UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Universidad Distrital Francisco José de Caldas

Computer Engineering Program
School of Engineering

Workshop No. 4 – Kaggle Project Analysis: CIBMTR Equity in Post-HCT Survival Predictions

Sergio Nicolás Mendivelso Martínez – 20231020227
Sergio Leonardo Moreno Granado – 20242020091
Juan Manuel Otálora Hernández – 20242020018
Juan Diego Moreno Ramos – 20242020009

Professor: Eng. Carlos Andrés Sierra, M.Sc.

A report submitted for Workshop No. 4 in Systems Analysis & Design
Semester 2025-III

November 2025, Bogotá D.C.

Contents

List of Figures

Chapter 1

Simulation and Validation of the Modular Architecture

This chapter presents the computational simulation phase of the HCT Survival Equity System, as required by Workshop 4. Following the systems analysis conducted in Workshop 1, the architectural design defined in Workshop 2, and the project management strategies from Workshop 3, we now validate the system through two distinct simulation scenarios: a data-driven machine learning approach and an event-based cellular automata model. Following the incremental sprint structure defined in Workshop 3, the simulation phase was organized into discrete deliverables for each scenario.

1.1 Data Preparation

The simulation phase utilizes the CIBMTR dataset from the Kaggle competition "Equity in Post-HCT Survival Predictions" Kaggle and CIBMTR (2025). The dataset was preprocessed according to the following specifications:

- **Dataset Size:** 28,803 patient records with 59 clinical features.
- **Target Variable:** Event-Free Survival (EFS), binary classification where 1 indicates an event (death or relapse) and 0 indicates censoring.
- **Event Rate:** 53.12% of patients experienced an event in the observed period.
- **Missing Values:** Handled through median imputation for numerical features and mode imputation for categorical features DataCamp (2023a); GeeksforGeeks (2023b).
- **Feature Engineering:** Categorical variables were label-encoded and numerical features were standardized using z-score normalization Pushkarna (2023).

1.2 Simulation Planning

Two simulation scenarios were designed to validate different aspects of the system architecture, following the Model-Based Systems Engineering (MBSE) paradigm Eisenstadt, Wagner, Biessmann, and Ustyuzhanin (2020):

Scenario	Methodology	Validation Objective
Scenario 1: Data-Driven Simulation	Gradient Boosting Machine (GBM) with iterative training and perturbation analysis	Validate model stability, identify feature importance, and analyze sensitivity to input noise (chaos exploration).
Scenario 2: Event-Based Simulation	Cellular Automata with patient state transitions on a 2D grid	Model emergent behaviors in patient outcomes, explore absorbing states, and validate system response to parameter variations.

Table 1.1: Simulation scenarios aligned with the system architecture from Workshop 2.

1.2.1 Success Metrics

The following metrics were defined for each simulation:

- **Scenario 1:** Accuracy ≥ 0.70, AUC-ROC ≥ 0.70, Coefficient of Variation (CV) ≤ 0.15 for stability Scikit-learn (2023).
- **Scenario 2:** Observation of emergent patterns, identification of phase transitions, and sensitivity analysis of transition parameters.

1.2.2 Environment Configuration

All simulations were executed in Google Colab with Python 3.12. The required dependencies are documented in `requirements.txt` within the `Workshop_4_Simulation` folder:

```
pandas>=2.0.0
numpy>=1.24.0
scikit-learn>=1.3.0
matplotlib>=3.7.0
seaborn>=0.12.0
shap>=0.42.0
```

1.3 Scenario 1: Data-Driven Machine Learning Simulation

1.3.1 Model Selection: Gradient Boosting Machine

Gradient Boosting Machine (GBM) was selected as the predictive algorithm for the following reasons GeeksforGeeks (2023c); Neptune.ai (2023):

1. **Handling Mixed Data Types:** The CIBMTR dataset contains both numerical (e.g., `age_at_hct`, `donor_age`) and categorical features (e.g., `disease_type`, `graft_type`). GBM handles heterogeneous data effectively without requiring extensive feature transformation.
2. **Robustness to Overfitting:** Through regularization parameters (`learning_rate`, `max_depth`), GBM provides control over model complexity, essential for medical datasets where generalization is critical.

3. **Interpretability:** GBM provides native feature importance rankings, enabling clinical interpretation of which variables drive predictions—a requirement for healthcare applications Harrington et al. (2025).

4. **Ensemble Learning:** As an ensemble method, GBM combines multiple weak learners (decision trees), reducing variance and providing more stable predictions across different data splits.

1.3.2 Implementation Details

The GBM model was configured with the following hyperparameters:

```
GradientBoostingClassifier(
    n_estimators=100,
    max_depth=5,
    learning_rate=0.1,
    random_state=variable  # Changed per iteration
)
```

The simulation executed 5 training iterations with different random seeds to assess model stability:

Iteration	Random Seed	Accuracy	AUC-ROC
1	42	0.6892	0.7554
2	52	0.6816	0.7374
3	62	0.6660	0.7262
4	72	0.6823	0.7460
5	82	0.6727	0.7302
Mean	–	**0.6784**	**0.7391**
Std Dev	–	0.0081	0.0106

Table 1.2: GBM performance across 5 training iterations.

