} \leftarrow Vitals(s_j, day_1)$
    **for** $v_k \in V_{baseline}$ **do**
      $v_k \leftarrow Normalize(v_k)$
      $v_k \leftarrow SortAscending(v_k)$
      $Mean_k \leftarrow mean(v_k)$
      $Means \leftarrow Means + Mean_k$
      $ObsinQ_1toQ4 \leftarrow count(v_k)$
      $DiscreteValue_L \leftarrow PPF(v_k, PPF_L)$
      $DiscreteValue_H \leftarrow PPF(v_k, PPF_H)$
      **for** $obs_i \in v_k$ **do**
        **if** $obs_i \leq DiscreteValue_L$ **then**
          $QObs_k \leftarrow QObs_k + obs_i$
          $ObsinQ_1 \leftarrow ObsinQ_1 + 1$
        **end if**
        **if** $obs_i \geq DiscreteValue_H$ **then**
          $QObs_k \leftarrow QObs_k + obs_i$
          $ObsinQ_4 \leftarrow ObsinQ_4 + 1$
        **end if**
      **end for**
      $ModMean_k \leftarrow mean(QObs_k)$
      $ModMeans \leftarrow ModMeans + ModMean_k$
      $ModSD_k \leftarrow standardDev(QObs_k)$
      $ModSDs \leftarrow ModSDs + ModSD_k$
      $QPercent_k \leftarrow QuantilePercentage(ObsinQ_1toQ4, ObsinQ_1, ObsinQ_4)$
      $QPercents \leftarrow QPercents + QPercent_k$
    **end for**
  **end for**
**end for**
$V_{Quantiles} \leftarrow Means + ModMeans + ModSDs + QPercents$
**return** $V_{Quantiles}$

for the percent point function (PPF), $PPF_H$ and $PPF_L$. The algorithm iterates through all patients in $P$ and for each ICU stay $S$, extracts the baseline vital sign features $V_{baseline}$ from the first day only, normalizes the observations in those features using the probability density function (PDF), and sorts them in ascending order. The algorithm next extracts two discrete values using the PPF function, $DiscreteValue_L$ and $DiscreteValue_H$ and uses these values as thresholds to extract observations that fall in low and high quantile ranges. For each vital sign in $V_{baseline}$, it calculates the following new features: the list of modified means *ModMeans*, the list of modified standard deviations *ModSDs*, and the list of quantile percentages *QPercents* of each baseline vital sign feature. Adding the new features to the baseline features $V_{Baseline}$ produces a new list $VQuantiles$ that achieves a better predictive power than would be obtained from only baseline vital sign features. The quantiles percentage, $Q_{percent}$, is calculated according to equation (1). $ObsinQ_1$ and $ObsinQ_4$ represent the vital sign observations that occur in the first and fourth quantiles, respectively.

$$Q_{Percent} = \frac{\#ObsinQ_1 + \#Obs.in\ Q_4}{\#ObsinQ_1toQ_4} \tag{1}$$

## IV. METHOD

### A. POPULATION SELECTION AND DATA EXTRACTION

We extracted hospital admission data from the publicly available ICU adult patient database, MIMIC-III (v1.4) [9]. The database is structured such that each hospital admission may contain one or more ICU stays and each ICU stay may span several days. On each day, a patient may have several recorded observations. We started with a total of 61,532 ICU stays (Figure 2) and extracted data recorded on only the first day of a patient's ICU stay, which resulted in a total of 45,254 unique ICU stays with demographic features (age, sex, height, and weight). We combined that with 59,241 ICU stay encounters that contained data pertaining to seven vital sign features (BT, HR, RR, SBP, DBP, SpO2, and GL). The total after merging was 44,626.

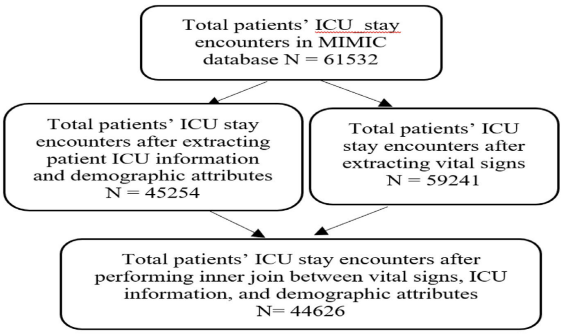



**FIGURE 2.** Population selection chart.

### B. MODEL CONCEPTUALIZATION AND VARIABLE SELECTION

Table 1 summarizes the rationale behind each model. The goal of the readmission model is to predict readmission risk, which might be an indicator that the current ICU type is not the best choice for the patient. The hospital system might decide to delay a patient discharge and increase the care level or move a patient from one ICU type to another. The outcome variable for the readmission model is a binary feature indicating whether a patient has been readmitted to the ICU (readmission = 1) or not (readmission = 0) within a single hospital admission. We define readmission as an incident in which a patient had been admitted to the ICU, discharged to the appropriate hospital population, and readmitted again to the ICU during a single hospital admission. If a patient had been admitted to the ICU and discharged once during a single hospital admission, that patient is not considered readmitted.

**TABLE 1.** Model conceptualization and variable selection.

| Model | Rationale | Predictors (Operationalization) | Clinical Outcome |
|---|---|---|---|
| Readmission risk | Predicting Readmission Risk<br>Call to action:<br>1. Delay discharge<br>2. Increase care level<br>3. ICU transfers | Demographics and vital signs from first day of ICU admission | A binary feature indicating whether a patient had been admitted to the ICU,<br>or discharged, then readmitted to the ICU during a single hospital admission. |
| Abnormality risk | Predicting patients at risk of showing undesired health conditions | Demographics and vital signs from first day of ICU admission | A binary feature indicating whether a patient is normal (normality=0) or abnormal (normality=1). Abnormality in our case is defined as the presence of one or<br>more of the following undesired health conditions in a patient record:<br>a. Mortality<br>b. Readmission to the ICU within a single hospital admission or<br>c. Prolonged ICU length-of- stay. |
| Next day vital sign reading | Predicting daily vital signs, thus providing early warnings about potential deterioration in a patient's condition based on prior physiological measurements in the baseline | Demographics and vital signs in reference to a time point in the baseline, where the baseline values could be measured on:<br>a. $Day_{i-1}$<br>b. $Day_{i-1}$ to $Day_k$<br>c. $Day_{i-1} + dev.errors$ | A numeric feature that indicates the mean of each vital sign reading on the next day of a patient's ICU stay in reference to a baseline. |

The goal of the abnormality model is to predict the pool of patients who risk showing undesired health conditions. For the abnormality model, the outcome is the abnormality of a patient, which is a binary feature indicating whether a patient is normal (normality $= 0$) or abnormal (normality $= 1$). We define abnormality as the presence of one or more of the following undesired health conditions in a patient record: death, readmission to the ICU, or prolonged ICU stay.

