applied sciences

MDPI

Article

Integrating Structured and Unstructured Data with BERTopic and Machine Learning: A Comprehensive Predictive Model for Mortality in ICU Heart Failure Patients

Shih-Wei Wu [1], Cheng-Cheng Li [2], Te-Nien Chien [2,*] and Chuan-Mei Chu [2]

[1] Department of Business Management, National Taipei University of Technology, Taipei 106, Taiwan; swu@ntut.edu.tw
[2] College of Management, National Taipei University of Technology, Taipei 106, Taiwan; chengchengli0006@gmail.com (C.-C.L.); sherchu66@gmail.com (C.-M.C.)
* Correspondence: tenienchien@gmail.com; Tel.: +886-2-2771-2171 (ext. 3403)

Abstract: Heart failure remains a leading cause of mortality worldwide, particularly within Intensive Care Unit (ICU)-patient populations. This study introduces an innovative approach to predicting ICU mortality by seamlessly integrating electronic health record (EHR) data with a BERTopic-based hybrid machine-learning methodology. The MIMIC-III database serves as the primary data source, encompassing structured and unstructured data from 6606 ICU-admitted heart-failure patients. Unstructured data are processed using BERTopic, complemented by machine-learning algorithms for prediction and performance evaluation. The results indicate that the inclusion of unstructured data significantly enhances the model's predictive accuracy regarding patient mortality. The amalgamation of structured and unstructured data effectively identifies key variables, enhancing the precision of the predictive model. The developed model demonstrates potential in improving healthcare decision-making, elevating patient outcomes, and optimizing resource allocation within the ICU setting. The handling and application of unstructured data emphasize the utilization of clinical narrative records by healthcare professionals, elevating this research beyond the traditional structured data predictive tools. This study contributes to the ongoing discourse in critical care and predictive modeling, offering valuable insights into the potential of integrating unstructured data into healthcare analytics.

Keywords: electronic health records; heart failure; BERTopic; machine learning; healthcare

check for updates

Citation: Wu, S.-W.; Li, C.-C.; Chien, T.-N.; Chu, C.-M. Integrating Structured and Unstructured Data with BERTopic and Machine Learning: A Comprehensive Predictive Model for Mortality in ICU Heart Failure Patients. *Appl. Sci.* **2024**, *14*, 7546. https://doi.org/10.3390/app14177546

Academic Editor: Zhi-Ting Ye

Received: 19 July 2024
Revised: 12 August 2024
Accepted: 21 August 2024
Published: 26 August 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Heart failure (HF), which is a prevalent cardiovascular ailment, leads to the majority of CVD and mortality resulting from the heart's inability to maintain the body's metabolism [1]. HF, which represents the end stage of various cardiovascular diseases, is a significant healthcare concern. The spread of COVID-19 worldwide has exacerbated matters, as HF has emerged as a severe complication that is associated with a poor prognosis and mortality from COVID-19 [2]. HF, a prominent contributor to mortality, affects approximately 6.2 million individuals in the United States, resulting in over 800,000 hospital admissions annually. Hospitalization for HF is associated with an alarming 36% annual mortality rate. Patients diagnosed with HF face an increased risk, with 30–50% experiencing mortality within five years of diagnosis [3]. Despite notable advancements in cardiac care, the prevalence of HF has substantially risen, contributing to persistent high mortality rates and a subsequent escalation in medical expenditures for this condition. Individuals with HF manifest identifiable symptoms, including shortness of breath, peripheral edema, ankle swelling, body fatigue, elevated jugular venous pressure, and pulmonary crackles attributed to either cardiac or noncardiac structural abnormalities . Despite improvements in cardiac care, the ongoing surge in HF cases underscores the urgent need for effective interventions to mitigate mortality rates and alleviate the associated economic burden.

The Electronic Health Records (EHRs) are a set of electronic personal health records, including medical records, ECG, medical imaging, etc. The EHR systems have been implemented all over the world due to their continuous up-to-date updates. The availability of electronic clinical data has increased dramatically due to developments in information technology and the use of EHRs. However, a wealth of overlooked patient-related information exists in EHRs, and such information, in part, helps predict patient outcomes [4]. EHRs are the digital version of paper charts. In the past, many researchers used EHR database data to predict patient mortality, hospital admission time, disease diagnosis, disease onset, etc., in order to prevent and intervene early diseases of patients that are critical to intensive care. As a fundamental risk-assessment tool, predictive models have been developed and applied in many areas of healthcare. Widespread implementation of EHRs has sparked a revolution bringing individual patient demographics, genetic profiles, medical, diagnostic records, laboratory results, and image data into electronic formats to facilitate access and use in "big data" research investigations. The feasibility of truly personalized and precision medicine depends on the management of the vast amount of available EHR data that are critical to the development of these models [5]. Accurate mortality prediction for ICU patients with heart failure (HF) is vital. It helps medical teams develop effective care plans, prioritize high-risk patients, and optimize resource allocation. It also provides crucial information for patients and their families about disease severity, aiding informed decision-making. Additionally, it supplies valuable data for HF research, advancing treatment methods and preventive measures. Overall, such predictions enhance patient care, resource management, patient education, and research efforts [6,7].

According to the World Federation of Societies of Intensive and Critical Care Medicine, an ICU is an organized system for the provision of care to critically ill patients that provides intensive and specialized medical and nursing care, an enhanced capacity for monitoring, and multiple modalities of physiologic organ support to sustain life during a period of life-threatening organ-system insufficiency [8]. In the United States, 40% of people die during hospitalization, and approximately 22% of patients spend their entire hospital stay in the ICU [9]. As a result, the healthcare system is under a heavy burden, which may affect the criteria for ICU admission, the interventions used, and the duration, all of which may affect the patient's prognosis [10]. The harnessing of big data for clinical and basic research analyses and applications to improve human well-being and health, such as the combination of big data and artificial intelligence, can assist physicians in diagnosing and treating diseases and improving the quality of care. Predictive models have been developed over the last few decades as important risk-assessment tools and are utilized in a variety of healthcare settings [11].

While traditional research in intensive-care settings has predominantly focused on quantitative variables to predict patient mortality, the utility of qualitative data is increasingly acknowledged. Current statistical models are primarily oriented towards manipulating quantitative data; however, significant challenges remain in standardizing and effectively utilizing qualitative data [12]. A wealth of patient information, particularly social and behavioral factors such as social isolation, stress, and psychological complexity, is stored in the unstructured data fields of the Electronic Health Records (EHRs). Recent advancements in artificial intelligence, especially through enhanced natural language-processing systems, have proven effective at extracting symptom information and differentiating between observed symptoms and those negated within EHR narratives [13,14]. The integration of AI into clinical medicine has broadened to include applications such as preclinical data processing, point-of-care diagnostic assistance, patient stratification, treatment decision-making, and early warning systems for primary and secondary prevention, reflecting a comprehensive approach to patient care. This capability not only enhances the accuracy and robustness of predictive algorithms but also underscores the importance of integrating both clinical and non-clinical factors related to patients' everyday lives in predicting outcomes [15,16].

However, processing the unstructured data that constitute medical big data takes a lot of time and money. This is especially true for the numerical portion, which is often an important component, presented in the bulk of clinical notes during treatment and hospitalization. By taking unstructured data as input, we are not using raw physiological data, but the perception and judgment of medical and nursing professionals in free-text annotations. These allow us to access higher-level concepts that are not present in physiological data. Egger et al. used the BERTopic to evaluate, alongside other topic modeling algorithms, to determine its effectiveness in analyzing Twitter data related to travel and the COVID-19 pandemic. Based on the results of the evaluation, BERTopic was found to be one of the most effective algorithms for this type of analysis [17]. Franco et al. proposed a machine learning-based approach to improve the classification of cognitive distortions in Arabic content over Twitter. The proposed approach enriches text representation by defining the latent topics within tweets using a transformer-based topic-modeling technique called BERTopic [18]. Both BERTopic and Latent Dirichlet Allocation (LDA) are topic-modeling algorithms that can convert text data into topic vectors. However, BERTopic uses the BERT model, while LDA uses a probabilistic model [19]. Modeling LDA for free text topics has some shortcomings. For instance, when LDA processes long texts, it splits them into multiple short texts, leading to information loss and dispersion. LDA only considers the frequency of word occurrences and lacks an understanding of contextual relationships among words. To use LDA, one needs to perform multiple cross-validation and parameter adjustments to obtain better results. Overall, BERTopic outperforms LDA in topic modeling, especially in long text and contextual understanding [20].

