Defensive Triage: Minimizing Insurance Liability via Bayesian Uncertainty Quantification in Emergency Queues

Shiming Wu

January 14, 2026

Abstract

Emergency Department (ED) overcrowding presents a dual crisis: it compromises patient safety and destabilizes the medical liability insurance market. High-cost malpractice claims, particularly those stemming from failure to diagnose ambiguous cases, drive up insurance premiums, thereby increasing healthcare overhead and limiting patient access to affordable care. Traditional triage protocols and deterministic Machine Learning (ML) models optimize for operational efficiency but ignore the fat-tail liability risks associated with diagnostic uncertainty. This paper proposes a novel "Defensive Triage" mechanism that integrates Bayesian Uncertainty Quantification into queueing logic. By prioritizing patients based on epistemic uncertainty, we effectively hedge against catastrophic misdiagnosis. Using a discrete-event simulation benchmarked against standard protocols, I demonstrate that this approach significantly reduces aggregate liability exposure. I argue that by lowering the frequency of high-severity claims, Defensive Triage allows insurers to stabilize premiums. This creates a virtuous economic cycle: reduced insurer risk lowers provider costs, ultimately expanding the capacity of the healthcare system to provide accessible, insured care to a broader population.

1 Introduction

Emergency Department (ED) overcrowding is a global public health crisis that leads to significant delays in treatment. While the clinical consequences—morbidity and mortality—are well documented, the financial ripple effects threaten the sustainability of the entire healthcare insurance ecosystem. From a legal and operational perspective, treatment delays introduce massive financial liability and direct medical costs. Specifically, Medical Malpractice claims related to "failure to diagnose" often arise when patients with ambiguous symptoms deteriorate while waiting in the queue.

Simultaneously, these same delays drive up health insurance expenditures, as a patient who could have been treated in a standard ward often requires expensive ICU escalation due to avoidable deterioration. These fat-tail risk events are financially catastrophic for hospitals, malpractice insurers, and health insurance providers alike.

The relationship between clinical risk management and insurance affordability is direct and critical. Within the malpractice sector, surging litigation costs force insurers to raise premiums to cover potential payouts, placing immense financial strain on healthcare providers and forcing hospitals to run on leaner margins or limit services. In the health insurance sector, the costs associated with treating complications from delayed care are passed directly to the consumer in the form of higher monthly premiums and increased deductibles. Consequently, the rising cost of both liability and health insurance becomes a significant barrier to entry for patients, leading to reduced coverage networks and a larger population of underinsured individuals. Therefore, reducing the frequency of catastrophic clinical events is not merely a legal defense strategy; it is a necessary step toward stabilizing insurance markets and ensuring equitable access to healthcare.

Current Machine Learning (ML) approaches to triage typically focus solely on maximizing clinical accuracy (e.g., predicting mortality). However, these models provide point estimates (e.g., Severity = 0.7) and ignore epistemic uncertainty—the model's lack of confidence in its own prediction. A model may assign a low risk score to a stable patient with high confidence, and a similar score to a complex, ambiguous patient with low confidence. Treating these predictions as identical is an operational failure. For the insurer, this lack of nuance results in hidden risks: the ambiguous patient is far more likely to trigger a million-dollar malpractice claim or a hundred-thousand-dollar ICU bill if left in the waiting room too long.

To address this, this research proposes a novel Defensive Triage mechanism. I implement a Bayesian Neural Network (BNN) that utilizes Monte Carlo (MC) Dropout to approximate the posterior distribution of clinical risk for each incoming patient. This methodology allows us to extract not only a predicted severity score but also a quantitative Risk Buffer representing the model's epistemic uncertainty. I then embed these dual metrics into a dynamic priority queue using an Upper Confidence Bound (UCB) optimization heuristic. By ranking patients according to a weighted severity score that takes into account both average severity and risk of severity, we explicitly prioritize patients based on the potential for catastrophic misclassification. This Defensive approach hedges against the most expensive risks by resolving clinical ambiguity early, effectively preventing both the legal triggers of malpractice and the medical triggers of high-cost ICU escalation.

