Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 9, 2025

MACHINE LEARNING APPROACH TO PREDICTING PATIENT OUTCOMES IN INTENSIVE CARE UNITS

Saira Khurram*[1], Wajeeha Rehman[2], Taha Malik[3]

[1]Roots IVY, Educational Complex Faisalabad.
[2]Department of Biomedical Sciences.
[3]BS Software Engineering 2012, Bahria University Islamabad, QA Project Lead at F3 Technologies
QA Project Lead at F3 Technologies

[3]tahamalik790@gmail.com

DOI:https://doi.org/10.5281/zenodo.17181780

Keywords
Outcomes, patients, Intensive Care Units, leading hospitals, Pakistan, application, machine learning models.

Article History
Received: 07 July 2025
Accepted: 07 September 2025
Published: 23 September 2025

Corresponding Author: *
Saira Khurram

Abstract
This study predicts the outcomes of patients in the Intensive Care Units (ICUs) of leading hospitals in Pakistan; this research was focused on the application of various machine learning models on the given data. For this, the study was given a database of 2850 patients in ICU from 5 hospitals of tertiary care for the duration of 2022-2024. The machine learning models, Random Forest, Support Vector Machines, Gradient Boosting, Neural Networks, and Logistic Regression were specifically tailored and tested on ICU mortality, length of stay of the patients in ICU and Mechanical Ventilation usage. The records include a range of age, sex, and vital signs, laboratory data, and clinical intervention within the first 48 hours of ICU admission Clinical endpoints include the ICU Severity Score, measuring and recording clinical intervention and other clinical parameters. The dataset undergoes data cleansing through multiple imputation, feature scaling, the SMOTE algorithm, and other class imbalance techniques. The models were tested through 10-fold cross validation where assessment metrics were precision, recall, score F1, and AUC ROC. The Gradient Boosting model performed best with an accuracy of 89.3 and AUC of 0.91 for mortality prediction. APACHE II score, lactate, and mechanical ventilation were the most feature predictors determined in the feature importance analysis models. To support the clinical decision support systems, clinical SHAP points were also incorporated into the model. The paper outlined the integration of predictive machine learning techniques to forecast ICU outcomes within Pakistan's healthcare system alongside decision support tools to optimize clinical workflows, resource allocation, and quality of care in the ICU.

INTRODUCTION

Most hospitals' advanced, and costly, units, Intensive Care Units, provide life-supporting treatment to patients who are critically ill or suffering from multiple organ failures. It is very complex and challenging to manage patients in an ICU, as providers have to make complex clinical decisions in a short time without a full understanding of the patient. Patients are systematically evaluated in order to provide the best possible treatment and care. Inadequate techniques to make predictions about patient outcomes stifle many of the clinical decisions that need to be taken, so clinical assessments are often made with an excessive dependence on the physician's abilities and professional experience, enormous clinical scoring systems, and arbitrary judgments (Okore, 2023). Constant patient monitoring incorporated with advanced information technologies, the proliferation of electronic health record systems, and the increase of ICU patients have all contributed to the increase in ICU patient information. This information is crucial to the growing fields of machine learning and predictive analytics, which seek to strengthen clinical decisions and optimize patient outcomes. In Pakistan, health care is challenged by a lack of adequate infrastructure, a lack of regionally appropriate medical care, and the need for tailor made approaches to the critical care of patients due to the imbalance in the priorities of health care providers and patients. This complex world of healthcare in Pakistan is a combination of a high patient to provider ration, resource scarcity, and differences in the quality of care provided at different healthcare facilities (Mani et al., 2024). In such situations, clinical outcome prediction is uncertain; however, enhancing the predictive accuracy can improve resource allocation and treatment effectiveness, healthcare expenditure, and patient care. Predictive analytics techniques like machine learning can circumvent the traditional clinical appraisal blind spots and provide answers to the problems stated above (van Beinum, 2021).

The application of machine learning to medical decision-making has, to an extent, revolutionized the field, as it can detect and predict trends embedded in intricate data sets without having to outline every expected response to an interaction. Furthermore, these models can process vast quantities of data simultaneously, containing many factors whose relationships give rise to countless patient outcome possibilities (Adlung, Cohen, Mor, & Elinav, 2021). The implementation of machine learning healthcare in many specialties has demonstrated beyond reasonable doubt the capacity to enhance diagnostic and treatment expectation accuracy, and the subsequent administrative actions applicable to the treatment's medical outcome. In intensive care medicine machine learning has been applied to predicting outcome indicators like mortality, length of stay, complications, and responses to treatment, in addition to clarifying the value of different clinical factors (Jayatilake & Ganegoda, 2021).

