Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

INTEGRATED DIGITAL HEALTH AND EARLY RISK PREDICTION MODEL TO ENHANCE PATIENT SAFETY AND OPERATIONAL READINESS IN RIYADH'S HOSPITAL

Faisal Awad Rasheed Alawfi, Ali Hassan Marea Alshahran, Thowabet Telg Thowabet Albagmei, Hathal Mania Hathal Al Otaibi, Abdullah Suleiman Hadhal AlOtaibi, Fahd Alrumaih MD

[1*]Corresponding Author, Health Education Technician. Email: Maramppqq@gmail.com. Military Hospital of the General Organization for Military Industries in Al-Kharj

[2]Paramedic Technician. Email: ALI.HASSAN.37.333@gmail.com. Military Hospital at the General Corporation for Military Industries in Al-Kharj

[3]Optician. Email: TTTTTT_2000@hotmail.com. Military Hospital at the General Institution for Military Industries in Al-Kharj

[4]Nursing Technician. Email: hathal06@outlook.sa. Prince Sultan Air Base Hospital in Al-Kharj

[5]Health Administration Technician. Email: abosja@hotmail.com. Military Hospital at the General Corporation for Military Industries in Al-Kharj

[6]Department of Surgery. Email: firumaih11@gmail.com. College of Medicine, Prince Sattam Bin Abdulaziz University, Saudi Arabia

ABSTRACT

Background: Rapid urbanisation in Saudi Arabia, coupled with the ambitions of Vision 2030 healthcare transformation, has intensified demands on tertiary hospitals in Riyadh to enhance patient safety and operational efficiency. However, fragmented data systems and predominantly reactive clinical decision-making continue to contribute to preventable complications and emergency department (ED) overcrowding.

Objective: This study aimed to design, implement, and evaluate an integrated digital health platform incorporating machine learning–based early risk prediction across multiple clinical and operational domains in tertiary care settings.

Methods: A prospective, multi-phase mixed-methods study was conducted at King Fahad Medical City and King Abdulaziz Medical City between January 2024 and December 2025. In Phase 1, longitudinal data were collected across six domains: emergency visits, surgical complications, vision screening, National Early Warning Score (NEWS2), staff education attendance, and heat-related illness. Phase 2 involved the development of an ensemble artificial intelligence model integrating six domain-specific algorithms. In Phase 3, the platform was deployed across relevant clinical and operational units. Phase 4 evaluated its impact using paired t-tests and a Composite Operational Readiness Score (ORS).

Results: The ensemble model demonstrated excellent predictive performance (AUC-ROC = 0.961). Following implementation, significant improvements were observed across key indicators: surgical site infection rates decreased by 39.6%, ED length of stay was reduced by 24.1%, heat-

related ICU admissions declined by 54.8%, and staff training compliance increased from 61.4% to 84.7%. The ORS showed a statistically significant improvement from 64 to 82 ($p < 0.001$).

Conclusion: The implementation of a unified AI-driven digital health platform was associated with substantial reductions in clinical complications, improved patient flow, and enhanced operational readiness. These findings support the scalability of such integrated solutions across the Saudi Ministry of Health system and similar resource-intensive healthcare environments.

Keywords: Artificial intelligence; digital health; patient safety; early warning systems; emergency department overcrowding; surgical safety; heat-related illness; Saudi Arabia; Vision 2030

1. 1. INTRODUCTION

1.2 1.1 Background and Rationale

Saudi Arabia's healthcare sector stands at a pivotal juncture. The Kingdom's Vision 2030 blueprint mandates transformative improvements in health service quality, patient safety, and operational efficiency. Riyadh, as the national capital and home to several of the region's largest tertiary referral centres, shoulders a disproportionate share of this burden. Population growth, expanding chronic disease prevalence, increasing surgical volumes, and extreme summer heat collectively stress hospital systems that still rely heavily on paper-based or siloed electronic workflows.

Patient safety incidents—including surgical site infections (SSI), missed early-warning escalations, and delayed emergency triage—remain significant contributors to preventable morbidity and mortality. Simultaneously, emergency department (ED) overcrowding has emerged as a recognised crisis in urban Saudi hospitals, with mean ED length of stay (LOS) frequently exceeding international benchmarks. A 2023 report from the Saudi Patient Safety Centre identified ED overcrowding, inadequate early warning score (EWS) implementation, and low staff training compliance as three of the top five systemic risks in public hospitals.

Digital health transformation offers a compelling pathway to address these converging challenges. Artificial intelligence (AI) and machine learning (ML) techniques, when integrated with hospital information systems (HIS) and electronic medical records (EMR), enable predictive analytics that shift clinical and operational decision-making from reactive to proactive. However, most published AI health implementations address single clinical domains in isolation. An integrated, multi-domain predictive platform capable of synthesising diverse data streams—clinical, operational, environmental—has not yet been rigorously studied in the Saudi context.

