Prediction of Hospital Associated Infections During Continuous Hospital Stays

Rituparna Datta[2], Methun Kamruzzaman[6], Eili Y. Klein[3], Gregory R. Madden[4], Xinwei Deng[5], Anil Vullikanti[1,2], Parantapa Bhattacharya[1]

[1]Biocomplexity Institute & Initiative, University of Virginia
[2]Department of Computer Science, University of Virginia
[3]Johns Hopkins University School of Medicine and Johns Hopkins Center for Health Security
[4]Division of Infectious Diseases & International Health, University of Virginia Health System
[5]Department of Statistics, Virginia Tech
[6]Systems Biology, Sandia National Laboratory

Abstract

The US Centers for Disease Control and Prevention (CDC), in 2019, designated Methicillin-resistant *Staphylococcus aureus* (MRSA) as a serious antimicrobial resistance threat. The risk of acquiring MRSA and suffering life-threatening consequences due to it remains especially high for hospitalized patients due to a unique combination of factors, including: comorbid conditions, immunosuppression, and antibiotic use, and risk of contact with contaminated hospital workers and equipment. In this paper, we present a novel generative probabilistic model, GenHAI, for modeling sequences of MRSA test results outcomes for patients during a single hospitalization. This model can be used to answer many important questions from the perspectives of hospital administrators for mitigating the risk of MRSA infections. Our model is based on the probabilistic programming paradigm, and can be used to approximately answer a variety of predictive, causal, and counterfactual questions. We demonstrate the efficacy of our model by comparing it against discriminative and generative machine learning models using two real world datasets.

Introduction

Complex workflows are associated with patient care in hospitals, even for seemingly routine conditions. Figure 1 shows the initial part of the workflow associated with a patient who comes into the Emergency Department. The patient is put on antibiotics and multiple procedures are done as part of diagnoses, and then moved to an ICU. There is often a significant risk of Healthcare-associated infections (HAIs), such as Methicillin-resistant *Staphylococcus aureus* (MRSA) during the patient's stay. HAIs lead to longer hospital stays, increased mortality (Weiner-Lastinger et al. 2020; Dantes et al. 2013), and billions of dollars in increased healthcare costs (Zimlichman et al. 2013). While regular antibiotics work for bacterial infections, more aggressive antibiotics are used for a patient with an HAI. Further, HAIs can be transmitted within the hospital, and patients with HAIs are isolated (Cui et al. 2024). Early detection and effective prediction of the risk of acquisition and severe outcomes from HAIs infections can help in judicious and targeted antibiotic administration and implementation of better isolation precautions. Therefore, clinicians constantly ask about the risk of their patients getting MRSA or other HAIs during their hospital stay; examples of such questions are indicated as Q1-Q4 in Figure 1. Some of these questions might be asked multiple times during the patient's stay, as clinicians pick the best treatments. This has motivated a lot of work on ML methods for risk prediction type problems using Electronic Health Record (EHR) data for MRSA (the focus here) and other HAIs, e.g., (Hartvigsen et al. 2018; Oh et al. 2018; Monsalve et al. 2015; Kamruzzaman et al. 2024; Chang et al. 2011; Dubberke et al. 2011; Na et al. 2015; Im et al. 2011; Fan et al. 2022; Liu et al. 2019; Huynh et al. 2022; Ambavane et al. 2023). These generally discriminative methods have shown that specific types of clinical predictions, e.g., risk of HAI infection within d days, can be done reasonably well, e.g, (Monsalve et al. 2015; Kamruzzaman et al. 2024).

Despite the success of EHR-based ML methods for some of HAI related questions that arise, their use in actual clinical practice is somewhat less; this is a significant contrast to other areas of clinical informatics, where ML has been used effectively, e.g., (Simpao et al. 2014; Nordo et al. 2019). While there are multiple reasons, an important issue is that the complex workflows in clinical practice related to HAIs require a decision-theoretic setting, in which risk prediction needs to be done in a variety of scenarios. As mentioned earlier, Q1-Q4 in Figure 1 are examples of questions related to HAIs which clinicians ask at different stages for their patients. Some of these questions (e.g., Q1, Q4) are counterfactual in nature, and arise when clinicians evaluate treatments (including antibiotic choice) and precautions. For instance, a clinician would use Q1, Q2 and Q3 to decide whether to order a an MRSA test, and if antibiotic treatments should be started or continued—these are key elements of antibiotic stewardship plans (CDC 2023; Cosgrove and Srinivasan 2023). Q4 arises when clinicians are deciding whether to keep a patient under contact precautions to reduce risk to others, or remove them, e.g., (Cui et al. 2024; Marshall, Richards, and McBryde 2013; Siegel et al. 2007).

Prior discriminative ML methods, e.g., (Hartvigsen et al. 2018), are developed and trained for a specific form of risk prediction, and cannot be used directly for such counterfactual settings, with robust uncertainty estimates. This could be done by learning a large collection of models,