My experimental results, derived from high-fidelity discrete-event simulations, validate the clinical and economic efficacy of this uncertainty-aware approach. I find that while a standard First-Come-First-Served (FCFS) policy leads to excessive rates of deterioration, even a Risk-Neutral AI triage (optimizing only for predicted mean) remains vulnerable to liability traps—ambiguous sepsis cases that the model misclassifies as low-risk. In contrast, my Defensive Triage mechanism achieves significant net savings by dramatically reducing the incidence of critical liability events and high-cost ICU escalations. Although this strategy slightly increases the wait times for clearly stable patients, the trade-off results in a more stable financial profile for insurers and a safer environment for patients. By lowering the aggregate cost of care and volatility of liability, this framework provides a pathway to lower premiums and broader healthcare coverage.

This paper integrates dynamic prioritization and queueing theory research. Traditional models often assume static priority levels (e.g., Triage Level 1–5). However, recent scholarship addresses the restless bandit problem, where the state of a job (patient health) deteriorates while waiting. Argon and Ziya [2009] established that when service times are unknown, optimal priority assignment must account for the distribution of holding costs. Building on this, Stanford et al. [2014] developed the mathematical foundation for Accumulating Priority Queues (APQ), a system where a customer's priority is modeled as a linear function of their wait time. This mechanism is critical in healthcare settings as it provides a rigorous method for managing the trade-off between clinical urgency and waiting time, ensuring that lower-priority patients do not suffer indefinite delays while their clinical risks potentially escalate. More recently, Diamant et al. [2018] and Dai and Shi [2019] utilized approximate dynamic programming (ADP) to optimize clinical pathways and inpatient overflow, respectively. While Diamant et al. [2018] explicitly models the stochastic nature of patient attrition, Dai and Shi [2019] demonstrates the efficacy of data-driven ADP policies in managing system-wide hospital capacity. While these models successfully account for biological deterioration, they typically assume perfect state information—that is, the system knows exactly how sick a patient is at any moment. My work challenges this assumption. I argue that in the Emergency Department, the patient's state is partially unobservable. My paper contributes to the literature by introducing epistemic uncertainty into the priority function, demonstrating that optimal queuing requires prioritizing not just the known sick, but the potentially catastrophic unknown cases.

This study contributes to literature of predictive analytics and the temporal displacement of care. A landmark study by Bardhan et al. [2020] introduced the concept of Temporal Displacement of Care (TDC). They argue that advanced analytics create value by enabling providers to predict disease progression and intervene earlier, effectively displacing care from a high-cost future state (e.g., ICU) to a lower-cost present state. Lin et al. [2017] further utilized Bayesian multitask learning to profile chronic care risks, providing granular risk strata for allocation decisions. While Bardhan et al. [2020] established the operational savings of TDC, my research extends this framework to the domain of risk management. I posit that TDC does not merely reduce hospital costs; it fundamentally reduces insurance liability. By specifically targeting ambiguous cases (high uncertainty), our Defensive Triage mechanism maximizes the economic value of information, offering a new theoretical link between information system analytics and malpractice insurance premiums.

This paper is related to literature about Clinical Disease Progression. The application of progression-based queueing is most mature in organ transplantation. Rizopoulos [2011] demonstrated that Joint Models (combining longitudinal biomarkers with survival analysis) significantly outperform static scores for liver transplant candidates. In the ED context, Henry et al. [2015] argues that manual scores often fail to capture non-linear deterioration in septic patients, necessitating continuous algorithmic monitoring. Medical studies typically optimize for clinical accuracy (e.g., Area Under the Curve) rather than system-level efficiency. Furthermore, transplant models rely on rich, longitudinal datasets that are unavailable in an acute ED setting. By utilizing a Bayesian Neural Network to quantify the uncertainty of single-point triage data, my paper operationalizes clinical progression models into a real-time decision support tool that aligns clinical safety with the financial constraints of the insurance ecosystem.

Finally, this research builds on literature about uncertainty quantification. Begoli et al. [2019] and Gal and Ghahramani [2016] argue that for AI to be safe in critical care, it must quantify what it does not know. Leibig et al. [2017] showed that leveraging uncertainty in diabetic retinopathy screening prevents diagnostic errors. My paper bridges the gap between these computer science methods and healthcare operations. While prior works use uncertainty to flag cases for human review, I propose using uncertainty as a ranking parameter in an automated queue. This research demonstrates that the Upper Confidence Bound (UCB) heuristic—typically used in reinforcement learning—can be adapted to minimize the fat-tail liability risks of ED operations.