Readmission and abnormality are different clinical goals. To measure abnormality, we use 3 metrics, one of which is the readmission flag. To measure readmission, we use only the readmission flag as a metric. The "within a single hospital admission" criterion applies to the readmission flag in both models.

The goal of the next-day vital sign measurements model is to predict ICU patients' physiological variables daily, thus providing early warnings about potential deterioration in a patient's condition based on prior physiological measurements in the baseline. We predict the mean of the vital sign readings for each vital sign on the next day.

We used the six main vital sign measurements (BT, HR, RR, SBP, DBP, and SpO2) along with GL and the five demographic attributes (age, sex, weight, and height) as predictor variables. Table 2 summarizes the cohort characteristics for the different variables used in these models. The vital sign measurements are calculated based on the first day of the patient's ICU stay.

### C. READMISSION MODEL

The readmission prediction reduces to a binary classification problem with two classes: non-readmitted (N = 41,597 stays) and readmitted (N = 3,029 stays). This results in data with imbalanced classes [19] with 93.21% of the patients in the non-readmitted class and only 6.79% in the readmitted class.

**TABLE 2.** Patient baseline characteristics.

| Variable | Operationalization | Value |
|---|---|---|
| Sex | N (%) | |
| | Male | 25,241 (56.56) |
| | Female | 19,385 (43.44) |
| Age | Mean (SD), years | 64.35 (16.87) |
| Height | Mean (SD), cm | 160.66 (11.76) |
| Weight | Mean (SD), kg | 80.45 (23.47) |
| HeartRate_mean | Mean (SD), beats/min | 85.99 (15.59) |
| sysbp_mean | Mean (SD), mmHg | 118.75 (16.90) |
| diasbp_mean | Mean (SD), mmHg | 60.47 (10.89) |
| RespRate_mean | Mean (SD), breaths/min | 18.93 (4.05) |
| Tempc_mean | Mean (SD), °C | 36.84 (0.62) |
| SpO2_mean | Mean, % | (97.27%) |
| Glucose_mean | Mean (SD), mg/dL | 138.74 (41.86) |

#### 1) SAMPLING APPROACHES

Several re-sampling approaches support the data with imbalanced classes. Two of the most popular ones are the under-sampling of the majority class and the over-sampling of the minority class. The readmitted patients in our case represent the minority class (3,029 samples), while the non-readmitted patients represent the majority class (41,597 samples). Under-sampling may result in discarding patient data that could cause model under-fitting, while oversampling generates new samples by duplicating samples in the minority class. The later may require more computational power to process the extra samples (83,194 samples compared to only 6,058). Thus, we elected to use the under-sampling approach. To overcome the limitation in this approach, we apply another approach that involves clustering patients in the majority class into equally sized groups before merging them back together. This mitigates the lack of variation limitation introduced by under-sampling (e.g., selecting patients that belong to a single ICU type or to a single hospital).

- Under-sampling of majority class re-sampling: in this approach, we kept the 3,029 patients in the minority class and randomly removed patients from the majority class to retain only 3,029 of the total 41,597 patients in that class.
- Clustering of majority class re-sampling: in this approach, we applied the k-means clustering algorithm to cluster the 41,597 patients in the majority class into equally sized groups (in this case, 5 clusters of 606 samples each) before merging them back together to obtain 3,029 patients (Figure 3). The reason for selecting 5 clusters is that after applying the elbow method, we observed that the WCSS value reduces significantly at 5 clusters.

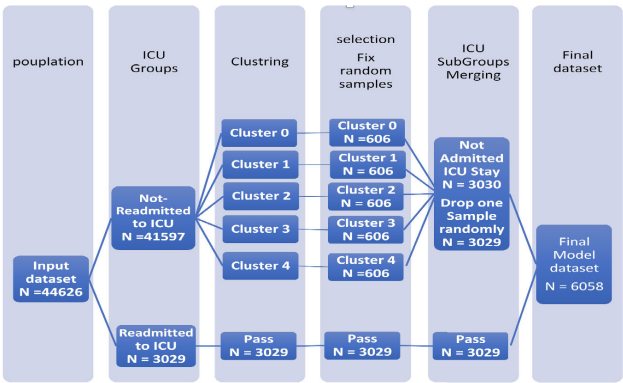



**FIGURE 3.** Clustering re-sampling.

### 2) CLASSIFICATION MODEL DEVELOPMENT

We built two variations of the readmission model, one using the baseline approach and one using the quantiles approach. We ran each model variation on both the imbalanced patient population before applying any re-sampling approaches and the balanced population after applying the re-sampling approaches described above.

We used supervised learning techniques because the model outcomes are labeled. In particular, we applied five common ML binary classification algorithms: 1) logistic regression (LR); 2) linear discriminant analysis (LDA); 3) random forest (RF); 4) k-nearest neighbors (KNN); and 5) support vector machine (SVM).

### 3) MODEL EVALUATION

For the imbalanced dataset, we randomly split the dataset into 75% training (N = 33,469) and 25% test set (N = 11,157). We then trained the readmission models using the training set and 10-fold cross-validation to avoid over-fitting and validated their performance using unseen data as a test set.

For the datasets resulting from the under-sampling and clustering approaches, we randomly split the datasets into training (N = 4,543) and test (N = 1,515) sets. We then trained the readmission models using the training sets and 10-fold cross-validation and measured the performance on an unseen test set from the same population. We used the accuracy along with 95% confidence interval (CI), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) as performance metrics.

### 4) HYPER-PARAMETER TUNING

We performed hyper-parameter tuning by running a grid search to select the best parameters in the different algorithms.

- For the under-sampling approach, we set the maximum number of features to consider for finding a good split in RF to four for models in both the baseline and quantiles approaches. We set the estimated number of trees in an RF to 300 for models in both the baseline and quantiles approaches. We used the radial basis function as a kernel type for SVM and set the penalty parameter of error, C, to 0.20 for models in the baseline approach and to 0.60 for models in the quantiles approach.
- For the clustering re-sampling approach, we set the estimated number of trees in an RF to 200 for models in both the baseline and the quantiles approaches and the number of features was selected automatically. We used the radial basis function as a kernel type for SVM, and set C to 0.10 in the baseline approach and to 0.50 in the quantiles approach.

## D. ABNORMALITY MODEL

The abnormality prediction reduces to a binary classification problem with two classes: normal (N = 19,620 stays) and abnormal (N = 25,006 stays). This results in a data with imbalanced classes, with 43.97% of ICU stays in the normal class compared to 56.04% of patients in the abnormal class. We randomly removed 5,386 samples from the abnormal patients' class to obtain a balanced classification with a total number of 39,240 ICU stays in both classes.