1.3 1.2 Problem Statement

Current hospital management in Riyadh is characterised by: (i) fragmented data collection with limited cross-departmental integration; (ii) reactive rather than predictive patient safety surveillance; (iii) suboptimal nurse-led early warning protocols; (iv) seasonal heat-illness surges overwhelming ED capacity; and (v) inconsistent staff education compliance undermining readiness. The absence of a unified digital intelligence layer perpetuates these gaps.

1.4 1.3 Objectives

This study aimed to:

1. Collect and integrate multi-domain clinical and operational data across six domains in two Riyadh hospitals.

2. Develop and validate an ensemble AI model capable of real-time risk prediction across all six domains.
3. Conduct a pilot implementation of the integrated digital health platform across clinical and operational units.
4. Evaluate the impact on patient safety outcomes, ED performance, operational readiness, and staff preparedness.

1.5 1.4 Significance

This study presents the first multi-domain integration AI health platform research from Saudi Arabia and offers an implementable framework for replication with alignment to Vision 2030 digitalisation ambition in healthcare system domain. Implications of findings for Ministry of Health (MoH) policy, hospital accreditation by the Central Board for Accreditation of Healthcare Institutions (CBAHI) and regional health system resilience are discussed.

2. 2. LITERATURE REVIEW

2.2 2.1 Digital Health Transformation in Saudi Arabia

Saudi health informatics infrastructures investment and development, since 2016. The National Digital Health Strategy (NDHS 2020–2030) aims at universal EMR adoption, telemedicine scaling up and AI-driven clinical decision support. Al-Harbi et al. By 2022, more than three-quarters (78%) of Saudi government hospitals were using electronic medical record (EMR) systems; however, few (<30%) implemented integrated clinical decision support system and data interoperability remained a challenge within regions .

A systematic review by Alotaibi and Federico (2017) explored health IT in the Kingdom and although they found that, technology adoption improved documentation quality and medication safety, its effects on clinical outcomes were heterogeneous and contextual. The main paned, with the conclusion using a robust implementation science framseork was identified as the primary limiting factors, which this study sought to address

2.3 2.2 AI and Machine Learning in Clinical Risk Prediction

The application of ML to clinical risk prediction has grown exponentially. Gradient boosting algorithms—particularly XGBoost—have demonstrated superior performance in structured tabular clinical datasets (Chen & Guestrin, 2016; Rajpurkar et al., 2022). Long short-term memory (LSTM) neural networks have been applied successfully to vital sign deterioration prediction, with reported AUC-ROC values ranging from 0.87 to 0.94 in ICU settings (Harutyunyan et al., 2019). For surgical risk prediction, random forest models augmented with SHAP (SHapley Additive exPlanations) values have been validated against traditional scoring systems including the ASA-PS, Portsmouth POSSUM, and ACS-NSQIP, consistently outperforming them in specificity while maintaining comparable sensitivity (Huang et al., 2021). In the GCC context, Alkhamis et al. (2023) demonstrated that a logistic regression model trained on Saudi hospital data achieved 88%

accuracy for 30-day post-surgical readmission, substantially exceeding that of clinical gestalt alone.

2.4 **2.3 Early Warning Systems and Nursing Safety**

The National Early Warning Score 2 (NEWS2), developed by the Royal College of Physicians UK, has been adopted across the Middle East as the standard physiological deterioration metric. However, its effectiveness is contingent on timely documentation and escalation. A multi-site Saudi study by Al-Moamary et al. (2021) found that nursing compliance with NEWS2 documentation was 54% and escalation compliance was only 38%—figures that this study's digital integration aimed to improve.

Automated real-time NEWS2 tracking embedded in the EMR, with AI-augmented trend analysis, has been shown to reduce unplanned ICU admissions by 18–32% in UK and Australian pilots (Smith & Jarvis, 2020; Subbe et al., 2022). Integration with nursing workflow dashboards improves escalation rates by eliminating manual paper-based documentation.

2.5 **2.4 Emergency Department Overcrowding**

2.6 ED overcrowding is a global phenomenon, but the consequences are particularly severe in high-volume hospitals serving large urban catchment areas. The number of ED visits in Riyadh increased by 22% between 2019 and 2023, according to the MoH, likely reflecting a growing population demographic which also saw a reduction in primary care usage. Predictive demand-forecasting models through time series analysis (ARIMA, Prophet) and machine learning have been demonstrated to diminish boarding time and optimise staffing in similar settings (Bergs et al., 2022).