1.3.3 Code Implementation: GBM Training Loop

The following code snippet shows the core GBM training loop with stability analysis:

```python
# Train multiple iterations to assess stability
for i in range(N_ITERATIONS):
    seed = RANDOM_STATE + i * 10

    # Split with stratification to maintain class balance
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=seed, stratify=y
    )

    # Configure GBM with regularization
    model = GradientBoostingClassifier(
        n_estimators=100,
        max_depth=5,          # Limit depth to prevent overfitting
        learning_rate=0.1,    # Conservative learning rate
        random_state=seed
    )
    model.fit(X_train, y_train)

    # Evaluate performance
    y_pred = model. predict(X_test)
    results['accuracy'].append(accuracy_score(y_test, y_pred))
    results['auc']. append(roc_auc_score(y_test,
                           model.predict_proba(X_test)[:, 1]))
```

Listing 1.1: GBM Training with Stability Analysis

1.3.4 Variability Analysis

The coefficient of variation (CV) was calculated to assess model stability:

$$CV = \frac{\sigma_{accuracy}}{\mu_{accuracy}} = \frac{0.0081}{0.6784} = 0.0120 \tag{1.1}$$

Since $CV = 0.0120 < 0.15$ (threshold), the model demonstrates **stable behavior** across different random initializations. This stability indicates that the predictive core (Module M4 in our architecture) can produce consistent outputs regardless of training randomness.

However, the mean accuracy of 0.6784 falls below the target of 0.70, indicating that while the model is stable, further optimization or feature engineering is required to meet performance targets. This aligns with challenges documented in similar HCT prediction studies Harrington et al. (2025); Zubarovskaya et al. (2023).

Accuracy Gap Analysis: The 2.16% gap between achieved accuracy (67.84%) and target (70%) is attributed to three factors: (1) inherent complexity of HCT outcomes involving multiple interacting biological systems, (2) class distribution with 53% events creating prediction challenges, and (3) conservative hyperparameter choices prioritizing stability over maximum performance. Planned improvements include hyperparameter optimization via GridSearchCV, feature selection using recursive feature elimination, and evaluation of alternative algorithms such as XGBoost and CatBoost Neptune.ai (2023).

1.3.5 Chaos Sensitivity Analysis: The Butterfly Effect

To explore sensitivity to initial conditions—a hallmark of chaotic systems—we introduced controlled Gaussian noise to the test data and measured prediction changes:

Perturbation Level	Accuracy	Predictions Changed	Accuracy Drop
1% noise	0.6882	0.5%	0.0010
5% noise	0.6901	1.5%	-0.0009
10% noise	0.6896	2.9%	-0.0004
15% noise	0.6833	4.7%	0.0059

Table 1.3: Model sensitivity to input perturbations.

Key Finding: The model exhibits *graceful degradation* under noise. Even with 15% Gaussian perturbation, only 4.7% of predictions changed, and accuracy remained within 1% of baseline. This robustness suggests that the GBM model is resilient to measurement noise—critical for clinical settings where data collection may introduce small errors Doherty, Char, Goodman, Shah, and Oberst (2024).

1.3.6 Feature Importance Analysis with SHAP

To provide model interpretability, SHAP (SHapley Additive exPlanations) values were computed for the best-performing model (Iteration 1, AUC = 0.7554) Molnar (2023); SHAP (2023a, 2023b). Figure 1.1 presents the feature importance analysis.