### 1) CLASSIFICATION MODEL DEVELOPMENT

We built two variations of the abnormality model, one using the baseline approach and the other using the quantiles approach. Since the model outcome is labeled, we ran supervised learning algorithms on each model variation. We applied the same five ML algorithms used in the readmission model: LR, LDA, RF, KNN, and SVM.

### 2) MODEL EVALUATION

To assess the goodness of fit in our classification models and validate them on an unseen test set from the same population, we compared the accuracy of the test set and the mean accuracy of the trained models along with corresponding 95% CI. We compared the sensitivity, specificity, PPV, and NPV on the test set. We also examined the difference in AUROC scores between the test and training sets.

### 3) HYPER-PARAMETER TUNING

For RF, we set the maximum number of features to consider for finding a good split in RF to four for the baseline and

quantiles approaches. We set the estimated number of trees in a random forest to 600 for both and the maximum number of features is four. For SVM, we used the radial basis function as a kernel type, and set the penalty parameter of error, C, to 0.20 in the baseline and to 0.60 in the quantiles approach.

### E. NEXT-DAY VITAL SIGN MODEL

Since the outcome variable for the next-day vital sign measurements prediction is numeric, the classification problem can be solved using a regression model. We built predictive models to predict the mean of the readings for each vital sign for the next day for each patient using only the baseline approach.

#### 1) SUB-POPULATION SELECTION

In this model, we considered vital signs for more than just the first day of the ICU stay. Here we extracted every day's vital signs after taking the mean of the readings for each sign. Also, since the goal is to predict the mean of each vital sign readings over several days, we included only patients who stayed in the ICU for more than one week. The total number of ICU stays corresponding to those patients was N = 7,438 (of the total N = 44,626 in Figure 2).

#### 2) REFERENCE TIME (BASELINE) SELECTION APPROACHES

We explored three variations of the next-day vital sign measurement prediction (the values considered for evaluating the model on the test set):

- Day-by-day approach: we considered readings from the most recent day to predict the readings for the following day. More formally, given $patient_j$, who stayed in the ICU for $k$ days, to predict vital sign measurements on $day_{k+1}$, we use vital sign measurements from $day_k$ as model inputs.
- Average measured approach: we considered long-term readings from all days in the most recent week to predict readings on the following day. More formally, given $patient_j$, who stayed in the ICU for $k$ days, to predict the vital sign measurements on $day_{k+1}$, we use the average of the measurements from $day_k$ to $day_1$ as model inputs. In our case, we considered measurements for 1 week.
- Error adjustment approach: we ignored readings from the distant past and focus on short-term readings from the most recent day along with any error introduced in predicting the readings for that day. More formally, given $patient_j$, who stayed in the ICU for $k$ days, to predict the vital sign measurements of $day_{k+1}$, we use the measurements of $day_k$ and the error between the actual and predicted values of $day_k$.

#### 3) CLASSIFICATION MODEL DEVELOPMENT

We used supervised ML algorithms since we rely on previously known vital sign measures. In particular, we used the extreme gradient boosting (XGB) and RF regression (RFR) algorithms to develop six next-day vital sign measure prediction models: three models that apply XGB on the three approaches discussed above and three models that apply RFR on those three approaches.

#### 4) MODEL EVALUATION

We split the ICU stay dataset (N = 7,438) into training (N = 5,578) and test (N = 1,860) sets, trained the models using the training set, and measured the performance on the test set.

We report the error between the predicted and actual values in the test set using the coefficient of variation of root mean squared error (CV-RMSE). We use the mean of dependent variable value to normalize the RMSE (CV = RMSE/ the mean of the real observations on the test set)

## V. RESULTS

### A. READMISSION MODEL

We present the results of the two patient population selection approaches in the readmission model.

#### 1) UNDER-SAMPLING OF MAJORITY CLASS

The accuracy of predicting readmission on the test set was 55.45% using RF and the quantiles approach and 55.25% using SVM and the quantiles approach. The highest improvement in accuracy of the readmission model using RF and the quantiles approach on the test set was 2.57%. RF also achieved the highest specificity (0.59), which indicates that the model using the RF algorithm and the quantiles approach can identify which patients will not be readmitted to the ICU better than the other algorithms. SVM achieved the highest sensitivity (0.56), which indicates that the model using the SVM algorithm and the quantiles approach can identify patients at risk of ICU readmission better than the other algorithms. The SVM algorithm using the quantiles approach also produced the highest AUROC (0.59) in predicting the ICU patients' readmission on the test set.

#### 2) CLUSTERING RE-SAMPLING APPROACH

The highest readmission model accuracy using the clustering re-sampling approach on the test set was 67.53% using the RF algorithm and the quantiles approach, while SVM and the quantiles approach achieved 64.10% accuracy (Table 3). When comparing the baseline and quantiles approaches, the improvement in accuracy on the test set was 6.03% using RF and 2.00% using SVM. This suggests that the clustering re-sampling approach achieved better improvement in model accuracy than did the under-sampling approach.

SVM achieved the highest sensitivity (0.758), which indicates that the model using SVM and the quantiles approach can identify patients at risk of readmission better than the other algorithms. RF, on the other hand, achieved the highest specificity (0.606), which indicates that the model using RF and the quantiles approach can identify patients who will not have ICU readmission risk better than the other algorithms.

We ran Cohen's kappa score function to express the level of agreement between the predicted and real outcome of interest

