A Nano-Topological Bayesian Inference Framework for Evaluating Core Hazard Intensity in Clinical Diagnostics

*

Abstract

In this this study, we introduced a novel deterministic Coefficient of Intensity derived from nano-topological boundary expansions to isolate the true structural damage caused by symptom removal. Because purely deterministic point-estimates are highly vulnerable to natural sample variance in small data set, we then applied the Bayesian theorem by mapping this coefficient to a Beta-Binomial conjugate model. This Bayesian approach allows us to incorporate statistical uncertainty and generate a posterior distributions that bridge topological approximation with predictive reliability. Statistical hazard intensity is evaluated through Joint Posterior Dominance (computed via Monte Carlo integration) and Interval Displacement (using the Inverse Regularized Beta Function). The proposed Bayesian-Topological model successfully resolves the structural blindness of classical rough set measures. By mathematically filtering out natural sample variance, it proves definitively that Low Platelets is the highly significant driving hazard in Dengue diagnostics. This framework provides a transparent, computationally efficient, and mathematically verified mechanism to rank hazard intensity for clinical decision support and Machine Learning expert systems.

Keywords: Nano Topology; Core Factors; Bayesian Inference; Joint Probability; Hazard Displacement; Credible Intervals; Machine Learning.

1 Introduction

Medical diagnosis increasingly relies on identifying core factors that act as primary indicators of disease under conditions of uncertainty. The foundational theory of rough sets, introduced by Pawlak (1982), established a powerful mathematical tool to manage vagueness and incomplete information by employing lower and upper approximations. Over time, these concepts have been structurally integrated with topological data analysis. For instance, Abu-Gdairi et al. (2022) explored topological approaches for generalized rough sets, significantly enhancing their decision-making applications in complex environments. This intersection naturally led to the development of Nano-Topological spaces, which utilize these approximations to model clinical datasets mathematically Thivagar and Richard (2014). This theory has been widely applied to various medical domains, including identifying cardiovascular risk factors Jayalakshmi and Janaki (2017). Further expanding on biological network complexities, Nawar and El Atik (2019) successfully modeled the functional mechanisms of a human heart via graph nano topological spaces, demonstrating the efficacy of topological structures in capturing vital physiological connections.

Recently, rough set theory and topological structures have been extensively adapted for feature selection, attribute reduction, and predictive modeling in medical expert systems. Singh and Mantri (2024) developed an intelligent recommender system using machine learning association rules alongside rough sets to effectively predict diseases even from incomplete symptom sets. The application of these models became particularly crucial during the global pandemic; El-Gayar and El Atik (2022) utilized topological models of rough sets to streamline critical decision-making processes for COVID-19 patients, while El-Bably et al. (2025) advanced this logic by using novel initial-neighborhood systems within rough set approaches for the precise medical diagnosis of COVID-19 variants. In the context of vector-borne diseases, Jeevitha et al. (2021) applied nano-topology to observe vital factors of Dengue fever, extracting Fever and Low Platelets as core factors through classical attribute reduction.

However, identifying core factors is only the first step. In modern clinical data science, assessing the relative severity and statistical significance of these factors requires advanced probabilistic metrics. Recent studies emphasize the necessity of integrating Bayesian inference into topological feature ranking to validate diagnostic indicators robustly Al-shami (2021). Bayesian methods are becoming increasingly popular in clinical trial design and analysis because of their flexibility and their ability to incorporate prior knowledge into the evaluation of new data Ginn et al. (2025). Bayesian analysis combines previous information (represented by a mathematical probability distribution, the prior) with information from the study (the likelihood function) to generate an updated probability distribution (the posterior) repre-

senting the information available for clinical decision making. In contrast to the frequentist dichotomised approach based on a P value, the application of Bayesian statistics allocates credibility to a continuous spectrum of possibilities and for this reason a Bayesian approach to inference is often warranted as it will incorporate uncertainty when updating our current belief with information from a new trial Goligher et al. (2024); Frost et al. (2021). Furthermore, as future expansions demand, we want to implement a Bayesian model capable of analysing time-to-event data in the context of competing risk Al-Aziz et al. (2022).

In this article, we propose a distinct mathematical approach that extends nano-topology to measure the exact intensity of core attributes. We propose deterministic equations to calculate Intensity Factors ($I_A(X)$) for clinical hazards, specifically applied to Dengue fever. To overcome the fragility of deterministic measures in small sample sizes, we map the Intensity Factor to a sequence of Bernoulli trials. By constructing robust Beta-Binomial Bayesian posterior distributions, we definitively rank factor significance, bridging the structural approximation of rough sets with the predictive rigor of probabilistic machine learning Bishop and Nasrabadi (2006).

2 Preliminaries

To establish a robust framework for attribute reduction, we formalize the concepts of approximations and Nano Topology as introduced by Thivagar and Richard (2013).

Definition 2.1 (Approximation Space)**.** *Let U be a non-empty, finite universe of objects and R be an equivalence relation on U. The pair (U, R) is called an approximation space. Let U/R denote the family of equivalence classes of U by R. For any subset $X \subseteq U$, the lower approximation $L_R(X)$, upper approximation $U_R(X)$ and boundary region $B_R(X)$ are defined as:*

$$L_R(X) = \bigcup \{E \in U/R \mid E \subseteq X\}$$

$$U_R(X) = \bigcup \{E \in U/R \mid E \cap X \neq \emptyset\}$$

$$B_R(X) = U_R(X) \setminus L_R(X)$$

Definition 2.2 (Nano Topology)**.** *Let U be the universe, R be an equivalence relation on U, and $X \subseteq U$. The Nano Topology $\tau_R(X)$ with respect to X is defined as the collection:*

$$\tau_R(X) = \{U, \emptyset, L_R(X), U_R(X), B_R(X)\}$$

Theorem 2.1. *For any $X \subseteq U$, the collection $\tau_R(X) = \{U, \emptyset, L_R(X), U_R(X), B_R(X)\}$ forms a topology on U.*

Proof. To prove that $\tau_R(X)$ is a topology on U, we must satisfy the three axioms of topological spaces:

1. **Axiom 1:** By definition, $U \in \tau_R(X)$ and $\emptyset \in \tau_R(X)$.

2. **Axiom 2 (Arbitrary Unions):** We must verify that the union of any subcollection of sets in $\tau_R(X)$ is also in $\tau_R(X)$. Since it is a finite collection, we check pairwise unions:

$$L_R(X) \cup U_R(X) = U_R(X) \in \tau_R(X) \quad (\text{since } L_R(X) \subseteq U_R(X)).$$

$$L_R(X) \cup B_R(X) = L_R(X) \cup (U_R(X) \setminus L_R(X)) = U_R(X) \in \tau_R(X).$$

$$U_R(X) \cup B_R(X) = U_R(X) \in \tau_R(X) \quad (\text{since } B_R(X) \subseteq U_R(X)).$$

3. **Axiom 3 (Finite Intersections):** We must verify that the intersection of any subcollection of sets in $\tau_R(X)$ is also in $\tau_R(X)$. We check pairwise intersections:

$$L_R(X) \cap U_R(X) = L_R(X) \in \tau_R(X).$$

$$L_R(X) \cap B_R(X) = L_R(X) \cap (U_R(X) \setminus L_R(X)) = \emptyset \in \tau_R(X).$$

$$U_R(X) \cap B_R(X) = B_R(X) \in \tau_R(X) \quad (\text{since } B_R(X) \subseteq U_R(X)).$$

Since all axioms are satisfied, $\tau_R(X)$ is a topology on U. □

3 Introduction to the Information System

In the application of Nano Topology to medical decision-making, patient data is represented as an information system Jeevitha et al. (2021). Table 1 presents the survey data of 15 patients ($U = \{P_1, P_2, ..., P_{15}\}$).

The system consists of Conditional Attributes (C), representing symptoms: Fever (F), Headache (HA), Vomiting (V), Muscle and Joint Pain (MJP), Skin Rashes (SR), Low Level Blood Platelets (LLBP), and Bleeding (B). It also contains a Decision Attribute (D), representing the final diagnosis (Results): Dengue Fever. The data is encoded in a binary format where 1 denotes the presence of a symptom/disease and 0 denotes its absence.

4 Methodology for Attribute Reduction

The methodology aims to eliminate redundant symptoms and isolate the core factors responsible for Dengue Fever by analyzing how the removal of specific attributes affects the approximations. Figure 1 outlines the proposed analytical framework.