In regards to Pakistani healthcare, the development of predictive models for ICU outcomes addresses a number of important clinical and administrative tasks. Better understanding and predicting the mortality risk can help in discussing necessary end-of-life arrangements with families, resource planning, and rationing. It can also assist physicians in flagging patients for high risk and might require more frequent and aggressive monitoring and treatment. Predicting length of stay helps in resource allocation by optimizing bed capacity. Predicting the need for mechanical ventilation is crucial in triaging patients and in the distribution of healthcare staff and equipment in resource limited systems where there is a shortage of ventilators (Mustafa et al., 2022).

The integration of machine learning into best practice medicine still requires intricate addressing of model understandability, relevancy, and pragmatic roadblocks to practice. Providers, and indeed all Stakeholders, will have to understand how predictive models generate advice, and which clinical variables will drive predictions the most (Sanchez-Martinez et al., 2022). The opacity of some machine learning models is a huge barrier to integration into the clinical workflow and requires evolved approaches to AI that are deliberately focused on model ascertainable and operational transparency. Integration of machine learning models into clinical practice requires uncompromised integration into set information systems, seamless workflows, and robust orientation and clinical instruction on model use (Lyu, Xu, Yang, & Liu, 2023).

The verification of machine learning models across different healthcare systems is vital for their generalizability and clinical application to different patient populations and care systems. Disease epidemiology, patient demographics, sociocultural context, and healthcare delivery systems of Pakistan will determine the expected performance, relevance, and validation of the model. The most predictive models will be the ones that best address the specific challenges of the Pakistani healthcare system and the corresponding opportunities in the intensive care unit (ICU) to ensure relevance and fidelity to the ecological and healthcare systems modeling (Pourhomayoun & Shakibi, 2021).

The potential results of machine learning technologies require multifarious considerations including algorithmic discrimination and bias, equity, privacy, and the range of clinical technology interfaces. Predictive models, algorithmic equity, and privacy crosstalk for protected clinician discretions should be tailored and tested. The predictive model bias particularly requires the most attention to the representativeness of the training data, its validation, and the processes for monitoring and evaluation model performance in diverse patient populations post-deployment (Brnabic & Hess, 2021).

Implementation of the machine learning approach to healthcare in the ICU is but one of many factors. There is also the machine's cost development, implementation, personnel training, and the clinically assessed resources savings. The machine learning investment in the ICU is significant, particularly in its initial stages. In comparison, decreased length of stay, decreased complications, improved resource allocation, and enhanced quality of care offer significant cost reductions in the long term (Kothinti, 2024).

This paper addresses the challenges and opportunities through the development and validation of customized Machine Learning (ML) models developed for the Pakistani ICU setting, utilizing the comprehensive clinical databases of leading tertiary care hospitals for the relevance and contextual applicability of the local healthcare system. This research aims to establish predictive ICU outcomes with machine learning techniques plausibly and accurately, while elucidating the essential clinical parameters predictive of healthcare outcomes in Pakistan.

Research Objectives

1. To derive and establish multiple ML models from the clinical datasets of tertiary care hospitals in Pakistan to independently forecast outcomes of ICU patient mortality, length of stay, and need for mechanical ventilation, as per model comparison frameworks for the best ICU patient outcome prediction generation.
2. To use explainable AI methods to identify the primary determinants of clinical outcomes in the ICU to allow their consideration as primary outcome predictors and classifiers of the outcome determinants patient demographics, vital signs, lab and clinical data, and ICU care and interventions in the context of Pakistan.
3. To evaluate the clinical relevance and real-world usability of machine learning models calibrated for outcome predictions in the ICU context of Pakistan, taking into account the diverse clinical context and patient population of the ICU for model performance assessment to validate model clinical utility for diverse healthcare systems.

Research Questions

1. Which machine learning algorithms attain the highest accuracy and clinically meaningful outcomes with respect to patients' prognostic outcomes (mortality, LOS, and Mechanical Ventilation) and Prognostication in ICU patients and how do they perform with clinical data repositories from tertiary level hospitals in Pakistan?
2. Which clinical variables in the Pakistan patient cohort are ICU outcomes most strongly associated with, and what are the relative contributions of demographic, clinical, and laboratory datasets in improving the model's predictive power and its overall accuracy?
3. How do the predictions made from machine learning models contrast with existing clinical scores and the assessments provided by the physicians with respect to accuracy, precision, and clinical relevance as decision support systems for the ICU setting in Pakistan?