**TABLE 3.** Readmission model results using the clustering re-sampling approach and different performance metrics.

| Method | ML Alg. | Training Set Accuracy Mean (SD) | Test Set Accuracy (95% CI, ) | Test Set Sensitivity (95% CI, ) | Test Set Specificity (95% CI, ) | Test Set PPV (95% CI, ) | Test Set NPV (95% CI, ) |
|---|---|---|---|---|---|---|---|
| Baseline Approach | LR | 0.613 (0.019971) | 0.607 (0.577-0.657) | 0.690 (0.655-0.746) | 0.527 (0.502-0.571) | 0.585 (0.556-0.634) | 0.657 (0.605-0.690) |
| | LDA | 0.613 (0.019164) | 0.609 (0.579-0.659) | 0.702 (0.666-0.759) | 0.519 (0.494-0.563) | 0.586 (0.557-0.634) | 0.643 (0.611-0.696) |
| | RF | 0.613 (0.022577) | 0.615 (0.585-0.666) | 0.683 (0.649-0.739) | 0.549 (0.522-0.595) | 0.595 (0.565- 0.644) | 0.642 (0.610-0.694) |
| | KNN | 0.573 (0.018208) | 0.547 (0.520-0.593) | 0.580 (0.551-0.628) | 0.516 (0.491-0.559) | 0.537 (0.511- 0.581) | 0.559 (0.532-0.606) |
| | SVM | 0.615 (0.013745) | 0.621 (0.590-0.672) | 0.721 (0.684-0.779) | 0.525 (0.499-0.569) | 0.595 (0.565- 0.644) | 0.660 (0.627- 0.714) |
| Quantiles Approach | LR | 0.609 (0.021899) | 0.619 (0.588-0.670) | 0.699 (0.664-0.756) | 0.542 (0.515-0.587) | 0.588 (0.567-0.645) | 0.670 (0.618-0.704) |
| | LDA | 0.609 (0.019180) | 0.622 (0.591-0.673) | 0.707 (0.671- 0.765) | 0.539 (0.513-0.584) | 0.591 (0.568-0.647) | 0.673 (0.623-0.709) |
| | RF | 0.655 (0.027100) | 0.675 (0.641-0.730) | 0.748 (0.709-0.808) | 0.606 (0.576-0.656) | 0.648 (0.615-0.701) | 0.731 (0.677- 0.771) |
| | KNN | 0.571 (0.029664) | 0.568 (0.540-0.615) | 0.597 (0.568-0.647) | 0.540 (0.514-0.585) | 0.557 (0.530-0.603) | 0.615 (0.552-0.629) |
| | SVM | 0.616 (0.018538) | 0.641 (0.609-0.694) | 0.758 (0.719-0.819) | 0.529 (0.503-0.573) | 0.609 (0.579-0.659) | 0.693 (0.658-0.750) |

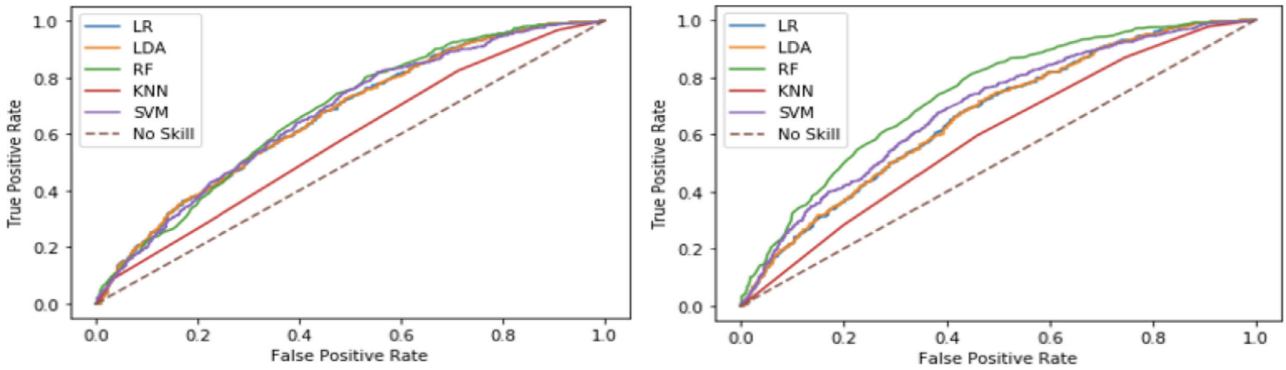



**FIGURE 4.** Comparison between the ROC curves for the different ML algorithms in the readmission model (with clustering re-sampling) using the baseline (left) and quantiles (right) approaches.

**TABLE 4.** Readmission model results using the clustering re-sampling approach and AUROC metric.

| Approach | ML Alg. | Training Set AUROC Mean (SD) | Test Set AUROC |
|---|---|---|---|
| Baseline | LR | 0.66566 (0.02682) | 0.66413 |
| | LDA | 0.66642 (0.02663) | 0.66415 |
| | RF | 0.65778 (0.02292) | 0.67330 |
| | KNN | 0.60113 (0.02491) | 0.57682 |
| | SVM | 0.66655 (0.01987) | 0.66634 |
| Quantiles | LR | 0.66443 (0.02443) | 0.66461 |
| | LDA | 0.66520 (0.02484) | 0.66447 |
| | RF | **0.71311 (0.02960)** | **0.73757** |
| | KNN | 0.59609 (0.03384) | 0.59292 |
| | SVM | 0.66283 (0.01999) | 0.68981 |

on the data set (observed and predicted for cases in the test set). We found that the highest score was 0.353 using the quantiles approach and the RF algorithm.

Table 4 shows the AUROC results of the readmission model on the training and test sets using the baseline and quantiles approaches and the five ML algorithms.

**TABLE 5.** Summary of the AUROC and accuracy metrics in the readmission model.

| Population Selection Solution | Accuracy (%) | AUROC | Improvement in AUROC (%) |
|---|---|---|---|
| Under-sampling (Reference) | 55.445 | 0.58563 | - |
| Clustering Re-sampling | 67.525 | 0.73757 | 25.944 |

Figure 4 depicts a comparison between the ROC curves for the five ML algorithms using the baseline (left) and quantiles (right) approaches. The figures show that RF had the best AUROC with both approaches, but it improved from 0.67 for the baseline approach to 0.74 using the quantiles approach.

Our two proposed population selection solutions show that the clustering re-sampling approach achieved the highest AUROC score compared to the under-sampling solution, indicating that the model using the clustering re-sampling approach was better at distinguishing between the positive and negative classes.

**TABLE 6.** Abnormality model results using the different performance metrics.

| Method | ML Alg. | Training Set Accuracy Mean (SD) | Test Set Accuracy (95% CI, ) | Test Set Sensitivity (95% CI, ) | Test Set Specificity (95% CI, ) | Test Set PPV (95% CI, ) | Test Set NPV (95% CI, ) |
|---|---|---|---|---|---|---|---|
| Baseline Approach | LR | 0.588753 (0.010767) | 0.59204 (0.588-0.600) | 0.594 (0.589-0.601) | 0.590 (0.586-0.598) | 0.594 (0.589-0.601) | 0.590 (0.586-0.598) |
| | LDA | 0.588923 (0.010611) | 0.59123 (0.587-0.599) | 0.592 (0.587-0.600) | 0.590 (0.586-0.598) | 0.593 (0.588-0.601) | 0.589 (0.585-0.597) |
| | RF | 0.640231 (0.008243) | 0.64097 (0.636-0.649) | 0.613 (0.608-0.621) | 0.669 (0.664-0.678) | 0.651 (0.646-0.659) | 0.632 (0.627-0.640) |
| | KNN | 0.563865 (0.006798) | 0.58797 (0.584-0.595) | 0.515 (0.511-0.521) | 0.662 (0.657-0.670) | 0.605 (0.601-0.613) | 0.575 (0.571-0.582) |
| | SVM | 0.641284 (0.004922) | 0.64128 (0.636-0.649) | 0.573 (0.568- 0.580) | 0.710 (0.705-0.719) | 0.666 (0.661-0.674) | 0.623 (0.618-0.631) |
| Quantiles Approach | LR | 0.623683 (0.010628) | 0.63221 (0.627-0.640) | 0.613 (0.608-0.621) | 0.651 (0.646- 0.660) | 0.637 (0.632-0.645) | 0.628 (0.623-0.636) |
| | LDA | 0.624057 (0.010506) | 0.63190 (0.627-0.640) | 0.610 (0.605-0.618) | 0.654 (0.649-0.662) | 0.638 (0.633-0.646) | 0.627 (0.622- 0.635) |
| | RF | 0.663982 (0.007873) | 0.67400 (0.669-0.683) | 0.657 (0.651-0.665) | 0.691 (0.686-0.700) | 0.680 (0.675-0.689) | 0.668 (0.663-0.677) |
| | KNN | 0.592423 (0.007988) | 0.60305 (0.598-0.611) | 0.491 (0.487-0.497) | 0.715 (0.709- 0.724) | 0.633 (0.628- 0.641) | 0.584 (0.580-0.592) |
| | SVM | 0.661332 (0.007580) | 0.66860 (0.663-0.677) | 0.614 (0.610-0.622) | 0.723 (0.717-0.732) | 0.689 (0.684- 0.698) | 0.652 (0.647-0.661) |