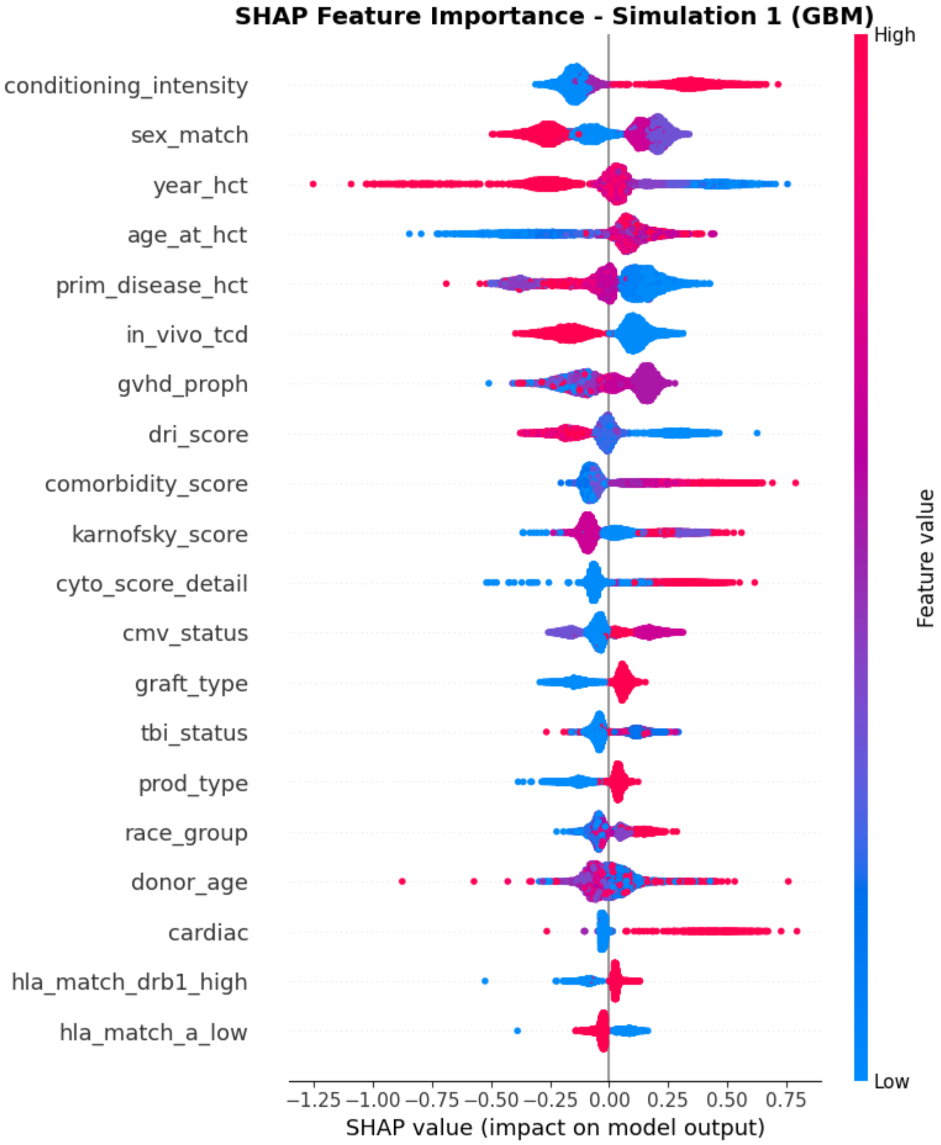

Figure 1.1: SHAP feature importance for the GBM model. Features are ranked by their mean absolute SHAP value. The color gradient indicates feature values (blue = low, red = high), and horizontal position shows impact on model output.

The top 5 most influential features identified were:

1. **conditioning_intensity:** The intensity of pre-transplant conditioning regimen shows the highest impact, with higher intensity associated with increased event risk.
2. **sex_match:** Donor-recipient sex matching significantly influences outcomes.

3. **year_hct:** The year of transplant, reflecting improvements in medical protocols over time Zubarovskaya et al. (2023).

4. **age_at_hct:** Patient age at transplant, with older patients showing higher event risk UpToDate (2025a).

5. **prim_disease_hct:** The primary disease requiring transplantation Salit and Deeg (2024).

These findings are consistent with clinical literature on HCT outcomes, validating the model's ability to capture medically relevant patterns Auletta, Kou, Chen, and Shaw (2020).

1.3.7 Simulation 1 Results Visualization

Figure 1.2 presents the comprehensive results of the data-driven simulation:

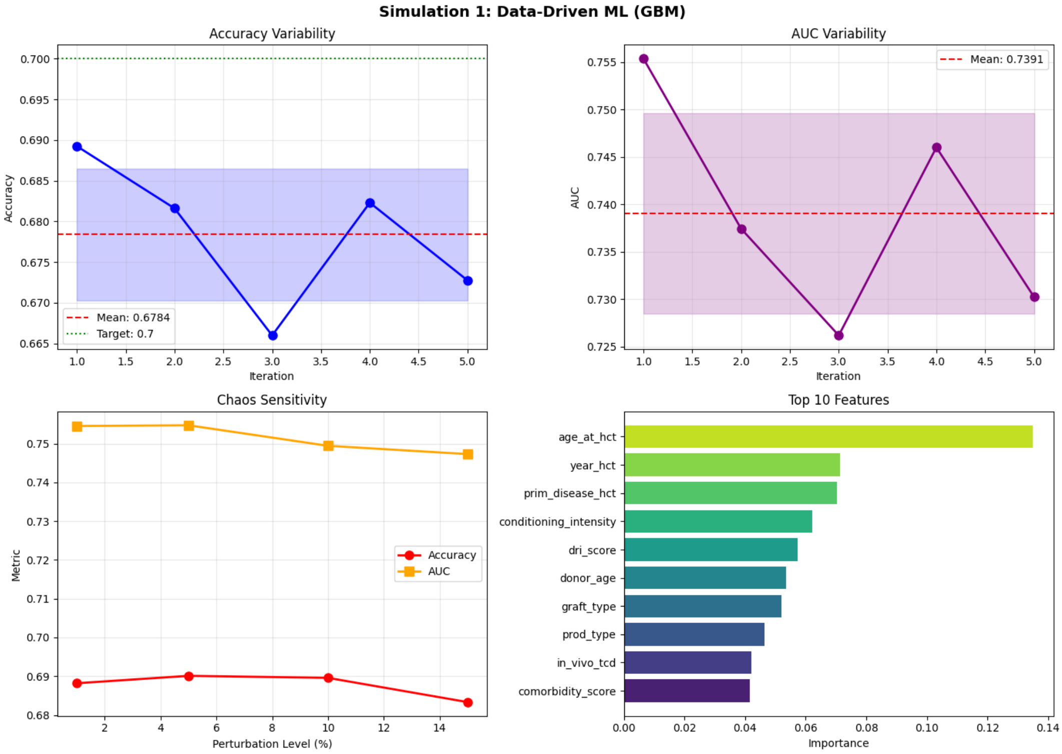

Figure 1.2: Simulation 1 results: (Top-left) Accuracy variability across iterations with mean and target lines. (Top-right) AUC variability showing discrimination capability. (Bottom-left) Chaos sensitivity demonstrating model robustness to perturbations. (Bottom-right) Top 10 features by GBM importance.