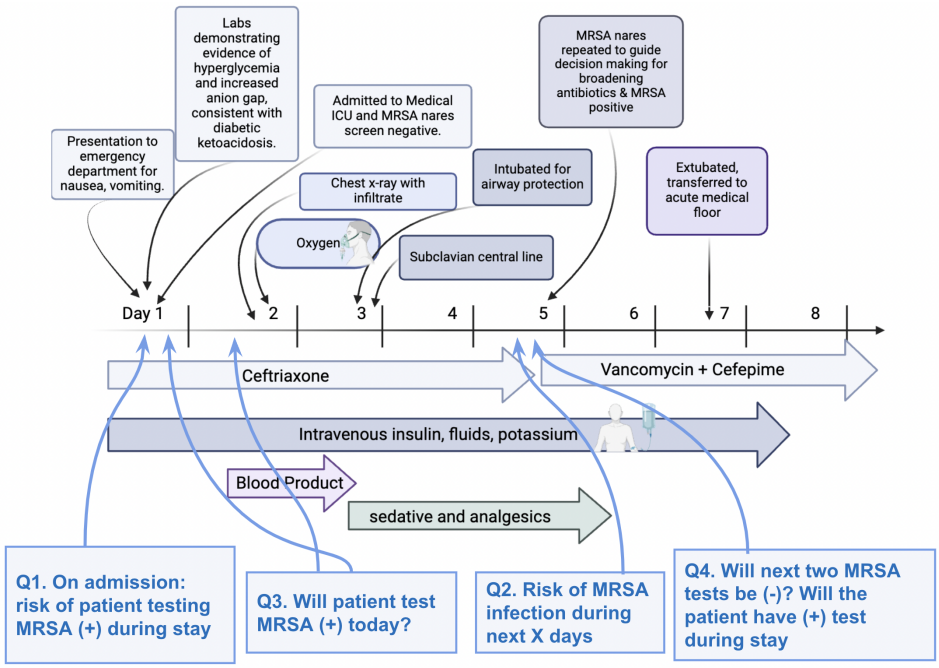


Figure 1: Examples of different questions related to infection risk for a patient during their hospital stay. Q1 arises on patient admission, while Q2, Q3, and Q4 are examples of questions clinicians ask during the patient's stay in the hospital.

but these become very hard to interpret, and ensure consistency, and there is usually limited data for learning multiple models, especially in clinical settings. Additionally, uncertainty estimation and statistical calibration become challenging with such methods. This is critical in clinical settings, where decisions must weigh not just predicted risk but also the confidence and reliability of those predictions.

Generative models using probabilistic programming (Gordon et al. 2014; van de Meent et al. 2021) provide a natural approach to handle the limitations of discriminative methods, especially for healthcare applications (Chen et al. 2021; Urteaga et al. 2021; Makar, Guttag, and Wiens 2018). They can provide very interpretable methods for modeling complex decision processes, while easily providing uncertainty estimates and statistical calibration. Generative models have supported diverse patient-level predictions, including menstrual cycle length (Urteaga et al. 2021) and probability of infection in a networked model (Makar, Guttag, and Wiens 2018). However, the problems considered in prior work are much simpler clinical questions; they do not contain conditionals and loops, which arise naturally in clinical practice.

Contributions (many details are presented in the Appendix)
(1) We present a novel generative sequence modeling problem that we call the *Hospital Test Sequence Modeling* (HTSM) problem. This is designed based on clinical problems arising from management of HAIs at the University of Virginia (UVA) and Johns Hopkins University (JHU) hospitals, which are serving two diverse populations.
(2) We present a novel modular interpretable probabilistic program, GenHAI (**Gen**erative Model for **H**ospital-**A**cquired **I**nfections), for modeling the generative process of MRSA test results. GenHAI can then be used to answer many important questions using probabilistic queries. The probabilistic program incorporates domain knowledge from clinical experts at UVA and JHU (Section); our results show that incorporating this domain knowledge (e.g., number of days on antibiotics, and ICU stays) helps in significantly improving the performance of GenHAI, and presents a template for other kinds of clinical questions.
(3) We evaluate GenHAI on EHR datasets from UVA and MIMIC-III (Johnson et al. 2016), and demonstrate its goodness of fit. We also compare it with state-of-the-art deep neural network-based generative sequence prediction models.
(4) We consider four case studies, each considering specific types of questions arising from clinical practice at the two hospitals, that can be answered using probabilistic queries and our probabilistic program. GenHAI gives clinically interpretable patient level estimates, which are not possible using standard supervised ML methods.

Related Works

We summarize the main threads of related work here; additional discussion is presented in the Appendix.

Prior work has developed prediction tools for MRSA and other HAIs using EHR data (Hartvigsen et al. 2018; Oh et al. 2018; Monsalve et al. 2015; Kamruzzaman et al. 2024; Chang et al. 2011; Dubberke et al. 2011; Na et al. 2015; Im et al. 2011; Fan et al. 2022; Liu et al. 2019; Huynh et al. 2022; Ambavane et al. 2023). While simple ML with careful feature engineering can capture some risks (Monsalve et al. 2015; Kamruzzaman et al. 2024; Fan et al. 2022), such

models are limited to narrow classification tasks and cannot support the broad queries clinicians face. Our work is also related to sequence modeling for clinical prediction, e.g., (Zhang and Yan 2023; Ekambaram et al. 2023; Gu and Dao 2023; Dao and Gu 2024; Nguyen et al. 2021).

Probabilistic machine learning and generative models have been proposed for many clinical tasks, e.g., (Chen et al. 2021; Urteaga et al. 2021; Makar, Guttag, and Wiens 2018). (Urteaga et al. 2021) develop a simple but flexible generative model for the prediction of menstrual cycle length using a hierarchical, Generalized Poisson-based generative model, which explicitly incorporates individual behaviors, and produces calibrated individual predictions. (Makar, Guttag, and Wiens 2018) show that a generative probabilistic model performs very well for the problem of predicting the activation of an individual in a networked epidemic process.

Background

As detailed in Section , EHRs capture comprehensive information about each patient's hospitalization, which we leverage in this work. A central element for our study is the MRSA testing data, consisting of two types: NARE tests (PCR-based) and culture tests. NARE tests are explicitly performed to detect MRSA, while culture tests may be done for a variety of clinical purposes and are not always intended as MRSA screenings. However, if MRSA is identified in a culture test, the result is documented in the patient's healthcare record. In our datasets, all NARE test results are available for each hospitalization record, whereas culture tests are recorded only when the result is MRSA-positive. We found this to be a common data issue when gathering MRSA test result data, based on our discussions with clinicians at UVA and JHU, as well as in the MIMIC-III dataset.