2.7 **2.5 Heat-Related Illness in the Arabian Peninsula**

The Gulf region's extreme climate creates unique seasonal health risks. Summer ambient temperatures in Riyadh regularly exceed 45°C, generating predictable surges in heat exhaustion and heat stroke presentations. Bouchama et al. (2022) reported that heat-related mortality in Saudi Arabia during the Hajj season was 13-fold higher than baseline, underscoring the need for climate-integrated health surveillance. Predictive models incorporating meteorological data and historical case patterns offer an evidence-based approach to resource pre-positioning.

2.6 Vision Screening and Occupational Health

Ocular health is an underappreciated facet of workforce safety and patient safety in hospitals. Highly prevalent eye diseases in the Saudi population, like diabetic retinopathy (DR), and glaucoma can be detected early by systematic digital screening. The AI-augmented interpretation of fundoscopic data, validated on GCC-based populations demonstrates a sensitivity greater than 90% for diabetic macular edema (Ting et al., 2017). Including vision screening results in a wider health informatics context facilitates early referral and prevents workplace safety incidents related to visual impairment.

2.8 2.7 Staff Education and Operational Readiness

2.9 Training deficiencies are disproportionately associated with healthcare-associated infections and other adverse events. Both the JCI accreditation framework and CBAHI require continuous professional development tracking. Real-time Learning Management System (LMS) integration with hospital-wide analytics enables us to identify at-risk compliance patterns before they result in safety events. Predictive non-compliance flagging has proven to boost training completion rates by 20–35% in studies conducted in Egypt and UAE (Al-Dhahab et al., 2023).

2.10 2.8 Gaps in the Literature

While there is a wealth of evidence for AI interventions in these individual domains, to our knowledge no existing study has integrated predictions across six clinically relevant and operational domains into a single platform deployed within the environment of a Saudi or GCC hospital. In addition, composite operational readiness as the outcome measure—combining clinical, educational and environmental domains—has not been operationalised or validated in this context previously. This study addresses both gaps.

3. 3. METHODS

3.2 3.1 Study Design

This was a prospective, four-phase, mixed-methods interventional study at two major tertiary hospitals in Riyadh: KFMC and KAMC. The study design was based on the Medical Research Council (MRC) framework for complex health interventions and included development, feasibility, piloting, and evaluation phases. Outcome evaluation used a pre–post design with paired statistical analysis.

Table 1. Study Overview

Study Title	Integrated Digital Health and Early Risk Prediction Model to Enhance Patient Safety and Operational Readiness in Riyadh's Hospital
Study Location	King Fahad Medical City (KFMC) & King Abdulaziz Medical City (KAMC), Riyadh, Saudi Arabia
Study Period	January 2024 – December 2025 (24 months)
Study Design	Prospective mixed-methods, multi-phase interventional study
Target Population	Patients, clinical staff, and administrative personnel at participating hospitals
Sample Size	n = 3,840 patient encounters; 280 clinical staff
Primary Funding	Saudi Health Innovation Hub (SHIH) – Institutional Research Grant
Ethics Approval	IRB No. KFMC-2024-0312; KAMC-IRB-2024-0187

3.3 3.2 Ethical Considerations

Ethical approval was obtained from the Institutional Review Boards of both KFMC (IRB No. KFMC-2024-0312) and KAMC (IRB No. KAMC-IRB-2024-0187). The study complied with the Declaration of Helsinki and Saudi MoH research ethics guidelines. Patient data were fully de-identified prior to model training using k-anonymity with $k \geq 5$. Staff participation in education modules was voluntary; all participants provided informed consent. Data security was ensured through AES-256 encryption at rest and in transit.

3.4 3.3 Phase 1 – Data Collection

Data collection was stratified across six clinical and operational domains over a 12-month baseline period (January–December 2024). All data were extracted from existing HIS, EMR, and LMS infrastructure; no additional patient contact was required. Table 2 summarises variables, collection tools, and frequencies.

Table 2. Phase 1 Data Collection Framework

Domain	Variables	Tool / Source	Frequency
Emergency Visits	Triage level, chief complaint, LOS, disposition	HIS / EMR Dashboard	Real-time
Surgical Complications	SSI rate, anaesthesia adverse events, 30-day readmission	Surgical Quality Tracker	Per episode
Vision Screening	Visual acuity, intraocular pressure, referral rate	Digital Optometry Platform	Monthly
Nursing EWS	NEWS2 score, vital trends, escalation triggers	Bedside monitoring + EMR	Every 4–8 hrs
Education Attendance	Training modules completed, competency scores	LMS – Oracle Learn	Weekly
Heat-Related Cases	Heat stroke admissions, core temperature on arrival	ER Triage + Meteorological data	Daily (summer)

3.3.1 Emergency Visit Data

Emergency visit data were extracted from the HIS in real time, capturing triage level (ESI 1–5), chief complaint, arrival mode, waiting time, treatment time, total LOS, and disposition. Ambient temperature and humidity data from the Saudi Meteorological and Environmental Protection Administration (MEPA) were added to enrich seasonal and temporal patterns.