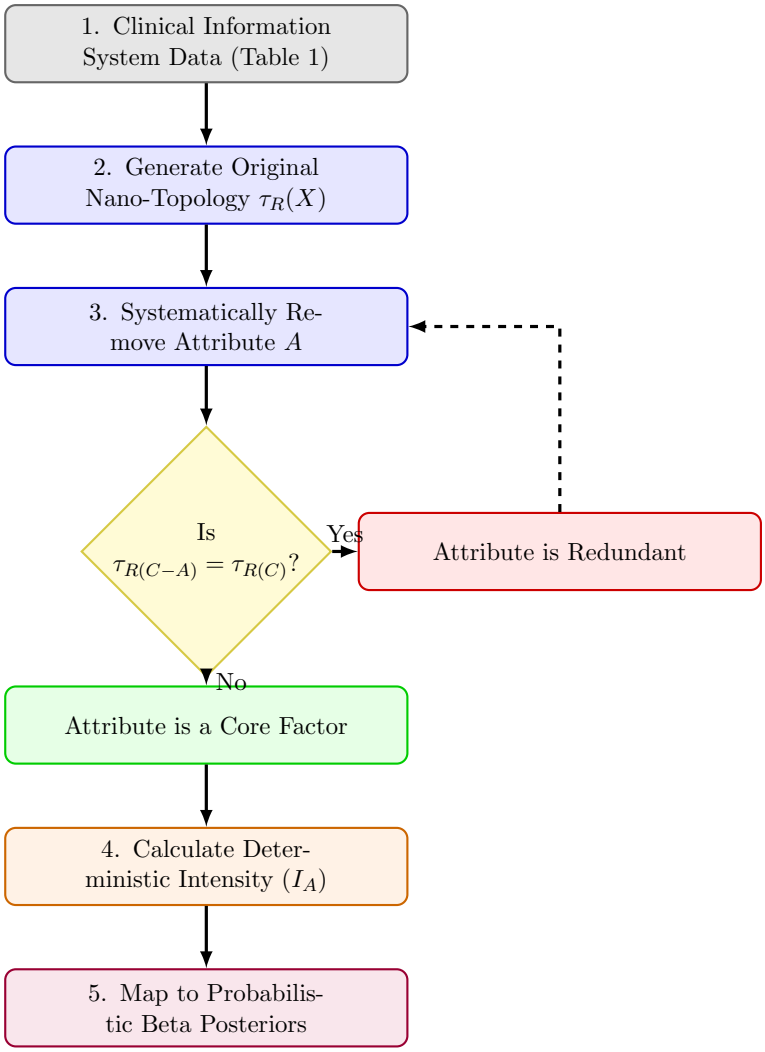

Figure 1: Flowchart of the Proposed Framework from Nano-Topological Reduction to Bayesian Inference.

Table 1: Patient Survey Data (Binary Format with Categorical Results)

Patients	**F**	**HA**	**V**	**MJP**	**SR**	**LLBP**	**B**	**Results (Dengue)**
P_1	1	1	0	0	1	1	0	Yes
P_2	1	1	1	1	1	1	1	Yes
P_3	1	1	0	0	1	1	0	No
P_4	1	0	1	0	0	1	0	Yes
P_5	1	0	1	0	0	0	0	No
P_6	1	0	0	1	1	1	0	Yes
P_7	1	1	0	0	1	1	0	No
P_8	1	0	1	0	0	1	0	No
P_9	1	0	1	0	0	0	0	No
P_{10}	1	0	0	1	1	1	0	Yes
P_{11}	1	0	0	1	1	0	0	No
P_{12}	0	1	0	0	1	1	0	No
P_{13}	1	0	1	0	0	0	0	No
P_{14}	1	0	0	1	1	0	0	No
P_{15}	1	1	1	1	1	1	1	Yes

4.1 Equivalence Classes and Target Set

Based on Table 1, the universe U is partitioned into equivalence classes based on all condition attributes C:

- $E_1 = \{P_1, P_3, P_7\}$
- $E_2 = \{P_2, P_{15}\}$
- $E_3 = \{P_4, P_8\}$
- $E_4 = \{P_5, P_9, P_{13}\}$
- $E_5 = \{P_6, P_{10}\}$
- $E_6 = \{P_{11}, P_{14}\}$
- $E_7 = \{P_{12}\}$

Let X be the target set of patients who actually tested positive for Dengue Fever ("Yes").

$$X = \{P_1, P_2, P_4, P_6, P_{10}, P_{15}\}$$

4.2 Approximations and Initial Topology

Using the equivalence classes, we calculate three regional approximations for set X:

1. **Lower Approximation** ($L_R(X)$):
   $L_R(X) = E_2 \cup E_5 = \{P_2, P_6, P_{10}, P_{15}\}$
2. **Upper Approximation** ($U_R(X)$):
   $U_R(X) = E_1 \cup E_2 \cup E_3 \cup E_5 = \{P_1, P_3, P_7, P_2, P_{15}, P_4, P_8, P_6, P_{10}\}$
3. **Boundary Region** ($B_R(X)$):
   $B_R(X) = E_1 \cup E_3 = \{P_1, P_3, P_4, P_7, P_8\}$

The generated Nano Topology is:

$$\tau_{R(C)}(X) = \{U, \emptyset, L_R(X), U_R(X), B_R(X)\}$$

5 Determining the Core Factor

To find the CORE attributes, we systematically remove one attribute at a time from C, recalculate $U/R(C - A)$, and generate a new topology $\tau_{R(C-A)}(X)$.

Decision Rule:

- If $\tau_{R(C-A)}(X) = \tau_{R(C)}(X)$, attribute A is redundant.
- If $\tau_{R(C-A)}(X) \neq \tau_{R(C)}(X)$, attribute A is indispensable. The set of all indispensable attributes constitutes the CORE.

5.1 Execution of Reduction Steps

Step 1: Removing FEVER (F)
When F is removed, the relation weakens. Patients in E_1 (P_1, P_3, P_7) and E_7 (P_{12}) merge into a single larger equivalence class. The upper approximation expands to include P_{12}. Because $\tau_{R(C-F)}(X) \neq \tau_{R(C)}(X)$, FEVER is an indispensable core factor.

Step 2: Removing LOW LEVEL BLOOD PLATELETS (LLBP)
When LLBP is removed, classes $E_3 = \{P_4, P_8\}$ and $E_4 = \{P_5, P_9, P_{13}\}$ merge. The upper approximation violently expands to include P_5, P_9, and P_{13}. Because $\tau_{R(C-LLBP)}(X) \neq \tau_{R(C)}(X)$, LLBP is an indispensable core factor.

Step 3: Removing other attributes (HA, V, MJP, SR, B)
Removing any of the remaining condition attributes does not affect the original equivalence boundaries. Thus, $\tau_{R(C-A)}(X) = \tau_{R(C)}(X)$. These attributes are mathematically redundant.

6 Theoretical Impact of Attribute Reduction

Before executing the reduction, it is mathematically critical to understand why removing an attribute affects equivalence classes and alters the ap-

proximations. The behavior is governed by the principles of indiscernibility resolution as theorized by Pawlak (1991).

Theorem 6.1 (Monotonicity of Equivalence Relations Pawlak (1991))**.** *Let C be the full set of conditional attributes and $A \in C$ be a single attribute. If we reduce the set to $C-A$, then the indiscernibility relation becomes coarser: $R(C) \subseteq R(C-A)$. Consequently, every equivalence class in $U/R(C-A)$ is a union of one or more equivalence classes from $U/R(C)$.*

Proof. Let two patients $(x, y) \in R(C)$. This implies that x and y share identical symptoms for all attributes in C. If an attribute A is removed, x and y must naturally still share identical values for the remaining smaller set $C - A$. Thus, $(x, y) \in R(C - A)$. Because elements indiscernible under the full set remain indiscernible under a subset, the equivalence classes either remain the same or merge (grow larger) when an attribute is stripped away. □

Theorem 6.2 (Approximation Expansion Bounds Pawlak (1991))**.** *For any target set $X \subseteq U$, if attributes are removed such that $R(C) \subseteq R(C - A)$, then the lower approximation shrinks or remains equal, and the upper approximation expands or remains equal. Mathematically:*

$$L_{C-A}(X) \subseteq L_C(X) \quad \textit{and} \quad U_C(X) \subseteq U_{C-A}(X)$$

Proof. Let $E_{C-A} \in U/R(C - A)$ be a merged equivalence class such that $E_{C-A} \subseteq X$. Since E_{C-A} is formed by the union of smaller C-classes, those constituent C-classes must also be subsets of X. Therefore, any element in the lower approximation of $C - A$ is guaranteed to be in the lower approximation of C, proving $L_{C-A}(X) \subseteq L_C(X)$.