Symbol	Description
$\mathbf{D}$	Dataset of hospitalization records
p_k	Patient corresponding to the k-th hospitalization record
m_k	Number of MRSA tests in the k-th record
$r_{k,i}$	Outcome of the i-th MRSA test for patient p_k
$t_{k,i}$	Test type of the i-th MRSA test (NARE or culture)
$d_{k,i}$	Time interval between tests $t_{k,i}$ and $t_{k,i+1}$
α_k	Admission-time features of patient p_k
$\beta_{k,i}$	Features of patient p_k at the time of test $t_{k,i}$
$\mathcal{D}_*$	Probabilistic sub-programs (e.g., $\mathcal{D}_{t_i}, \mathcal{D}_{r_i}, \mathcal{D}_{d_i\|+}$)
$\mathcal{D}_{\beta_{[ab]}}$,$\mathcal{D}_{\beta_{[icu]}}$, $\mathcal{D}_{\beta_{[dia]}}$	Distributions for days on antibiotics (30d), in ICU (7d), dialysis (7d)
$\mathcal{D}_{d_i\|+,-}$	Inter-test delay distribution when previous result is positive/ negative
$\mathcal{D}_{\text{cont}}$	Bernoulli continuation distribution (decides if sequence continues)

Table 1: Table of notation

Here we formally describe the assumptions made about the patient healthcare care records and the overall modeling problem; the notation we use is summarized in Table 1. We assume that the dataset $\mathbf{D}$ consists of n hospitalization records. Let p_k ($k \in [1 \dots n]$) be the patient corresponding to the k^{th} hospitalization record. Additionally, let $\{r_{k,i}\}$ and $\{t_{k,i}\}$ ($i \in [1 \dots m_k]$) be the sequence of observable test outcomes and the corresponding test type (NARE test or culture test) in the k^{th} hospitalization record. Here m_k is the number of MRSA test results in the k^{th} hospitalization record. Further, let $d_{k,i}$ be the time duration in between tests $t_{k,i}$ and $t_{k,i+1}$, where $i \in [1 \dots m_k - 1]$. Also, let $\boldsymbol{\alpha}_k$ be the vector representing the set of features of the patient p_k that are observable at admission time, and let $\boldsymbol{\beta}_{k,i}$ be a vector representing the set of observable features of the patient p_k at the time when the test $t_{k,i}$ was done. We assume that each hospital sequence is independent of every other sequence.

Problem Statement. The objective of the *Hospital Test Sequence Modeling* (HTSM) is to learn the joint conditional probability distribution:

$$P(\mathbf{D}) = \prod_{k=1}^{n} P(\{r_{k,i}\}, \{t_{k,i}\}, \{d_{k,i}\}, \{\boldsymbol{\beta}_{k,i}\} \mid \boldsymbol{\alpha}_k) \quad (1)$$

Note that we do not model $\boldsymbol{\alpha}_k$ which represents the patient's characteristics available at the time of admission — such as gender, age, past MRSA test results, etc. — but consider it as an input to our model. Given the distribution $P(\mathbf{D})$ all of the questions posed in Section can then be answered using probabilistic queries using the model.

Modeling Test Result Sequences

We take a generative modeling approach for solving the HTSM problem. In particular, we describe the generative process using a modular probabilistic program. A probabilistic program is generative model that takes the form of an imperative or functional program with support for sampling random variables and conditioning the values of those variables to observations (Gordon et al. 2014). Probabilistic programs are Bayesian models. They are inherently more interpretable and allow for easy incorporation of domain knowledge from experts compared to deep neural network-based models. They are, however, generally harder to use due to the computational complexity of training them.

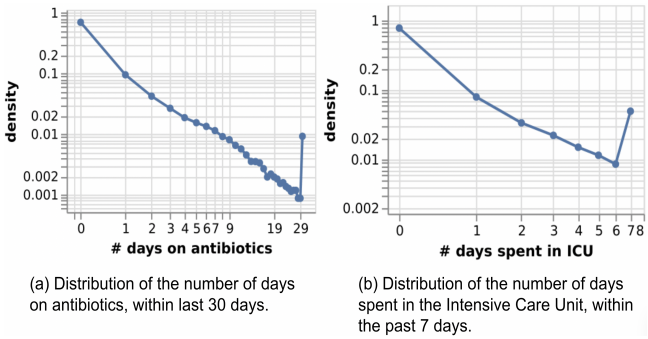
To alleviate the issue of computational complexity, the probabilistic program used here is a modular one, composed of a number of smaller probabilistic programs that can be trained independently.

Overview. We use a probabilistic program to model the generative process for hospital test sequences. The program uses a number of sub-programs: $\mathcal{D}_{\boldsymbol{\beta}_i[\text{ab}]}$, $\mathcal{D}_{\boldsymbol{\beta}_i[\text{icu}]}$, $\mathcal{D}_{\boldsymbol{\beta}_i[\text{dia}]}$, $\mathcal{D}_{t_1}$, $\mathcal{D}_{r_i}$, $\mathcal{D}_{\text{cont}}$, $\mathcal{D}_{d_i|-}$, $\mathcal{D}_{d_i|+}$; these are denoted by $\mathcal{D}_*$ for convenience, and are described below. Each sub-program takes the form of a Bayesian model with a Generalized Linear Model (GLM) like design.