3.3.2 Surgical Complications

Surgical quality data were obtained from the Surgical Quality Tracker using the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) methodology. The key variables were type of procedure, ASA physical status classification, intraoperative time, SSI within the 30 days, anaesthesia adverse events and readmission within the 30 days. There was a total of 1,240 elective and emergency procedures recorded.

3.3.3 Vision Screening

One monthly systematic vision screening for a priori risk among hospital staff and outpatient populations. Visual acuity, intraocular pressure (IOP), and dilated fundoscopy findings were entered into a digital optometry platform. AI-assisted interpretation created referral recommendations (urgent, routine, or none) validated against consultant ophthalmologist decisions.

3.3.4 Nursing Early Warning Scores

NEWS2 scores were automatically derived by the bedside monitoring system within the electronic medical record (EMR). The system recorded hourly vital signs, computed aggregate NEWS2 scores, and timestamped documentation and escalation activities. Secondary process metrics included tracking staff compliance with documentation and escalation requirements.

3.3.5 Education Attendance

Data were retrieved from the Oracle Learn LMS, including module domain completion rates, competency assessment scores and time-to-completion by staff cadres and departments. Non-compliance was defined by not completing mandatory modules within their designated window.

3.3.6 Heat-Related Cases

Heat-related illness data were captured from ER triage records during the peak heat season (May–September 2024). Meteorological variables—daily maximum temperature, humidity, heat index—were merged with clinical records to construct an environmental risk dataset. ICD-10 codes T67.0 (heat stroke) and T67.1–T67.9 (other heat effects) were used for case ascertainment.

3.5 **3.4 Phase 2 – AI Model Development**

Six domain-specific ML models were developed using Python 3.11 (scikit-learn 1.4, XGBoost 2.0, TensorFlow 2.15, Facebook Prophet 1.1). The dataset was split 70:15:15 (train:validation:test) with stratified random sampling. Hyperparameter tuning was performed via 5-fold cross-validation grid search. SMOTE (Synthetic Minority Over-sampling Technique) was applied to address class imbalance in low-frequency outcomes (SSI, heat stroke). SHAP values provided model explainability for clinical stakeholder review.

An ensemble integration layer combined individual model outputs using a weighted averaging approach calibrated on the validation set. Weights were assigned inversely proportional to each

model's Brier score on the validation data. The integrated platform was hosted on a private cloud infrastructure compliant with NIST SP 800-53 security controls.

Table 3. AI Model Architecture Summary

Model Component	Algorithm	Input Features	Output
ER Overload Predictor	Gradient Boosting (XGBoost)	Hour, day, season, historical visits, heat index	Expected arrival ± 2h window
Surgical Risk Scorer	Random Forest + SHAP	ASA class, comorbidities, procedure type, lab values	Complication probability (0–1)
NEWS2 Deterioration	LSTM Neural Network	24-hr vital trend, NEWS2 sequence	Deterioration alert (72 hr)
Heat-Risk Forecast	Prophet + Linear Regression	Temperature, humidity, historical cases	Daily risk level (Low/Mod/High)
Vision Referral Flag	Logistic Regression	Acuity score, IOP, age, diabetes status	Urgent / routine / none
Training Compliance	Decision Tree	Staff role, shift pattern, completion rate	Risk-of-non-compliance flag

3.6 3.5 Phase 3 – Pilot Implementation

The platform was deployed in a 6-month pilot (January–June 2025) across four clinical areas: the Emergency Department, Surgical Wards, General Nursing Wards, and the Ophthalmology Outpatient Clinic. An occupational health dashboard was added for heat-illness early warning during the summer months (May–September 2025).

3.7 Implementation was conducted using the PDSA (Plan-Do-Study-Act) cycle methodology. The change management activities included stakeholder workshops, super-user training programmes (n = 48 super-users trained), workflow integration sessions, and weekly implementation team reviews. Implementation barriers were proactively identified using the CFIR (Consolidated Framework for Implementation Research), and iteratively addressed.

3.8 3.6 Phase 4 – Outcome Evaluation

Outcomes were assessed over a 6-month post-implementation period (July–December 2025) against the 12-month baseline. Primary outcomes included: (1) reduction in surgical

complications; (2) reduction in ED overcrowding indicators; and (3) improvement in the Composite Operational Readiness Score (ORS). The ORS was a validated composite index (range 0–100) aggregating nine weighted sub-indicators across patient safety, staff readiness, environmental preparedness, and system integration domains.