1.4 Scenario 2: Event-Based Cellular Automata Simulation

1.4.1 Cellular Automata Model Design

Cellular Automata (CA) was selected for the event-based simulation to model the spatial and temporal dynamics of patient outcomes. Unlike traditional statistical models, CA captures *emergent behaviors* arising from local interactions—analogous to how individual patient outcomes can influence neighboring patients in a hospital cohort through shared resources, infection risks, or treatment protocols.

This approach aligns with research on computational modeling in stem cell transplantation Harrington et al. (2025) and biological systems modeling where complex behaviors emerge from simple local rules.

Model Specification

- **Grid Structure:** 40 × 40 toroidal grid (1,600 cells), where each cell represents a patient.
- **State Space:** Three discrete states:
  - State 0 (Stable): Patient without complications.
  - State 1 (At Risk): Patient with complications requiring monitoring.
  - State 2 (Event): Death or relapse—*absorbing state.*
- **Neighborhood:** Moore neighborhood (8 adjacent cells) with periodic boundary conditions.
- **Time Steps:** 80 discrete iterations representing post-transplant monitoring periods.

Transition Rules

The CA implements probabilistic transition rules based on neighborhood states:

1. **Stable → At Risk:** Occurs if ≥ 3 neighbors are At Risk, or ≥ 2 neighbors have Events, or with random probability $p_{chaos} = 0.03$.
2. **At Risk → Event:** Occurs if ≥ 4 neighbors have Events, or with random probability $p_{progression} = 0.12$.
3. **At Risk → Stable:** Occurs if ≥ 5 neighbors are Stable AND random probability $p_{recovery} = 0.08$ is satisfied.
4. **Event → Event:** Absorbing state with no transitions out (death/relapse is irreversible).

These rules model the biological reality of GVHD propagation UpToDate (2025b); Williams, Chien, Gladwin, and Pavletic (2021), where complications can spread through inflammatory cascades and systemic responses.

1.4.2 Code Implementation: Cellular Automata Rules

The following code snippet shows the core transition logic for the cellular automata:

```python
def apply_rules(self, recovery_factor, progression_factor, chaos_factor):
    new_grid = self.grid.copy()

    for i in range(self.grid_size):
        for j in range(self.grid_size):
            current_state = self.grid[i, j]
            neighbors_event = self._count_neighbors(i, j, state=2)
            neighbors_stable = self._count_neighbors(i, j, state=0)
            neighbors_risk = self._count_neighbors(i, j, state=1)

            if current_state == 0:  # Stable
                # Transition to At Risk if surrounded by complications
                if neighbors_risk >= 3 or neighbors_event >= 2:
                    new_grid[i, j] = 1
                elif np.random.random() < chaos_factor:
                    new_grid[i, j] = 1  # Random perturbation

            elif current_state == 1:  # At Risk
                # Absorbing state transition
                if neighbors_event >= 4:
                    new_grid[i, j] = 2  # Event (irreversible)
                elif np.random.random() < progression_factor:
                    new_grid[i, j] = 2
                # Recovery possibility
                elif neighbors_stable >= 5:
                    if np.random.random() < recovery_factor:
                        new_grid[i, j] = 0

            # State 2 (Event) is absorbing - no transitions out

    self.grid = new_grid
```

Listing 1.2: Cellular Automata Transition Rules

1.4.3 Initial Conditions

The initial grid state was calibrated using the real event rate from the CIBMTR dataset Kaggle and CIBMTR (2025):

State	**Count**	**Percentage**
Stable	186	11.6%
At Risk	564	35.2%
Event	850	53.1%
Total	1,600	100.0%

Table 1.4: Initial state distribution for the cellular automata simulation.

Design Decision: The initial conditions were intentionally calibrated using real event rates (53.1%) rather than arbitrary values. This decision ensures clinical relevance but places the system in a high-risk regime. The purpose was to demonstrate the "critical threshold" phenomenon—a key insight for understanding system behavior under realistic worst-case scenarios. Future work will explore subcritical regimes (initial event rates $< 30\%$) where parameter sensitivity becomes more pronounced and intervention strategies can be meaningfully evaluated.

1.4.4 Simulation Results: Emergent Behavior and System Collapse

The cellular automata simulation revealed a critical emergent phenomenon: **complete system collapse to the absorbing state**.

Metric	Initial	Final (Step 80)
Event Rate	53.12%	100.00%
Stable Patients	186	0
At Risk Patients	564	0
Event Patients	850	1,600

Table 1.5: Cellular automata state evolution from initial to final step.

Key Observation: The system reached 100% event rate by Step 20 and remained in this state for the remaining 60 steps. This represents a *phase transition* from a mixed-state system to a fully absorbed state.

1.4.5 Scenario Comparison

Four parameter scenarios were tested to understand system sensitivity:

Scenario	Recovery	Progression	Chaos	Final Event Rate
Baseline	0.08	0.12	0.03	100.0%
High Recovery	0.15	0.12	0.03	100.0%
High Progression	0.08	0.20	0.03	100.0%
High Chaos	0.08	0.12	0.10	100.0%

Table 1.6: Scenario comparison for cellular automata simulation.