Consider the case of a generic sub-program $\mathcal{D}_Y$ which is used to sample the random variable Y given the input parameters $\mathbf{x}$. Subprogram $\mathcal{D}_Y$ can be defined using the output distribution $\mathbb{D}_Y$, parameters $\boldsymbol{\theta}_Y$, link function ℓ_Y, and a distribution on the model parameters. The random variable Y is sampled from $\mathcal{D}_Y$ as follows:

$$Y \sim \mathbb{D}_Y(\theta_i, \theta_d); \quad \theta_d = \ell_Y(\mathbf{w}^T \cdot \mathbf{x} + c); \quad \theta_i, \mathbf{w}, c \sim \mathbb{Q}(\boldsymbol{\theta}_Y)$$

Here θ_i and θ_d represent the input independent and input dependent parameters of $\mathbb{D}_Y$, $(\mathbf{w}, c)$ are the linear model parameters, and $\mathbb{Q}(\boldsymbol{\theta}_Y)$ represents the posterior distribution of the model parameters $(\theta_i, \mathbf{w}, c)$, conditioned on data.



(a) Distribution of the number of days on antibiotics, within last 30 days.

(b) Distribution of the number of days spent in the Intensive Care Unit, within the past 7 days.

Figure 2: Distribution of features observable at test time.

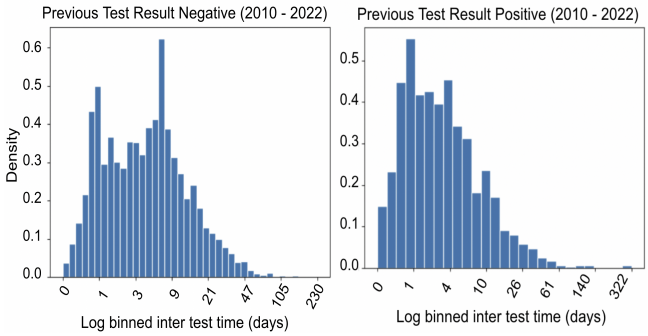


Figure 3: Distribution of inter test time

Each of the individual sub-programs $\mathcal{D}_*$ is trained using *Stochastic Variational Inference (SVI)* by maximizing the *Evidence Lower Bound (ELBO)* objective (Kingma and Welling 2014; Ranganath, Gerrish, and Blei 2014) We use Multivariate Normal distributions as variational approximation of the posterior distributions $\mathbb{Q}_*$.

Algorithm GenHAI (informal). The generative process simulates a patient's MRSA test sequence during hospitalization. The simulation of the program starts by sampling admission-time features (α) and initial test-time features (β_1). At each step, it
(i) samples test type, $t_i \sim \mathcal{D}_{t_i}(\cdot)$ and result r_i (with $r_i = 1$ deterministically if t_i is culture, otherwise $r_i \sim \mathcal{D}_{r_i}(\cdot)$)
(ii) draws inter-test delay (d_i) using a Log-Normal mixture if the previous result (r_{i-1}) was negative, or a single Log-Normal otherwise
(iii) updates test-time covariates (β_i) via $\mathcal{D}_{\beta[ab]}, \mathcal{D}_{\beta[icu]}$, $\mathcal{D}_{\beta[dia]}$, modeling recent antibiotics (30d), ICU stays (7d), and dialysis status. The first two are parameterized as truncated Negative Binomial distributions to account for overdispersed count data with hard upper bounds, while $\mathcal{D}_{\beta[dia]}$ is modeled as a Bernoulli random variable.
(iv) applies a Bernoulli continuation rule ($\mathcal{D}_{\text{cont}}$) and repeats until termination, yielding a synthetic test sequence conditioned on α.

The complete program expands this workflow into modular sub-programs $\mathcal{D}_*$ with explicit distributional forms and inference details.

Our model simulates a patient's sequence of MRSA tests during hospitalization. It initializes patient features such as days on antibiotics, ICU days, and dialysis status (Figure 2). At each step, the model samples the test type (culture or NARE), result, and inter-test delay, which depends on the previous result. Negative tests follow a mixture of Log-Normal delays, while positives use a single Log-Normal (Figure 3). The sequence ends via a Bernoulli stopping rule; additional details in Appendix.

Implementation. The probabilistic program is implemented in Python. The individual sub-programs $\mathcal{D}_*$ were created using the Pyro probabilistic programming language (Bingham et al. 2019), uses the PyTorch (Paszke et al. 2019) deep neural network library for auto-differentiation and speeding up computations on hardware accelerators.

Experiments

Datasets. For this study, we have used two datasets for modeling MRSA test result sequences: (1) UVA Dataset, and (2) the MIMIC-III v1.4 dataset (Johnson et al. 2016). An overview of the dataset statistics and feature construction details is provided in Appendix.
UVA Dataset. The UVA Dataset is derived from Electronic Health Records at the UVA, and is Institutional Review Board (IRB)-restricted for authorized research use. It contains over 27K patients, 37K hospitalizations, and spans 2012–2022. Each hospitalization includes:
1. On-admission data. – demographics (age, gender, race, ethnicity) and visit details (reason, type, source), used for admission-time features $\boldsymbol{\alpha}$. **2. Clinical event data.** – procedures (ICU stays), medications (e.g., antibiotics), comorbidities, and lab tests (including MRSA), used for test-time features $\boldsymbol{\beta}$, test types t_i, and results r_i. **3 MRSA test data** – clinical cultures (MRSA-positive only) and surveillance NARE tests (both positive and negative).
The MIMIC-III v1.4 dataset. The MIMIC-III dataset contains deidentified hospitalization records from over 40,000 patients at Beth Israel Deaconess Medical Center (2001–2012), including demographics, vitals, and labs, and is publicly available for clinical research.