The objective of this study was to develop a comprehensive prediction model for mortality in ICU heart-failure patients by integrating both structured and unstructured data. Leveraging artificial intelligence, specifically BERTopic, for the analysis of unstructured data, this approach shows significant promise for enhancing the accuracy of predictive models. By combining structured and unstructured data through machine-learning methods, this study aims to equip healthcare professionals with more precise mortality-prediction tools, ultimately improving treatment outcomes and medical decision-making. The incorporation of unstructured data into health research facilitates a deeper understanding of patient vital signs, personal backgrounds, and diagnostic reports from medical staff. This research endeavors to introduce a more comprehensive and insightful methodology to medical research, thereby contributing to the ongoing advancement of predictive models in healthcare.

2. Materials and Methods

In our study, we initiated data collection by obtaining quantitative information from patients in ICU experiencing congestive HF, including vital signs and basic details. Textual data were sourced from post-admission medical records. Subsequently, we performed data filtering and preprocessing to ensure data quality and consistency. For textual data analysis, BERTopic technology was employed to generate topics, enhancing the comprehension of patients' textual descriptions. Leveraging six commonly used machine-learning methods, namely GaussianNB, AdaBoost, Bagging, Gradient Boosting, MLP Classifier, LightGBM, and XGBoost, we integrated quantitative and textual data to predict the mortality rate of ICU patients with HF. Finally, model performance was evaluated using five metrics: precision, recall, F1-score, accuracy, and AUROC. Our research framework, combining diverse data analysis and machine-learning techniques, aims to establish a predictive model for assessing the mortality rate of ICU patients with congestive HF, offering valuable insights for clinical decision-making and treatment strategies. The research framework is shown in Figure 1.

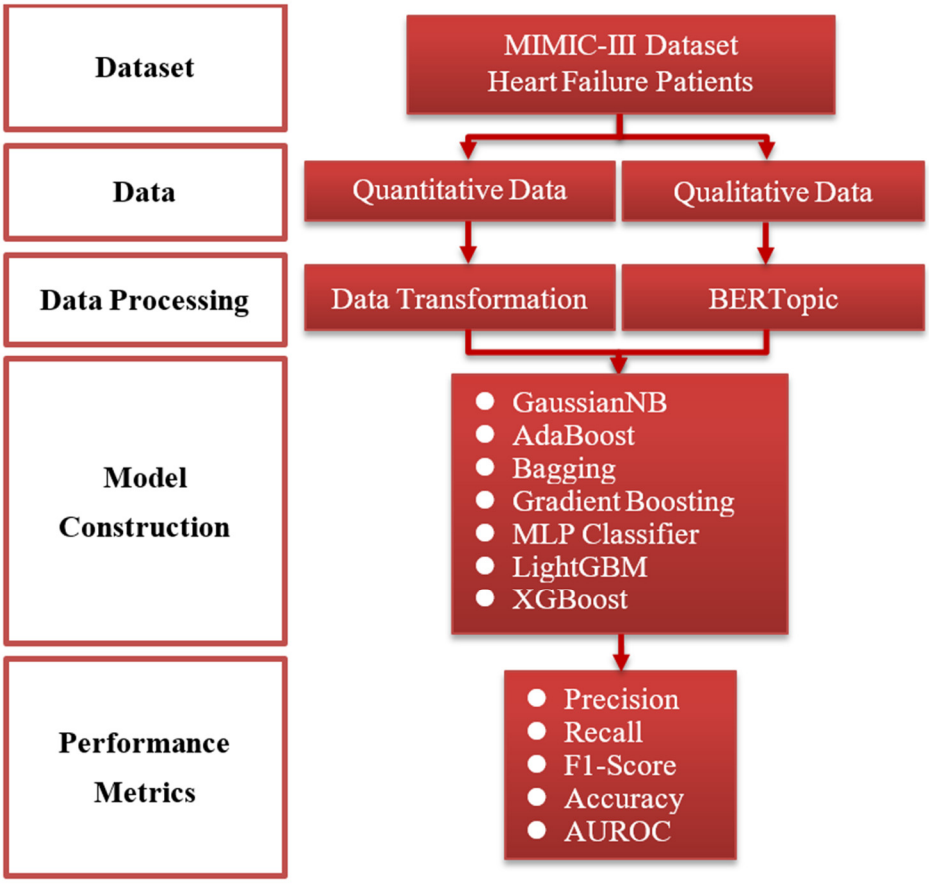


Figure 1. The detailed process of data extraction.

2.1. Data Collection and Preprocessing

Our investigation utilized the Medical Intensive Care Information Mart (MIMIC-III) database, renowned for its extensive clinical data capturing patients admitted to Beth Israel Deaconess Medical Center in Boston, MA. Acknowledged as a principal repository of de-identified public data for research purposes [21], MIMIC-III encompasses 26 CSV-formatted tables meticulously documenting various aspects of patient data during ICU treatment. This comprehensive dataset encompasses laboratory test results, demographic attributes, microbiological findings, hemodynamic parameters, treatment regimens, and fluid balance, among other parameters. The wealth of clinical information embedded in MIMIC-III renders it an invaluable asset for medical research. Within the scope of this study, we leveraged MIMIC-III to analyze 46,520 patient-specific data elements and 58,976 admission-related data points, encompassing vital signs, medication records, laboratory metrics, and observational entries. Ethical approval for accessing the MIMIC-III database was secured through the completion of the National Institutes of Health (NIH) online course, successful passage of the Human Research Participant Protection Examination, and the submission of an access request (certificate number: 35628530). These measures were implemented to ensure the incorporation of appropriate safeguards for all data employed in this research study.

This study aimed to prognosticate mortality among patients with HF by leveraging physiological variables sourced from the MIMIC-III database. The selection of predictor variables was informed by prior investigations focused on mortality prediction in HF patients. The considered variables encompassed vital indicators, such as heart rate, respiratory rate, diastolic and systolic blood pressure, body temperature, oxygen saturation, inspired oxygen fraction, blood urea nitrogen, creatinine, mean blood pressure, blood glucose, white blood cells, red blood cells, prothrombin time, International Normalized Ratio, platelets, Glasgow Coma Scale (GCS) scores for eyes, movement, and language [7,22]. Additionally, the inclusion of comprehensive social and environmental context was augmented

by incorporating patient vital signs and personal background information, including age, gender, insurance type, marital status, religion, and race (as detailed in Appendix A). By exclusively considering patients aged 16 years and above, our study diligently avoided potential information leakage and ensured experimental uniformity by utilizing solely the initial ICU admission data for each patient, excluding subsequent admissions. To underscore the early predictive capacity, input to the predictive model was derived from data collected within the first 24 h of patient admission to the ICU. Addressing the challenges posed by imbalanced datasets, particularly prevalent in numerous application domains, is imperative for robust machine-learning outcomes. Mitigating this imbalance is crucial, and Synthetic Minority Over-Sampling Technique (SMOTE) emerges as a potent strategy for rectifying class-distribution disparities [23]. Through the generation of synthetic profiles, SMOTE effectively balances class distributions, enhancing prediction performance. In this study, we applied various SMOTE ratios to address data skewness and create balanced datasets.

2.2. BERTopic

BERTopic represents an innovative approach to topic modeling, leveraging the power of BERT (Bidirectional Encoder Representations from Transformers), a pre-trained language model, to extract latent themes from textual data [24]. Specifically designed for large datasets, this algorithm has demonstrated its efficacy in various natural language-processing tasks. BERTopic employs both BERT and c-TF-IDF to derive interpretable topics [18]. Figure 2 below illustrates the BERTopic Algorithm workflow, providing a visual representation of the process described.

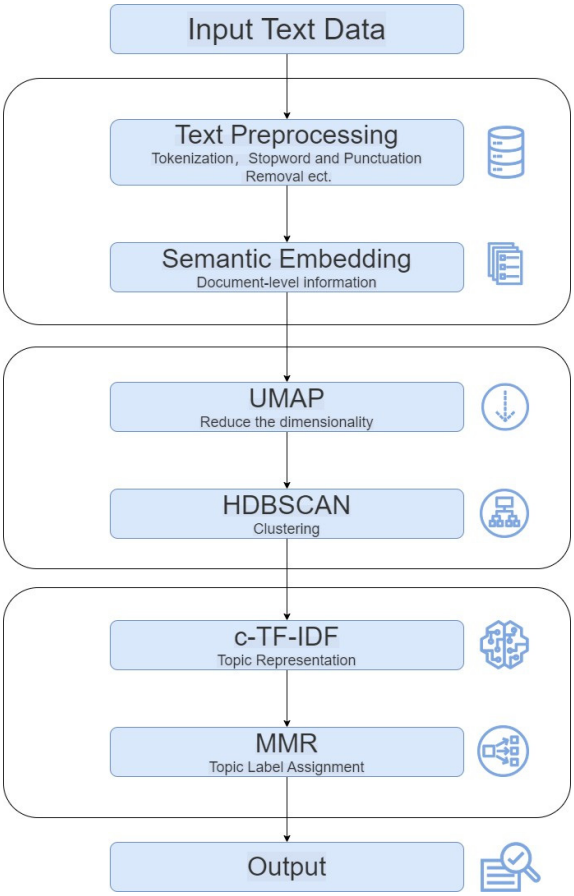


Figure 2. The BERTopic Algorithm workflow.