Significance of the Study

In the Pakistan context, as well as in countries with comparable healthcare infrastructures, the predictive analytic techniques and the enhancement of clinical decision support systems presented herein to derive actionable insights from the data available at ICU setting contributes to the advancement of critical care medicine. The lack of literature machine learning in ICU in developing countries serves as the motivation for this study, in hopes to shed light on the feasibility and impact of such systems and technologies in resource-limited settings. The integration of regionally validated models into predictive analytics frameworks will improve clinical decision making, resource allocation and patient care in intensive care unit in

Pakistan. More broadly, these models will contribute to the growing literature on the use of machine learning in the health care system. This research will help inform policies concerning the acceptance of health care technologies, the development of clinical practice guidelines, and the critical care quality improvement initiatives in the domain of cross border critical medicine. This study provides the foundation for research into the impact of Artificial Intelligence technologies on the health care system in Pakistan. More specifically, it seeks to equip health care leaders, clinicians, and technology developers with the knowledge necessary to enhance critical care analytics using artificial intelligence.

Literature Review

The advent of electronic health records and constant surveillance data, along with the need to improve clinical decision support in complex ICU environments, has resulted in an exponential increase of the application of machine learning to the field of critical care medicine, albeit with a time lag. The pioneering efforts in this area were based on clinical outcome prediction models and traditional statistics, hinged on a small set of clinical parameters. The complex set of machine learning algorithms we know today was developed alongside improvements in hardware and software, and with an understanding that ICU patient outcomes stem from multifaceted systems of interrelated clinical variables which are poorly captured by conventional statistical models (Hong et al., 2022).

Research involving outcome prediction in the ICUs has laid the groundwork for machine learning in medicine by demonstrating data-driven approaches augmenting conventional clinical score systems. In many comparative studies where machine learning approaches are juxtaposed to the conventional severity scoring systems, it has been consistently proven that the sophisticated systems outperform the traditional models and establish cause and effect relationships which the traditional models fail to. Such studies, therefore, reinforce the assertion that in feature extraction and data preprocessing, alongside the machine learning principles specific to healthcare in model validation, there is a need to work with an interdisciplinary approach (Bai, Gu, & Tang, 2025).

The seemingly disparate and often conflicted values in outcomes predicted in an Intensive Care Unit are mirrored in the forecast techniques formulated in the various specialized facets of critical care medicine. For example, the unprecedented such case of hospital readmission predicted by classifiers seems to show utility across numerous other sectors of healthcare, in particular, the Random Horest and Gradient Boosting classifiers whose effectiveness is attributed to their competence in heterogeneous data streams, their bypass of data loss filtration, and offering measures of relative importance in an explainable and actionable manner to healthcare practitioners. Furthermore, in the realm of chaotic and complex systems, Support Vector Machines have proven to be very adept and neural networks are also beginning to show promise in unlocking intricate patterns and associations in multifactorial clinical data sets (Chakilam, 2022).

While focusing on machine learning's application on predicting mortality in ICU settings, intensive research has shown the superiority of advanced algorithms relative to all other methods. The research has also highlighted the fact that machine learning, when compared to traditional severity scoring systems, has proven improvements in discrimination, calibration, and clinical utility. This research has delved into the importance of model validation, representative training datasets, and advanced algorithms' use in the clinic. The research suggests that optimal and effective mortality prediction models cannot compete on the measures of accuracy and specificity within the domain of clinical relevance and applicability (Chen, 2024). Length of stay prediction is yet another major area of machine learning application in the ICU, where the models have proven useful for enhancing and optimizing the allocation of resources, planning, and managing the costs of healthcare associated with discharge decisions. This research has illustrated the multiple and varied clinical, social, and systemic healthcare factors that determine the length of stay in different settings (Aslam, Aslam, Aslam, Aslam, & Aslam, 2025a). Studies show that deep learning methods and ensemble techniques used in predicting length of stay are able to capture and predict the intricacies of the problem in terms of health service's utilization (Aslam, Aslam, Aslam, Aslam, & Aslam, 2025b).

Machine learning is a new approach that models clinical activities in Intensive Care Units (ICU), particularly in predicting need for mechanical ventilation. These models understood the relief they could provide in treatment planning and resource apportionment and allocated. Advanced studies indicate that clinical outcomes could be improved enormously with prompt resource and care planning. Predicting clinical actions is not easy. These actions depend on physician preferences, institution customs, and available resources (Aslam et al., 2025b).

In healthcare, the importance of machine learning outcome features and its proper use in predicting outcomes of specific events represent the last mile for model accuracy and the need for retrieval of model outcomes that are relied upon and the need for retrieval of model outcomes that are relied upon for action. The need for clinical algorithms is also driven by the attempts to bias and limitations of healthcare algorithms capture by techniques like SHAP, LIME, and permutation importance (Saqlain, Gao Xiaoling, & Hussain). This work proves the need for clinical validation of machine learning insights in healthcare systems is of primary importance. The synthesis of machine learning and clinical practice is captured with the emphasis on model outcome, need for timely action, and physician engagement.