Baselines. We compare our probabilistic program (GenHAI) with recent state-of-the-art deep learning models for sequence prediction, including Crossformer (Zhang and Yan 2023), TSMixer (Ekambaram et al. 2023), Mamba (Gu and Dao 2023; Dao and Gu 2024), and TPAMTL (Nguyen et al. 2021) (Details in Appendix). These baselines were selected for their relevance, availability, and performance on time-series tasks. Table 2 compares GenHAI with baseline models, showing that while standard sequence models achieve reasonable accuracy, GenHAI provides competitive performance with better calibration and interpretability.

Training. All models were trained on machines equipped with four Nvidia Tesla V100 GPUs, two 20-core Intel Xeon Gold 6230 CPUs, and 380 GB of RAM. Each of the sub-programs used in Algorithms were trained using Stochastic Variational Inference (SVI) (Kingma and Welling 2014; Ranganath, Gerrish, and Blei 2014) using the Pyro probabilistic programming language (Bingham et al. 2019).

Model	Accuracy	Precision	Recall	F1 Score	AUROC	AUPRC
Mamba	0.749	0.876	0.834	0.855	0.820	0.727
Crossformer	0.809	0.650	0.639	0.644	0.825	0.606
TSMixer	0.758	0.817	0.758	0.786	0.886	0.790
TPAMTL	0.873	0.697	0.297	0.416	0.689	0.258
GenHAI-$\mathcal{D}_{r_i}$	0.931	0.690	0.745	0.714	0.858	0.635

Table 2: Comparing performance of baselines with $\mathcal{D}_{r_i}$ for MRSA test result prediction task for NARE tests.

Results. We evaluate the negative log-likelihood (NLL) and perplexity of each GenHAI sub-program ($\mathcal{D}_*$) on the UVA and MIMIC-III datasets, using an 80:20 train–test split. Across both datasets, the sub-programs achieve consistently low NLL and perplexity values, ranging from 0.135–2.33 (UVA) and 0.018–2.006 (MIMIC-III) in NLL, and 1.019–10.283 (UVA) and 1.027–7.433 (MIMIC-III) in perplexity, indicating that the generative components capture realistic temporal dynamics of patient MRSA testing sequences. This validation confirms that the model components are statistically calibrated and provide reliable likelihood estimates for observed sequences. Such calibration is essential, since the overall framework depends on these subprograms to support predictive and counterfactual queries. Unlike discriminative models that output only point estimates, GenHAI produces full posterior distributions, enabling uncertainty quantification. This ability to report both risk and confidence makes GenHAI well-suited for clinical decision-making, where uncertainty directly influences isolation, antibiotic use, and retesting. Here, the goal is not classification accuracy but generative fit, ensuring that probabilistic reasoning on realistic patient-level distributions rather than heuristic approximations.

Table 2 compares the performance of the four deep neural network-based baselines for the NARE test result prediction task. Notably, TPAMTL (Nguyen et al. 2021) is the only model specifically designed for healthcare EHR data and evaluated on MIMIC-III using a probabilistic multitask learning framework. Despite its simplicity, GenHAI-$\mathcal{D}_{r_i}$ achieves about 6.6% higher accuracy than the next best model while maintaining competitive recall and F1.

Case Studies

In this section, we present answers to the real-world questions posed. Although GenHAI's simulation (Section) can directly answer the probabilistic queries via rejection sampling, this approach is highly inefficient. Thus, for answering questions in case studies 2–4, we create variants of GenHAI. The variants, however, utilize the same pre-trained sub-programs (i.e., $\mathcal{D}_*$).

Case Study 1 (Admission Risk). *A 70-year-old patient is admitted from a long-term care facility. The infection-control nurse wants to know: Should we place this patient in isolation on admission?*

A1: Let $\mathbb{I}_A(\{r_i\})$ be the indicator function that returns 1 if any of the results are positive. Then the answer to question Q1 can be obtained as:

$$P_A = \int \mathbb{I}_A(\{r_i\}) \approx \mathbb{E}[\mathbb{I}_A(\{r_i\})]$$

Using GenHAI, we can generate sample sequences $\{r_i\}$ which can be used in the above equation to estimate the relevant likelihood. The system shows a higher risk for patients transferred from facilities compared to community admissions (Figure 4a), supporting proactive isolation and early testing for this patient group.

Case Study 2: (Future Risk with Extended Stay). *A patient has been hospitalized for several weeks, and the past MRSA test result is known. Assuming this person will be in the hospital for at least X more days, what is the likelihood that they will test positive for MRSA in the next X days?*

A2: To efficiently answer this question, we create a variant of GenHAI, that runs the stochastic simulation with the additional test result as input and runs the simulation forward. A detailed description of the variant can be found in the Appendix. In addition to using the *cont* variable to stop the simulation, the variant algorithm also keeps track of the time passed and stops if the cumulative simulated duration crosses X days. Figure 4b shows how the likelihood changes with an increase in X for a few selected representative patients. As can be observed from the figure, there is a general trend that the likelihood increases with increasing duration of stay in the hospital.

Case Study 3 (Immediate Retest). *An ICU patient had their last MRSA test X days ago. What's the chance of a positive result, if tested today?*

A3: The above question is an example of a causal question with interventions. In particular, it requires that we ignore sampling the inter-test time delay and use X days as the delay. Additionally, since the question requires the answer to a single test result, the one to be done "right now" the variant of GenHAI need only execute an iteration of the test result generation.