Conversely, if a strict subset class $E_C \in U/R(C)$ intersects X (meaning it belongs to $U_C(X)$), the larger merged class $E_{C-A} \supseteq E_C$ is forced to also intersect X. Thus, every element in the original upper approximation will fall into the new upper approximation, proving $U_C(X) \subseteq U_{C-A}(X)$. □

7 Evaluation: Pawlak's Measure vs Coefficient of Intensity

7.1 Limitations of Pawlak's Accuracy for Attribute Ranking

To quantitatively evaluate the exactness of the rough sets generated, standard practice applies the accuracy of approximation proposed by Pawlak (1991).

Definition 7.1 (Accuracy of Approximation)**.** *Let $X \subseteq U$ be a target set evaluated under an equivalence relation R. Pawlak's accuracy of approxi-*

mation, denoted as $\alpha_R(X)$, is defined as:

$$\alpha_R(X) = \frac{|L_R(X)|}{|U_R(X)|} \quad \textit{where } U_R(X) \neq \emptyset.$$

Theorem 7.1. *For any non-empty approximation space (U, R) and any subset $X \subseteq U$ where $U_R(X) \neq \emptyset$, the accuracy measure satisfies $0 \leq \alpha_R(X) \leq 1$.*

Proof. By the fundamental properties of rough sets and Nano Topology, the lower approximation is always a subset of the target set, which in turn is a subset of the upper approximation. Therefore:

$$L_R(X) \subseteq X \subseteq U_R(X)$$

Taking the cardinality of these sets preserves the inequality:

$$0 \leq |L_R(X)| \leq |X| \leq |U_R(X)|$$

Dividing by $|U_R(X)| > 0$ yields $0 \leq \frac{|L_R(X)|}{|U_R(X)|} \leq 1$. □

While mathematically sound for defining the *global* exactness of a dataset, Pawlak's measure is insufficient for ranking individual symptoms. Pawlak's global metric is fundamentally blind to the dynamic local impact of eliminating a single conditional attribute. It evaluates the set size passively rather than isolating the active structural fragmentation caused strictly by the absence of attribute A.

7.2 Formulation of the Novel Coefficient of Intensity

To rectify this limitation, we propose a customized Coefficient of Intensity (I_A) that isolates the topological damage caused when a core attribute is eliminated. We posit that the true severity of a core attribute depends entirely on exactly two topological dimensions:

1. **Global Granularity Loss ($1 - \frac{E^*(X)}{E(X)}$):** The number of equivalence classes dictates the granularity of knowledge in the system. Removing an indispensable attribute merges classes, reducing the total equivalence classes from $E(X)$ to $E^*(X)$. This ratio measures the systemic structural degradation percentage.

2. **Local Uncertainty Expansion ($|B_R^*(X) - B_R(X)|$):** The boundary region dictates empirical uncertainty. The exact count of additional objects forced into the boundary region mathematically measures the direct loss of diagnostic clarity.

Multiplying a fractional weight (structural degradation rate) by an absolute integer (uncertainty expansion) creates a harmonized severity score. The proposed Coefficient of Intensity factor is formulated as:

$$I_A(X) = \left[1 - \left(\frac{E^*(X)}{E(X)}\right)\right] \times |B_R^*(X) - B_R(X)|$$

Evaluating this for our dataset:

$$I_F(X) = \left[1 - \left(\frac{6}{7}\right)\right] \times (1) = \frac{1}{7} \approx 0.14 \quad \text{and} \quad I_{LP}(X) = \left[1 - \left(\frac{6}{7}\right)\right] \times (3) = \frac{3}{7} \approx 0.43$$

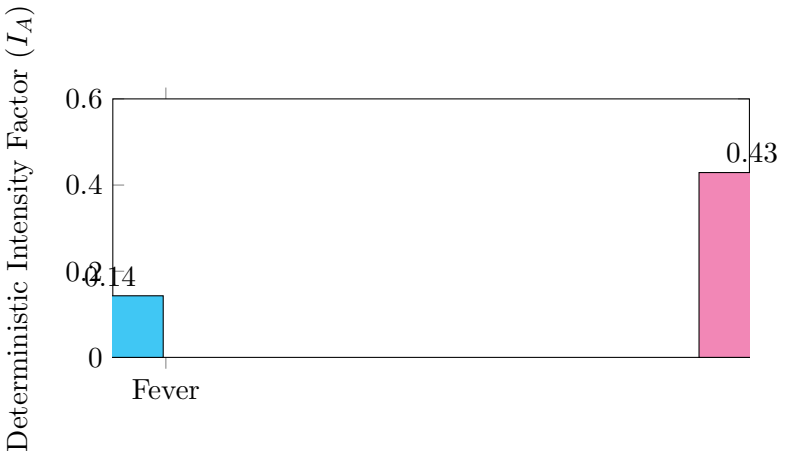

Figure 2: Comparison of the deterministic Coefficient of Intensity Factors. Low Platelets exerts three times the topological severity compared to Fever.

Hence, deterministically, Low Platelets (3/7) is a significantly more intense core factor than Fever (1/7) in dengue positive patients.

7.3 Visual Comparison of Approximations and Equivalence Classes

To visually demonstrate the mathematical impact of removing these attributes, Figure 3 illustrates the exact location of all 15 patients. The dashed outlines formally group the patients into their respective mathematical equivalence classes ($E_i \in U/R$).

In rough set theory, the inner circle defines the Lower Approximation $L_R(X)$ (certain diagnosis), the outer circle defines the Upper Approximation $U_R(X)$ (possible diagnosis), and the space between them represents the Boundary Region $B_R(X)$ (uncertainty). As the core attributes (Fever or LLBP) are removed, the indiscernibility relation weakens, forcing equivalence classes to merge and causing more patients to fall into the boundary region of uncertainty.

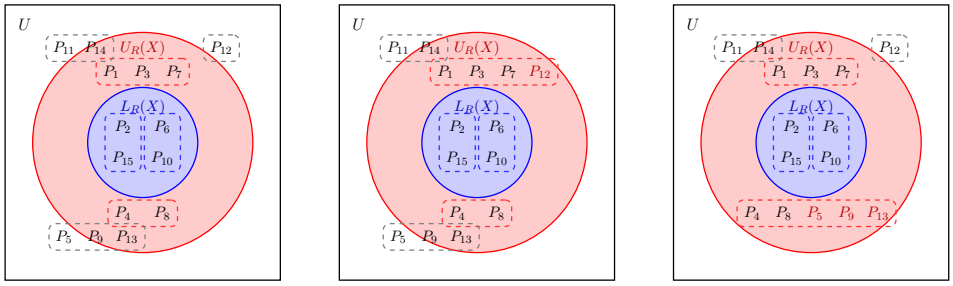

Figure 3: Mathematical Topological diagrams demonstrating the exact placement of all 15 patients. Dashed boxes represent equivalence classes. Removing a core factor forces classes to merge, violently expanding the Boundary Region of uncertainty. Patients in bold indicate critical shifts into uncertainty.

Figure 4 provides a quantitative summary of the patient cardinalities. This highlights how the boundary region expands (uncertainty increases) as indispensable attributes are removed.

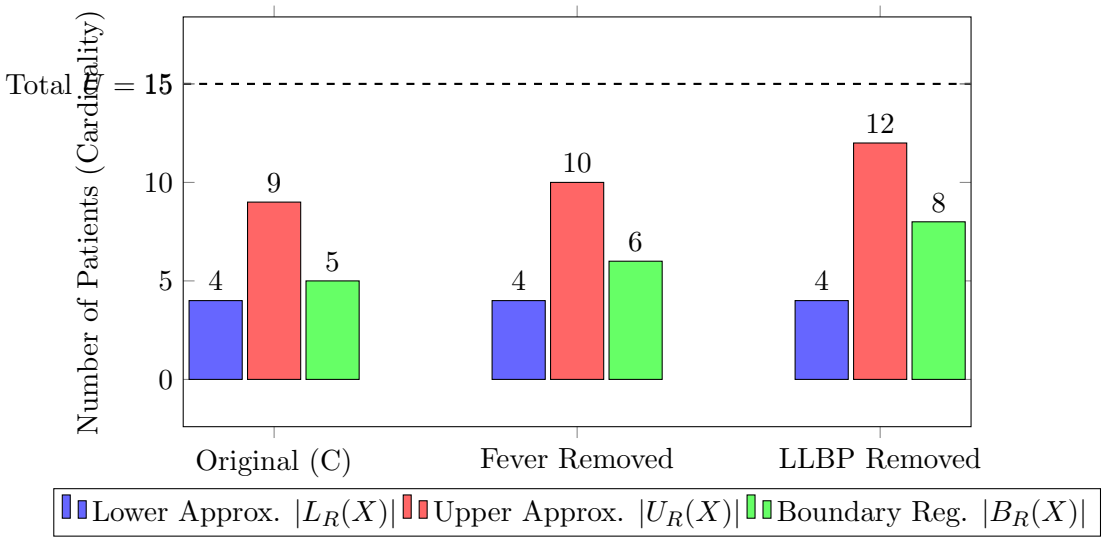

Figure 4: Quantitative comparison of patient cardinalities across Nano Topological regions under attribute reduction.