Table 5 summarizes the AUROC and accuracy results in each of the two solutions and the corresponding improvement in AUROC for the clustering re-sampling approach over the under-sampling approach.

### B. ABNORMALITY MODEL

The accuracy of predicting abnormality on the test set with the quantiles approach was 67.40% using the RF algorithm and 66.86% using the SVM algorithm. Table 6 shows the abnormality model performance on both the training and test sets using the baseline and quantiles approaches and the different ML algorithms.

The best improvement in model accuracy on the test set was 3.30% using RF, while it was 2.73% using SVM. The RF algorithm achieved the highest sensitivity (0.657) using the quantiles approach, which indicates that the model using the RF algorithm and the quantiles approach can identify abnormal patients better than the other algorithms. SVM achieved the highest specificity (0.723), which indicates that the model using the SVM algorithm and the quantiles approach can identify normal patients better than the other algorithms.

We found that the highest Cohen's kappa score was 0.348 using the quantiles approach and the RF algorithm.

Table 7 shows the AUROC results of the abnormality model on both the training and test sets using the baseline and quantiles approaches for the different ML algorithms. Figure 5 shows the ROC curves for the algorithms in the baseline approach and the quantiles approach, respectively. The RF algorithm using the quantiles approach produced the highest AUROC (0.74).

### C. NEXT DAY VITAL SIGN MODEL

We present the detailed results for the RFR algorithm since it out-performed XGB across all three approaches. The bottom rows of Tables 8, 9,and 10 show the mean error for each vital sign using the CV-RMSE metric after applying the RF Regression algorithm using the three approaches for next-day vital sign measurements.

**TABLE 7.** Abnormality model results using the AUROC metric.

| Method | ML Alg. | Training Set AUROC Mean (SD) | Test Set AUROC |
|---|---|---|---|
| Baseline Approach | LR | 0.632925 (0.012502) | 0.62599 |
| | LDA | 0.632921 (0.012481) | 0.62600 |
| | RF | 0.691979 (0.009579) | 0.68441 |
| | KNN | 0.614272 (0.009072) | 0.61444 |
| | SVM | 0.693203 (0.008545) | 0.68854 |
| Quantiles Approach | LR | 0.680267 (0.007371) | 0.68321 |
| | LDA | 0.680126 (0.007335) | 0.68322 |
| | RF | **0.723461 (0.008198)** | **0.73794** |
| | KNN | 0.620266 (0.007058) | 0.63874 |
| | SVM | 0.722019 (0.006709) | 0.73327 |

Figure 6 depicts a visual representation of the error comparison curves using the three approaches and the RF Regression algorithm for the seven vital signs.

Overall, the RFR algorithm produced less error, on average, in all three approaches compared to XGB. While the average errors for the three approaches are very close, the average measured approach produced the highest error among all three. The error adjustment approach achieved the lowest mean error in diastolic BP, respiration rate, body temperature, SpO2, and glucose level, while the day-by-day approach achieved the lowest mean error in heart rate and systolic blood pressure.

These findings indicate that RFR performed better than XGB, and that the error adjustment approach performed best while the average measured approach performed worst among all three approaches. This may indicate two things: i) the distant past does not help much in predicting values

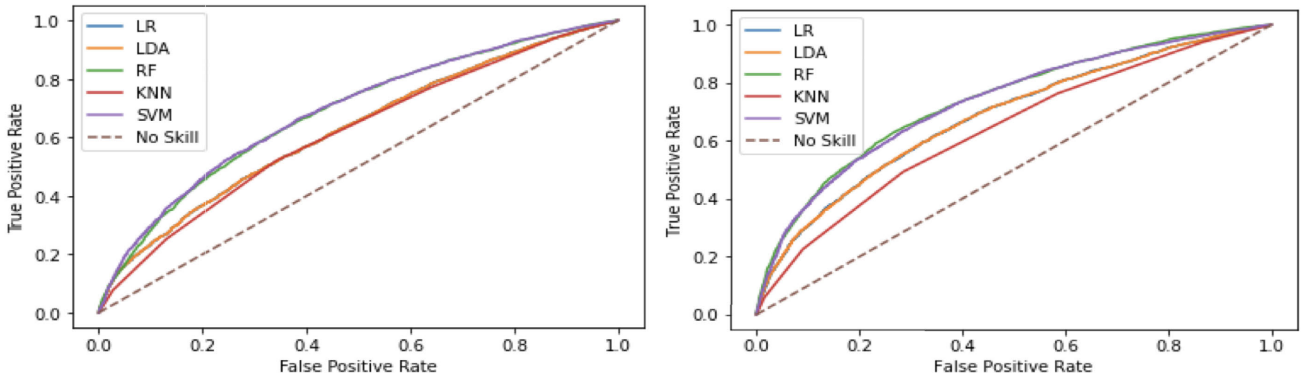



**FIGURE 5.** Comparison of the ROC curves in the abnormality model using the baseline (left) and quantiles (right) approaches.