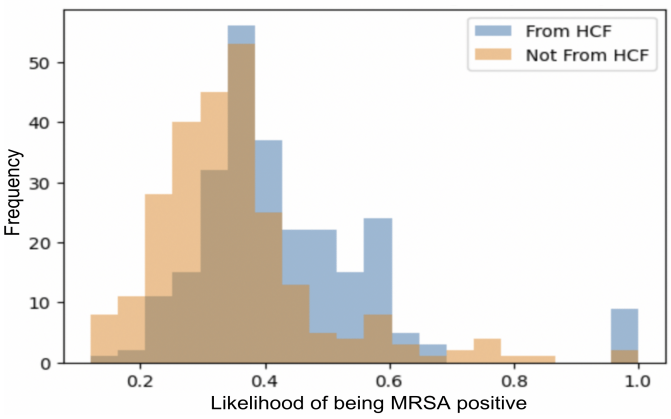
Figure 4c shows how the likelihood of the testing positive changes for a few selected representative patients. We can observe, that for the selected time period, for some of the patients, the likelihood of testing positive increase since the last test result. However, for a subset of them, the likelihood stays constant, illustrating the variability captured by the method.

Case Study 4 (Discontinuing Precautions). *A patient previously MRSA-positive has just returned a negative nasal swab. Physicians ask: can we safely discontinue isolation? What is the likelihood that the subsequent NARES test will both be negative and they will not have any positive MRSA?*

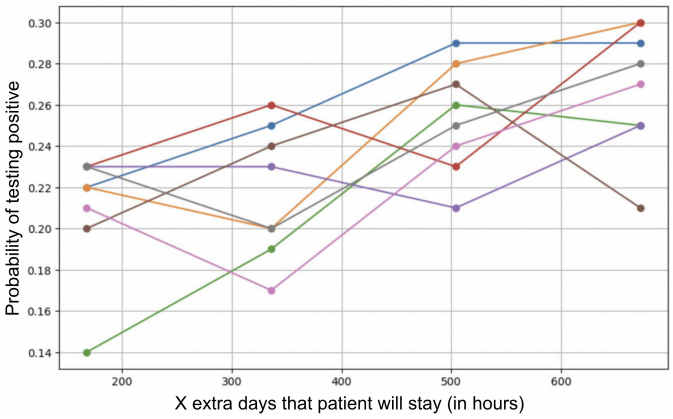
A4: The question posed above is particularly important from the perspective of hospital-acquired infection risk management. Patients who test positive for MRSA are generally kept in isolation. This precautionary measure, while important from a MRSA spread risk management perspective is however, bad for the isolated patient's mental health and expensive for the hospital administrators. The methodology presented in the current study can be used to evaluate the risk of moving a patient off precautionary isolation (Figure 4d).

Conclusion

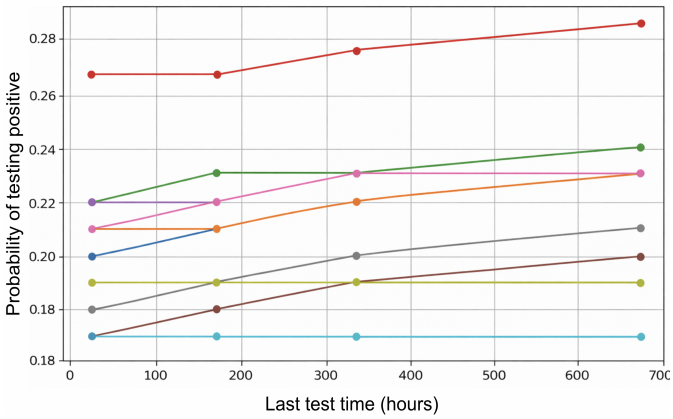
Hospital associated infections (HAIs) pose a significant burden for healthcare institutions. Here, we present a new approach, GenHAI, for HAI risk prediction using generative



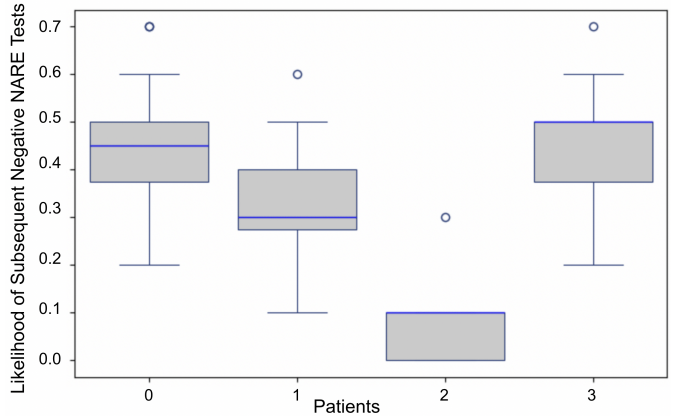
(a) Distribution of MRSA-positive risk during stay by admission source(From HCF vs. Not From HCF)



(b) Change in MRSA-positive risk with the increasing number of additional days spent in the hospital (selected patients)



(c) Change in MRSA-positive risk in the subsequent test with increasing delay between previous and subsequent tests.



(d) Variation in the likelihood of patients staying MRSA negative during their hospitalization period for selected patients.

Figure 4: Four clinical decision scenarios supported by GenHAI: (a) admission risk, (b) risk with longer stay, (c) immediate retest probability, and (d) discontinuing isolation.

sequence modeling based on a novel modular probabilistic program, guided by clinical domain experts. All prior methods for HAI risk prediction are based on discriminative ML methods, and are not directly useful in clinical practice. In contrast, GenHAI provides a powerful framework for supporting questions related to HAIs that arise naturally during complex clinical workflows; further, GenHAI gives better performance than prior methods in many metrics, since it explicitly incorporates clinical workflows. Our methods are more interpretable than prior approaches, which build methods for capturing such workflows, and highlight the utility of probabilistic programming for clinical applications.