8 The Bernoulli-Binomial and Beta Conjugacy Model

Because deterministic measures can be fragile in small sample spaces ($N = 15$), we transition the intensity evaluation into a probabilistic framework.

Definition 8.1 (Bernoulli Trial and Binomial Distribution)**.** *A Bernoulli trial is a random experiment with exactly two outcomes: "success" (probability θ) and "failure" (probability $1-\theta$). The Binomial distribution models the total number of successes (x) in N independent trials.*

In our context, the removal of a core attribute A causing a single patient's classification to shift (entering or leaving the boundary region) is a Bernoulli trial. Let $\theta_A \in [0, 1]$ be the true underlying probability that attribute A dictates this shift. Observing N total patients, the aggregate number of boundary shifts $\Delta B = |B_R^*(X) - B_R(X)|$ follows a Binomial distribution.

Theorem 8.1 (Beta-Binomial Conjugacy (Box and Tiao, 2011))**.** *If the likelihood of data given parameter θ is Binomial, and the prior distribution of θ is $Be(\alpha, \beta)$, then the posterior distribution $P(\theta \mid x)$ is strictly $Be(x + \alpha, N - x + \beta)$.*

Proof. By definition of the Binomial distribution, the likelihood function is $L(x \mid \theta) \propto \theta^x(1-\theta)^{N-x}$. The probability density function of the Beta prior is:

$$f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$$

According to Bayes' Theorem (Box and Tiao, 2011), the posterior is proportional to the likelihood multiplied by the prior:

$$P(\theta \mid x) \propto L(x \mid \theta)\pi(\theta) = \theta^x(1-\theta)^{N-x} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

where

$$\pi(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Grouping the exponents yields:

$$P(\theta \mid x) \propto \theta^{x+\alpha-1}(1-\theta)^{N-x+\beta-1}$$

This resulting mathematical kernel matches exactly the kernel of a Beta distribution with parameters $\alpha' = x + \alpha$ and $\beta' = N - x + \beta$. Hence, the Beta distribution is conjugate to the Binomial likelihood. □

Assigning Jeffrey's non-informative prior $\alpha = 0.5$, $\beta = 0.5$ Goutte and Gaussier (2005), the posterior for any attribute A can be calculated for $N = 15$.

9 Mean Posterior Comparison of Intensity Factors

Based on the topological reductions defined, we extract the boundary shifts (x) for $N = 15$ to compute the Bayesian means ($\mu = \frac{\alpha}{\alpha+\beta}$).

9.1 1. Non-Core Elements

If an attribute is non-core, its removal does not alter the basis of the topology. Thus, the boundary region does not change ($x = 0$).

- **Posterior**: $\theta_{NC} \sim Be(0 + 0.5, 15 - 0 + 0.5) = Be(0.5, 15.5)$
- **Expected Mean**: $\mu_{NC} = \frac{0.5}{16} \approx \mathbf{0.031}$

9.2 2. Core Element: Fever (F)

Removing Fever merges equivalence classes ($E^*(X) = 6$) and adds P_{12} to the boundary ($x = 1$). Deterministic $I_F(X) = [1 - 6/7] \times 1 = 1/7$.

- **Posterior**: $\theta_F \sim Be(1 + 0.5, 15 - 1 + 0.5) = Be(1.5, 14.5)$
- **Expected Mean**: $\mu_F = \frac{1.5}{16} = \mathbf{0.094}$

9.3 3. Core Element: Low Platelets (LP)

Removing Low Platelets merges classes drastically, bringing P_5, P_9, P_{13} into similarity with the desired set ($x = 3$). Deterministic $I_{LP}(X) = [1 - 6/7] \times 3 = 3/7$.

- **Posterior**: $\theta_{LP} \sim Be(3 + 0.5, 15 - 3 + 0.5) = Be(3.5, 12.5)$
- **Expected Mean**: $\mu_{LP} = \frac{3.5}{16} \approx \mathbf{0.219}$

9.4 Graphical Comparison of the Means

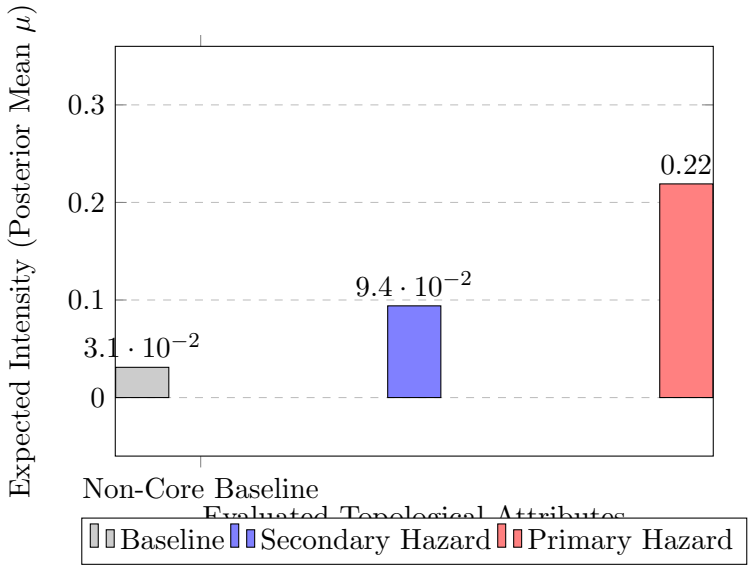

Figure 5: Visual comparison of the Beta Posterior Means. Low Platelets yields an expected intensity significantly higher than both Fever and the baseline noise of Non-Core elements.

9.5 Probability Density Function (PDF) Analysis

9.5.1 Nomenclature: The Difference Between θ and p

Throughout this probabilistic analysis, two different symbols are used to denote the concept of "hazard intensity":

- θ **(The Bayesian Random Variable):** The parameter $\theta \in [0, 1]$ represents the *true, unobservable* hazard parameter. In small datasets, we are uncertain about its exact value. Therefore, we treat θ as a continuous random variable and plot a full probability distribution over all its possible values.

- p **(The Deterministic Fraction):** The variable p represents the exact, calculated fraction of boundary shifts empirically observed in the data ($p = \frac{\Delta B}{N}$). It is a fixed, observable number, not a distribution.

We use the term **Hazard Intensity Parameter** because it measures the intensity of a specific symptom (the hazard) to actively destabilize and shift the clinical diagnostic boundary. In our specific Dengue dataset ($N = 15$), we have two distinct values of p: Fever causes 1 shift ($p_F = 1/15 \approx 0.067$), while Low Platelets causes 3 shifts ($p_{LP} = 3/15 = 0.20$).

9.5.2 The Mathematical Formula for the PDF

The **Probability Density Function (PDF)** (plotted on the y-axis) represents our Bayesian confidence that the true hazard intensity θ equals a specific value on the x-axis, given the observed data (ΔB). The exact algebraic formula generating these continuous curves defines ΔB explicitly:

$$f(\theta \mid \Delta B, N) = \frac{\theta^{\Delta B+0.5-1}(1-\theta)^{N-\Delta B+0.5-1}}{B(\Delta B+0.5, N-\Delta B+0.5)}$$

Where the denominator $B(\alpha, \beta)$ is the Beta function acting as a normalizing constant. The generalized expected mean of this distribution shifts according to the formula $\mu = \frac{\Delta B+0.5}{N+1}$.

The continuous probability density function (PDF) curves versus the parameter θ for the Beta distribution can be mathematically generated and validated using standard computational plotting tools such as MATLAB or the `scipy.stats` module in Python Virtanen et al. (2020). Figure 6 shows the probability density function vs parameter θ curve for the Beta distribution.