**TABLE 8.** CV-RMSE results using the day-by-day approach and the RFR algorithm.

| | Heart Rate | Systolic BP | Diastolic BP | Respiration Rate | Body Temp | SpO2 | Glucose Level |
|---|---|---|---|---|---|---|---|
| Day 2 | 12.21 | 9.30 | 11.10 | 16.98 | 1.46 | 1.93 | 25.01 |
| Day 3 | 11.28 | 8.88 | 10.29 | 15.46 | 1.32 | 1.53 | 23.30 |
| Day 4 | 10.95 | 8.86 | 9.95 | 15.02 | 1.27 | 1.69 | 24.45 |
| Day 5 | 9.94 | 8.77 | 10.13 | 15.02 | 1.20 | 1.39 | 20.81 |
| Day 6 | 9.87 | 8.59 | 10.14 | 13.84 | 1.19 | 1.40 | 19.93 |
| Day 7 | 10.19 | 8.48 | 9.93 | 14.15 | 1.19 | 1.49 | 20.66 |
| Arerage CV-RMSE | 10.74 | 8.81 | 10.25 | 15.08 | 1.27 | 1.57 | 22.36 |

**TABLE 9.** CV-RMSE results using the average measured approach and the RFR algorithm.

| | Heart Rate | Systolic BP | Diastolic BP | Respiration Rate | Body Temp | SpO2 | Glucose Level |
|---|---|---|---|---|---|---|---|
| Day 2 | 12.21 | 9.30 | 11.10 | 16.98 | 1.46 | 1.93 | 25.01 |
| Day 3 | 12.16 | 9.29 | 10.79 | 16.83 | 1.39 | 1.58 | 23.35 |
| Day 4 | 12.07 | 9.36 | 10.59 | 16.26 | 1.35 | 1.75 | 25.23 |
| Day 5 | 10.59 | 9.30 | 10.65 | 16.33 | 1.26 | 1.50 | 22.38 |
| Day 6 | 10.55 | 8.99 | 10.60 | 14.97 | 1.24 | 1.43 | 20.42 |
| Day 7 | 10.94 | 9.08 | 10.45 | 14.94 | 1.24 | 1.60 | 20.88 |
| Arerage CV-RMSE | 11.42 | 9.22 | 10.69 | 16.05 | 1.32 | 1.63 | 22.88 |

**TABLE 10.** CV-RMSE results using the error adjustment approach and the RFR algorithm.

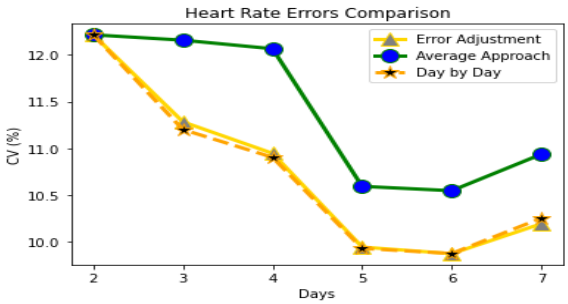
| | Heart Rate | Systolic BP | Diastolic BP | Respiration Rate | Body Temp | SpO2 | Glucose Level |
|---|---|---|---|---|---|---|---|
| Day 2 | 12.21 | 9.30 | 11.10 | 16.98 | 1.46 | 1.93 | 25.01 |
| Day 3 | 11.20 | 8.98 | 10.39 | 15.54 | 1.32 | 1.49 | 23.45 |
| Day 4 | 10.90 | 8.90 | 9.95 | 15.08 | 1.28 | 1.82 | 24.82 |
| Day 5 | 9.93 | 8.76 | 10.17 | 15.03 | 1.20 | 1.40 | 20.88 |
| Day 6 | 9.88 | 8.60 | 10.18 | 13.92 | 1.20 | 1.43 | 20.12 |
| Day 7 | 10.25 | 8.47 | 9.92 | 14.17 | 1.21 | 1.92 | 20.74 |
| Arerage CV-RMSE | 10.73 | 8.84 | 10.28 | 15.12 | 1.28 | 1.66 | 22.50 |

for the next day vital sign measurements as much as the near future does; ii) considering the error of the previous day helps reduce the prediction error for the next day.

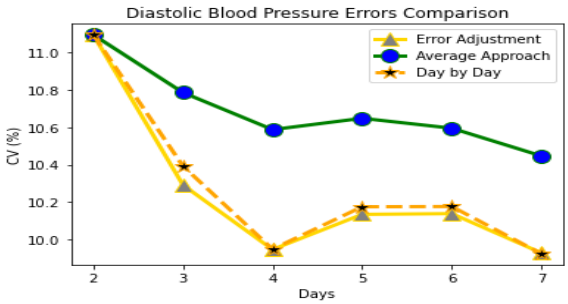
## VI. DISCUSSION

There are other approaches for sampling, which we could have applied in the readmission model to solve the data with imbalanced classes. For instance, the synthetic minority over-sampling technique (SMOTE) [20] is one method for solving data with imbalanced classes. SMOTE will over-sample the minority class by generating synthetic instances. However, SMOTE suffers from the problem of over-generalization because it propagates the minority class region and readmits it to the ICU class without considering the majority class, which is not-readmitted to the ICU class. However, we built our solution (the clustering re-sampling approach) to solve the data with imbalanced classes in the readmission model using k-means to maintain a variety of the population without duplication and without using under- or over-sampling techniques.

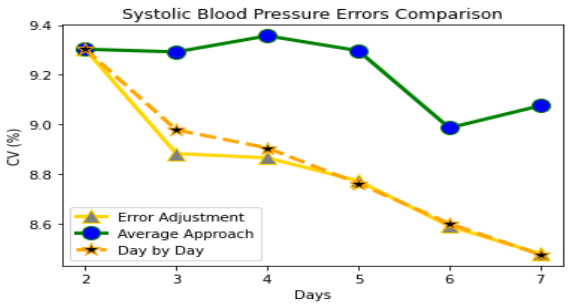
In the abnormality model, we relied on three criteria to define abnormality. It would be worthwhile to explore the impact of adding other relevant abnormality indicators such as cardiac problems and organ disorders. The patient abnormality model might improve if we included more models in the IICUPM module.