The remainder of this paper is structured as follows. Section 2 details the problem formulation and the mathematical framework, including the Bayesian methodology. Section 3 presents the simulation results, comparing the proposed method against standard benchmarks. Finally, Section 4 discusses the implications and concludes.

2 Problem Formulation and Methodology

Standard triage protocols prioritize patients based on observed vital signs. However, patients with noisy or incomplete data often receive low-priority scores despite carrying hidden risks. In insurance terms, these patients represent Fat Tail risks. If a low-priority patient collapses due to unseen deterioration, the malpractice liability cost (C_{legal}) is orders of magnitude higher than the cost of treating a stable patient slightly later.

Let $\mathcal{Q}$ be the set of patients in the waiting room. The objective is to determine a scheduling policy π that minimizes the Total Expected Economic Loss (J), defined as the sum of operational costs and catastrophic liability exposure.

The proposed Defensive Triage system operates in two stages: (1) Probabilistic Disease Progression Modeling, and (2) Value-Based Queue Optimization.

2.1 Stage 1: Bayesian Prediction

I utilize the MIMIC-IV-ED database schema to define the input space. For a patient i, let $\mathbf{x}_i$ be the vector of vital signs recorded at triage:

$$\mathbf{x}_i = [\text{HR}, \text{SBP}, \text{Temp}, \text{O2}_{\text{sat}}, \text{RespRate}]^T \tag{1}$$

The components of this vector correspond to the standard MIMIC 'triage' table columns and are defined as follows:

- **HR:** Heart Rate (beats per minute).
- **SBP:** Systolic Blood Pressure (mmHg).
- **Temp:** Body Temperature (°F or °C).
- **$\mathbf{O2_{sat}}$:** Peripheral Oxygen Saturation (%).
- **RespRate:** Respiratory Rate (breaths per minute).

I define the ground truth severity, y_i, using the weighted NEWS2 (National Early Warning Score) framework. To predict this severity, I employ a Bayesian Neural Network (BNN). Unlike standard frequentist networks that learn a single set of fixed weights, a Bayesian approach assumes the weights $\mathbf{W}$ are random variables described by a posterior distribution given the training data $\mathcal{D}$.

Our goal is to compute the posterior predictive distribution for a new patient's risk y^*, which requires integrating over all possible weight configurations:

$$p(y^*|\mathbf{x}^*, \mathcal{D}) = \int p(y^*|\mathbf{x}^*, \mathbf{W})\, p(\mathbf{W}|\mathcal{D})\, d\mathbf{W} \tag{2}$$

This integral is analytically intractable for deep neural networks due to the high dimensionality of $\mathbf{W}$. To resolve this, I utilize Monte Carlo (MC) Dropout as an approximate Bayesian inference technique.

As demonstrated by Gal and Ghahramani [2016], training a neural network with dropout is mathematically equivalent to performing Variational Inference. Specifically, it approximates the true posterior $p(\mathbf{W}|\mathcal{D})$ with a simpler variational distribution $q(\mathbf{W})$ (a mixture of Gaussians) by minimizing the Kullback-Leibler (KL) divergence between them. Consequently, performing stochastic forward passes with dropout active at inference time is equivalent to sampling from the approximate posterior distribution of the weights.

For patient i, I perform T stochastic forward passes. I then approximate the predictive moments as follows:

$$\hat{\mu}_i \approx \frac{1}{T} \sum_{t=1}^{T} f^{\mathbf{W}_t}(\mathbf{x}_i) \tag{3}$$

$$\hat{\sigma}_i^2 \approx \frac{1}{T} \sum_{t=1}^{T} \left(f^{\mathbf{W}_t}(\mathbf{x}_i) - \hat{\mu}_i \right)^2 \tag{4}$$

Here, $\hat{\mu}_i$ represents the estimated clinical severity, while $\hat{\sigma}_i$ quantifies the *epistemic uncertainty*—the model's lack of knowledge regarding the patient's condition.

While the architecture of my model is designed for the real-world MIMIC-IV-ED dataset, this study utilizes a controlled discrete-event simulation to validate the theoretical bounds of the Defensive Triage hypothesis. I generate synthetic patient streams that statistically mirror the vital sign distributions found in MIMIC-IV. Patients are categorized into three latent classes: "Stable", "Classic Sepsis", and "Ambiguous Sepsis".