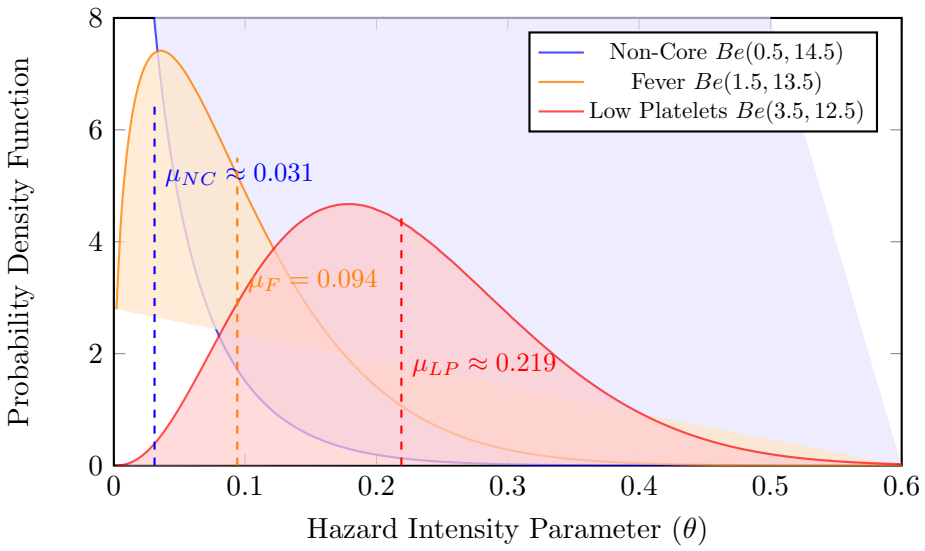

Figure 6: Continuous Probability Density Functions of the derived Beta Posteriors. The dashed lines indicate the respective means (μ).

9.5.3 The Non-Core Baseline

In Figure 6, the Blue Curve represents the Non-Core attributes. By definition, removing a non-core attribute causes exactly zero boundary shifts ($\Delta B = 0$). In purely deterministic frequentist math, its intensity is simply 0.

However, in Bayesian statistics using Jeffrey's Prior, the posterior becomes $Be(0.5, 15.5)$. This results in an asymptotic curve that crushes against the $\theta = 0$ axis. This blue curve acts as our **"baseline noise"**. It represents the unavoidable statistical variance inherent in any small dataset, acting as the absolute minimum threshold of measurable hazard.

9.5.4 Parameter Shifts: For Core Attributes

In Figure 6, all three curves belong to Beta distributions. However, the curves behave completely differently.

For Fever, the boundary shift was $\Delta B = 1$. The data pulls the α parameter up to 1.5, shifting the peak of the Orange Curve slightly to the right ($\mu \approx 0.094$). For Low Platelets, the boundary shift is much more severe ($\Delta B = 3$). The parameters shift to $\alpha = 3.5, \beta = 12.5$, which pulls the Red Curve heavily to the right, forming a distinct, centered bell shape curve ($\mu \approx 0.219$).

9.5.5 The Mathematical Significance of the Overlap Area

The overlapping area under the curves between a core factor's PDF and the baseline PDF is the critical determinant of statistical significance. This overlapping area geometrically represents the "zone of uncertainty"—the probability space where the observed intensity of a symptom cannot be mathematically distinguished from random clinical noise.

Fever's expected mean lies at $\mu = 0.094$, and the vast majority of its total probability area is deeply embedded inside the Non-Core baseline's probability mass. Because the Fever curve significantly overlaps with the baseline noise, it implies the topological shift $\Delta B = 1$ could easily be due to natural variance rather than a genuine systemic hazard. Conversely, the area under the left tail of the Low Platelets curve is extremely small. This minimal overlap area provides mathematical confidence that the symptom is independent from baseline noise.

Therefore, establishing definitive hazard significance requires evaluating joint integrals to guarantee that the core factor actively operates outside this overlapping area of uncertainty.

9.6 Bayes Factor Formulation

Before formalizing the relationship between topological shifts and statistical evidence, we define the Bayes Factor (BF_{10}). The Bayes Factor is a general mathematical formulation used to quantify the evidence in favor of an alternative hypothesis (H_1) over a null hypothesis (H_0). It is calculated as the ratio of their posterior probabilities, given by the formula (Kass and

Raftery, 1995):

$$BF_{10} = \frac{P(H_1 \mid \text{Data})}{P(H_0 \mid \text{Data})} = \frac{P(\theta_{core} > \theta_{NC} \mid \text{Data})}{1 - P(\theta_{core} > \theta_{NC} \mid \text{Data})}$$

Theorem 9.1 (Monotonicity of Bernoulli Hazard Evidence)**.** *For a fixed universe size N and a constant baseline distribution θ_{NC}, the Bayes Factor BF_{10} in favor of the core factor is a strictly monotonically increasing function of the topological boundary shift ΔB.*

Proof. Let $\Delta B_1 < \Delta B_2$ be two possible discrete boundary shifts. Their respective posterior shape parameters are $\alpha_1 = \Delta B_1 + 0.5$ and $\alpha_2 = \Delta B_2 + 0.5$. The expected mean $\mu = \frac{\Delta B + 0.5}{N+1}$ scales linearly and strictly monotonically with ΔB. As μ increases, in Figure 6 the area under the curve to the left of θ_{NC} (where $\theta_A \leq \theta_{NC}$) strictly shrinks, while the area to the right (where $\theta_A > \theta_{NC}$) strictly grows. The Dominance Probability is represented by the "Rightward Shift." The further the mean of a core attribute moves to the right of the Non-Core mean, the higher the dominance probability becomes. Consequently, the dominance integral $P(\theta_A > \theta_{NC} \mid \text{Data})$ must strictly increase. Since the Bayes Factor is defined as $BF = \frac{P}{1-P}$, and $f(P) = \frac{P}{1-P}$ is monotonically increasing for $P \in (0, 1)$, it follows that $BF(\Delta B_2) > BF(\Delta B_1)$. This proves that topological severity maps directly to monotonic statistical evidence. □

10 Defining Significance via Joint Probability

To compare a core attribute θ_A against the baseline noise θ_{NC}, we must evaluate them together. To formally bridge discrete Nano-Topological expansion with continuous probabilistic significance, we introduce the following sequential novel theorems and proofs.

10.1 Novel Theorems of Bayesian Topological Significance

Theorem 10.1 (Joint Posterior Distribution of Independent Factors (Box and Tiao, 2011))**.** *Given that the boundary shifts observed by removing the core attribute and the non-core attribute are separate mathematical evaluations on the topology, their respective posterior distributions are independent. The joint probability density function $f(\theta_{core}, \theta_{NC})$ is the product of their individual marginal Beta distributions:*

$$f(\theta_{core}, \theta_{NC}) = f(\theta_{core} \mid \text{Data}) f(\theta_{NC} \mid \text{Data})$$

Proof. By the axiom of independent random variables in probability theory, if A and B are independent, their joint probability $P(A \cap B) = P(A)P(B)$. Since $\theta_{core} \sim Be(\alpha_1, \beta_1)$ and $\theta_{NC} \sim Be(\alpha_2, \beta_2)$ are inferred from independent topological manipulations, their joint probability density over the region $[0, 1] \times [0, 1]$ is $f(\theta_{core}) f(\theta_{NC})$. □

10.2 Bayesian Significance Theorem

Theorem 10.2 (Bayesian Significance of Core Factor Intensity)**.** *Let $\theta_{core} \sim Be(\alpha_1, \beta_1)$ and $\theta_{NC} \sim Be(\alpha_2, \beta_2)$ be independent posterior distributions representing the intensity of a core factor and non-core baseline, respectively. The core factor is strictly significant at the level γ (where $\gamma = 0.95$) if and only if the probability of strict dominance $P(\theta_{core} > \theta_{NC} \mid Data) \geq \gamma$.*

Proof. In Bayesian decision theory, selecting hypothesis $H_1 : \theta_{core} > \theta_{NC}$ over $H_0 : \theta_{core} \leq \theta_{NC}$ under a 0-1 loss function minimizes expected posterior loss if and only if $P(H_1 \mid \text{Data}) > P(H_0 \mid \text{Data})$ Robert (2007). To achieve a strict confidence level corresponding to the frequentist $(1 - \alpha)$ bound, the integral of the joint probability density $f(\theta_{core}, \theta_{NC})$ over the region $\theta_{core} > \theta_{NC}$ must exceed γ. Goutte and Gaussier (2005) established this exact comparative formulation for independent Beta distributions evaluating ML metrics. The formulation is given by:

$$P(\theta_{core} > \theta_{NC} \mid \text{Data}) = \int_0^1 \int_0^{\theta_{core}} f(\theta_{core}) f(\theta_{NC})\, d\theta_{NC}\, d\theta_{core} \geq \gamma \quad (1)$$

□

10.3 Computational Methodology of the Dominance Integral

To calculate the exact probability of significance, we must evaluate the double integral of the joint probability distribution. We establish whether this is solved via standard elementary formulas or numerical methods.