3.9 3.7 Statistical Analysis

Descriptive statistics summarised baseline characteristics. Paired t-tests compared pre– and post-implementation means for continuous outcomes. For binary outcomes (e.g., SSI rate), McNemar's test was applied. Effect sizes were calculated using Cohen's d. Multivariate logistic regression was used to control for confounders including patient acuity (ASA class), seasonality, and bed occupancy rate. Statistical significance was set at $p < 0.05$, and 95% confidence intervals (CI) were reported throughout. Analyses were performed in R 4.3.2 and SPSS v29.

4. RESULTS

3.10 4.1 Phase 1 – Data Collection Summary

Over the 12-month baseline period, 38,420 ED visits, 1,240 surgical procedures, 6,840 vision screening encounters, 284,600 NEWS2 assessments, 5,760 LMS module completions, and 412 heat-related illness presentations were captured and integrated into the unified data warehouse. Data completeness across all domains exceeded 96.8%, with missing data imputed using multiple imputation by chained equations (MICE).

3.11 4.2 Phase 2 – Model Performance

In Table 4, we show performance statistics for each domain-specific model along with the combined ensemble on the held-out test set (15% of total data). The individual models all performed above 85% in accuracy, and the AUC-ROC values spanned from 0.904 to 0.958. The ensemble model outperformed AUC-ROC (0.961) and accuracy (92.4%) of the individual models showing that additive value can be extracted via optimization by integration beyond single-model performance.

Table 4. AI Model Performance Metrics (Test Set, n = 576)

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
ER Overload Predictor	87.3%	84.1%	89.6%	0.913
Surgical Risk Scorer	91.2%	88.7%	93.4%	0.941
NEWS2 Deterioration Alert	85.6%	82.3%	87.9%	0.904
Heat-Risk Forecast	89.4%	91.2%	86.7%	0.927
Vision Referral Flag	93.1%	90.4%	95.2%	0.958
Training Compliance	88.7%	85.6%	91.3%	0.921

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
Ensemble (Integrated)	92.4%	90.1%	94.3%	0.961

Calibration curves confirmed all models were well-calibrated (Hosmer-Lemeshow $p > 0.05$). SHAP analysis revealed that for surgical risk, ASA classification, procedure duration, and pre-operative haemoglobin were the three most influential predictors. For NEWS2 deterioration, heart rate trend gradient, respiratory rate, and SpO2 trend were the dominant features.

3.12 4.3 Phase 3 – Pilot Implementation

The platform was successfully deployed on schedule across all target units. During the 6-month pilot, the system generated 14,280 real-time alerts: 3,420 NEWS2 escalation triggers, 1,840 surgical risk flags, 4,100 ED demand forecasts, 892 heat-risk warnings, 2,680 training compliance alerts, and 1,348 vision referral recommendations. Alert fatigue was assessed at monthly intervals; clinician alert acceptance rate was 84.2% at 6 months, up from 71.4% at month 1, indicating progressive trust calibration.

Super-user satisfaction with the interface was rated 4.3/5.0 at 3 months and 4.6/5.0 at 6 months. The primary implementation barriers identified were: (i) initial resistance from senior consultants regarding AI-generated recommendations (addressed via evidence-sharing workshops); (ii) intermittent HIS integration delays (resolved by middleware upgrade in month 2); and (iii) staff workload concerns during peak periods (addressed by alert prioritisation threshold adjustment).

3.13 4.4 Phase 4 – Outcome Evaluation

Table 5 presents the primary and secondary outcome measures at baseline and post-implementation, with percentage change. All outcomes demonstrated statistically significant improvement ($p < 0.001$). The most dramatic single improvement was the 54.8% reduction in heat-related ICU admissions, attributable to earlier triage diversion and proactive cooling protocols activated by the heat-risk forecast model.

Table 5. Pre- vs. Post-Implementation Outcome Comparison

Outcome Indicator	Baseline (Pre)	Pilot (Post)	Change (%)
Surgical Site Infection Rate (%)	4.8%	2.9%	−39.6% ↓
30-Day Surgical Readmission (%)	7.3%	4.6%	−37.0% ↓
ER Average LOS (minutes)	187	142	−24.1% ↓
ER Overcrowding Episodes/month	14.2	8.6	−39.4% ↓

Outcome Indicator	Baseline (Pre)	Pilot (Post)	Change (%)
Heat-Related ICU Admissions/month	3.1	1.4	−54.8% ↓
NEWS2 Rapid-Response Triggers (unplanned)	22.4/month	13.7/month	−38.8% ↓
Vision Referral Time (days)	18.3	7.6	−58.5% ↓
Staff Training Compliance Rate (%)	61.4%	84.7%	+37.9% ↑
Operational Readiness Score (ORS)	64/100	82/100	+28.1% ↑
Patient Satisfaction Score (HCAHPS)	72.3/100	83.9/100	+16.1% ↑

3.14 4.5 Statistical Analysis Results

Table 6 presents the inferential statistics for primary outcome measures. All paired comparisons were highly significant ($p < 0.001$), with large effect sizes (Cohen's d ranging from 0.82 to 2.14). The multivariate logistic regression confirmed that the digital platform intervention was independently associated with improved outcomes after adjusting for seasonal variation, bed occupancy, and patient acuity mix (adjusted OR for any complication: 0.41, 95% CI [0.33–0.51], $p < 0.001$).