The 'Ambiguous Sepsis' class is the focal point of this study; these patients present with compensated vital signs (e.g., normal blood pressure but low temperature) that deterministic models often misclassify as stable. The simulation allows us to inject these specific liability traps at known rates to stress-test the algorithms. In future work, this data generation module will be replaced with direct sampling from the MIMIC-IV-ED 'triage' and 'vitalsign' tables to validate the model's performance on historical patient records.

To establish a ground truth metric for patient deterioration within the simulation, I map the MIMIC variables to the National Early Warning Score 2 (NEWS2). The aggregate severity score $y_{news} \in [0, 20]$ is defined as:

$$y_{news}(\mathbf{x}_i) = f_{RR} + f_{SpO_2} + f_{SBP} + f_{HR} + f_{Temp} \tag{5}$$

Each function maps a continuous variable to a discrete integer risk score $\{0, 1, 2, 3\}$ based on Royal College of Physicians guidelines (e.g., $f_{SpO_2}(v) = 3$ if $v \leq 91$). I define a Critical Event (Liability Trigger) as a binary outcome where $y_{news} \geq 7$.

2.2 Stage 2: Value-Based Queue Optimization

I model the ED as a single-server dynamic priority queue, characterized in Kendall's notation as an $M/G/1$ system. This framework is defined as follows:

- **M (Markovian):** Represents the arrival process, assuming that patients arrive according to a Poisson process with independent, exponentially distributed inter-arrival times.
- **G (General):** Denotes a general service time distribution. In our context, this accounts for the fact that treatment duration is not fixed but is a stochastic variable dependent on the patient's clinical severity (e.g., septic patients require longer stabilization times than stable patients).
- **1 (Single Server):** Indicates a single-resource constraint, representing the physician or treatment bay for which patients compete.

The decision variable in this system is the Scheduling Policy π. I aim to find the optimal policy π^* that minimizes the Total Expected Economic Loss (J):

$$\pi^* = \arg\min_{\pi} \sum_{i \in \mathcal{Q}} \left(C_{op}(w_i, s_i) + C_{liab}(w_i) \right) \tag{6}$$

2.2.1 Cost Components

The economic impact of the triage policy is quantified through two primary loss channels representing direct medical expenditures and indirect legal risk. The first component, operational cost (C_{op}), represents the direct medical loss incurred by the health insurer. The cost for patient i is modeled as the base treatment cost plus a potential escalation premium:

$$C_{op}(w_i) = C_{base} + \alpha(w_i) \cdot \mathbb{I}(\text{Deterioration}_i) \cdot (C_{ICU} - C_{base}) \tag{7}$$

In this formulation, C_{base} is approximately \$1,500, representing the cost of standard emergency department care and ward-based stabilization, while C_{ICU} is approximately \$50,000, reflecting the resources required for critical care intervention in cases of septic shock or acute failure. The term $\mathbb{I}(\text{Deterioration}_i)$ is an indicator function active for high-risk patients with a ground truth NEWS2 score of seven or greater. The coefficient $\alpha(w_i) = \min(1.0, w_i/300)$ models the linear accumulation of complications over a five-hour clinical window, where w_i denotes the waiting time.

The escalation term $\alpha(w_i) \cdot \mathbb{I}(\cdot) \cdot (C_{ICU} - C_{base})$ is grounded in three core principles of health economics and operations management. First, it accounts for the asymmetric cost of delay, acknowledging that clinical delays do not impact all patients equally. While the treatment cost for a stable patient remains relatively constant regardless of wait time, a deteriorating patient experiences a significant increase in resource requirements. Second, this formulation mathematically operationalizes the concept of temporal displacement of care (TDC) as described by Bardhan et al. [2020]. By accepting a marginal operational cost, such as slightly increasing the wait time of a stable patient, the system can intervene earlier in a high-risk patient's trajectory, effectively displacing care from a high-cost future state to a lower-cost present state. Finally, from an insurance standpoint, this represents the direct medical loss. The triage algorithm acts as a hedging mechanism for the insurer, minimizing the probability of incurring an expensive critical care claim when a more efficient allocation of resources could have stabilized the patient at a fraction of the cost.