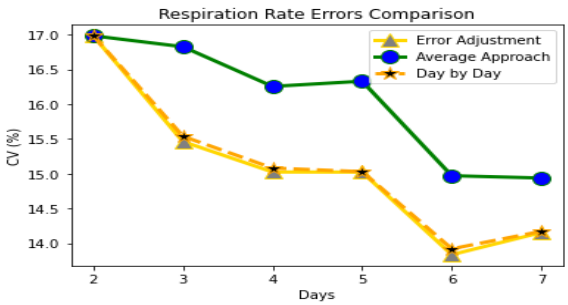

(a) Error in predicting next day vital sign, heart rate



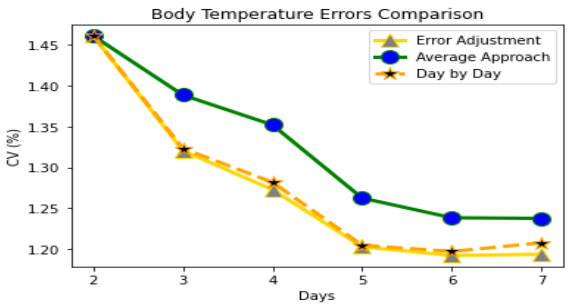

(b) Error in predicting next day vital sign, systolic blood pressure



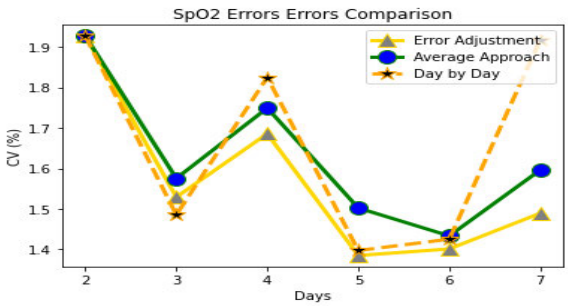

(c) Error in predicting next day vital sign, diastolic blood pressure



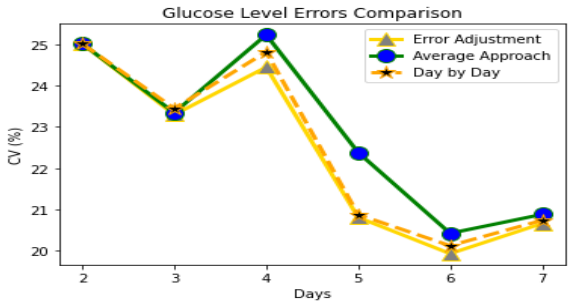

(d) Error in predicting next day vital sign, respiration rate




(e) Error in predicting next day vital sign, body temperature




(f) Error in predicting next day vital sign, SpO2




(g) Error in predicting next day vital sign, glucose level

**FIGURE 6. Comparison between the error produced by the three next-day vital sign measurement prediction approaches using the RFR algorithm for the seven vital signs.**

### 1) QUALITATIVE COMPARISON WITH OTHER APPROACHES

We compared the performance of our models to those of other researchers. For the readmission model, we achieved 67.53% accuracy and an AUROC of 0.74 using only seven vital sign features, four demographic attributes, and 21 features engineered from those original features. Other researchers used excessively more features to achieve similar or comparable accuracy. For instance, Lin *et al.* [11] used 17 patient EHR events, four demographic features, and the International Classification of Diseases, Ninth Revision (ICD-9) code to achieve an AUROC of 0.79. We predict patient readmission without prior knowledge about their medical conditions

or diagnoses. Fialho *et al.* [12] used six monitoring signals, 15 laboratory tests, and urine output to build a model with data with imbalanced classes and achieved an AUROC of 0.72. We used data from only 11 features from the first day of patients' ICU stay and we provide two solutions to handle the class imbalance problem. We include imbalanced techniques to produce a balanced classification with the same total number of samples in both the readmitted and non-readmitted patients' classes. Moreover, we achieved an AUROC of 0.73757. Rajkomar *et al.* [7] developed a DL model to predict 30-day unplanned hospital readmissions and achieved an AUROC between 0.75 and 0.76. In our study, we predict ICU readmission during the same hospitalization rather than readmission after hospital discharge.

Pakbin *et al.* [15] developed different imbalanced models for predicting ICU readmission at various time points using patient data before ICU discharge. They used all data available from MIMIC-III, including ICD-9 admission diagnosis codes. They achieved an AUROC of 0.76 for risk of ICU readmission after discharge at 72 hours. They achieved an AUROC of 0.84 for risk of ICU readmission within a single hospital admission. In our study, we built a balanced ICU readmission model with fewer features.

#### 2) LIMITATIONS

Our approach, like other ML approaches, has a generalizability limitation. In this study, we trained our models based on only the MIMIC database, which represents a single hospital in Boston, MA. Had we applied the models to various patient data from different demographic backgrounds and locations, we could have obtained different results. Also, all prognostic models developed in clinical settings are prone to producing false positives, and our approach is no exception. The method often used to mitigate that in clinical settings is by involving a human-in-the loop by usually having a critical care specialist manually provide a qualitative review of the false positive alerts.

#### 3) CLINICAL IMPLICATIONS

The focus of this work is to extend the IICUPM module functionalities by incorporating three more prediction models. After exploring the best available approach for achieving our clinical goals (i.e., abnormality, readmission, and next day vital sign risk stratification and prediction), we use our findings as a way to choose the best approaches in our IRPM framework. We incorporated our prediction models into a cloud-based Intelligent Remote Patient Monitoring (IRPM) framework. By integrating the predictive functionalities of the IICUPM module into existing decision support systems already used in clinical workflows we may provide significant practical implications for cost reduction and quality of care improvement.

#### 4) TIME COMPLEXITY

Our feature engineering algorithm runs in $\theta(P.V_{base}.S.v_k)$ time since it requires four nested loops: a loop through all patients ($P$); a loop through the ICU Stays $S$ of each patient; a loop through each vital sign in $V_{base}$; and a loop through each observation $Obs_i$ of each vital sign $v_k$. Thus, the algorithm heavily depends on the population size and the number of observations for each vital sign. However, our algorithm has several assumptions. First, we always focus on only seven vital signs, $V_{base}$ is always constant. Also, the number of ICU stays in our case is always the same, on average. Therefore, the time complexity can be minimized to $\theta(P.v_k)$. Finally, we mitigate the overhead caused by the number of vital sign observations by pre-processing observations within each vital sign feature (Section III-C).

## VII. CONCLUSION

Machine learning approaches applied to clinical datasets offer great promise for the delivery of personalized medicine for targeted treatment of human disease. Building on top of our recent development of an intelligent ICU patient monitoring (IICUPM) framework, we provide three reproducible risk stratification ML models. Our findings indicate that we can build balanced prediction models for ICU patient readmission and abnormality with better accuracy using a combination of ML and a quantiles approach that relied on only vital signs. To avoid inaccurate results and poor accuracy in the readmission model, we proposed two solutions for the data with imbalanced classes: one that uses the under-sampling method and one that uses the clustering re-sampling method. We also provide three approaches for selecting predictors of next-day vital sign measurements in reference to a baseline. We built regression models using two different classification algorithms to each approach. In general, we found that the error adjustment approach performed best while the average measured approach performed the worst. The result indicates that, generally, using the most recent vital sign measurements achieves the least error especially when we account for previous errors. In addition to providing three transparent and reproducible ML models, this work contributes a feature engineering algorithm that can be deployed in different critical care settings.

## CONFLICT OF INTEREST

No conflicts of interest were detected.

## ACKNOWLEDGMENT

The authors would like to thank the UTHSC Office of Scientific Writing for editing of the manuscript.

## REFERENCES

[1] A. Shaban-Nejad, M. Michalowski, N. Peek, J. S. Brownstein, and D. L. Buckeridge, "Seven pillars of precision digital health and medicine," *Artif. Intell. Med.*, vol. 103, Mar. 2020, Art. no. 101793.