Research Methodology

The study has examined on using machine learning techniques to focus on the imitative outcomes of patients in the Intensive Care Units of several hospitals in Pakistan. The researchers have adopted a quantitative and retrospective approach in developing and validating a model on a dataset of 2850 patients from five tertiary care hospitals spanning a 2-year period from 2022 to 2024. These hospitals are Services hospital Lahore, Pakistan Institute of Medical Sciences Islamabad, Lady Reading Hospital Peshawar, Civil Hospital Karachi, and Aga Khan University Hospital Karachi. The construction of this patient data sets lasted from 2022 to 2024. The data deck included demographic, clinical and laboratory data, real-time vital monitoring, patient history, diagnosis and clinical therapeutics, ICU hours, and level II and SOFA scores, real-time monitoring patient databases, with ward data, hours of the ICU and patient history with clinical diagnosis and management, and the patient monitoring data base. The first and foremost outcome of the study was the ICU mortality and the secondary outcomes included ventilator dependence and the duration of stay in the ICU. The data to be used was pruned and raw data sets were filled using various other data sets and other imputation methods, normalization and standardization, and class imbalance was treated with the SMOTE algorithm. A range of machine learning techniques were used and discussed. Their scope encompasses Random Forest, Support Vector Machines, Gradient Boosting, Neural Networks, and Logistic Regression, which were executed through the scikit-learn and TensorFlow Python libraries. Model training on each algorithm was accomplished with the use of 10-fold cross validation and all models were subjected to performance assessment on specific parameters such as accuracy, precision, recall, F1-score, and the area under the ROC curve. To enhance the explainability of the models, the feature importance analysis was conducted with the use of SHAP (SHapely Additive exPlanations) values to ensure clinical relevancy and transparency within the Pakistani healthcare context.

Results and Data Analysis

Quantitative Analysis

The comprehensive analysis of machine learning models for predicting ICU patient outcomes revealed significant insights into the effectiveness of different algorithmic approaches and the clinical factors most strongly associated with patient prognosis in Pakistani healthcare settings. The dataset preprocessing and model development process successfully addressed common challenges associated with clinical data including missing values, class imbalance, and feature scaling requirements.

Table 1: Dataset Characteristics and Patient Demographics

Characteristic	Overall (n=2850)	Survivors (n=2223)	Non-survivors (n=627)	P-value
Age (years)	58.4 ± 16.2	56.1 ± 15.8	65.3 ± 16.1	<0.001
Gender (Male %)	62.3	61.8	64.1	0.312
APACHE II Score	18.7 ± 8.4	16.2 ± 7.1	26.8 ± 7.9	<0.001
SOFA Score	8.2 ± 4.6	7.1 ± 4.0	11.8 ± 4.8	<0.001
ICU Length of Stay (days)	6.8 ± 8.2	7.2 ± 8.6	5.6 ± 6.8	0.002
Mechanical Ventilation (%)	45.2	38.9	67.5	<0.001
Chronic Conditions (%)	74.1	71.8	82.3	<0.001

The dataset characteristics revealed significant differences between survivors and non-survivors across multiple clinical variables, providing evidence for the potential discriminatory power of machine learning models. Non-survivors were significantly older with higher severity scores (APACHE II and SOFA), increased rates of mechanical ventilation, and higher prevalence of chronic conditions. These patterns aligned with established clinical understanding while providing a foundation for machine learning model development. The overall mortality rate of 22.0% was consistent with reported ICU mortality rates in similar healthcare settings, supporting the representativeness of the study population.

Table 2: Machine Learning Model Performance for Mortality Prediction

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	AUC-ROC	AUC-PR
Gradient Boosting	89.3	84.7	78.2	0.814	0.912	0.756
Random Forest	87.8	82.1	75.6	0.787	0.895	0.723
Neural Network	86.4	80.3	73.8	0.769	0.881	0.698
Support Vector Machine	85.1	79.6	71.2	0.751	0.874	0.681
Logistic Regression	82.7	76.4	68.9	0.724	0.847	0.634
APACHE II (baseline)	79.2	71.8	65.3	0.685	0.798	0.567

The comparative analysis of machine learning algorithms demonstrated superior performance compared to traditional clinical scoring systems, with the Gradient Boosting model achieving the highest overall performance across multiple evaluation metrics. All machine learning approaches significantly outperformed the APACHE II baseline, with improvements in accuracy ranging from 3.5% to 10.1%. The Gradient Boosting algorithm achieved the highest AUC-ROC value of 0.912, indicating excellent discrimination capability between survivors and non-survivors. The consistent performance advantages across different metrics suggested robust model development and validation approaches while highlighting the potential clinical utility of these advanced analytical techniques.