The second component is the liability risk (C_{liab}), which captures the catastrophic fat tail risk associated with medical malpractice:

$$C_{liab}(w_i) = \mathbb{P}(\text{Critical Event}_i | w_i) \cdot L_{penalty} \tag{8}$$

In this expression, $L_{penalty}$ is set to \$1,000,000. A critical event is defined as a case where a patient with a ground truth NEWS2 score of seven or greater experiences a wait time w_i exceeding 180 minutes. This threshold represents the point at which a delay constitutes a breach of the standard of care, triggering significant legal and settlement risks that destabilize malpractice insurance premiums.

2.2.2 Approximation of the Optimal Policy

The objective defined in Equation (8) constitutes a dynamic scheduling problem with stochastic transitions and a continuous state space (the posterior distributions of all waiting patients). This is classified as a Restless Multi-Armed Bandit (RMAB) problem, as the state of waiting patients (their health) continues to evolve ("deteriorate") even when they are not being served.

Finding the exact optimal policy π^* requires solving the Bellman equation for the value function $V(\mathbf{S}_t)$:

$$V(\mathbf{S}_t) = \min_{a \in \mathcal{A}} \mathbb{E}\left[C(\mathbf{S}_t, a) + \gamma V(\mathbf{S}_{t+1})\right] \tag{9}$$

However, due to the high dimensionality of the state space $\mathbf{S}_t$ (the vector of vital signs and uncertainty estimates for all N patients in the queue), obtaining an exact analytical solution is computationally intractable (PSPACE-hard). Consequently, standard optimization techniques cannot be applied directly.

To overcome this intractability, I adopt an Index Policy approach. I decompose the global optimization problem into a set of independent indices (priority scores) for each patient. I approximate the optimal policy π^* using an Upper Confidence Bound (UCB) heuristic.

In Reinforcement Learning, UCB is used to balance exploration and exploitation. In our context, I adapt it to balance expected severity (mean risk) against potential catastrophic outcome (epistemic uncertainty). I define the Defensive Priority Score (S_i) as:

$$S_i(t) = \hat{\mu}_i + \lambda \cdot \hat{\sigma}_i \cdot \tau(t) \tag{10}$$

Where:

- $\hat{\mu}_i$ is predicted clinical severity (The "Exploitation" term).
- $\hat{\sigma}_i$ is epistemic uncertainty (The "Risk Buffer" / "Exploration" term).
- λ is the Liability Coefficient.
- $\tau(t) = 1 + \gamma(t - t_{arrival})$: A linear aging factor to prevent starvation of low-risk patients.

Since we cannot solve for the function π^* analytically, my "optimization" step consists of finding the optimal scalar value for λ.

- When $\lambda = 0$, the policy acts as a standard Risk-Neutral AI, optimizing only for the most likely outcome.
- As λ increases, the policy places a "shadow price" on uncertainty, effectively hedging against the Fat Tail liability costs (C_{liab}) defined in Equation (12).

In my experimental results (Section 5), I simulate the system across a range of λ values to empirically determine the setting that minimizes the Total Expected Economic Loss (J).

3 Results

To evaluate the efficacy of the proposed Defensive Triage mechanism, I conducted a simulation over 50 runs, each representing 48 hours of high-intensity ED operations. I compared three distinct queuing strategies to understand the trade-offs between operational efficiency and liability protection.

The first benchmark is the **First-Come-First-Served (FCFS)** strategy. This represents a naive queue where patients are treated strictly in the order of arrival, regardless of their vital signs. While this approach satisfies a basic definition of fairness regarding wait times, it performs poorly in terms of economic loss and patient safety. Because high-risk patients wait behind stable patients, the rate of deterioration is high. As shown in Table 1, this strategy incurs the highest Liability Cost and the highest number of Critical Events, demonstrating why triage is mandatory in modern healthcare.

The second benchmark is the **Standard AI Triage** (where $\lambda = 0$). This strategy sorts patients based solely on the predictive mean ($\hat{\mu}_i$) generated by the neural network. This represents the current state in predictive healthcare. The model prioritizes patients who are clearly sick. This drastically reduces Operational Costs compared to FCFS because sick patients are treated before they require ICU admission. However, this "Risk Neutral" approach fails to account for "Ambiguous" cases. Because the model is not confident in these predictions, their mean scores are moderate, causing them to wait. Consequently, a significant number of these ambiguous cases deteriorate, leading to substantial Liability Costs.

