Candida auris Risk Intelligence:

A Hybrid Epidemiological–Machine Learning Framework for ICU-Level and Individual Risk Assessment

Abstract

Candida auris is an emerging multidrug-resistant fungal pathogen responsible for severe healthcare-associated outbreaks, particularly in intensive care units (ICUs). Traditional surveillance methods often lack real-time predictive capability and fail to translate population-level infection dynamics into actionable individual-level risk insights. This study presents a hybrid risk intelligence framework that integrates a compartmental epidemiological model with machine learning to predict infection growth dynamics and assess both ICU-level and individual-level risk. Differential equations derived from published *Candida auris* transmission models are used to generate mechanistically meaningful features, which are then processed by a supervised regression model to estimate the rate of infection growth. The predicted growth rate is further mapped to categorical ICU risk levels and adjusted using individual exposure factors such as ICU stay duration, healthcare worker contact intensity, and immune status. The proposed system provides an interpretable, extensible, and clinically relevant decision-support tool for infection control planning and risk stratification.

Keywords

Candida auris, Epidemiological Modeling, Machine Learning, ICU Risk Assessment, Infection Dynamics, Clinical Decision Support

1. Introduction

Healthcare-associated infections caused by *Candida auris* have emerged as a major global concern due to their high mortality rates, environmental persistence, and resistance to multiple antifungal agents. Outbreaks are particularly prevalent in intensive care units (ICUs), where patients are highly vulnerable and frequent contact with healthcare workers facilitates pathogen transmission.

Conventional infection surveillance approaches are largely retrospective and descriptive, offering limited predictive power. Mathematical epidemiological models provide mechanistic understanding of transmission dynamics but often require complex numerical solvers and

are difficult to apply in real-time clinical environments. Conversely, machine learning models offer predictive capability but may lack biological interpretability when applied in isolation.

This work addresses these limitations by proposing a **hybrid epidemiological–machine learning framework** that leverages the strengths of both approaches. The system predicts the infection growth rate within an ICU and translates this prediction into clinically interpretable ICU-level and individual-level risk categories.

---

2. Related Work

Previous studies on *Candida auris* transmission have primarily focused on compartmental models capturing patient colonization, healthcare worker contamination, and environmental persistence. These models employ systems of differential equations to describe transitions between epidemiological states. While effective for theoretical analysis, such models are often impractical for real-time deployment.

Recent advances in healthcare machine learning have demonstrated success in predicting infection risk using electronic health records and statistical features. However, purely data-driven approaches may fail to respect biological constraints and can be difficult to justify in clinical decision-making contexts.

Hybrid modeling approaches that integrate mechanistic models with machine learning have shown promise in infectious disease forecasting. This study builds on this paradigm by embedding epidemiological structure directly into the feature engineering process for machine learning prediction.

---

3. Mathematical Model and Theoretical Foundation

3.1 Compartmental Structure

The proposed system is grounded in a compartmental model that divides the ICU population into the following epidemiological states:

- **PU**: Uncolonized patients
- **PC**: Colonized patients
- **PI**: Infected patients
- **HC**: Contaminated healthcare workers
- **HU**: Uncontaminated healthcare workers

These compartments capture the primary transmission pathways of *Candida auris* within ICU environments.

3.2 Differential Equations

Based on established *Candida auris* transmission models, the following differential equations describe the system dynamics:

Uncolonized Patients

$$\frac{dPU}{dt} = \Lambda - \beta_2 HC \cdot PU - d_1 PU + \phi(PC + PI)$$

This equation accounts for patient admission, colonization via healthcare workers, discharge or death, and recovery.

Colonized Patients

$$\frac{dPC}{dt} = \beta_2 HC \cdot PU - (d_2 + \phi + \sigma)PC$$

Colonization arises from contact with contaminated healthcare workers and decreases due to discharge, recovery, or progression.

Healthcare Worker Contamination

$$\frac{dHC}{dt} = \beta_1 PC \cdot HU - \lambda HC$$

Healthcare workers become contaminated through contact with colonized patients and are decontaminated via hygiene practices.

3.3 Infection Growth Rate

The primary target of prediction in this study is the **infection growth rate**:

$$\frac{dPI}{dt}$$

This quantity represents the rate at which infected individuals increase within the ICU and serves as a proxy for outbreak intensity.

---

4. Machine Learning Integration

4.1 Feature Engineering

Instead of directly learning from raw epidemiological parameters, the system computes the derivatives of key compartments using the above equations. These derivatives encode biologically meaningful dynamics and are used as input features for the machine learning model.

The final feature vector includes:

- Epidemiological parameters (transmission rates, clearance rates)
- Derived differential terms ((dPU/dt), (dPC/dt), (dHC/dt))

This hybrid feature construction preserves mechanistic interpretability while enabling data-driven learning.

4.2 Model Architecture

A supervised regression model is trained to predict the infection growth rate (dPI/dt). The model learns nonlinear relationships between epidemiological parameters and infection growth dynamics.

The trained model is deployed within a Flask-based API for real-time inference.

---

5. Risk Classification Framework

5.1 ICU-Level Risk Assessment

The predicted infection growth rate is mapped to categorical ICU risk levels:

- **LOW**: Minimal infection growth
- **MODERATE**: Sustained transmission risk
- **HIGH**: Rapid outbreak escalation

These thresholds are selected to reflect clinically meaningful distinctions in infection pressure.

5.2 Individual-Level Risk Assessment

To extend ICU-level predictions to individual patients, the system incorporates exposure modifiers:

- Duration of ICU stay
- Healthcare worker contact intensity
- Immune status

A scoring mechanism adjusts the baseline ICU risk to derive an **individual risk category** (LOW, MODERATE, HIGH). This approach bridges population-level dynamics with patient-specific vulnerability.

---

6. System Implementation

The backend system is implemented using a Python-based web framework. It performs the following steps:

1. Receives epidemiological and individual exposure inputs
2. Computes differential equation–based features
3. Predicts infection growth rate using the trained model
4. Classifies ICU and individual risk levels
5. Returns structured risk intelligence for downstream use

The design ensures modularity, interpretability, and extensibility.

---

7. Results and Interpretation

The system successfully demonstrates:

- Sensitivity to epidemiological parameter changes
- Distinction between ICU-level and individual-level risk
- Stable and interpretable risk classification behavior

Predictions reflect realistic infection dynamics and respond appropriately to changes in exposure duration, contact intensity, and immune status.

---

8. Discussion

The proposed framework illustrates how mechanistic epidemiological theory can be combined with machine learning to create practical clinical decision-support tools. By grounding predictions in differential equation dynamics, the system avoids the pitfalls of purely data-driven approaches while retaining predictive flexibility.

Limitations include dependence on parameter accuracy and the assumption of homogeneous mixing within ICU compartments. Future work may incorporate temporal modeling, environmental reservoirs, and adaptive learning from real-world data.

---

9. Conclusion

This study presents a hybrid epidemiological–machine learning system for *Candida auris* risk intelligence that predicts ICU infection growth and translates it into individual patient risk. By integrating compartmental modeling with machine learning, the system achieves interpretability, clinical relevance, and predictive capability. The framework offers a scalable foundation for advanced infection surveillance and decision support in healthcare environments.

References

1. Schelenz et al., *Candida auris: An Emerging Pathogen*, Clinical Microbiology Reviews
2. Chowdhary et al., *Multidrug-resistant Candida auris*, Lancet Infectious Diseases
3. Keeling & Rohani, *Modeling Infectious Diseases in Humans and Animals*