Path to Deployment. GenHAI uses standard data recorded in EHRs, such as admission covariates, antibiotic exposure, ICU stays, and treatments such as dialysis, with feature construction and preprocessing described in Appendix. Our results on the UVA and MIMIC-III EHR datasets show that GenHAI can be easily adapted to most EHR systems, in spite of differences in how this information is coded. GenHAI can be easily integrated with current workflows supported in EHR systems such as EPIC. Individual patient level risk scores and uncertainty bands at key decision points, including admission screening, ongoing risk assessment, retest prioritization, and de-isolation, can be easily computed and made available to clinicians. Governance spans infection prevention and ID clinicians, nursing leadership, IT/EHR teams, and oversight bodies (IRB, HIPAA). Its modular structure supports site-specific retraining and extension to other HAIs without altering the program.

Validation Plan. Deployment can follow a safe pathway of silent replay, pilot, and gradual scale-up, with ongoing calibration monitoring to demonstrate real-world impact.

Limitations and Future Work. GenHAI reflects practices from two US hospitals, but needs to be redesigned for other hospitals, using their specific workflows and patient populations. Its modular design improves interpretability but reduces expressiveness by avoiding shared latent variables; future work will relax this constraint to capture richer dependencies. Current inference uses SVMs, which scale well but only approximate posteriors—alternative approaches (e.g., probabilistic circuits) may improve calibration. Extending GenHAI to evolving hospital policies and multi-institutional datasets will enhance robustness and clinical applicability.

Acknowledgments

This research is partially supported by NSF grants CCF-1918656, CNS-2317193, DMS-2311187 and DMS-2436319, DTRA award HDTRA1-24-R-0028, CDC MIND U01CK000589, National Institutes of Health K23AI163368, and the University of Virginia Prominence-to-Preeminence (P2PE) STEM Targeted Initiatives Fund, SIF176A Contagion Science.

References

Ambavane, A.; Litkiewicz, M.; Sardesai, A.; Teloian, D.; Obi, E. N.; and Yang, J. 2023. 2814. The Impact of Novel Antibiotics on Curbing Antimicrobial Resistance and Infection Spread a Dynamic Modeling Study. In *Open Forum Infectious Diseases*, volume 10, ofad500–2425. Oxford University Press US.

Bingham, E.; Chen, J. P.; Jankowiak, M.; Obermeyer, F.; Pradhan, N.; Karaletsos, T.; Singh, R.; Szerlip, P.; Horsfall, P.; and Goodman, N. D. 2019. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1): 973–978.

CDC. 2023. Core Elements of Antibiotic Stewardship. https://www.cdc.gov/antibiotic-use/hcp/core-elements/index.html. Accessed: 2025-08-17.

Chang, Y.-J.; Yeh, M.-L.; Li, Y.-C.; Hsu, C.-Y.; Lin, C.-C.; Hsu, M.-S.; and Chiu, W.-T. 2011. Predicting Hospital-Acquired Infections by Scoring System with Simple Parameters. *PloS one*, 6: e23137.

Chen, I. Y.; Joshi, S.; Ghassemi, M.; and Ranganath, R. 2021. Probabilistic machine learning for healthcare. *Annual review of biomedical data science*, 4(1): 393–415.

Cosgrove, S. E.; and Srinivasan, A. 2023. Antibiotic stewardship: a decade of progress. *Infectious Disease Clinics*, 37(4): 659–667.

Cui, J.; Heavey, J.; Lin, L.; Klein, E. Y.; Madden, G. R.; Sifri, C. D.; Lewis, B.; Vullikanti, A. K.; and Prakash, B. A. 2024. Modeling relaxed policies for discontinuation of methicillin-resistant Staphylococcus aureus contact precautions. *Infection Control & Hospital Epidemiology*, 1–6.

Dantes, R.; Mu, Y.; Belflower, R.; Aragon, D.; Dumyati, G.; Harrison, L. H.; Lessa, F. C.; Lynfield, R.; Nadle, J.; Petit, S.; Ray, S. M.; Schaffner, W.; Townes, J.; Fridkin, S.; and for the Emerging Infections Program–Active Bacterial Core Surveillance MRSA Surveillance Investigators. 2013. National Burden of Invasive Methicillin-Resistant Staphylococcus aureus Infections, United States, 2011. *JAMA Internal Medicine*, 173(21): 1970–1978.

Dao, T.; and Gu, A. 2024. Transformers are SSMs: Generalized Models and Efficient Algorithms Through Structured State Space Duality. In *International Conference on Machine Learning (ICML)*.

Dubberke, E. R.; Yan, Y.; Reske, K. A.; Butler, A. M.; Doherty, J.; Pham, V.; and Fraser, V. J. 2011. Development and Validation of a Clostridium difficile Infection Risk Prediction Model. *Infection Control & Hospital Epidemiology*, 32(4): 360–366.

Ekambaram, V.; Jati, A.; Nguyen, N.; Sinthong, P.; and Kalagnanam, J. 2023. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23. ACM.

Fan, S.; Lin, J.; Wu, S.; Mu, X.; and Guo, J. 2022. Random forest model can predict the prognosis of hospital-acquired Klebsiella pneumoniae infection as well as traditional logistic regression model. *Plos one*, 17(11): e0278123.

Gordon, A. D.; Henzinger, T. A.; Nori, A. V.; and Rajamani, S. K. 2014. Probabilistic programming. In *Future of Software Engineering Proceedings*, 167–181.

Gu, A.; and Dao, T. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv preprint arXiv:2312.00752*.