The third strategy is my proposed **Defensive AI Triage** (where $\lambda = 2.0$). By incorporating the uncertainty term ($\hat{\sigma}_i$) into the priority score, the system boosts the priority of ambiguous cases. Effectively, the system identifies that it lacks information about these patients and treats them earlier to resolve the uncertainty.

The selection of $\lambda = 2.0$ is grounded in both statistical and actuarial logic. Statistically, the UCB score $\hat{\mu} + 2\hat{\sigma}$ serves as a proxy for the 95th percentile of the risk distribution. In a clinical setting, this "defensive" stance ensures that the system accounts for the upper bound of potential deterioration. $\lambda = 2.0$ represents an optimal hedging strategy; given that the liability penalty ($L_{penalty} = \$1M$) is several orders of magnitude higher than the base treatment cost ($C_{base} = \$1.5K$), the system places a high "shadow price" on uncertainty to prevent catastrophic fat-tail events. As indicated in Table 1, this results in a slight increase in

Operational Cost compared to the Standard AI, but this is far outweighed by the dramatic reduction in Liability Cost. The Defensive strategy successfully prevents the majority of fat-tail critical events, resulting in the lowest Total Cost among all strategies.

Table 1: Comparative Results: Standard vs. Defensive Triage

Metric	**FCFS**	**Standard AI**	**Defensive AI**
Liability Cost	$17,660,000	$12,460,000	$10,500,000
Operational Cost	$1,014,569	$1,178,689	$819,565
Total Cost	**$18,674,569**	**$13,638,689**	**$11,319,565**
Critical Events (Avg)	17.66	12.46	10.50

Notes: This table compares costs across three benchmarks. FCFS relies on arrival time; Standard AI relies on predicted severity ($\lambda = 0$); Defensive AI incorporates uncertainty ($\lambda = 2.0$).

4 Conclusion and Discussion

This paper presents a theoretical and simulation-based framework for "Defensive Triage", bridging the gap between clinical operations, medical ethics, and actuarial risk management. I have demonstrated that standard predictive models, which optimize for point-estimate accuracy, are insufficient for minimizing the multi-dimensional financial risks inherent in the Emergency Department. By failing to account for epistemic uncertainty, standard models leave hospitals, malpractice insurers, and health insurance providers exposed to catastrophic fat-tail risks arising from ambiguous cases that deteriorate during wait times.

The broader economic implication of these findings extends across the entire healthcare insurance spectrum. My results suggest that by incorporating an uncertainty penalty into the priority score, the Defensive Triage algorithm effectively mitigates the dual drivers of healthcare inflation: malpractice litigation and high-acuity medical spending. While this approach incurs a marginal increase in wait times for clearly stable patients, the reduction in catastrophic outcomes is substantial. For health insurers, the prevention of avoidable clinical deterioration directly translates to fewer high-cost ICU escalations and shorter total hospital stays. For malpractice insurers, the resolution of clinical ambiguity early in the care pathway reduces the frequency of failure to diagnose claims.

By lowering the aggregate cost of both liability and medical intervention, Defensive Triage helps break the cycle of rising healthcare costs. Lower premiums for providers and insurers reduce the overall financial burden on the healthcare system, creating an environment where insurance coverage can be expanded to more patients at a reasonable price point. In this framework, the stability of the insurance market becomes an engine for healthcare equity.

In summary, this research argues that the goals of patient safety, clinical excellence, and financial sustainability are not mutually exclusive. By designing algorithms that explicitly hedge against uncertainty, I protect the patient from medical error, the hospital from legal volatility, and the insurance market from avoidable medical expenditures. This fosters a more resilient and accessible healthcare system for all.

Future work will focus on two primary areas. First, I will transition from synthetic simulation to the full MIMIC-IV-ED dataset, validating the Bayesian disease progression model against high-fidelity historical logs to ensure the identified "uncertainty signatures" match real-world patient profiles. Second, I aim to extend the queuing model to a multi-server environment ($M/G/k$) to better simulate the complexities of staffing constraints and resource allocation in live hospital settings, further refining the economic cost-benefit analysis of defensive scheduling.