The algorithm comprises three key stages for generating a topic distribution across a collection of documents. In the initial stage, a similarity matrix is constructed among the documents through a pre-trained transformer model. The second stage involves dimensionality reduction in the similarity matrix using Uniform Manifold Approximation and Projection (UMAP). Subsequently, the third stage encompasses clustering the documents with HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), treating resulting clusters as topics. Notably, BERTopic shares similarities with Top2Vec, offering continuous rather than discrete topic modeling. Consequently, the stochastic nature of the model may yield varied outcomes upon repeated modeling. Additionally, recognizing the potential emergence of a substantial number of topics, BERTopic facilitates exploration through an interactive inter-topic distance map [17]. Figure 2 delineates the workflow of the BERTopic Algorithm.

2.3. Machine Learning

Our dataset was organized and divided into two parts for training and testing the model, with 80% and 20% of the data, respectively. We developed an HF patients' mortality prediction model using seven commonly used machine-learning algorithms. The selected machine-learning algorithms include GaussianNB, AdaBoost, Bagging, Gradient Boosting, MLP Classifier, LightGBM, and XGBoost. All data-mining tasks were performed using the Python programming language. GaussianNB is a probabilistic classification technique commonly used in machine learning. It assumes that each feature or predictor has an independent ability to predict the output variable [25]. By combining the predictions of all features, the algorithm assigns a probability of the dependent variable belonging to each group. The final classification is determined by selecting the group with the highest probability. MLP Classifier (Multilayer Perceptron) is a feed-forward artificial neural network that consists of multiple layers of interconnected neurons. It learns and predicts data using principles inspired by the human nervous system. MLPs are suitable for tasks involving classification and prediction with various feature sets [26]. In our study, the MLP neural network had five layers, namely an input layer, three hidden layers, and an output layer. Each hidden layer had 13 neuron nodes. We used regular activation functions and Adam optimizers for training and weight optimization. The normalization parameter was set to 10^{-5}.

The five methods that follow are also integral to ensemble learning, a technique that combines multiple models to improve predictive performance and robustness. AdaBoost is an adaptive ensemble learning method that focuses on misclassified samples from previous classifiers. It is sensitive to noisy and outlier data [27]. The algorithm trains a base classifier and assigns higher weights to incorrectly classified samples. This process is iterated with subsequent classifiers. The iterations continue until a stopping condition is met or the error rate becomes sufficiently small. In our implementation using the Python scikit-learn library, we set the maximum number of iterations to 50, while keeping other hyperparameters at their default values. Bagging, also known as bootstrap aggregating, is an ensemble-learning algorithm widely used in machine learning. It aims to improve accuracy and stability, while reducing variance and overfitting by combining multiple predictors [28]. In our implementation using the Python scikit-learn library, we constructed an ensemble classifier consisting of 500 DecisionTreeClassifiers. Each classifier was trained on a maximum subset of 100 samples obtained through bootstrapping, where sampling is performed with replacement. We used default values for other hyperparameters. Gradient Boosting is another ensemble-learning algorithm that can enhance the accuracy of different classification models. It utilizes the negative gradient information of the model's loss function to train models with poor predictive accuracy [29]. These trained models are then accumulated to obtain the final predictive model. Gradient Boosting has been widely used for feature selection, classification, and regression tasks. LightGBM is a Gradient Boosting framework that utilizes an improved histogram algorithm. It divides continuous feature values into K intervals and selects split points from these intervals, significantly accelerating prediction

speed and reducing memory consumption, without sacrificing prediction accuracy [30]. LightGBM has been widely applied for feature selection, classification, and regression tasks. Notable techniques used in LightGBM include Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). In our study, GOSS retained features with high gradients while randomly sampling features with low gradients to maintain consistency with the original feature distribution. EFB was used to bundle mutually exclusive features and reduce the number of features. XGBoost is a scalable tree-boosting system that provides an optimized implementation of the Gradient Boosting framework. It is known for its ability to handle missing data efficiently, its flexibility, and its ability to assemble weak prediction models into a more accurate one [31]. XGBoost generates a series of decision trees during training, where each tree builds upon the previous one by minimizing the loss function gradient. By combining multiple decision trees, a predictive model is obtained. XGBoost can handle missing values by incorporating a default orientation for missing values in each tree node and learning the best orientation from the data.

2.4. Performance Evaluation

Our study rigorously assessed the impact of integrating both quantitative and textual data on the predictive accuracy of mortality in ICU patients through the application of five widely acknowledged evaluation metrics in binary-classification tasks. These metrics serve as key indicators of model effectiveness and precision:

- *Precision* (*PPV*): Denoting the accuracy of positive predictions, precision is calculated as the ratio of true positives (*TP*s) to the sum of true positives (*TP*s) and false positives (*FP*s).

$$Precision = PPV = \frac{TP}{TP + FP} \tag{1}$$

- *Recall* (sensitivity, *TPR*): Recall quantifies the model's ability to correctly identify positive cases and is calculated as the ratio of true positives (*TP*s) to the sum of true positives (*TP*s) and false negatives (*FN*s).

$$Recall = TPR = \frac{TP}{TP + FN} \tag{2}$$

- *F1-score*: This is a comprehensive metric combining precision and recall, calculated as the harmonic mean of the two, offering a balanced measure of overall performance.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

- *Accuracy*: *Accuracy* assesses the overall correctness of predictions and is calculated as the ratio of the sum of true positives (*TP*s) and true negatives (*TN*s) to the total number of samples in the test dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

- AUROC (Area Under the Receiver Operating Characteristic Curve): AUROC is a common diagnostic metric that provides a comprehensive measure of classifier performance by plotting the relationship between the true-positive rate (TPR) and the false-positive rate (FPR) across various classification thresholds.

By employing these evaluation metrics, this study aimed to provide a thorough and standardized evaluation of predictive models that integrates both quantitative and text data, specifically in the context of mortality prediction for ICU patients.

3. Results

This study utilized data from the MIMIC-III database for preprocessing, ultimately focusing on 6606 patients with HF and their ICU admission records. Table 1 presents the demographic characteristics of the study cohort. The average age of the patients was 70.32 years, with a standard deviation of 13.03 years. Gender distribution revealed that 55.07% were male. In terms of insurance type, a majority of patients (70.91%) were covered by Medicare, followed by private insurance (21.47%).

Table 1. Selected patient demographic information.