Hartvigsen, T.; Sen, C.; Brownell, S.; Teeple, E.; Kong, X.; and Rundensteiner, E. A. 2018. Early Prediction of MRSA Infections using Electronic Health Records. In *HEALTHINF*.

Huynh, P. K.; Setty, A. R.; Tran, Q. M.; Yadav, O. P.; Yodo, N.; and Le, T. Q. 2022. A domain-knowledge modeling of hospital-acquired infection risk in Healthcare personnel from retrospective observational data: A case study for COVID-19. *Plos one*, 17(11): e0272919.

Im, G. Y.; Modayil, R. J.; Feuerman, M.; Lin, C. T.; Balani, A. R.; Katz, D. S.; and Grendell, J. H. 2011. A prediction model of disease severity in Clostridium difficile-associated disease. *Gastroenterology*, 5(140): S–361.

Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L. W.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L. A.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3: 160035.

Kamruzzaman, M.; Heavey, J.; Song, A.; Bielskas, M.; Bhattacharya, P.; Madden, G.; Klein, E.; Deng, X.; Vullikanti, A.; et al. 2024. Improving Risk Prediction of Methicillin-Resistant Staphylococcus aureus Using Machine Learning Methods With Network Features: Retrospective Development Study. *JMIR AI*, 3(1): e48067.

Kingma, D.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *Conference proceedings: papers accepted to the International Conference on Learning Representations (ICLR) 2014*, 1–14. Accepted author manuscript.

Liu, X.; Zou, N.; Zhu, D.; and Wang, D. 2019. Influencing factors analysis and modeling of hospital-acquired infection in elderly patients. *Journal of Combinatorial Optimization*, 37: 248–270.

Makar, M.; Guttag, J.; and Wiens, J. 2018. Learning the Probability of Activation in the Presence of Latent Spreaders. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

Marshall, C.; Richards, M.; and McBryde, E. 2013. Do active surveillance and contact precautions reduce MRSA acquisition? A prospective interrupted time series. *PloS one*, 8(3): e58112.

Monsalve, M.; Pemmaraju, S.; Johnson, S.; and Polgreen, P. 2015. Improving Risk Prediction of Clostridium Difficile Infection Using Temporal Event-Pairs. In *2015 International Conference on Healthcare Informatics*, 140–149.

Na, X.; Martin, A. J.; Sethi, S.; Kyne, L.; Garey, K. W.; Flores, S. W.; Hu, M.; Shah, D. N.; Shields, K.; Leffler, D. A.; et al. 2015. A multi-center prospective derivation and validation of a clinical prediction tool for severe Clostridium difficile infection. *PloS one*, 10(4): e0123405.

Nguyen, A. T.; Jeong, H.; Yang, E.; and Hwang, S. J. 2021. Clinical Risk Prediction with Temporal Probabilistic Asymmetric Multi-Task Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10): 9081–9091.

Nordo, A. H.; Levaux, H. P.; Becnel, L. B.; Galvez, J.; Rao, P.; Stem, K.; Prakash, E.; and Kush, R. D. 2019. Use of EHRs data for clinical research: historical progress and current applications. *Learning health systems*, 3(1): e10076.

Oh, J.; Makar, M.; Fusco, C.; McCaffrey, R.; Rao, K.; Ryan, E.; Washer, L.; West, L.; Young, V.; Guttag, J.; Hooper, D.; Shenoy, E.; and Wiens, J. 2018. A Generalizable, Data-Driven Approach to Predict Daily Risk of Clostridium difficile Infection at Two Large Academic Health Centers. *Infection Control & Hospital Epidemiology*, 39: 425–433.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Ranganath, R.; Gerrish, S.; and Blei, D. 2014. Black box variational inference. In *Artificial intelligence and statistics*, 814–822. PMLR.

Siegel, J. D.; Rhinehart, E.; Jackson, M.; Chiarello, L.; Committee, H. C. I. C. P. A.; et al. 2007. 2007 guideline for isolation precautions: preventing transmission of infectious agents in health care settings. *American journal of infection control*, 35(10): S65.

Simpao, A. F.; Ahumada, L. M.; Gálvez, J. A.; and Rehman, M. A. 2014. A review of analytics and clinical informatics in health care. *Journal of medical systems*, 38: 1–7.

Urteaga, I.; Li, K.; Shea, A.; Vitzthum, V. J.; Wiggins, C. H.; and Elhadad, N. 2021. A Generative Modeling Approach to Calibrated Predictions: A Use Case on Menstrual Cycle Length Prediction. In Jung, K.; Yeung, S.; Sendak, M.; Sjoding, M.; and Ranganath, R., eds., *Proceedings of the 6th Machine Learning for Healthcare Conference*, volume 149 of *Proceedings of Machine Learning Research*, 535–566. PMLR.

van de Meent, J.-W.; Paige, B.; Yang, H.; and Wood, F. 2021. An Introduction to Probabilistic Programming. https://arxiv.org/abs/1809.10756.

Weiner-Lastinger, L. M.; Abner, S.; Edwards, J. R.; Kallen, A. J.; Karlsson, M.; Magill, S. S.; Pollock, D.; See, I.; Soe, M. M.; Walters, M. S.; and et al. 2020. Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network, 2015–2017. *Infection Control & Hospital Epidemiology*, 41(1): 1–18.

Zhang, Y.; and Yan, J. 2023. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *International Conference on Learning Representations*.

Zimlichman, E.; Henderson, D.; Tamir, O.; Franz, C.; Song, P.; Yamin, C. K.; Keohane, C.; Denham, C. R.; and Bates, D. W. 2013. Health Care–Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. *JAMA Internal Medicine*, 173(22): 2039–2046.