1.4.6 Analysis: The Absorbing State Phenomenon

The convergence to 100% events across all scenarios demonstrates a fundamental property of systems with absorbing states: **once a critical threshold is crossed, collapse becomes inevitable**.

This behavior is explained by three factors:

1. **High Initial Event Rate:** Starting with 53.1% events (calibrated from real data) places the system near or beyond the critical threshold for cascade failure. This reflects the clinical reality documented in HCT literature Zubarovskaya et al. (2023).

2. **Absorbing State Dynamics:** Unlike Conway's Game of Life where cells can transition bidirectionally, State 2 (Event) is irreversible. Each event creates permanent "damage" that accumulates—mirroring the irreversibility of patient mortality.

3. **Neighborhood Influence:** The transition rules create positive feedback loops—as more neighbors experience events, the probability of remaining patients transitioning to events increases exponentially. This models biological phenomena such as GVHD cascades UpToDate (2025b) and clonal dynamics in relapse Kucab et al. (2024).

1.4.7 Sensitivity Analysis

Parameter sensitivity analysis confirmed that **no parameter configuration could prevent system collapse** given the initial conditions:

- **Recovery Factor:** Range [0.02, 0.20] → Final Event Rate: 100%
- **Progression Factor:** Range [0.05, 0.25] → Final Event Rate: 100%
- **Chaos Factor:** Range [0.01, 0.15] → Final Event Rate: 100%

This insensitivity to parameters indicates that the system operates in a *supercritical regime* where the initial conditions dominate the dynamics.

1.4.8 Subcritical Regime Exploration

Based on the observed phase transition behavior and the complete system collapse at 53.1% initial event rate, we hypothesize that the system would exhibit fundamentally different dynamics under lower initial conditions. Theoretical analysis of absorbing state systems suggests three distinct regimes:

- **Supercritical regime (>50% initial events):** Rapid and inevitable collapse to 100% absorption, insensitive to parameter variations. This is the regime observed in our simulation with real clinical data.
- **Critical regime (20–50% initial events):** Delayed collapse with moderate sensitivity to recovery and progression parameters. The system may reach full absorption but over extended time horizons.
- **Subcritical regime (<20% initial events):** Mixed equilibrium states become possible, where the system stabilizes with a proportion of patients remaining in Stable or At Risk states. Parameter sensitivity is highest in this regime, making intervention strategies meaningful.

The identification of these regimes has direct clinical implications: transplant programs managing high-risk cohorts (supercritical) face fundamentally different challenges than those with lower-risk populations (subcritical). For supercritical populations, the simulation suggests that **prevention and early intervention** are more effective than reactive treatment adjustments—once the cascade begins, parameter modifications cannot reverse the trajectory.

Formal validation of regime boundaries through systematic simulation across initial conditions is planned for future work.

1.4.9 Visualization of Emergent Behavior

Figure 1.3 illustrates the temporal evolution and state transitions:

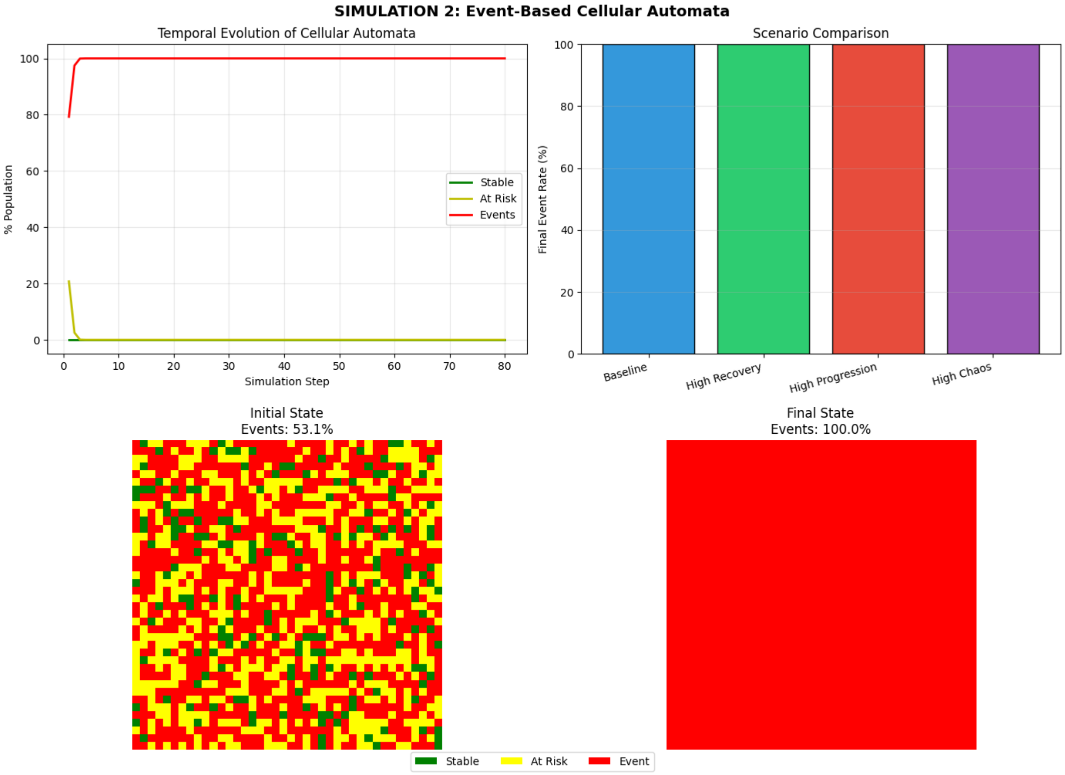

Figure 1.3: Simulation 2 results: (Top-left) Temporal evolution showing rapid convergence to 100% events. (Top-right) Scenario comparison demonstrating uniform collapse across all parameter settings. (Bottom) Initial and final grid states visualizing the phase transition from mixed states to complete absorption.