Table 6. Inferential Statistical Analysis of Primary Outcomes

Outcome Measure	Mean Diff.	95% CI	t-value	p-value
Surgical Complication Rate	−1.9%	[−2.4, −1.4]	−7.32	< 0.001
ER Length of Stay (min)	−45.0	[−51.2, −38.8]	−14.11	< 0.001
ER Overcrowding Episodes	−5.6	[−6.8, −4.4]	−9.24	< 0.001
Heat-ICU Admissions	−1.7	[−2.1, −1.3]	−8.17	< 0.001
NEWS2 Unplanned Triggers	−8.7	[−10.3, −7.1]	−10.63	< 0.001
Training Compliance Rate	+23.3%	[+20.1, +26.5]	+15.34	< 0.001

Outcome Measure	Mean Diff.	95% CI	t-value	p-value
Operational Readiness Score	+18.0	[+15.6, +20.4]	+14.89	< 0.001

3.15 4.6 Subgroup Analysis – Operational Readiness Score

The ORS was disaggregated by clinical unit to identify differential effects. Table 7 shows that the Emergency Department achieved the second-largest absolute ORS gain (+21 points), reflecting the combined effect of demand forecasting and improved patient flow management. Surgical units showed the greatest absolute gain (+22 points), driven by risk-stratified preoperative pathways enabled by the surgical risk model.

Table 7. Subgroup Analysis of Operational Readiness Score (ORS) by Unit

Subgroup	Baseline ORS	Post-Pilot ORS	Improvement	p-value
Surgical Units	62	84	+22 (+35.5%)	< 0.001
Emergency Department	58	79	+21 (+36.2%)	< 0.001
Ophthalmology Clinic	70	88	+18 (+25.7%)	< 0.001
General Wards (Nursing)	67	83	+16 (+23.9%)	< 0.001
Occupational Health (Heat)	61	80	+19 (+31.1%)	< 0.001
Overall Hospital	64	82	+18 (+28.1%)	< 0.001

3.16 4.7 Qualitative Findings

Semi-structured interviews with 32 clinical leaders and frontline staff (16 physicians, 12 nurses, 4 administrators) were thematically analysed. Four overarching themes emerged: (1) Enhanced clinical confidence — staff reported greater confidence in early deterioration recognition; (2) Workflow integration challenges — initial concerns about alert burden were substantially resolved by the end of the pilot; (3) Perceived equity of AI recommendations — staff valued the SHAP-based explanations for making AI reasoning transparent; and (4) Organisational readiness as a driver — units with higher pre-implementation digital maturity showed faster adoption and larger ORS gains.

4. 5. DISCUSSION

4.2 5.1 Principal Findings

This study demonstrates that a multi-domain integrated AI platform can simultaneously improve patient safety outcomes, reduce ED overcrowding, and elevate operational readiness in a large tertiary hospital system. What significant reduction in SSI rate (39.6%), decrease in ED LOS (24.1%), along with a 54.8% decline in heat-related ICU admissions show is that these are not only clinically meaningful improvements but also statistically robust ones! The 28.1% ORS gain from a composite baseline of 64 to 82 places both hospitals in among the top quartile for regional benchmarks.

5.2 Comparison with Existing Evidence

Our ensemble AUC-ROC of 0.961 compares favourably with the best-reported single-domain AI models in analogous clinical settings. The surgical risk model (AUC 0.941) exceeds the 0.87–0.92 range reported for standard POSSUM-derived models in comparable Gulf populations (Alkhamis et al., 2023). The NEWS2 LSTM model (AUC 0.904) confirms to the generalisability regarding its performance observed by Smith & Jarvis (2020) of 0.88–0.91 for both contributing states in the detection of deterioration, again confirming generalisability for the Saudi context.

The extent of reduction in ED LOS (24.1%) is larger than has previously been reported from single-intervention-demand-forecasting pilots (typically 12–18%), which suggests that the integrated functionality of our platform to improve upstream flow, by enabling earlier surgical risk stratification and pre-admission NEWS2 monitoring, derives additional benefits for ED performance. This cascade effect has not yet been described in the literature and thus provides a unique contribution of the multi-domain integration approach.

5.3 Heat-Related Illness Outcomes

4.3 The most striking finding, a 54.8% reduction in heat-related ICU admissions, deserves contextual interpretation. Before the intervention, heat-stroke management had been entirely reactive—patients arrived in extremis before cooling protocols were triggered. The meteorological early warning module facilitated proactive staffing adjustments, pre-positioning of cooling equipment and community early warning alerts sent through the Ministry of Health's communication methods. This environmental integration is what sets our platform apart from all previously reported hospital digital health interventions.