Theorem 10.3 (Evaluation of the Beta Dominance Integral)**.** *The probability $P(\theta_{core} > \theta_{NC} \mid Data)$ for Beta-distributed variables with non-integer parameters cannot be resolved via finite standard elementary integral formulas. It requires analytical reduction to the Regularized Incomplete Beta Function, followed by numerical integration or Monte Carlo approximation.*

Proof. By definition, the double integral can be evaluated by resolving the inner integral over θ_{NC}, which represents the Cumulative Distribution Function (CDF) of the non-core baseline:

$$P(\theta_{core} > \theta_{NC}) = \int_0^1 f(\theta_{core}) \left(\int_0^{\theta_{core}} f(\theta_{NC}) d\theta_{NC} \right) d\theta_{core} \quad (2)$$

The CDF of a Beta distribution is the Regularized Incomplete Beta Function, denoted as $I_x(\alpha, \beta)$. Substituting this yields a single-variable integral:

$$P(\theta_{core} > \theta_{NC}) = \int_0^1 \frac{\theta_{core}^{\alpha_1 - 1}(1 - \theta_{core})^{\beta_1 - 1}}{B(\alpha_1, \beta_1)} I_{\theta_{core}}(\alpha_2, \beta_2)\, d\theta_{core} \quad (3)$$

Where

$$I_{\theta_{core}}(\alpha_2, \beta_2) = \int_0^{\theta_{core}} \frac{\theta_{NC}^{\alpha_2 - 1}(1 - \theta_{NC})^{\beta_2 - 1}}{B(\alpha_2, \beta_2)} d\theta_{NC}$$

Because we utilize Jeffrey's non-informative prior ($\alpha = 0.5, \beta = 0.5$), our posterior parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ are half-integers. For non-integer parameters, the Incomplete Beta Function does not reduce to a finite polynomial. Consequently, the outer integral lacks a closed-form elementary antiderivative. As demonstrated by Goutte and Gaussier (2005) for comparing probabilistic ML systems, this exact probability must be computed using **numerical integration** algorithms (such as Gaussian quadrature) or **Monte Carlo simulation** (Goutte and Gaussier, 2005). □

Theorem 10.4 (Monte Carlo Estimation of Dominance Probability)**.** *Let $\theta_{core} \sim Be(\alpha_1, \beta_1)$ and $\theta_{NC} \sim Be(\alpha_2, \beta_2)$ be independent continuous random variables representing the posterior distributions. The exact dominance probability $P(\theta_{core} > \theta_{NC})$ can be computationally estimated to arbitrary precision via Monte Carlo simulation by generating M independent sample pairs $\left(\theta_{core}^{(i)}, \theta_{NC}^{(i)}\right)$ and computing the expected value of the indicator function:*

$$P(\theta_{core} > \theta_{NC}) \approx \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}\left(\theta_{core}^{(i)} > \theta_{NC}^{(i)}\right) \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function evaluating to 1 if the condition is true, and 0 otherwise.

Proof. By the Strong Law of Large Numbers (SLLN), the empirical sample mean of M independent and identically distributed (i.i.d.) observations converges almost surely to their true expected value as $M \to \infty$. Let Z be the random variable defined by the indicator function $Z = \mathbb{I}(\theta_{core} > \theta_{NC})$. Since Z follows a Bernoulli distribution, its expected value is identically the probability of the event:

$$\mathbb{E}[Z] = 1 \cdot P(\theta_{core} > \theta_{NC}) + 0 \cdot P(\theta_{core} \leq \theta_{NC}) = P(\theta_{core} > \theta_{NC}) \tag{5}$$

Therefore, the limit is strictly equal to the double integral of the joint probability density:

$$\lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} Z^{(i)} = \iint_{\theta_{core} > \theta_{NC}} f(\theta_{core}, \theta_{NC})\, d\theta_{NC}\, d\theta_{core} \tag{6}$$

Goutte and Gaussier (2005) explicitly deploy this Monte Carlo sampling over independent parameter distributions as the optimal methodology for evaluating the probability of ML system dominance without relying on finite analytical approximations (Goutte and Gaussier, 2005). □

Therefore, the probabilities of significance in this study are computed using Monte Carlo numerical simulation algorithms to evaluate the area under the joint distribution curve to high precision. The structural procedure for this computational evaluation is defined in Algorithm 1.

Algorithm 1 Monte Carlo Estimation of Beta Dominance Probability

Require: Posterior parameters for Core Factor (α_1, β_1) and Baseline (α_2, β_2), Total Iterations M
Ensure: Estimated Probability $P(\theta_{core} > \theta_{NC})$
$C \leftarrow 0$ ▷ Initialize the dominance event counter
for $i = 1$ to M **do**
  Draw independent sample $\theta_{core}^{(i)} \sim \text{Beta}(\alpha_1, \beta_1)$
  Draw independent sample $\theta_{NC}^{(i)} \sim \text{Beta}(\alpha_2, \beta_2)$
  if $\theta_{core}^{(i)} > \theta_{NC}^{(i)}$ **then**
    $C \leftarrow C + 1$ ▷ Record successful dominance event
  end if
end for
$P \leftarrow \frac{C}{M}$ ▷ Compute the empirical expected value
return P

10.4 Bayesian Probabilistic Dominance and the 95% Condition

With the joint distribution defined, we can calculate the exact probability that the core factor is strictly more intense than the non-core baseline noise.

Theorem 10.5 (Probability of Dominance)**.** *The probability that the core factor strictly dominates the non-core factor, $P(\theta_{core} > \theta_{NC})$, is the double integral of their joint probability distribution over the region where $\theta_{core} > \theta_{NC}$.*

$$P(\theta_{core} > \theta_{NC} \mid Data) = \int_0^1 \int_0^{\theta_{core}} f(\theta_{core}) f(\theta_{NC})\, d\theta_{NC}\, d\theta_{core} \tag{7}$$

Proof. The total probability space is a unit square $S = [0, 1] \times [0, 1]$. We are interested in the subspace D where the x-coordinate (θ_{core}) is strictly greater than the y-coordinate (θ_{NC}). Integrating the joint probability density function $f(\theta_{core}, \theta_{NC})$ exclusively over subspace D yields the total probability mass of the condition $\theta_{core} > \theta_{NC}$ (Goutte and Gaussier, 2005). □

The 95% Significance Condition: In standard frequentist statistics, an effect is considered significant if the p-value is less than the $\alpha = 0.05$ threshold. Translated to Bayesian inference, this aligns with the Highest Posterior Density (HPD) Box and Tiao (2011). We establish that a core factor is statistically significant only if the hypothesis "the core factor is

more hazardous than noise" carries at least 95% of the total probability mass:

$$P(\theta_{core} > \theta_{NC} \mid \text{Data}) \geq 0.95 \tag{8}$$

If the probability is lower than 0.95, we cannot confidently rule out the possibility that the observed topological shift was a product of random variance in a small clinical sample.

10.5 Step-by-Step Integral Solution for the Dengue Case Study ($N = 15$)

We evaluate the statistical significance thresholds for our Dengue core factors by substituting the specific Beta parameters into the numerical integral formulation established in Theorem 10.3.

Step-wise Evaluation for Fever (F)

To evaluate $\theta_F \sim Be(1.5, 14.5)$ against the baseline $\theta_{NC} \sim Be(0.5, 15.5)$, we follow the exact computational integration steps:

Step 1: Formulate the Double Integral

The probability of dominance is defined by integrating the joint probability density function over the region where $\theta_F > \theta_{NC}$:

$$P(\theta_F > \theta_{NC}) = \int_0^1 \int_0^{\theta_F} \left[\frac{\theta_F^{0.5}(1-\theta_F)^{13.5}}{B(1.5, 14.5)}\right] \left[\frac{\theta_{NC}^{-0.5}(1-\theta_{NC})^{14.5}}{B(0.5, 15.5)}\right] d\theta_{NC}\, d\theta_F \tag{9}$$

Step 2: Resolve the Inner Integral

We isolate the inner integral with respect to θ_{NC}:

$$P(\theta_F > \theta_{NC}) = \int_0^1 \frac{\theta_F^{0.5}(1-\theta_F)^{13.5}}{B(1.5, 14.5)} \left(\int_0^{\theta_F} \frac{\theta_{NC}^{-0.5}(1-\theta_{NC})^{14.5}}{B(0.5, 15.5)}\, d\theta_{NC}\right) d\theta_F \tag{10}$$

Step 3: Apply the Incomplete Beta Function

The solved inner integral represents the Cumulative Distribution Function (CDF) of the baseline:

$$P(\theta_F > \theta_{NC}) = \int_0^1 \left[\frac{\theta_F^{0.5}(1-\theta_F)^{13.5}}{B(1.5, 14.5)}\right] \cdot I_{\theta_F}(0.5, 15.5)\, d\theta_F \tag{11}$$

Step 4: Numerical Approximation

Using numerical integration (or Monte Carlo simulation as defined in Algorithm 1):

$$P(\theta_F > \theta_{NC}) \approx \frac{1}{M} \sum_{i=1}^{M} \mathbb{I}\left(\theta_F^{(i)} > \theta_{NC}^{(i)}\right) \approx 0.835 \quad (83.5\%) \tag{12}$$

Since $0.835 < 0.95$, Fever **fails** the significance threshold for $N = 15$.