Table 3: Feature Importance Rankings for Top Predictive Variables

Rank	Feature	Importance Score	Clinical Category	95% CI
1	APACHE II Score	0.184	Severity Score	(0.171-0.197)
2	Lactate Level	0.156	Laboratory Value	(0.142-0.170)
3	Mechanical Ventilation	0.143	Clinical Intervention	(0.129-0.157)
4	Age	0.121	Demographics	(0.108-0.134)
5	SOFA Score	0.118	Severity Score	(0.105-0.131)
6	Glasgow Coma Scale	0.094	Neurological Assessment	(0.082-0.106)
7	Creatinine Level	0.087	Laboratory Value	(0.075-0.099)

8	Mean Arterial Pressure	0.079	Vital Sign	(0.067-0.091)
9	Vasopressor Use	0.072	Clinical Intervention	(0.061-0.083)
10	Platelet Count	0.065	Laboratory Value	(0.054-0.076)

The feature importance analysis revealed that established clinical severity scores maintained their predictive value within machine learning frameworks while identifying additional clinical variables that contributed significantly to outcome prediction. The dominance of APACHE II scores in the feature importance rankings validated the clinical relevance of the machine learning approach while demonstrating that advanced algorithms could effectively utilize established clinical knowledge. Laboratory values, particularly lactate levels and creatinine, emerged as highly important predictors, reflecting the significance of metabolic and renal dysfunction in critical illness prognosis. The inclusion of clinical interventions such as mechanical ventilation and vasopressor use highlighted the importance of treatment intensity indicators in outcome prediction.

Table 4: Model Performance Across Hospital Settings

Hospital	Sample Size	Mortality Rate (%)	Model Accuracy (%)	AUC-ROC	Calibration Score
Aga Khan University Hospital	687	18.3	91.2	0.924	0.032
Services Hospital Lahore	612	24.1	88.7	0.901	0.041
PIMS Islamabad	578	23.6	89.1	0.908	0.038
Lady Reading Hospital	534	25.2	87.9	0.895	0.046
Civil Hospital Karachi	439	19.8	90.4	0.918	0.035
Overall Performance	2850	22.0	89.3	0.912	0.037

The cross-hospital validation analysis demonstrated robust model performance across different healthcare settings, with accuracy rates ranging from 87.9% to 91.2% and consistently high AUC-ROC values. The variation in mortality rates across hospitals reflected different patient populations, case-mix complexity, and institutional practices, while the consistent model performance suggested good generalizability of the machine learning approach. The calibration scores indicated well-calibrated probability estimates across all hospital settings, supporting the clinical utility of the predictive models for risk stratification and decision-making. The superior performance at university hospitals may reflect better data quality, standardized protocols, and case-mix factors that influenced both patient outcomes and model performance.

Table 5: Length of Stay Prediction Results

Prediction Category	Actual Mean LOS (days)	Predicted Mean LOS (days)	Mean Absolute Error	R^2 Score
Short Stay (≤3 days)	2.1 ± 0.8	2.3 ± 1.1	0.7	0.73
Medium Stay (4-7 days)	5.4 ± 1.2	5.2 ± 1.6	1.2	0.68
Long Stay (8-14 days)	10.8 ± 2.1	10.3 ± 2.8	2.1	0.65
Extended Stay (>14 days)	23.6 ± 8.4	21.9 ± 9.2	4.3	0.58
Overall Performance	6.8 ± 8.2	6.5 ± 8.7	2.1	0.67

The length of stay prediction analysis demonstrated reasonable performance for shorter stays with decreased accuracy for extended stays, reflecting the increased uncertainty and complexity associated with longer ICU admissions. The machine learning models achieved better predictive performance for patients with shorter expected stays, likely due to more predictable clinical trajectories and fewer confounding variables. The prediction accuracy for extended stays was limited by the complex interactions of medical factors, social determinants, and healthcare system variables that influence discharge decisions. The overall R^2 score of 0.67 indicated that the models explained a substantial portion of length of stay variance while highlighting opportunities for improvement through additional feature engineering and model refinement.

Table 6: Mechanical Ventilation Prediction Performance

Metric	Day 1 Prediction	Day 2 Prediction	Day 3 Prediction	Overall Performance
Accuracy (%)	82.4	85.7	88.2	85.4
Sensitivity (%)	78.9	83.1	86.4	82.8
Specificity (%)	84.6	87.2	89.1	86.9
PPV (%)	76.3	80.5	83.7	80.2
NPV (%)	86.1	89.1	91.2	88.8
AUC-ROC	0.847	0.881	0.906	0.878

The mechanical ventilation prediction results demonstrated improving performance with longer prediction horizons, suggesting that clinical deterioration patterns became more apparent over time and could be effectively captured by machine learning algorithms. The high negative predictive values across all time periods indicated strong capability for identifying patients unlikely to require mechanical ventilation, which could support resource planning and patient triage decisions. The improving specificity over time suggested that the models became better at distinguishing patients who would not require ventilation support, potentially reducing unnecessary preparation and resource allocation. The overall performance metrics supported the clinical utility of these predictive models for ventilator management and ICU resource planning.