	Overall	Dead at ICU	Alive at ICU
General			
Number	6606 (100%)	902 (13.65%)	5704 (86.35%)
Age	70.32 ± 13.03	72.81 ± 12.70	69.92 ± 13.08
Gender (male)	3638 (55.07%)	3153 (55.28%)	485 (53.77%)
Insurance type			
Self-pay	27 (0.41%)	5 (0.55%)	22 (0.39%)
Government	94 (1.42%)	8 (0.89%)	86 (1.51%)
Medicare	4684 (70.91%)	700 (77.61%)	3984 (69.85%)
Medicaid	383 (5.8%)	35 (3.88%)	348 (6.1%)
Private	1418 (21.47%)	154 (17.07%)	1264 (22.16%)
Ethnicity			
Asian	121 (1.83%)	19 (2.11%)	102 (1.79%)
Black/African	574 (8.69%)	42 (4.66%)	532 (9.33%)
Hispanic/Latino	155 (2.35%)	19 (2.11%)	136 (2.38%)
White	4718 (71.42%)	638 (70.73%)	4080 (71.53%)
Other	1038 (15.71%)	184 (20.40%)	854 (14.97%)
Marital status			
Married	3284 (49.71%)	445 (49.33%)	2839 (49.77%)
Single	1197 (18.12%)	145 (16.08%)	1052 (18.44%)
Divorced	436 (6.60%)	51 (5.65%)	385 (6.75%)
Separated	70 (1.06%)	9 (1.00%)	61 (1.07%)
Widowed	1282 (19.41%)	171 (18.96%)	1111 (19.48%)
Unknown	337 (5.1%)	81 (8.98%)	256 (4.49%)
Religion			
Christianity	4725 (71.53%)	638 (70.73%)	4087 (71.65%)
Judaism	574 (8.69%)	42 (4.66%)	532 (9.33%)
Islam	159 (2.41%)	20 (2.22%)	139 (2.44%)
Buddhism	122 (1.85%)	19 (2.11%)	103 (1.81%)
Other/unknown	1026 (15.53%)	183 (20.29%)	843 (14.78%)
Variable value			
Heart rate	85.60 ± 15.49	85.03 ± 15.15	89.04 ± 16.98
Respiratory rate	19.51 ± 4.18	19.31 ± 4.01	20.74 ± 4.94
Diastolic blood pressure	57.10 ± 12.15	57.57 ± 12.15	54.35 ± 11.73
Systolic blood pressure	115.01 ± 19.11	115.70 ± 19.04	110.98 ± 19.01
Temperature	98.18 ± 1.42	98.21 ± 1.38	98.02 ± 1.60
Oxygen saturation	97.00 ± 2.24	97.09 ± 2.02	96.51 ± 3.25
Fractional inspired oxygen	15.55 ± 26.57	16.12 ± 26.69	12.73 ± 25.81
Blood urea nitrogen	33.53 ± 24.12	31.73 ± 22.70	44.50 ± 29.12
Creatinine	1.76 ± 2.06	1.72 ± 2.13	1.99 ± 1.57
Mean blood pressure	76.97 ± 12.45	76.98 ± 11.50	76.93 ± 17.14
Glucose	146.11 ± 46.21	144.65 ± 45.09	154.76 ± 51.53
White blood cell	12.63 ± 11.94	12.21 ± 7.27	15.15 ± 26.12
Red blood cell	3.54 ± 0.52	3.54 ± 0.52	3.50 ± 0.55
Prothrombin time	16.61 ± 6.74	16.43 ± 6.63	17.66 ± 7.24
International Normalized Ratio	1.65 ± 1.00	1.61 ± 0.94	1.87 ± 1.28
Platelets	216.09 ± 95.72	217.45 ± 93.55	207.80 ± 107.65
GCS eye	3.32 ± 0.86	3.38 ± 0.80	2.90 ± 1.05
GCS motor	5.33 ± 1.11	5.41 ± 1.01	4.88 ± 1.49
GCS verbal	3.44 ± 1.67	3.54 ± 1.64	2.80 ± 1.74
Outcomes			
Hospital LOS (days) [Q1–Q3]	13.12 [6.03–16.05]	12.79 [6.09–15.80]	14.61 [5.30–19.15]
ICU LOS (days) [Q1–Q3)	5.80 [1.95–6.24]	5.40 [1.90–5.79]	8.18 [2.40–10.41]

The ethnicity distribution demonstrated a predominantly White population (71.42%), followed by Black/African (8.69%). Marital status varied, with 49.71% being married, and 18.12% single. Religious affiliation indicated that the majority of patients identified with

Christianity (71.53%), followed by Judaism (8.69%). These demographic details provide a comprehensive overview of the study population and are summarized in Table 1, offering essential insights for the subsequent analyses conducted in this research. Table 1 also presents the mean values of the predictive variables utilized in this study for ICU-admitted HF patients.

3.1. Prediction of the Mortality

In our study, we employed a 10-fold cross-validation method to construct a predictive model for HF patients, forecasting outcomes at 3 days, 30 days, and 365 days. The model utilized data collected within the 24 h preceding admission to the ICU, in conjunction with physicians' treatment plans and test records, including the 365-day mortality rate. The dataset was split, with 80% used for model training and 20% for testing, followed by comprehensive statistical analyses to assess performance. The results of these analyses, detailing the AUROC scores for various classifiers with and without the inclusion of qualitative data, are presented in Table 2.

Table 2. AUROC of different classifiers.

Terms	Method	Without Qualitative Data	With Qualitative Data
3 Days	GaussianNB	0.6349 ± 0.0222	0.6691 ± 0.0255
	AdaBoost	0.7666 ± 0.0127	0.7708 ± 0.0143
	Bagging	0.7425 ± 0.0111	0.7522 ± 0.0168
	Gradient Boosting	0.7725 ± 0.0071	0.7816 ± 0.0171
	MLP Classifier	0.7492 ± 0.0047	0.7573 ± 0.0073
	LightGBM	0.7935 ± 0.0059	0.8218 ± 0.0100
	XGBoost	0.7986 ± 0.0063	0.8228 ± 0.0117
30 Days	GaussianNB	0.6391 ± 0.0066	0.6616 ± 0.0137
	AdaBoost	0.7335 ± 0.0090	0.7469 ± 0.0113
	Bagging	0.7316 ± 0.0054	0.7541 ± 0.0031
	Gradient Boosting	0.7385 ± 0.0141	0.7580 ± 0.0066
	MLP Classifier	0.7237 ± 0.0119	0.7390 ± 0.0184
	LightGBM	0.7436 ± 0.0081	0.7599 ± 0.0060
	XGBoost	0.7389 ± 0.0132	0.7553 ± 0.0027
365 Days	GaussianNB	0.6311 ± 0.0049	0.6524 ± 0.0112
	AdaBoost	0.6872 ± 0.0206	0.7047 ± 0.0036
	Bagging	0.7096 ± 0.0178	0.7128 ± 0.0033
	Gradient Boosting	0.6905 ± 0.0250	0.6919 ± 0.0077
	MLP Classifier	0.6721 ± 0.0144	0.6793 ± 0.0035
	LightGBM	0.6310 ± 0.0057	0.6508 ± 0.0055
	XGBoost	0.6003 ± 0.0045	0.6125 ± 0.0150

Table 2 presents the AUROC values obtained from our study, showcasing the performance of six distinct machine-learning methods. To assess the impact of incorporating BERTopic topics on predicting HF-patient mortality in the ICU, we compared models using only quantitative data with those utilizing both quantitative and BERTopic data within the initial 24 h of ICU admission. Our findings reveal that integrating textual discharge summaries for new patients enhances the accuracy of mortality-rate predictions. Figure 3 displays the AUROC curves for various machine-learning methods, highlighting the effect of incorporating qualitative data on predictive performance. Figure 4 further illustrates the improvement in model applicability achieved through the integration of textual and quantitative data. The inclusion of qualitative data results in a substantial enhancement of AUROC scores, demonstrating that the models exhibit superior effectiveness in predicting mortality across different machine-learning methods and temporal intervals.