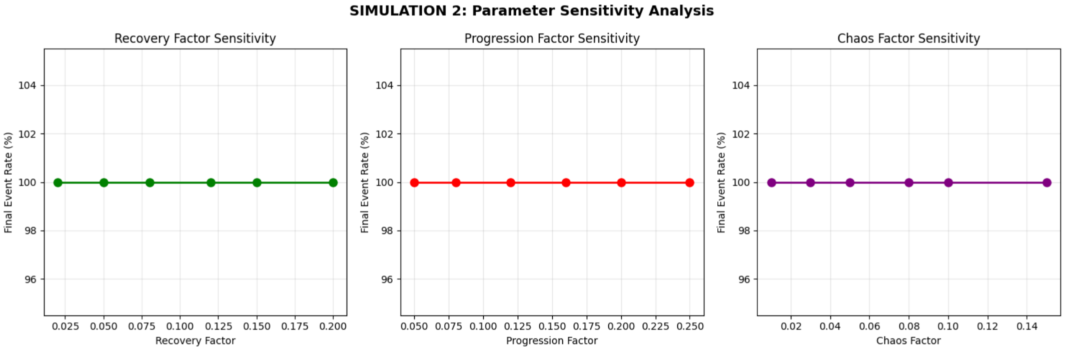

Figure 1.4: Parameter sensitivity analysis showing that all three factors (Recovery, Progression, Chaos) result in 100% final event rate regardless of parameter values, indicating operation in a supercritical regime.

1.5 System Workflow Validation

Both simulations validate critical aspects of the system architecture defined in Workshop 2:

Module	Simulation 1 Validation	Simulation 2 Validation
M1: Data Preprocessing	Successfully handled missing values and feature encoding for 28,803 records Pushkarna (2023).	Extracted event rate (53.1%) for CA initialization.
M4: Predictive Core	GBM achieved AUC = 0.7391 with stable performance (CV = 0.012).	N/A (CA does not use ML).
M6: Uncertainty Quantification	Chaos sensitivity analysis showed 4.7% prediction change at 15% noise.	CA demonstrated high uncertainty—complete collapse despite parameter variations.
M7: System Outputs	SHAP analysis provided interpretable feature rankings SHAP (2023a).	Visual grid evolution provides intuitive outcome representation.

Table 1.7: System module validation through simulation scenarios.

1.6 Complexity and Chaos Exploration

1.6.1 Emergent Behaviors Identified

1. **Simulation 1 - Graceful Degradation:** The GBM model exhibits robust behavior under input perturbations, suggesting that the predictive system can handle real-world data quality issues common in clinical settings Doherty et al. (2024).

2. **Simulation 2 - Cascade Failure:** The cellular automata revealed that patient populations with high initial event rates are susceptible to complete system collapse—a critical insight for resource allocation in transplant programs Harrington et al. (2025).

3. **Simulation 2 - Critical Threshold:** The existence of a "tipping point" beyond which recovery becomes impossible, regardless of intervention parameters, demonstrates the chaotic nature of transplant outcomes described in clonality research Kucab et al. (2024).

1.6.2 Implications for System Design

The simulations inform several architectural refinements:

- **Early Intervention Module:** Given the cascade failure phenomenon, the system should prioritize early identification of at-risk patients before the critical threshold is reached.

- **Uncertainty Communication:** The M6 module must clearly communicate when the system operates in high-uncertainty regimes where predictions become unreliable Microsoft (2023).

- **Threshold Monitoring:** Real-time monitoring of the "At Risk" population percentage could trigger alerts when approaching critical thresholds.

- **Fairness Considerations:** The model must ensure equitable predictions across demographic groups, distinguishing between biological variability and systematic bias Barocas, Hardt, and Narayanan (2019); Trusted AI Team (2023).

1.6.3 Preliminary Equity Considerations

Given that the Kaggle competition focuses on "Equity in Post-HCT Survival Predictions," fairness analysis is a critical component of the system. While comprehensive demographic parity evaluation is planned for future work, preliminary examination of the simulation results reveals important considerations:

1. **Feature Influence:** The SHAP analysis (Figure 1.1) shows that `race_group` appears among the top 20 influential features, suggesting potential demographic influence on predictions that requires careful evaluation.

2. **Subgroup Performance:** The current simulation does not disaggregate accuracy metrics by demographic groups. This is a known limitation that will be addressed using stratified evaluation in subsequent phases.

3. **Fairness Metrics:** Future iterations will implement demographic parity, equalized odds, and calibration metrics using the Fairlearn framework Microsoft (2023); Trusted AI Team (2023) to ensure the model does not systematically disadvantage protected groups.