Table 7: Model Performance by Patient Subgroups

Subgroup	Sample Size	Mortality Rate (%)	Model Accuracy (%)	AUC-ROC	Calibration
Age <50 years	876	14.2	91.7	0.932	0.028
Age 50-70 years	1247	22.8	88.9	0.908	0.039
Age >70 years	727	29.4	87.2	0.894	0.048
Male patients	1776	21.6	89.8	0.916	0.035
Female patients	1074	22.7	88.4	0.904	0.041
Medical admissions	1823	25.1	88.7	0.905	0.042
Surgical admissions	1027	16.8	91.1	0.925	0.031

The subgroup analysis revealed variations in model performance across different patient populations, with consistently superior performance in younger patients and surgical admissions compared to older patients and medical admissions. The age-related performance differences likely reflected the increased complexity and comorbidity burden in older patients, while the superior performance in surgical patients may be attributed to more predictable clinical trajectories and standardized perioperative care protocols. The gender-based performance differences were minimal, suggesting that the models provided equitable predictive performance across male and female patients. These findings supported the generalizability of the machine learning approach while highlighting the importance of considering patient-specific factors in clinical implementation.

Qualitative Evaluation

The clinical reasoning accompanying the SHAP analysis on the Machine Learning models outputs and their corresponding clinical reasoning, especially the clinical features weights per patient, was insightful. Their reasoning was that the models managed to grasp clinical associations and, in addition, some useful new associations that could add to the knowledge on ICU patients' prognoses. The SHAP values for some patients illustrated how some clinical features could positively or negatively change the corresponding dying chances and death probability, hence provided realistic explanatory grounds that could be used for operational clinical reasoning.

The reasoning on the scope of the decisions implemented by the models showed that the algorithms managed to find complex clinical intervariable interactions that would be impossible to uncover by standard methods. For example, the interaction of age and severity score showed that there were advanced age nonlinear relationships, in which advanced age was associated with different mortality risks with different levels of organ dysfunction. The models also showed that some lab driven clinical actions were dependent on the patients' physiological condition, demonstrating that certain clinical interventions were effective.

In terms of the why prediction accuracy varies over time, the models performed significantly better when clinical information was gathered in the first 24 to 48 hours following ICU admission as opposed to the time of admission. This suggested that the early clinical response to ICU treatment was prognostically meaningful and improved the prediction accuracy. The models were able to identify patients whose clinical courses were much different than what was anticipated at admission, which also enabled early intervention for patients who were expected to have the worse outcomes.

The analysis of the prediction intervals also showed the presence of prediction uncertainties that were appropriately tiered and of clinical interest in the event that the model requires improvement. Confidence in the prediction of models, which was ascertained at a high level, positively correlated with prediction accuracy and could be used to inform clinical decisions. Inversely, predictions which were established at the lowest confidence bore implications for clinical decisions, thus necessitating a prolonged surveillance period. Clinical relevance can be enhanced when the prediction model builder takes rational steps when it comes to assumption reliability.

Discussion

The evaluation of different approaches using machine learning to predict outcomes within Intensive Care Units (ICUs) certainly helps improve the clinically driven decision making and quality of care seems to be the focus of primary healthcare in Pakistan. It is apparent that the machine learning methods of prediction in the intensivist setting outperform the classical clinical scoring methods. This emphasizes the clinical relevance of advanced ICUs and specialized healthcare facilities while illustrating the importance of feature selection, model validation, and metric scrutiny. Also, rather encouraging is the machine learning model prediction proficient to dose correct a variety of care facilities models which vary in resource configuration and patient demographic profiles. Within clinical practice, predictive models can be advanced using the identification of clinical predictors, which validate established clinical constructs, and provide a new dimension of their use within clinical practice and quality improvement. This is particularly the case for larger sets of clinical and clinical laboratory case study data. This, in turn, makes the machine learning driven clinical intervention variables such as mechanical ventilation and use of vasopressors highly relevant. This, however, indicates the necessity of tackling the constructs the models oversimplify, both in patient outcomes and clinical decisions, implementing artificial intelligence in the healthcare domain, and the intricacy, and non-linearity of the clinical conditions.

There is a specific population these models of machine learning disregard, and this is substantiated by the outcome as well as the patient subgroup performance gaps. Advanced surgical admissions, paired with younger patient cohorts, flag this age group as a possible target for focused implementation deployment, maximizing potential real-world impact and downstream benefits, while retaining understanding, for the complex patients the other model use cases still indicate the model's need for additional refinement and validation in cross population adaptability, and real-world application appropriateness.