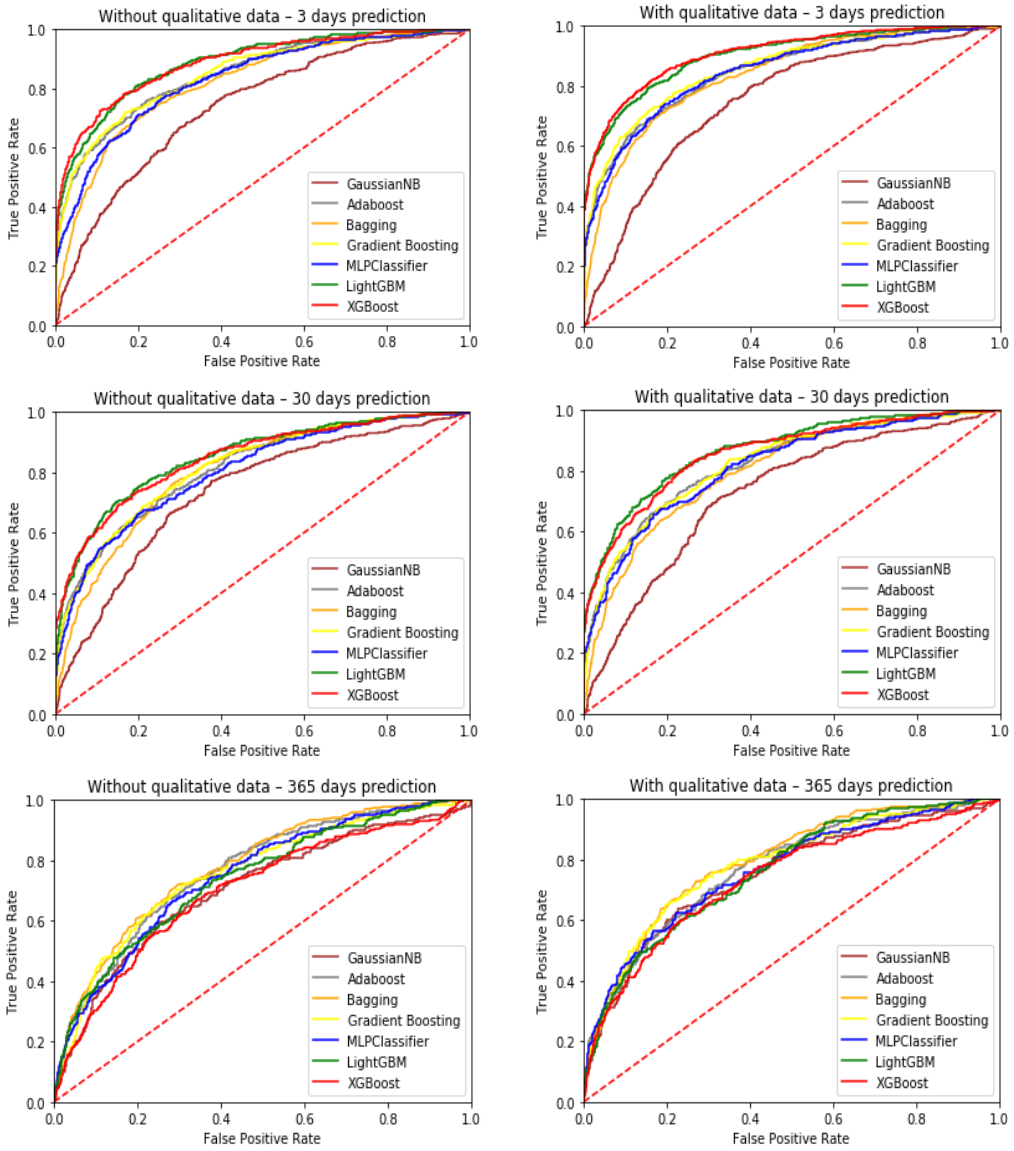


Figure 3. ROC curves for the different classifiers.

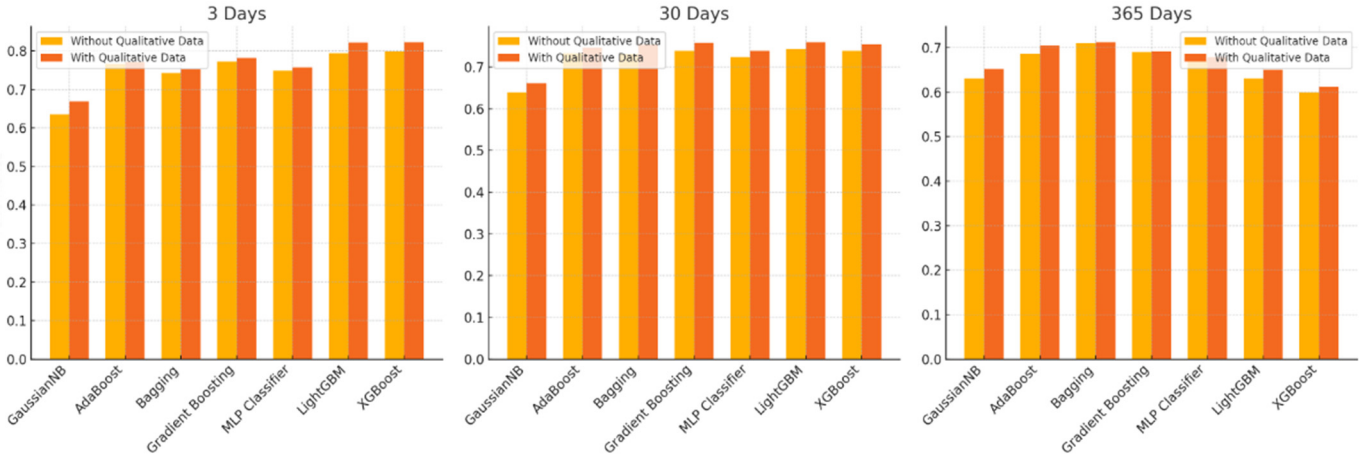


Figure 4. AUROC of machine-learning methods with and without qualitative data.

This study shows that integrating quantitative and textual data significantly improves mortality predictions for HF patients in the ICU. Models using both data types, particularly XGBoost, LightGBM, and Bagging, outperformed those using only quantitative data, as depicted in Figure 3. The inclusion of BERTopic-processed textual data enhances predictive accuracy by incorporating higher-order concepts from clinical observations and judgments. This integration underscores the value of combining different data forms for precise mortality predictions in ICU settings. The detailed diagnostic precision, recall, F1-score, and accuracy of different classifiers with and without qualitative data are presented in Table 3.

Table 3. Diagnostic precision, recall, F1-score, and accuracy.

Terms	Dataset	Method	Precision	Recall	F1-Score	Accuracy
3 Days	Without qualitative data	GaussianNB	0.4697 ± 0.0217	0.4390 ± 0.0548	0.4524 ± 0.0348	0.7307 ± 0.0095
		AdaBoost	0.5628 ± 0.0119	0.7263 ± 0.0248	0.6340 ± 0.0145	0.7763 ± 0.0078
		Bagging	0.5143 ± 0.0110	0.7166 ± 0.0247	0.5986 ± 0.0123	0.7551 ± 0.0065
		Gradient Boosting	0.5800 ± 0.0102	0.7246 ± 0.0210	0.6440 ± 0.0061	0.7958 ± 0.0045
		MLP Classifier	0.5644 ± 0.0099	0.6776 ± 0.0156	0.6156 ± 0.0041	0.7843 ± 0.0033
		LightGBM	0.7237 ± 0.0084	0.6753 ± 0.0164	0.6984 ± 0.0051	0.8514 ± 0.0018
		XGBoost	0.7617 ± 0.0140	0.6691 ± 0.0194	0.7120 ± 0.0048	0.8620 ± 0.0016
	With qualitative data	GaussianNB	0.4360 ± 0.0086	0.6197 ± 0.0854	0.5100 ± 0.0364	0.7032 ± 0.0056
		AdaBoost	0.5504 ± 0.0153	0.7599 ± 0.0299	0.6382 ± 0.0184	0.7876 ± 0.0073
		Bagging	0.5339 ± 0.0203	0.7278 ± 0.0346	0.6157 ± 0.0220	0.7638 ± 0.0098
		Gradient Boosting	0.5812 ± 0.0216	0.7546 ± 0.0333	0.6563 ± 0.0219	0.7995 ± 0.0108
		MLP Classifier	0.5591 ± 0.0528	0.7185 ± 0.0316	0.6259 ± 0.0204	0.7855 ± 0.0251
		LightGBM	0.7636 ± 0.0291	0.7224 ± 0.0224	0.7419 ± 0.0153	0.8693 ± 0.0067
		XGBoost	0.7918 ± 0.0240	0.7115 ± 0.0240	0.7491 ± 0.0162	0.8760 ± 0.0071
30 Days	Without qualitative data	GaussianNB	0.4029 ± 0.0169	0.4342 ± 0.0329	0.4166 ± 0.0078	0.7542 ± 0.0087
		AdaBoost	0.4259 ± 0.0136	0.6919 ± 0.0180	0.5272 ± 0.0151	0.7590 ± 0.0047
		Bagging	0.4041 ± 0.0018	0.7191 ± 0.0174	0.5174 ± 0.0060	0.7394 ± 0.0052
		Gradient Boosting	0.4411 ± 0.0163	0.6870 ± 0.0282	0.5372 ± 0.0199	0.7701 ± 0.0088
		MLP Classifier	0.4387 ± 0.0384	0.6503 ± 0.0094	0.5228 ± 0.0262	0.7684 ± 0.0221
		LightGBM	0.6127 ± 0.0169	0.5747 ± 0.0135	0.5931 ± 0.0150	0.8468 ± 0.0048
		XGBoost	0.6481 ± 0.0150	0.5498 ± 0.0279	0.5947 ± 0.0218	0.8547 ± 0.0031
	With qualitative data	GaussianNB	0.3906 ± 0.0229	0.5271 ± 0.0461	0.4478 ± 0.0263	0.7633 ± 0.0129
		AdaBoost	0.4479 ± 0.0291	0.7101 ± 0.0159	0.5491 ± 0.0259	0.7691 ± 0.0084
		Bagging	0.4410 ± 0.0232	0.7404 ± 0.0194	0.5522 ± 0.0159	0.7622 ± 0.0091
		Gradient Boosting	0.4720 ± 0.0265	0.7135 ± 0.0205	0.5674 ± 0.0168	0.7847 ± 0.0062
		MLP Classifier	0.4473 ± 0.0166	0.6904 ± 0.0695	0.5411 ± 0.0230	0.7690 ± 0.0147
		LightGBM	0.6627 ± 0.0326	0.5945 ± 0.0087	0.6265 ± 0.0176	0.8598 ± 0.0032
		XGBoost	0.6976 ± 0.0344	0.5721 ± 0.0115	0.6279 ± 0.0070	0.8658 ± 0.0040
365 Days	Without qualitative data	GaussianNB	0.3116 ± 0.0081	0.4043 ± 0.0254	0.3514 ± 0.0064	0.7859 ± 0.0115
		AdaBoost	0.2953 ± 0.0206	0.6027 ± 0.0316	0.3963 ± 0.0252	0.7486 ± 0.0125
		Bagging	0.2968 ± 0.0147	0.6715 ± 0.0295	0.4116 ± 0.0197	0.7373 ± 0.0096
		Gradient Boosting	0.3001 ± 0.0259	0.6051 ± 0.0375	0.4012 ± 0.0313	0.7425 ± 0.0158
		MLP Classifier	0.2672 ± 0.0169	0.6101 ± 0.0185	0.3714 ± 0.0190	0.7171 ± 0.0136
		LightGBM	0.3916 ± 0.0207	0.3484 ± 0.0210	0.3676 ± 0.0068	0.8363 ± 0.0052
		XGBoost	0.4083 ± 0.0279	0.2349 ± 0.0056	0.2980 ± 0.0114	0.8386 ± 0.0029
	With qualitative data	GaussianNB	0.2959 ± 0.0165	0.4962 ± 0.0570	0.3687 ± 0.0123	0.7959 ± 0.0283
		AdaBoost	0.2990 ± 0.0049	0.6541 ± 0.0048	0.4104 ± 0.0045	0.7514 ± 0.0030
		Bagging	0.2962 ± 0.0108	0.6861 ± 0.0145	0.4135 ± 0.0086	0.7421 ± 0.0071
		Gradient Boosting	0.2936 ± 0.0111	0.6232 ± 0.0123	0.3991 ± 0.0128	0.7517 ± 0.0056
		MLP Classifier	0.2621 ± 0.0049	0.6533 ± 0.0325	0.3737 ± 0.0019	0.7283 ± 0.0219
		LightGBM	0.4238 ± 0.0138	0.3854 ± 0.0126	0.4034 ± 0.0076	0.8432 ± 0.0029
		XGBoost	0.4072 ± 0.0285	0.2666 ± 0.0306	0.3219 ± 0.0301	0.8459 ± 0.0038