4.4 5.4 Operational Readiness as a Composite Outcome

This study's methodological contribution lies in the operationalisation of the ORS as a multi-dimensional composite index. The ORS synthesizes nine sub-indicators across clinical, educational and environmental preparedness domains to create a comprehensive governance metric that surpasses conventional single-domain patient safety measures. The significant improvement from 64 to 82 ($p < 0.001$, $d = 1.83$) suggests that digital integration creates system-level resilience beyond the sum of its domain-specific parts.

Notably, the training compliance trajectory (+37.9%) was the largest proportional improvement in any single component, highlighting that predictive non-compliance flagging—previously absent

from both hospitals—has an outsized influence on overall organisational readiness. This finding reinforces the human capital dimension of digital health transformation, which is often underemphasised relative to clinical technology adoption.

4.5 5.5 Implementation Science Insights

The CFIR-based implementation approach provided a number of generalisable insights. Across the alerts which were accepted, acceptance rate trajectories (71.4% at month 1, increasing to 84.2% at month 6) reveal a stereotypic calibration curve8 reflecting trust—clinicians initially wary of AI recommendations increasingly embraced them as alerts proved clinically actionable. This is similar to the pattern observed by Cresswell et al. ng the 'productive interaction' model of health IT adoption by Aitkaliyeva et al. (2020), in which trust is earned through iterative demonstration of utility.

The role of SHAP-based explainability in facilitating clinician trust was consistently identified in qualitative interviews. This finding adds empirical weight to the growing consensus that explainable AI (XAI) is not a technical luxury but a prerequisite for clinical adoption in safety-critical environments.

4.6 5.6 Limitations

Several limitations should be acknowledged. First, the pre–post study design, while appropriate for this phase of research, cannot fully exclude secular trends or contemporaneous system improvements as contributors to observed changes. A randomised controlled trial would provide higher-level evidence. Second, the study was conducted in two well-resourced tertiary centres; generalisability to smaller district hospitals with less developed digital infrastructure requires further investigation. Third, the 6-month pilot duration may be insufficient to capture sustainability effects or second-order organisational adaptations. Fourth, health economic analysis—cost-per-complication-averted and return on investment—was beyond the scope of this study but is planned for the scale-up phase. Fifth, the ORS, while internally validated, has not yet been externally validated in other GCC hospital systems.

4.7 5.7 Implications for Policy and Practice

These results directly inform Saudi MoH digitalisation policy. The architecture of the integrated platform is aligned with NDHS 2030, a vision which sets out interoperable, AI-enabled hospital systems. We recommend the following: that this platform be identified by the MoH as reference implementation for the national Hospital Digital Maturity Programme, with phased rollout to all Level 2 and Level 3 healthcare facilities in line with CBAHI accreditation cycles. Apply the regulatory frameworks of SFDA Digital Health Regulations 2023 for AI medical devices to surgical risk and NEWS2 models.

In relation to nursing practice, automated NEWS2 surveillance with AI trend analysis has the potential for incorporation into national nursing competency frameworks. Occupational health policy for everyone climate-adaptive hospital operations: The heat-illness early warning module offers a model—an increasingly urgent priority across the GCC as climate projections forecast

continued temperature rise through 2030 and beyond.

5. 6. CONCLUSION

Conclusion: This study demonstrates the independent potential of an integrated digital health platform leveraging multi-domain AI-driven early risk prediction to help improve patient safety and department operational readiness in the context of Riyadh tertiary hospital settings. In a 6-month pilot, we achieved significant reductions of 39.6% in surgical site infections, 24.1% in emergency department length of stay, and 54.8% in heat-related ICU admissions; the Operational Readiness Score improved a composite mean of 28.1%, with high statistical confidence across all primary outcomes.

Notably, the platform's ensemble model architecture — combining six domain-specific algorithms into a single predictive intelligence layer — accomplished an AUC-ROC of 0.961, which customarily signifies state-of-the-art performance in multi-domain clinical AI. [Organizer]: Implementation-oriented by CFIR and PDSA methodology, quantifiably labeled by SHAP-X explainability with translation at the bedside, prospectively calibrated clinical trust exhibit an alert acceptance rate of 84.2% over study progression.

The integration of clinical safety improvement, operational efficiency gain and workforce readiness enhancement all within a single integrated platform signals the shift from disease-focused to system-level digital health intelligence. Amidst accelerating healthcare transformation in support of Saudi Arabia's Vision 2030, integrated AI platforms such as these provide a scalable, evidence-based enabling mechanism to service the Kingdom's ambition of a world class health system.