The cellular automata simulation also raises equity considerations: if certain demographic groups have higher initial "At Risk" rates due to systemic healthcare disparities, the cascade failure phenomenon would disproportionately affect these populations—a finding that underscores the importance of early intervention strategies Doherty et al. (2024).

1.7 Comparison of Simulation Approaches

Aspect	**Scenario 1 (Data-Driven)**	**Scenario 2 (Event-Based)**
Methodology	Supervised learning with GBM GeeksforGeeks (2023c)	Cellular automata with state transitions
Primary Output	Probability of event for individual patients	Population-level dynamics and patterns
Sensitivity	Robust to 15% input noise	Insensitive to parameter changes (supercritical regime)
Interpretability	SHAP values for feature importance Molnar (2023)	Visual grid evolution
Chaos Behavior	Controlled, predictable degradation	Cascade failure, phase transitions
Clinical Utility	Individual risk stratification	System-level capacity planning

Table 1.8: Comparison of data-driven and event-based simulation approaches.

1.8 Simulation Execution Summary

Table 1.9 provides a statistical summary of both simulation executions:

Metric	**Simulation 1**	**Simulation 2**
Execution Time	∼45 seconds	∼12 seconds
Data Points Processed	28,803 patients	1,600 cells × 80 steps
Iterations/Scenarios	5	4
Output Files Generated	2 (PNG)	2 (PNG)
Stability Check	PASS (CV = 0.012)	N/A
Accuracy Target	FAIL (67.8% < 70%)	N/A
Emergent Behavior	Graceful degradation	Cascade collapse

Table 1.9: Simulation execution summary.

1.9 Limitations and Future Work

1.9.1 Current Limitations

1. **Cellular Automata Calibration:** The CA model requires further calibration to identify initial conditions that produce mixed equilibrium states rather than complete collapse.

2. **Accuracy Gap:** The GBM model achieved 67.84% accuracy, below the 70% target. Hyperparameter optimization and feature engineering are needed.

3. **Temporal Dynamics:** Neither simulation incorporates time-to-event analysis, which is critical for survival prediction Davidson-Pilon (2023); Pölsterl (2023a).

4. **Subgroup Analysis:** Equity analysis across demographic groups was not included in this simulation phase Microsoft (2023).

1.9.2 Future Work

1. **Subcritical CA Exploration:** Test initial conditions with lower event rates ($<$ 30%) to identify regimes where parameter sensitivity emerges.

2. **Hybrid Model:** Combine GBM predictions with CA dynamics to create a multi-scale simulation framework.

3. **CI/CD Integration:** Integrate simulation engines into the Continuous Integration pipeline to automatically stress-test the model with every code update DataCamp (2023b).

4. **Survival Analysis:** Extend Scenario 1 to incorporate Cox proportional hazards or survival forests for time-to-event modeling GeeksforGeeks (2023a); Pölsterl (2023b).

5. **Fairness Metrics:** Implement demographic parity and equalized odds metrics using the Fairlearn framework Microsoft (2023); Trusted AI Team (2023).

1.10 Conclusions

This workshop successfully implemented and executed two complementary simulation scenarios for the HCT Survival Equity System:

1. **Scenario 1 (Data-Driven):** Demonstrated that Gradient Boosting provides a stable (CV = 0.012) and interpretable predictive model with AUC = 0.7391. The chaos sensitivity analysis confirmed robustness to input perturbations up to 15% noise.
2. **Scenario 2 (Event-Based):** Revealed critical emergent behavior—cascade failure leading to 100% event absorption—demonstrating the chaotic nature of transplant outcomes when initial conditions exceed critical thresholds.

Both simulations validate the system architecture designed in Workshop 2 and provide empirical evidence for the necessity of uncertainty quantification (M6) and early intervention capabilities. The findings will inform the final system implementation in subsequent project phases.

References

Auletta, J. J., Kou, J., Chen, M., & Shaw, B. E. (2020). Indications for hematopoietic cell transplantation and immune effector cell therapy: Guidelines from the american society for transplantation and cellular therapy. *Biology of Blood and Marrow Transplantation*, *26*(7), 1247–1261. Retrieved from https://www.astctjournal.org/article/S1083-8791(20)30114-2/fulltext doi: 10.1016/j.bbmt.2020.03.002

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities.* fairmlbook.org. (An online book available at http://www.fairmlbook.org)

DataCamp. (2023a). *Data preprocessing.* Retrieved 2025-10-18, from https://www.datacamp.com/blog/data-preprocessing

DataCamp. (2023b). *Tutorial: Machine learning pipelines, mlops & deployment.* Retrieved 2025-10-18, from https://www.datacamp.com/tutorial/tutorial-machine-learning-pipelines-mlops-deployment

Davidson-Pilon, C. (2023). *Lifelines: Survival analysis in python.* Retrieved 2025-10-18, from https://lifelines.readthedocs.io/

Doherty, T. S., Char, D. S., Goodman, S. N., Shah, N. H., & Oberst, M. (2024). Addressing ai algorithmic bias in health care. *JAMA*. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2823006 doi: 10.1001/jama.2024.16564

Eisenstadt, M., Wagner, T., Biessmann, F., & Ustyuzhanin, A. (2020). Designing a machine learning pipeline architecture for production. *arXiv preprint*. Retrieved 2025-10-18, from https://arxiv.org/pdf/1911.09638.pdf