Table 3 shows that, by using data collected within 24 h of ICU admission for HF patients, our model can predict mortality rates at 3 and 30 days with up to 70% accuracy.

LightGBM and XGBoost performed best, achieving the highest accuracy (LightGBM, 0.8693; XGBoost, 0.8760) and F1-scores (LightGBM, 0.7419; XGBoost, 0.7491). Integrating quantitative and qualitative data enhances mortality prediction accuracy, aiding clinicians in making informed decisions to improve ICU patient outcomes. Figure 5 graphically represents the accuracy of machine-learning methods with and without qualitative data across three different timeframes: 3 days, 30 days, and 365 days. These charts are designed to visually emphasize the improvement in predictive accuracy when qualitative data are incorporated. By comparing the performance across different periods, the charts demonstrate how the integration of diverse data forms significantly enhances the models' ability to predict mortality in ICU settings more accurately and reliably.

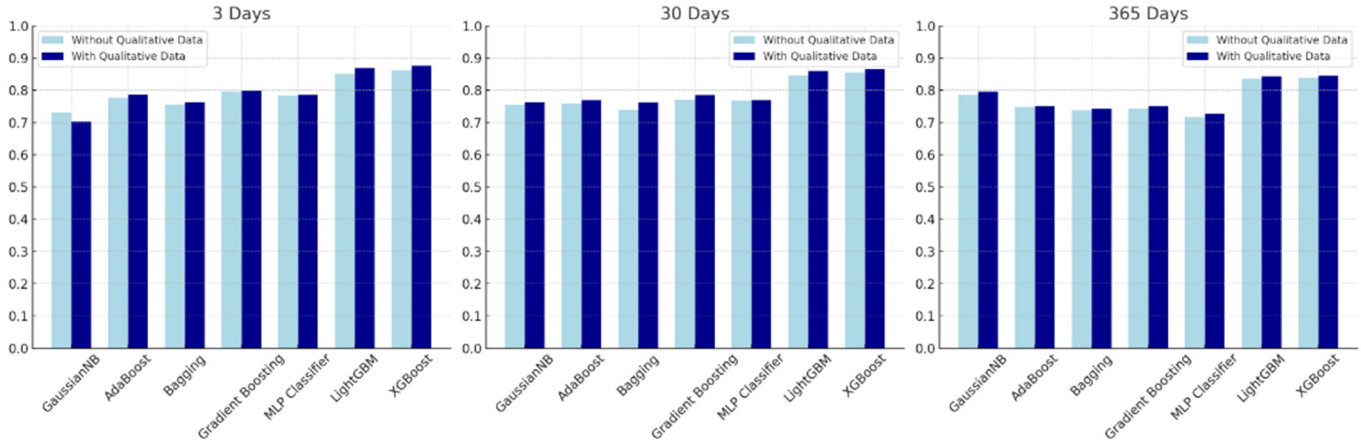


Figure 5. Accuracy of machine-learning methods with and without qualitative data.

3.2. Text Data Analysis

Within the MIMIC-III database, the NOTEEVENTS table comprises numerous text records for each patient's ICU hospitalization. For our investigation, we concentrated on medical-record transcripts and discharge summaries, specifically targeting test reports of HF patients within the ICU to predict mortality. We analyzed discharge abstracts sourced from the MIMIC-III database and generated a word cloud, a succinct visual representation highlighting key patient characteristics. The objective is to extract clinical insights supporting the effective management of HF patients in the ICU. Figure 6 depicts the resultant word cloud, providing an intuitive snapshot of prominent features.



Figure 6. Word cloud for ICU HF patients' medical records in MIMIC-III.

In our research, we employed advanced algorithmic techniques inspired by the works of Abuzayed et al. [32] to preprocess and analyze non-quantitative data from the MIMIC-

III database. Our approach involved refining medical-record annotations, eliminating duplicates, and filtering out unidentified or non-alphabetic sentences. Leveraging the BERTopic method, we generated six fundamental topics that encapsulate key themes in the dataset. These topics, along with their corresponding keywords, were instrumental in constructing predictive models for clinical outcomes in patients. The topics cover critical aspects such as cardiac health assessment, patient summary, valvular heart conditions, respiratory examination procedures, peripheral artery interventions, and holistic patient care and monitoring. The selection of these topics was optimized using the Grid Search method, and their comprehensive understanding is crucial for improving patient care and healthcare decision-making. Table 4 summarizes the six topics and their corresponding keywords generated by the BERTopic method in this study.

Table 4. Topics and keywords for dataset.

No.	Topic	Keywords
Topic 1	Cardiac health assessment	Chest, heart, pulmonary, ventricular, failure, cardiac, artery, edema, acute, renal
Topic 2	Patient summary and care transition	Identifier, dictated, MedQuist, summary, failure, tube, discharged, facility, rehabilitation, allergies
Topic 3	Valvular heart conditions	Aortic, mitral, capsule, INR, aorta, cardiac, chest, ventricular, regurgitation
Topic 4	Respiratory examination procedures	Bronchoscopy, tracheal, tube, stent, tracheostomy, trachea, capsule, inhalation, findings, bid
Topic 5	Peripheral artery interventions	Femoral, bypass, graft, artery, incision, surgery, iliac, wound
Topic 6	Patient care and monitoring	Home, capsule, care, service, family, service, facility, disposition, dictated

3.3. Feature Importance

Our study calculated feature-importance scores using the Gini index to rank variables in the predictive model. Higher scores indicate a greater contribution to the decision-tree value. Table 5 shows the importance of variables for ICU patients with heart failure, using data collected within 24 h of admission. The analysis includes both quantitative data and qualitative data processed by BERTopic, comparing predictions for 3-, 30-, and 365-day periods.

According to Table 5, the prediction of mortality in HF patients admitted to the ICU involves several important variables. Platelets (X_{17}), glucose (X_{12}), and white blood cells (X_{13}) are consistently identified as significant variables for predicting mortality at 3 days, 30 days, and 365 days. Specifically, in the short-term 3-day mortality prediction, platelets (X_{17}), glucose (X_{12}), and age (X_1) are relatively important factors. On the other hand, for 30-day and 365-day mortality prediction, respiratory rate (X_4) and platelets (X_{17}) emerge as relatively important variables.