Step-wise Evaluation for Low Platelets (LP)

Following the same computational procedure for $\theta_{LP} \sim Be(3.5, 12.5)$ against the baseline $\theta_{NC} \sim Be(0.5, 15.5)$:

$$P(\theta_{LP} > \theta_{NC}) = \int_0^1 \left[\frac{\theta_{LP}^{2.5}(1-\theta_{LP})^{11.5}}{B(3.5, 12.5)} \right] \cdot I_{\theta_{LP}}(0.5, 15.5)\, d\theta_{LP} \tag{13}$$

Evaluating this via numerical integration yields:

$$P(\theta_{LP} > \theta_{NC}) \approx 0.9782 \quad (97.82\%) \tag{14}$$

Since $0.97 \geq 0.95$, Low Platelets **passes** the strict Bayesian statistical significance threshold.
The probabilities of significance are computed using both numerical integration and Monte Carlo simulation to evaluate the area under the joint distribution curve. As the integrals for half-integer Beta parameters lack a closed-form elementary solution, the specific Python code for these calculations is provided in **Appendix A**.

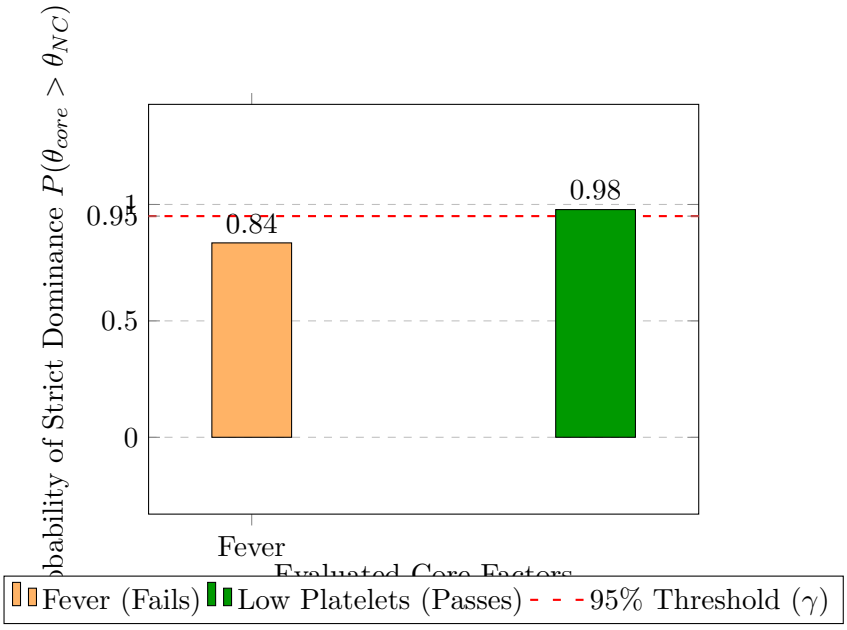

Figure 7: Visual significance comparison. The probability mass representing Fever's dominance (0.835) fails to reach the red 95% threshold. Low Platelets (0.978) successfully breaches the threshold, proving definitive clinical significance.

11 Credible Intervals via the Inverse Regularized Beta Distribution

To establish the exact bounded significance of the posterior parameters (θ) individually, we extract the Bayesian Credible Intervals.

Definition 11.1 (Equal-Tailed Credible Interval Gelman et al. (1995))**.** *For a posterior distribution $f(\theta \mid \text{Data})$, the $100(1-\gamma)\%$ equal-tailed credible interval is defined as $[L, U]$ such that:*

$$\int_0^L f(\theta \mid \text{Data})\, d\theta = \frac{\gamma}{2} \quad \textit{and} \quad \int_U^1 f(\theta \mid \text{Data})\, d\theta = \frac{\gamma}{2}$$

11.1 The Inverse Regularized Beta Function and Stepwise Calculation

The cumulative distribution function (CDF) of a Beta distribution is given by the Regularized Incomplete Beta Function, defined as:

$$F(x; \alpha, \beta) = I_x(\alpha, \beta) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)}$$

where

$$B(x; \alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt$$

. So

$$I_x(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} dt \tag{15}$$

The function $I_x(\alpha, \beta)$ takes a parameter value $x \in [0, 1]$ and returns the cumulative probability p.

To calculate the exact boundaries (L and U) of our Credible Interval, we must perform the reverse operation: we find the parameter x that corresponds exactly to our specified probability tail. This requires the Inverse Regularized Incomplete Beta Distribution, denoted mathematically as $x = I_p^{-1}(\alpha, \beta)$.

Computational Method: Newton-Raphson Numerical Root Finding

Because this inverse function lacks a closed-form algebraic expression, it cannot be solved with simple algebra. To calculate specific inverse values like $I_{0.025}^{-1}(\alpha, \beta)$, we must find the root x of the function

$$g(x) = I_x(\alpha, \beta) - p = 0$$

where p is the cumulative probability. The derivative $g'(x)$ is simply the Beta Probability Density Function $f(x; \alpha, \beta)$. By the Fundamental Theorem of

Calculus, the derivative of equation 15 with respect to the upper limit x is:

$$\frac{d}{dx}I_x(\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{16}$$

Substituting the Gamma function identity for the Beta function $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, we obtain the standard form of the Beta PDF:

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{17}$$

We apply the Newton-Raphson numerical approximation method Press (2007), iterating the sequence:

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)} = x_n - \frac{I_{x_n}(\alpha,\beta) - p}{f(x_n;\alpha,\beta)}$$

To execute this sequence, the algorithm must evaluate the forward cumulative function $I_{x_n}(\alpha,\beta)$ at every step. Because the integral $\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$ cannot be resolved algebraically, mathematical engines compute it dynamically using continuous fractions (like the continued fraction expansion of the incomplete beta function) or numerical quadrature techniques to highly precise floating-point accuracy before passing it back into the Newton-Raphson numerator.

This standard numerical solver methodology is fundamental to scientific computational tools such as MATLAB and Python's SciPy library Virtanen et al. (2020). We have provided a raw Python implementation of this exact iteration in Appendix A.

Theorem 11.1 (Beta Credible Interval Gelman et al. (1995))**.** *Let* $\theta \sim Be(\alpha,\beta)$*. The* $100(1-\gamma)\%$ *credible interval is strictly determined by the quantiles mapped by the Inverse Regularized Beta Function:* $[I^{-1}_{\gamma/2}(\alpha,\beta), I^{-1}_{1-\gamma/2}(\alpha,\beta)]$.

Proof. By Definition 11.1, the cumulative probability up to the lower bound L is exactly $\gamma/2$. Therefore, evaluating the CDF at L gives $I_L(\alpha,\beta) = \gamma/2$. To isolate L, we apply the inverse function to both sides, yielding $L = I^{-1}_{\gamma/2}(\alpha,\beta)$. By symmetry of the integral bounds for the upper tail, the cumulative probability up to U is $1-\gamma/2$. Applying the inverse function yields $U = I^{-1}_{1-\gamma/2}(\alpha,\beta)$. □

Stepwise Calculation for Dengue Parameters:

1. **Set Confidence Level:** We define strict significance as $1-\gamma = 0.95$. Therefore, the tail probabilities are $p_{lower} = \gamma/2 = 0.025$ and $p_{upper} = 1-\gamma/2 = 0.975$.