[2] G. Gutierrez, "Artificial intelligence in the intensive care unit," *Crit. Care*, vol. 24, no. 1, p. 101, 2020, doi: 10.1186/s13054-020-2785-y.

[3] A. Shaban-Nejad, H. Mamiya, A. Riazanov, A. J. Forster, C. J. O. Baker, R. Tamblyn, and D. L. Buckeridge, "From cues to nudge: A knowledge-based framework for surveillance of healthcare-associated infections," *J. Med. Syst.*, vol. 40, no. 1, pp. 1–12, Jan. 2016.

[4] T. J. Pollard and L. A. Celi, "Enabling machine learning in critical care," *ICU Manage. Pract.*, vol. 17, no. 3, p. 198, 2017.
[5] D. Plant and A. Barton, "Machine learning in precision medicine: Lessons to learn," *Nature Rev. Rheumatol.*, vol. 17, no. 1, pp. 5–6, Jan. 2021.
[6] K. Alghatani, N. Ammar, A. Rezgui, and A. Shaban-Nejad, "Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation," *JMIR Med. Informat.*, vol. 9, no. 5, May 2021, Art. no. e21347.
[7] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digit. Med.*, vol. 1, no. 1, p. 18, 2018.
[8] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, May 2016, Art. no. 160035.
[9] T. Pollard and A. Johnson, III. (2016). *The MIMIC-III Clinical Database*. [Online]. Available: https://doi.org/10.13026/c2xw26
[10] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
[11] Y.-W. Lin, Y. Zhou, F. Faghri, M. J. Shaw, and R. H. Campbell, "Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0218942.
[12] A. S. Fialho, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, and S. N. Finkelstein, "Data mining using clinical physiology at discharge to predict ICU readmissions," *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13158–13165, Dec. 2012.
[13] E. K. Shin, R. Mahajan, O. Akbilgic, and A. Shaban-Nejad, "Sociomarkers and biomarkers: Predictive modeling in identifying pediatric asthma patients at risk of hospital revisits," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–5, Dec. 2018.
[14] S. N. Golmaei and X. Luo, "DeepNote-GNN: Predicting hospital readmission using clinical notes and patient network," in *Proc. 12th ACM Conf. Bioinf., Comput. Biol., Health Informat.*, Aug. 2021, pp. 1–9.
[15] A. Pakbin, P. Rafi, N. Hurley, W. Schulz, M. H. Krumholz, and J. B. Mortazavi, "Prediction of ICU readmissions using data at patient discharge," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 4932–4935.
[16] N. Momenzadeh, H. Hafezalseheh, M. R. Nayebpour, M. Fathian, and R. Noorossana, "A hybrid machine learning approach for predicting survival of patients with prostate cancer: A SEER-based population study," *Informat. Med. Unlocked*, vol. 27, Jan. 2021, Art. no. 100763.
[17] L. Akter, M. Raihan, M. Raihan, M. Sarker, M. Ghosh, N. Alvi, and Ferdib-Al-Islam, "Breast cancer risk prediction using different clustering techniques," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer, 2022, pp. 191–203.
[18] O. Lo, L. Fan, W. Buchanan, and C. Thuemmler, "Technical evaluation of an e-health platform," in *Proc. ADIS Int. Conf. e-Health*, Lisbon, Portugal, Jul. 2012, pp. 21–28.
[19] G. Menardi and N. Torelli, "Training and assessing classification rules with imbalanced data," *Data Mining Knowl. Discovery*, vol. 28, no. 1, pp. 92–122, Jan. 2014.
[20] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2009, pp. 475–482.

**KHALID ALGHATANI** (Member, IEEE) received the Ph.D. degree in computer science from the New Mexico Institute of Mining and Technology, Socorro, NM, USA. He is currently an Artificial Intelligence Specialist at King Fahad Medical City, Saudi Arabia. His research interests include artificial intelligence, medical informatics, and data analytics.

**NARIMAN AMMAR** (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from Wayne State University, Detroit, MI, USA. She is currently a Postdoctoral Research Scholar at the UTHSC-OAK-Ridge National Laboratory (ORNL), Center for Biomedical Informatics, Department of Pediatrics, The University of Tennessee Health Science Center (UTHSC), Memphis, TN, USA. Before joining UTHSC, she was a fixed-term Assistant Professor with the Department of Computer Science, Birzeit University, Palestine. Her primary research interests include designing explainable artificial intelligence digital solutions for healthcare using techniques from the data science and computer science domains, including data engineering, machine learning, graph representation learning, and data privacy.

**ABDELMOUNAAM REZGUI** received the Ph.D. degree in computer science from Virginia Tech, Blacksburg, VA, USA. He is currently an Associate Professor with the School of Information Technology, Illinois State University, Normal, IL, USA. Research conducted by him and his students was invited for presentation at selective events organized by major organizations, including NSF, Google, and Siemens. He has authored or coauthored over 100 articles in top journals and conferences, including the IEEE Transactions on Big Data (TBD), the IEEE Transactions on Knowledge and Data Engineering (TKDE), *ACM Transactions on Internet Technology* (TOIT), the IEEE Transactions on Parallel and Distributed Systems (TPDS), the IEEE Internet Computing, the IEEE Security and Privacy, the IEEE ICDE, the IEEE IC2E, and the IEEE Cloud Computing. His research has been funded by several institutions, including NASA and NSF. His research interests include health informatics, data analytics, graph algorithms, networking, cloud computing, and big data. He regularly serves on the program committees of several major conferences, including the IEEE Big Data, the IEEE BigDataSE, the IEEE GLOBECOM, the IEEE CloudNet, the IEEE LCN, and IoTBDS. He has also been a Proposal Reviewer and a Panelist of NSF and QNRF. He has been the track chair, a keynote speaker, or a tutorial presenter at several international conferences. He is on the Editorial Board of several journals, including *Big Data Analytics* (Springer).

**ARASH SHABAN-NEJAD** received the M.Sc. and Ph.D. degrees in computer science from Concordia University, Montreal, Canada, and the M.P.H. degree from the University of California at Berkeley, Berkeley, CA, USA. He is currently an Associate Professor with the UTHSC-OAK-Ridge National Laboratory (ORNL), Center for Biomedical Informatics, and the Department of Pediatrics, The University of Tennessee Health Science Center (UTHSC), Memphis, TN, USA. He is also an Adjunct Faculty at The Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee at Knoxville, Knoxville, TN, USA. Before coming to UTHSC, he was a Postdoctoral Fellow at the McGill Clinical and Health Informatics Group, McGill University, Montreal, QC, Canada. He accrued additional training at the Harvard School of Public Health, Boston, MA, USA. His primary research interests include population health intelligence, epidemiologic surveillance, explainable AI and big-data semantic analytics using tools and techniques from artificial intelligence, knowledge representation, and semantic web.