Furthermore, the inclusion of qualitative data and the application of the BERTopic method allow for a more comprehensive analysis of mortality prediction in ICU patients with HF. When qualitative data are integrated, platelets (X_{17}), glucose (X_{12}), and temperature (X_7) are found to be relatively more important variables. In addition, thematic variables derived from BERTopic analysis, such as the Cardiac Health Assessment (Topic 1) and Respiratory Examination Procedures (Topic 3), are identified as important factors that should be considered. These findings highlight the significance of incorporating both quantitative and qualitative data into mortality prediction for ICU patients with HF. By utilizing the BERTopic method to conduct topic modeling on qualitative data, this study reveals that qualitative information from ICU patients contains valuable insights that can effectively support clinicians in making critical clinical decisions.

Table 5. Selected ten important variables.

	Feature Importance	3 Days	30 Days	365 Days
Without qualitative data	1	X_{17}	X_4	X_4
	2	X_{12}	X_{17}	X_8
	3	X_1	X_7	X_{17}
	4	X_7	X_{13}	X_7
	5	X_4	X_{12}	X_{12}
	6	X_{13}	X_3	X_5
	7	X_6	X_1	X_{13}
	8	X_9	X_8	X_3
	9	X_5	X_5	X_1
	10	X_8	X_{10}	X_6
With qualitative data	1	X_{17}	X_{12}	X_7
	2	TOPIC 3	X_7	X_8
	3	TOPIC 1	X_{17}	X_1
	4	X_{12}	X_4	X_{13}
	5	X_1	X_3	X_3
	6	X_7	X_{14}	X_{12}
	7	X_6	TOPIC 1	X_4
	8	X_8	X_8	X_{17}
	9	X_9	X_1	TOPIC 1
	10	X_{13}	TOPIC 2	X_5

4. Discussion

4.1. Principal Findings

Our research presents a notable contribution to the field through the application of BERTopic technology in analyzing EHR text data, generating thematic variables relevant to mortality prediction. The integration of both quantitative and textual clinical records allowed us to develop machine-learning models specifically designed for predicting the mortality rate of HF patients in the ICU. We further explored the importance of variable generation through machine-learning methods, offering insights into crucial factors influencing patient mortality and facilitating the identification of potential preventive measures in medical practice.

For the construction of our predictive model targeting HF patients, a 10-fold cross-validation method was employed to forecast outcomes at intervals of 3 days, 30 days, and 365 days. The model incorporated data from the 24 h preceding ICU admission, including physicians' treatment plans and test records. The rigorous evaluation of model performance through various statistical analyses, presented in Table 2 with AUROC values, highlighted the effectiveness of six machine-learning methods. The evaluation of incorporating BERTopic topics on mortality prediction revealed that the inclusion of textual discharge summaries significantly improved the accuracy of mortality rate predictions. Our study demonstrated the effectiveness of models integrating both quantitative and textual data, outperforming those relying solely on quantitative data. Notably, XGBoost, LightGBM, and Bagging emerged as superior models for predicting mortality rates at different time intervals. This integration offers a comprehensive approach to mortality prediction, providing valuable support for clinicians in making informed decisions to enhance patient outcomes in the ICU setting. The variable importance analysis, as depicted in Table 5, underscored the significance of both quantitative and qualitative data in mortality prediction for HF patients in the ICU. The application of the BERTopic method revealed valuable insights within qualitative data, offering crucial support for clinicians in critical decision-making [33]. The comparison of variable importance across different prediction periods emphasized the nuanced factors influencing patient outcomes.

Our study strongly suggests that the integration of both quantitative and qualitative data significantly improves the precision of mortality prediction for ICU patients with HF.

Leveraging BERTopic for topic modeling on qualitative data enhances our understanding of patient conditions, providing clinicians with nuanced insights for improved decision-making and ultimately contributing to enhanced patient care.

4.2. Limitations

Our study is subject to several important limitations that warrant consideration. The primary constraint lies in the utilization of patient data sourced from the MIMIC-III database, which reflects a specific patient demographic and healthcare context. Given that the data are predominantly sourced from large urban areas, the generalizability of our findings to ICU patients in smaller healthcare institutions remains uncertain. To enhance the robustness of our conclusions, future studies should strive to include data from diverse settings, including rural or smaller medical facilities.

Furthermore, the absence of high-quality nursing records for a significant subset of patients poses a potential limitation, particularly in extrapolating the model's performance to other intensive care environments. Relying on doctor-recorded information during patient consultations, although valuable for disease and treatment research, introduces inherent challenges, such as common spelling errors and inaccuracies. These issues may compromise the quality of interpretation and subsequent predictions [34]. The lack of external validation for our prediction model is another notable limitation, impacting the overall reliability and credibility of the proposed model. Subsequent research endeavors should prioritize the inclusion of robust external validation procedures to fortify the standing of the model.

In addition, it is crucial to note that our study exclusively focused on clinical data collected within the initial 24 h after ICU admission, along with a doctor's written summary. This temporal constraint hinders the comprehensive evaluation of changes occurring throughout the entire ICU admission period. A more exhaustive compilation of clinical records and test reports is essential for a thorough understanding and may exert a substantial influence on predicted outcomes. Researchers conducting future studies should remain cognizant of these limitations and actively address these challenges to propel advancements in the field.

Moreover, it is imperative to highlight that our study specifically targeted the prediction of mortality rates among heart-failure patients in the intensive care unit. Future research endeavors should extend their focus to encompass various diseases or conditions, allowing for a more nuanced exploration of predictive modeling across different patient populations [35,36]. Investigating diverse patient groups is likely to reveal distinct patterns and factors influencing outcomes, contributing to a richer understanding of the complexities involved. This expansion in scope can pave the way for tailored predictive models that cater to the unique characteristics of specific medical conditions, thereby advancing the applicability and effectiveness of such models in clinical practice.

5. Conclusions

This study introduces an innovative approach to predicting ICU mortality in patients with HF by seamlessly integrating EHR data with a BERTopic-based hybrid machine-learning methodology. The inclusion of unstructured data processed through natural language-processing technology significantly enhances the model's predictive accuracy for patient mortality. The merging of structured and unstructured data identifies key variables that contribute to the accuracy of predictive models. This research provides valuable insights into the potential synergies between machine learning and natural language processing for accurately predicting mortality in ICU patients with heart failure.

For future research, we recommend extending the investigation in four key directions. First, a more detailed classification and analysis of patient data based on different medical departments or diseases will enhance the specificity of predictive models. Second, collecting patient data from various healthcare settings—including outpatient, emergency, inpatient, and ICU—will facilitate more comprehensive model construction and evaluation,

thus enhancing the model's generalizability. Third, integrating diverse types of unstructured data, such as hospital workflow, patient needs, and social media content, along with seeking input from a wider range of healthcare professionals, can further improve the model's predictive accuracy and provide a more comprehensive understanding of patient conditions. Fourth, considering that heart failure manifests as a dynamic clinical condition characterized by varying symptoms and treatment responses over time, future studies should employ time-series techniques to capture these changes. By collecting longitudinal data and applying a time-series analysis, researchers can perform a more precise and nuanced analysis of heart-failure progression. This approach will enable the development of predictive models that better reflect the fluctuating nature of heart failure, ultimately improving patient outcomes and optimizing healthcare resource allocation. Finally, subsequent investigations into heart failure will employ a multidimensional analytical framework to meticulously examine the complex interdependencies between cardiac dysfunction and additional physiological systems, such as the renal, endocrine, and neurohormonal functions. This integrated approach aims to enhance the precision of predictive models and enrich the comprehension of the multifaceted pathophysiology of heart failure, ultimately aiding in the creation of targeted and more efficacious therapeutic strategies.

Author Contributions: Conceptualization, S.-W.W. and C.-C.L.; data curation, S.-W.W., C.-C.L. and T.-N.C.; formal analysis, S.-W.W., C.-C.L. and T.-N.C.; methodology, S.-W.W., C.-C.L. and T.-N.C.; supervision, S.-W.W., C.-C.L. and T.-N.C.; writing—original draft, C.-C.L., T.-N.C. and C.-M.C.; writing—review and editing, C.-C.L., T.-N.C. and C.-M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to sincerely thank the editor and reviewers for their kind comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Selected Variable.