Conclusion

This study machine learning in the Pakistani healthcare context to predict the outcome and the timing of discharge for patients admitted into the ICU. The difference in performance between advanced techniques and clinical judgment suggests a triage judgment improvement, the availability of which ultimately improves care and resource use in resource-poor settings. The findings are also supported by the meticulous evaluation in several hospitals and patient cohorts, which strengthens the use of machine learning in critical care.

Considerable advances have been made in the specialization vocabulary concerning machine learning in healthcare. The therapeutic boundaries of a decade ago have been patently exceeded. Rational decision making defined within the confines of a crystal-clear problem space is now recognized as the central function of the aligned streams of evidence constituting what is now termed Constructivist Paradox. Deep learning frameworks support these systems by composing a synthesis of the decision boundaries constituting a defined answer. Such systems typically produce a plurality of results from which the user is expected to apprehend the predominant ideology which centers the set of outcomes from the suite of models constituting the machine learning paradigm.

These phenomena have come to prominence simultaneously with the diversification of machine learning models. The transformation of complex problems has joined with constructionist thinking in attempts to address shortcomings within the real-world application of healthcare. The generated systems attribute a distinct brand of constructivist supervision to the policy context surrounding the elastic boundaries of the research process. Large scale datasets in a condition of unsupervised learning pose computationally real-world elusive complexities to these models where the domain of application is healthcare.

These elements show a newly emerging shared intangible structure because of the distributed logic infrastructure with machine learning identifiers dominant in the purpose. These systems now fall within the range of accuracy of trained medical practitioners, sidelining basic tier health practitioners. The design elements from which the autonomy surface is extricated now, in many respects, render the operators as surrogates as to the models which operate in a scaled machine learning loop laptop systems for a decoupled approach to deployment of softer collaboration. Within the range of predictive agnosticity in the agnostic querying of elements for a blend of machine learning models, a neck-to-neck confrontation ensues with the golden tier practitioners.

From the back-end of the motifs, there is a range of models which indicate an agnosticity to the machine learning model across a health care boundary. To the practitioners managing the model interfaces monitoring layer, these outcomes stem from decision frames as bounded by induction within the learning tables. Each mechanism diverse in flavor has an agnostic domain which practitioners with a low to moderate ability to deploy machine learning models regard as default config. Edge cases are an encompassing default which silently surfaced in streams of shamans plugged in as core to the designed control elements.

Predictions for research assets are derived from the core stoic layer in an investment model. These surfaces are bounded practitioner models. Framing elements of the practitioner domain enrich the repositories these interfaces cull from. With the pilot systems designed under tiered parameters and gradient targets, shifting in the value fabric for dominant compute frameworks tiered targets propagate energies that let the models steer to optimal outputs. Stitched machine learning components in hospital systems are rendered as supple interfaces. These surfaces anchor policy flexibility.

Recommendations

When implementing machine learning algorithms to predict patient outcomes in the ICU, healthcare organizations need to first refine the organization's system architecture surrounding the data collection system, focusing on real-time reporting and collection of all relevant data. Second, within the bounds of the institution's capabilities and resources, the machine learning algorithms are to be tailored. Lastly, the machine learning algorithms need to be validated across all relevant patient populations. There are other implementation challenges, such as embedding the algorithms in clinical workflows, ease of use interfaces, training the clinicians, and reporting real-time outcomes of the algorithms. There are still other governance issues, such as the clinical AI ethics, workflows, and the governance frameworks on the continuous improvement of clinical outcomes and patient decision support.

Authors' Notes on Request

[1]**Saira Khurram** is an emerging Research Scholar from Pakistan, gaining recognition both nationally and internationally for her contributions in Biological Sciences, Psychology, and Education. She has worked extensively in the private sector, serving at Roots IVY Educational Complex as a Program Leader and Subject Administrator, where she has led international qualifications including IGCSE, O-Level, A-Level, AQA, and Pearson Edexcel Biology. Alongside her teaching and leadership, she has consistently advanced her research career, producing impactful publications in renowned international journals.

Saira Khurram's research portfolio demonstrates excellence and global relevance. She has more than 25 publications and Some of her most cited publications include:

- Springer – Wireless Personal Communications
- ScienceDirect – Chemosphere
- EBSCOhost – Personal and Ubiquitous Computing
- Journal of Political Studies and Analysis (JPSA) – Pakistan
- IEEE Xplore
- Dialogues SSR
- Journal of Political Studies and Analysis (JPSA) – Article 176

Her academic influence is also reflected through Google Scholar citations, with contributions recognized by the Higher Education Commission (HEC) of Pakistan:

- HEC-recognized publication 1
- HEC-recognized publication 2
- HEC-recognized publication 3

Her Springer publication in Wireless Personal Communications is particularly notable, holding a strong ranking internationally and being cited by scholars across multiple countries. Through her teaching, curriculum leadership, and high-impact research, Saira Khurram continues to bridge the gap between academic scholarship and global innovation.