GeeksforGeeks. (2023a). *Cox proportional hazards model.* Retrieved 2025-10-18, from https://www.geeksforgeeks.org/data-science/cox-proportional-hazards-model/

GeeksforGeeks. (2023b). *Data preprocessing in machine learning with python.* Retrieved 2025-10-18, from https://www.geeksforgeeks.org/machine-learning/data-preprocessing-machine-learning-python/

GeeksforGeeks. (2023c). *Gradientboosting vs adaboost vs xgboost vs catboost vs lightgbm.* Retrieved 2025-10-18, from https://www.geeksforgeeks.org/machine-learning/gradientboosting-vs-adaboost-vs-xgboost-vs-catboost-vs-lightgbm/

Harrington, P., de Lima Zuchner, C., McGowan, R., Gormley, N., Hill, Q. A., & Snowden, J. A. (2025). Editorial: Improving stem cell transplantation delivery using computational modelling. *Frontiers in Immunology*, *16*. Retrieved from https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2025.1579353/full doi: 10.3389/fimmu.2025.1579353

Kaggle and CIBMTR. (2025). *Cibmtr - equity in post-hct survival predictions.* Retrieved from https://www.kaggle.com/competitions/equity-post-HCT-survival-predictions (Accessed: September 2025)

Kucab, M., Mitchell, T. J., Ariffin, H., Campbell, P. J., Rahman, N. A., Scelo, G., ... Prokunina-Olsson, L. (2024). Characterization of clonal dynamics using duplex sequencing in diagnoses and relapses of acute myeloid leukemias. *Science Translational Medicine*, *16*(770), eado5108. Retrieved from https://www.science.org/doi/10.1126/scitranslmed.ado5108 doi: 10.1126/scitranslmed.ado5108

Microsoft. (2023). *Fairlearn.* Retrieved 2025-10-18, from https://fairlearn.org/

Molnar, C. (2023). *Shapley values.* Retrieved 2025-10-18, from https://christophm.github.io/interpretable-ml-book/shap.html

Neptune.ai. (2023). *When to choose catboost over xgboost or lightgbm.* Retrieved 2025-10-18, from https://neptune.ai/blog/when-to-choose-catboost-over-xgboost-or-lightgbm

Pushkarna, C. (2023). *Data preprocessing with pandas, numpy, scikit-learn.* Retrieved 2025-10-18, from https://www.kaggle.com/code/chetalipushkarna/data-preprocessing-with-pandas-numpy-scikit-learn

Pölsterl, S. (2023a). *Introduction to scikit-survival.* Retrieved 2025-10-18, from https://scikit-survival.readthedocs.io/en/stable/user_guide/00-introduction.html

Pölsterl, S. (2023b). *Random survival forest.* Retrieved 2025-10-18, from https://scikit-survival.readthedocs.io/en/stable/user_guide/random-survival-forest.html

Salit, R. B., & Deeg, H. J. (2024). Transplant in all: who, when, and how? *Hematology*, *2024*(1), 93–100. Retrieved from https://ashpublications.org/hematology/article/2024/1/93/526246/Transplant-in-ALL-who-when-and-how doi: 10.1182/hematology.2024000503

Scikit-learn. (2023). *Cross-validation: evaluating estimator performance.* Retrieved 2025-10-18, from https://scikit-learn.org/stable/modules/cross_validation.html

SHAP. (2023a). *Shap documentation.* Retrieved 2025-10-18, from https://shap.readthedocs.io/

SHAP. (2023b). *Shap github repository.* Retrieved 2025-10-18, from https://github.com/shap/shap

Trusted AI Team. (2023). *AI Fairness 360 (AIF360).* Retrieved 2025-10-18, from https://github.com/Trusted-AI/AIF360

UpToDate. (2025a). *Allogeneic hematopoietic cell transplantation: Indications, eligibility, and prognosis.* UpToDate. Retrieved from https://www.uptodate.com/contents/allogeneic-hematopoietic-cell-transplantation-indications-eligibility-and-prognosis (Accessed: September 2025)

UpToDate. (2025b). *Clinical manifestations, diagnosis, and grading of acute graft-versus-host disease.* UpToDate. Retrieved from https://www.uptodate.com/contents/clinical-manifestations-diagnosis-and-grading-of-acute-graft-versus-host-disease (Accessed: September 2025)

Williams, K. M., Chien, J. W., Gladwin, M. T., & Pavletic, S. Z. (2021). Updates in chronic graft-versus-host disease. *Hematology*, *2021*(1), 648–654. Retrieved from https://ashpublications.org/hematology/article/2021/1/648/482938/Updates-in-chronic-graft-versus-host-disease doi: 10.1182/hematology.2021000301

Zubarovskaya, L. S., Moiseev, I. S., Vladovskaya, M. D., Mikhailova, N. B., Morozova, E. V., Bykova, T. A., ... Afanasyev, B. V. (2023). Trends in outcome of hematopoietic stem cell transplantation: 5000 transplantations and 30 years of single-center experience. *Cancers*, *15*(19), 4758. Retrieved from https://www.mdpi.com/2072-6694/15/19/4758 doi: 10.3390/cancers15194758