References

Argon, N. T., & Ziya, S. (2009). Priority assignment under imperfect information on service times. *Manufacturing & Service Operations Management*, 11(4), 674-693.

Bardhan, I., Chen, H., & Karahanna, E. (2020). Chronic disease management: How IT and analytics create healthcare value through the temporal displacement of care. *MIS Quarterly*, 44(1), 1-28.

Begoli, E., Bhattacharya, T., & Kusnezov, D. (2019). The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence*, 1(1), 20-23.

Dai, J. G., & Shi, P. (2019). Inpatient overflow: An approximate dynamic programming approach. *Manufacturing & Service Operations Management*, 21(4), 894-911.

Diamant, A., Milner, J., & Quereshy, F. (2018). Dynamic patient scheduling for multi-appointment health care programs. *Production and Operations Management*, 27(1), 58-79.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning* (pp. 1050-1059).

Henry, K. E., Hager, D. N., Pronovost, P. J., & Saria, S. (2015). A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, 7(299), 299ra122.

Leibig, C., Allken, V., Ayhan, M. S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scientific Reports*, 7(1), 1-14.

Lin, Y. K., Chen, H., Brown, R. A., & Li, S. H. (2017). Healthcare predictive analytics for risk profiling in chronic care: A Bayesian multitask learning approach. *MIS Quarterly*, 41(2), 473-495.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3), 819-829.

Royal College of Physicians. (2017). *National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS*. London: RCP.

Stanford, D. A., Taylor, P., & Ziedins, I. (2014). Waiting time distributions in the accumulating priority queue. *Queueing Systems*, 77(3), 297-330.

A National Early Warning Score 2 (NEWS2) Reference

The ground truth severity labels used in this study are derived from the National Early Warning Score 2 (NEWS2), established by the Royal College of Physicians (RCP) in 2017. The system aggregates scores from seven physiological parameters to quantify acute illness severity.

A.1 Scoring System

Table 2 details the exact thresholds used to calculate the component scores. The total score for patient i is calculated as:

$$S_{news} = \sum_{j=1}^{7} s_j \tag{11}$$

where s_j corresponds to the score assigned to vital sign j.

Table 2: NEWS2 Scoring Thresholds (Scale 1)

Physiological Parameter	**Score 3**	**Score 2**	**Score 1**	**Score 0**
Respiration Rate (bpm)	≤ 8 or ≥ 25	21 – 24	9 – 11	12 – 20
SpO$_2$ Scale 1 (%)	≤ 91	92 – 93	94 – 95	≥ 96
Air or Oxygen?	–	Oxygen	–	Air
Systolic BP (mmHg)	≤ 90 or ≥ 220	91 – 100	101 – 110	111 – 219
Pulse (bpm)	≤ 40 or ≥ 131	111 – 130	41 – 50 or 91 – 110	51 – 90
Consciousness	–	–	–	Alert
	CVPU (Confusion, Voice, Pain, Unresponsive)			
Temperature (°C)	≤ 35.0	≥ 39.1	35.1 – 36.0 or 38.1 – 39.0	36.1 – 38.0

Note: SpO$_2$ Scale 1 is used for patients without hypercapnic respiratory failure. CVPU represents New Confusion, Response to Voice, Response to Pain, or Unresponsive.

A.2 Clinical Risk Thresholds and Liability Triggers

The NEWS2 system maps aggregate scores to clinical response levels. In the context of this study, these thresholds define the boundary between "routine care" and "catastrophic liability risk".

Table 3 illustrates why a score of ≥ 7 is utilized as the binary target Y_i for the Bayesian Neural Network.

Table 3: Clinical Risk Categories and Associated Liability

Total Score	**Clinical Risk**	**RCP Recommended Response**	**Liability Profile**
0 − 4	Low	Ward-based monitoring (4–6 hourly).	Minimal
Score of 3 (in any single parameter)	Low-Medium	Urgent review by ward-based doctor.	Low
5 − 6	Medium	**Urgent Response.** Hourly observation. Review by clinician with acute care skills.	Moderate
$\geq$ **7**	**High**	**Emergency Response.** Continuous monitoring. Immediate transfer to Critical Care/ICU.	**Critical** (Catastrophic Risk)