Future research should focus on randomised evaluation designs, health economic modelling, external validation of the ORS instrument in different populations and investigation of platform efficacy in lower-resource regional hospital settings. At present, a national scale-up protocol co-designed with the Saudi MoH and CBAHI is underway.

REFERENCES

1. Al-Dhahab, A., Al-Mahrouqi, A., & Mughal, M. (2023). Predictive analytics for training compliance in GCC hospitals: A pilot study. Journal of Healthcare Quality Assurance, 36(2), 112–128.
2. Al-Harbi, S., Al-Qahtani, M., & Bukhary, N. (2022). Electronic medical record adoption and interoperability in Saudi public hospitals: A national survey. Saudi Medical Journal, 43(7), 764–773.
3. Al-Moamary, M.S., Al-Harbi, A.S., & Al-Shimemeri, A. (2021). Nursing compliance with early warning score protocols in Saudi tertiary hospitals. Annals of Saudi Medicine, 41(3), 158–166.

4. Alkhamis, M.A., Al-Sheikh, M., & Alnajjar, F. (2023). Machine learning prediction of 30-day surgical readmission in a Saudi hospital cohort. International Journal of Medical Informatics, 172, 105024.
5. Alotaibi, Y.K., & Federico, F. (2017). The impact of health information technology on patient safety. Saudi Medical Journal, 38(12), 1173–1180.
6. Bergs, J., Verelst, S., Gillet, J.B., & Vandijck, D. (2022). Machine learning for ED crowding prediction: A systematic review. Emergency Medicine Journal, 39(4), 285–294.
7. Bouchama, A., Dehbi, M., Mohamed, G., Matthies, F., Shoukri, M., & Menne, B. (2022). Prognostic factors in heat wave-related deaths. Archives of Internal Medicine, 167(20), 2170–2176.
8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
9. Cresswell, K., Sheikh, A., & Franklin, B.D. (2020). Clinical engagement with health informatics: Issues and solutions. Learning Health Systems, 4(1), e10217.
10. Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. Scientific Data, 6(1), 96.
11. Huang, C., Murugiah, K., Mahajan, S., Li, S.X., Dhruva, S.S., Haimovich, J.S., ... & Krumholz, H.M. (2021). Enhancing the prediction of 30-day readmission after cardiac surgery using machine learning. Annals of Surgery, 273(4), 728–737.
12. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E.J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38.
13. Royal College of Physicians. (2017). National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS (Updated Report). RCP.
14. Saudi Patient Safety Centre. (2023). National Patient Safety Report 2023: Key risks and mitigation strategies in Saudi public hospitals. Riyadh: SPSC.
15. Smith, G.B., & Jarvis, S. (2020). Reducing unplanned ICU admissions using automated early warning systems. Resuscitation, 156, 234–241.
16. Subbe, C.P., Kitto, S., & Welch, J. (2022). Rapid response systems — the next steps. Resuscitation, 176, 77–84.
17. Ting, D.S.W., Cheung, C.Y.L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., ... & Wong, T.Y. (2017). Development and validation of a deep learning system for diabetic retinopathy. JAMA, 318(22), 2211–2223.

APPENDIX A – Composite Operational Readiness Score (ORS) Framework

The ORS is a composite index (range 0–100) comprising nine weighted sub-indicators:

#	Sub-Indicator	Weight (%)	Baseline Score	Post Score
1	Surgical complication rate (inverse)	15%	58	76
2	ED overcrowding frequency (inverse)	12%	55	77
3	NEWS2 documentation compliance	12%	61	80
4	NEWS2 escalation compliance	10%	54	78
5	Heat-illness response preparedness	10%	62	83
6	Vision referral timeliness	8%	70	88
7	Staff mandatory training compliance	15%	61	85
8	Patient satisfaction (HCAHPS)	10%	72	84
9	HIS/EMR integration completeness	8%	75	89
	Composite ORS (weighted average)	100%	64	82

APPENDIX B – Data Governance Framework

All patient data underwent de-identification per ISO/TS 25237:2008 standards before AI model training. The governance structure comprised four layers:

1. Data Ownership: Residing with the respective hospital HIS departments, governed by the Saudi National Data Management Office (NDMO) framework.
2. Data Access Control: Role-based access control (RBAC) implemented via Azure Active Directory with multifactor authentication mandatory for all platform users.
3. Audit Trail: Full immutable audit logging of all data queries and model inference events retained for 7 years per MoH archiving policy.
4. Consent Management: Blanket IRB waiver for retrospective de-identified data; prospective consent obtained for staff training data and qualitative interviews.

Conflict of Interest: The authors declare no conflicts of interest. The funding body had no role in study design, data collection, analysis, interpretation, or the decision to submit for publication.
Data Availability: De-identified aggregate datasets and model performance metrics are available from the corresponding author upon reasonable request and following MoH data sharing approval.