[2]**Wajeeha Rehman** is an aspiring biomedical sciences student with a strong record of research, leadership, and community engagement. She has volunteered over 30 hours with WWF, launched an anti-littering passion project (2021), and founded the CAS initiative "Her Health Hub" to promote women's health. Her academic experiences include Biomedizone Camps (2023 & 2024), Research Cohort 101 by Tabeebs, a virtual gynecology education series, and her first sponsored research project on household air quality and microbial growth. She has also attended the NIBGE workshop and the 9th Belt and Road Teenager Maker Camp, China.

Wajeeha's scholarly engagement extends to membership in the New York Academy of Sciences and research collaborations with NGOs. She earned second place in the 'Women in Tech' essay competition for her article on technology and endometriosis care. Beyond science, she is a voice-over artist for Cambridge International Urdu books, showcasing her versatility. Through her blend of research, advocacy, and creativity, Wajeeha stands out as a socially conscious innovator dedicated to advancing health, science, and community impact.

REFERENCES

Adlung, L., Cohen, Y., Mor, U., & Elinav, E. (2021). Machine learning in clinical decision making. *Med, 2*(6), 642-665.

Aslam, M. W., Aslam, S., Aslam, N., Aslam, T., & Aslam, J. (2025a). A COMPREHENSIVE REVIEW OF WEARABLE HEALTH DEVICES: ADVANCES, CHALLENGES, AND FUTURE DIRECTIONS.

Aslam, M. W., Aslam, S., Aslam, N., Aslam, T., & Aslam, J. (2025b). FINANCIAL FEASIBILITY OF IMPLEMENTING SMART SAFETY TECHNOLOGIES IN ELECTRICAL ENGINEERING PROJECTS: A REVIEW OF CURRENT STATUS AND FUTURE PROSPECTS. *Spectrum of Engineering Sciences, 3*(3), 557-567.

Bai, Y., Gu, B., & Tang, C. (2025). Enhancing real-time patient monitoring in intensive care units with deep learning and the internet of things. *Big Data*.

Brnabic, A., & Hess, L. M. (2021). Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making. *BMC medical informatics and decision making, 21*(1), 54.

Chakilam, C. (2022). Integrating Machine Learning and Big Data Analytics to Transform Patient Outcomes in Chronic Disease Management. *Journal of Survey in Fisheries Sciences, 9*(3), 118-130.

Chen, M. (2024). *Massive data processing and explainable machine learning in neonatal intensive care units.* Université de Rennes,

Hong, N., Liu, C., Gao, J., Han, L., Chang, F., Gong, M., & Su, L. (2022). State of the art of machine learning–enabled clinical decision support in intensive care units: literature review. *JMIR medical informatics, 10*(3), e28781.

Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of healthcare engineering, 2021*(1), 6679512.

Kothinti, R. R. (2024). Deep learning in healthcare: Transforming disease diagnosis, personalized treatment, and clinical decision-making through AI-driven innovations. *World Journal of Advanced Research and Reviews, 24*(2), 2841-2856.

Lyu, Y., Xu, Q., Yang, Z., & Liu, J. (2023). Prediction of patient choice tendency in medical decision-making based on machine learning algorithm. *Frontiers in Public Health, 11*, 1087358.

Mani, R. K., Bhatnagar, S., Butola, S., Gursahani, R., Mehta, D., Simha, S., . . . Deodhar, J. (2024). Indian society of critical care medicine and Indian association of palliative care expert consensus and position statements for end-of-life and palliative care in the intensive care unit. *Indian Journal of Critical Care Medicine: Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine, 28*(3), 200.

Mustafa, Z. U., Tariq, S., Iftikhar, Z., Meyer, J. C., Salman, M., Mallhi, T. H., . . . Seaton, R. A. (2022). Predictors and outcomes of healthcare-associated infections among patients with COVID-19 admitted to intensive care units in Punjab, Pakistan; findings and implications. *Antibiotics, 11*(12), 1806.

Okore, L. A. (2023). *Experiences of Critical Care Nurses in End-of-life Care to Adult Patients in the Critical Care Unit, Kenyatta National Hospital.* University of Nairobi,

Pourhomayoun, M., & Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart health, 20*, 100178.

Sanchez-Martinez, S., Camara, O., Piella, G., Cikes, M., González-Ballester, M. Á., Miron, M., . . . Bijnens, B. (2022). Machine learning for clinical decision-making: challenges and opportunities in cardiovascular imaging. *Frontiers in cardiovascular medicine, 8*, 765693.

Saqlain, M., Gao Xiaoling, D. S. A. S., & Hussain, M. Journal of Population Therapeutics & Clinical Pharmacology.

van Beinum, A. (2021). *Reimagining death: A posthuman analysis of life support and the concept of death in the intensive care unit.* Carleton University,