	Feature	Item ID	Table
1	Age		Admissions
2	Gender		Admissions
3	DSP (diastolic blood pressure)	8441, 220180	Chart events
4	SBP (systolic blood pressure)	455, 220179	Chart events
5	Temperature	676, 223762, 678, 223761	Chart events
6	OS (oxygen saturation)	220277, 646	Chart events
7	BUN (blood urea nitrogen)	225624, 781	Chart events
8	Creatinine	220615, 791, 1525	Chart events
9	MBP (non-invasive blood pressure)	220181,	Chart events
10	Glucose	807, 225664, 811	Chart events
11	WBC (white blood cell)	1127, 861, 1542, 220546	Chart events
12	RBC (red blood cell)	833	Chart events
13	PT (prothrombin time)	227465, 1286, 824	Chart events
14	INR (International Normalized Ratio)	815, 1530, 227467	Chart events

Table A1. *Cont.*

	Feature	Item ID	Table
15	Platelets	828	Chart events
16	GCS eye	184, 220739	Chart events
17	GCS motor	454, 223901	Chart events
18	GCS verbal	723, 223900	Chart events
19	Insurance		Admissions
20	Religion		Admissions
21	Marital Status		Admissions
22	Ethnicity		Admissions

References

1. Misumi, K.; Matsue, Y.; Nogi, K.; Fujimoto, Y.; Kagiyama, N.; Kasai, T.; Kitai, T.; Oishi, S.; Akiyama, E.; Suzuki, S.; et al. Derivation and validation of a machine learning-based risk prediction model in patients with acute heart failure. *J. Cardiol.* **2023**, *81*, 531–536. [CrossRef]
2. Li, J.L.; Liu, S.R.; Hu, Y.D.; Zhu, L.F.; Mao, Y.J.; Liu, J.L. Predicting Mortality in Intensive Care Unit Patients with Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study. *J. Med. Internet Res.* **2022**, *24*, e38082. [CrossRef]
3. Mpanya, D.; Celik, T.; Klug, E.; Ntsinjana, H. Predicting in-hospital all-cause mortality in heart failure using machine learning. *Front. Cardiovasc. Med.* **2023**, *9*, 1032524. [CrossRef]
4. Abedi, V.; Avula, V.; Razavi, S.M.; Bavishi, S.; Chaudhary, D.; Shahjouei, S.; Wang, M.; Griessenauer, C.J.; Li, J.; Zand, R. Predicting short and long-term mortality after acute ischemic stroke using EHR. *J. Neurol. Sci.* **2021**, *427*, 117560. [CrossRef]
5. Guo, A.X.; Pasque, M.; Loh, F.; Mann, D.L.; Payne, P.R.O. Heart Failure Diagnosis, Readmission, and Mortality Prediction Using Machine Learning and Artificial Intelligence Models. *Curr. Epidemiol. Rep.* **2020**, *7*, 212–219. [CrossRef]
6. Kedia, S.; Bhushan, M. Prediction of mortality from heart failure using machine learning. In Proceedings of the 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 24–25 June 2022.
7. Chiu, C.-C.; Wu, C.-M.; Chien, T.-N.; Kao, L.-J.; Li, C.; Jiang, H.-L. Applying an improved stacking ensemble model to predict the mortality of ICU patients with heart failure. *J. Clin. Med.* **2022**, *11*, 6460. [CrossRef]
8. Marshall, J.C.; Bosco, L.; Adhikari, N.K.; Connolly, B.; Diaz, J.V.; Dorman, T.; Fowler, R.A.; Meyfroidt, G.; Nakagawa, S.; Pelosi, P. What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine. *J. Crit. Care* **2017**, *37*, 270–276. [CrossRef]
9. Romano, M. S The Role of Palliative Care in the Cardiac Intensive Care Unit. *Healthcare* **2019**, *7*, 30. [CrossRef]
10. Haase, N.; Plovsing, R.; Christensen, S.; Poulsen, L.M.; Brochner, A.C.; Rasmussen, B.S.; Helleberg, M.; Jensen, J.U.S.; Andersen, L.P.K.; Siegel, H.; et al. Characteristics, interventions, and longer term outcomes of COVID-19 ICU patients in Denmark—A nationwide, observational study. *Acta Anaesthesiol. Scand.* **2020**, *65*, 68–75. [CrossRef]
11. Kim, J.Y.; Yee, J.; Park, T.I.; Shin, S.Y.; Ha, M.H.; Gwak, H.S. Risk Scoring System of Mortality and Prediction Model of Hospital Stay for Critically Ill Patients Receiving Parenteral Nutrition. *Healthcare* **2021**, *9*, 853. [CrossRef]
12. Zhang, D.D.; Yin, C.C.; Zeng, J.C.; Yuan, X.H.; Zhang, P. Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 280. [CrossRef]
13. Albashayreh, A.; Bandyopadhyay, A.; Zeinali, N.; Zhang, M.; Fan, W.; Gilbertson White, S. Natural Language Processing Accurately Differentiates Cancer Symptom Information in Electronic Health Record Narratives. *JCO Clin. Cancer Inform.* **2024**, *8*, e2300235. [CrossRef]
14. Noaeen, M.; Amini, S.; Bhasker, S.; Ghezelsefli, Z.; Ahmed, A.; Jafarinezhad, O.; Abad, Z.S.H. Unlocking the power of EHRs: Harnessing unstructured data for machine learning-based outcome predictions. In Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023.
15. Adlung, L.; Cohen, Y.; Mor, U.; Elinav, E. Machine learning in clinical decision making. *Med* **2021**, *2*, 642–665. [CrossRef]
16. Molloy-Paolillo, B.; Mohr, D.; Levy, D.R.; Cutrona, S.L.; Anderson, E.; Rucci, J.; Helfrich, C.; Sayre, G.; Rinne, S.T. Assessing Electronic Health Record (EHR) Use during a Major EHR Transition: An Innovative Mixed Methods Approach. *J. Gen. Intern. Med.* **2023**, *38*, 999–1006. [CrossRef]
17. Egger, R.; Yu, J.N. A Topic Modeling Comparison between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Front. Sociol.* **2022**, *7*, 886498. [CrossRef]
18. Alhaj, F.; Al-Haj, A.; Sharieh, A.; Jabri, R. Improving Arabic Cognitive Distortion Classification in Twitter using BERTopic. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 854–860. [CrossRef]
19. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [CrossRef]
20. Venugopalan, M.; Gupta, D. An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. *Knowl.-Based Syst.* **2022**, *246*, 108668. [CrossRef]
21. Johnson, A.E.W.; Pollard, T.J.; Shen, L.; Lehman, L.W.H.; Feng, M.L.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L.A.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 160035. [CrossRef] [PubMed]

22. Guo, W.Q.; Peng, C.N.; Liu, Q.; Zhao, L.Y.; Guo, W.Y.; Chen, X.H.; Li, L. Association between base excess and mortality in patients with congestive heart failure. *ESC Heart Fail.* **2021**, *8*, 250–258. [CrossRef] [PubMed]
23. Blagus, R.; Lusa, L. Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models. *BMC Bioinform.* **2015**, *16*, 363. [CrossRef] [PubMed]
24. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K.; Assoc Computat, L. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Conference of the North-American-Chapter of the Association-for-Computational-Linguistics—Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019.
25. Pushpakumar, R.; Prabu, R.; Priscilla, M.; Renisha, P.; Prabu, R.T.; Muthuraman, U. A Novel Approach to Identify Dynamic Deficiency in Cell using Gaussian NB Classifier. In Proceedings of the 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 22–24 June 2022.
26. Windeatt, T. Accuracy/diversity and ensemble MLP classifier design. *IEEE Trans. Neural Netw.* **2006**, *17*, 1194–1211. [CrossRef] [PubMed]
27. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-class adaboost. *Stat. Its Interface* **2009**, *2*, 349–360. [CrossRef]
28. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
29. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
30. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 52.
31. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
32. Abuzayed, A.; Al-Khalifa, H. BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. In Proceedings of the 5th Conference on AI in Computational Linguistics (ACLing), Electr Network, Online, 4–5 June 2021.
33. Uncovska, M.; Freitag, B.; Meister, S.; Fehring, L. Rating analysis and BERTopic modeling of consumer versus regulated mHealth app reviews in Germany. *NPJ Digit. Med.* **2023**, *6*, 115. [CrossRef]
34. Caicedo-Torres, W.; Gutierrez, J. ISeeU2: Visually interpretable mortality prediction inside the ICU using deep learning and free-text medical notes. *Expert Syst. Appl.* **2022**, *202*, 117190. [CrossRef]
35. Ye, Z.X.; An, S.Y.; Gao, Y.X.; Xie, E.M.; Zhao, X.C.; Guo, Z.Y.; Li, Y.K.; Shen, N.; Ren, J.Y.; Zheng, J.A. The prediction of in-hospital mortality in chronic kidney disease patients with coronary artery disease using machine learning models. *Eur. J. Med. Res.* **2023**, *28*, 33. [CrossRef]
36. Kasim, S.; Malek, S.; Aziz, M.F.; Ibrahim, K.S. Machine learning to predict in-hospital mortality risk among heterogenous STEMI patients with diabetes. *Eur. Heart J.* **2022**, *43*, ehab849.176. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.