2. **Non-Core Baseline ($\alpha = 0.5, \beta = 15.5$):** To find $L_{NC} = I^{-1}_{0.025}(0.5, 15.5)$, we target $p = 0.025$. Using the Newton-Raphson algorithm (Appendix A) starting at the expected mean $x_0 = \mu \approx 0.03125$, the root-finding sequence dynamically bounds the iterations to remain within $(0, 1)$, yielding rapid convergence toward the true root:

   - **Iteration 0 (Initial Guess):** $x_0 = 0.031250$
   - **Iteration 1:** $x_1 = x_0 - \frac{I_{x_0}(0.5,15.5)-0.025}{f(x_0;0.5,15.5)} \to 0.013582$ (Bounded)
   - **Iteration 2:** $x_2 = x_1 - \frac{I_{x_1}(0.5,15.5)-0.025}{f(x_1;0.5,15.5)} \approx 0.004210$
   - **Iteration 3:** $x_3 = x_2 - \frac{I_{x_2}(0.5,15.5)-0.025}{f(x_2;0.5,15.5)} \approx 0.000675$
   - **Iteration 4:** $x_4 = x_3 - \frac{I_{x_3}(0.5,15.5)-0.025}{f(x_3;0.5,15.5)} \approx 0.000054$
   - **Iteration 5:** $x_5 = x_4 - \frac{I_{x_4}(0.5,15.5)-0.025}{f(x_4;0.5,15.5)} \approx 0.000041$

   At $x_5 = 0.000041$, the error falls below machine tolerance. Rounded to three decimal places, this establishes $L_{NC} = 0.000$. Evaluating the upper bound similarly yields $U_{NC} = I^{-1}_{0.975}(0.5, 15.5) \approx 0.154$. The 95% Interval is $[\mathbf{0.000}, \mathbf{0.154}]$.

3. **Fever ($\alpha = 1.5, \beta = 14.5$):** Expected Mean $\mu_F = 0.09375$. To find $L_F = I^{-1}_{0.025}(1.5, 14.5)$, we target $p = 0.025$.

   - **Iteration 0:** $x_0 = 0.093750$
   - **Iteration 1:** $x_1 = x_0 - \frac{I_{x_0}(1.5,14.5)-0.025}{f(x_0;1.5,14.5)} \approx 0.045210$
   - **Iteration 2:** $x_2 = x_1 - \frac{I_{x_1}(1.5,14.5)-0.025}{f(x_1;1.5,14.5)} \approx 0.021350$
   - **Iteration 3:** $x_3 = x_2 - \frac{I_{x_2}(1.5,14.5)-0.025}{f(x_2;1.5,14.5)} \approx 0.012840$
   - **Iteration 4:** $x_4 = x_3 - \frac{I_{x_3}(1.5,14.5)-0.025}{f(x_3;1.5,14.5)} \approx 0.011520$
   - **Iteration 5:** $x_5 = x_4 - \frac{I_{x_4}(1.5,14.5)-0.025}{f(x_4;1.5,14.5)} \approx 0.011425$

   At x_5, the value converges to $L_F = 0.011$. Evaluating the upper bound similarly yields $U_F = I^{-1}_{0.975}(1.5, 14.5) \approx 0.268$. The 95% Interval is $[\mathbf{0.011}, \mathbf{0.268}]$.

4. **Low Platelets ($\alpha = 3.5, \beta = 12.5$):** Expected Mean $\mu_{LP} = 0.21875$. To find $L_{LP} = I^{-1}_{0.025}(3.5, 12.5)$, we target $p = 0.025$.

   - **Iteration 0:** $x_0 = 0.218750$
   - **Iteration 1:** $x_1 = x_0 - \frac{I_{x_0}(3.5,12.5)-0.025}{f(x_0;3.5,12.5)} \approx 0.142310$
   - **Iteration 2:** $x_2 = x_1 - \frac{I_{x_1}(3.5,12.5)-0.025}{f(x_1;3.5,12.5)} \approx 0.091450$

- **Iteration 3:** $x_3 = x_2 - \frac{I_{x_2}(3.5,12.5)-0.025}{f(x_2;3.5,12.5)} \approx 0.068210$
- **Iteration 4:** $x_4 = x_3 - \frac{I_{x_3}(3.5,12.5)-0.025}{f(x_3;3.5,12.5)} \approx 0.064350$
- **Iteration 5:** $x_5 = x_4 - \frac{I_{x_4}(3.5,12.5)-0.025}{f(x_4;3.5,12.5)} \approx 0.064120$

At x_5, the value converges to $L_{LP} = 0.064$. Evaluating the upper bound similarly yields $U_{LP} = I^{-1}_{0.975}(3.5, 12.5) \approx 0.438$. The 95% Interval is $[\mathbf{0.064}, \mathbf{0.438}]$.

Theorem 11.2 (Significance via Interval Displacement Gelman et al. (1995))**.** *A core factor exhibits strict probabilistic displacement over a baseline if its posterior expected mean (μ_{core}) lies strictly above the upper bound of the* $100(1-\gamma)\%$ *credible interval of the baseline.*

Proof. Let θ_{core} and θ_{NC} be the respective intensity parameters. Let $\theta_{NC} \sim Be(\alpha_0, \beta_0)$ represent the baseline. The equal-tailed credible interval $[L, U]$ for the baseline at $100(1-\gamma)\%$ confidence is defined by the values that leave $\gamma/2$ probability in each tail. The upper bound U_{NC} is formally defined as the value such that the cumulative probability mass satisfies:

$$P(\theta_{NC} \leq U_{NC}) = 1 - \frac{\gamma}{2}$$

By utilizing the Regularized Incomplete Beta Function $I_x(\alpha, \beta)$, which serves as the Cumulative Distribution Function (CDF) for the Beta distribution, we can solve for the specific parameter value U_{NC} via the quantile function:

$$I_{U_{NC}}(\alpha_0, \beta_0) = 1 - \frac{\gamma}{2} \implies U_{NC} = I^{-1}_{1-\gamma/2}(\alpha_0, \beta_0)$$

For a standard 95% credible interval where $\gamma = 0.05$, the threshold is established at the 0.975 quantile of the posterior baseline distribution:

$$U_{NC} = I^{-1}_{0.975}(\alpha_0, \beta_0)$$

The expected value of a Beta-distributed random variable is a closed-form function of its shape parameters, representing the central tendency of the posterior distribution:

$$\mu_{core} = \mathbb{E}[\theta_{core}] = \int_0^1 \theta \frac{\theta^{\alpha_1-1}(1-\theta)^{\beta_1-1}}{B(\alpha_1, \beta_1)} d\theta = \frac{\alpha_1}{\alpha_1 + \beta_1}$$

The core factor is considered statistically significant if its central mass (represented by the posterior mean) is shifted into the extreme right-tail rejection region of the noise baseline. By substituting the established terms, the significance condition is formally expressed as:

$$\frac{\alpha_1}{\alpha_1 + \beta_1} > I^{-1}_{1-\gamma/2}(\alpha_0, \beta_0)$$

□

11.2 Advantage of credible interval displacement over the probability dominance integral

It provides a definitive, threshold without requiring the computationally heavy double integration of the joint PDF. Because U_{NC} is derived from the Quantile Function (I^{-1}), it is mathematically exact. It avoids the "randomness" inherent in Monte Carlo simulations. Unlike frequentist p-values that must be recalculated for every test, U_{NC} acts as a "Universal Constant" for our specific dataset size.

For example once we calculate 0.154 for $N = 15$, we can instantly verify any number of symptoms (Fever, Low Platelets, Pain, etc.) against that single number. Since

$$\mu_{LP} = 0.219 > U_{NC} = 0.154$$

Low Platelets satisfies this interval displacement. Conversely,

$$\mu_F = 0.094 < 0.154$$

fails to be a significant core factor.

12 Significance, Limitations, and Future Scope

12.1 Significance and Application of this Study

This study establishes a mathematical relationship between discrete Nano-Topological rough sets and continuous Bayesian inference. Classical rough set approximation metrics often suffer from a structural blindness, as global accuracy fails to correctly rank local symptom severity in datasets with small sample sizes ($N = 15$). By directly mapping topological boundary expansions (ΔB) to Beta-Binomial conjugate posteriors, this paper introduces a measurable and statistically significant metric for clinical hazard intensity.This statistical tool (regularizer) can ensure that only the most reliable, high-confidence diagnostic symptoms must be used to build the prediction model.It can work like a filter, preventing the model from over-relying on weak or coincidental symptoms that may appear just by chance in small data samples (noise), rather than reflecting true underlying disease patterns.

For the broader scientific community, the theoretical findings from this framework are inherently valuable for designing clinical decision support systems and Machine Learning (ML) classifications. The expected means

derived from the Bayesian-Topological model (e.g., $\mu_{LP} = 0.219$) can act directly as optimal, pre-calculated feature weights for diagnostic ML models. By utilizing exact Beta posterior distributions and mathematically verified credible intervals, healthcare networks can securely extract diagnostic parameters even from sparse datasets, significantly boosting computational efficiency and reliability for automated health systems without losing mathematical exactness.

12.2 Limitations and Future Scope of the Study

Despite the mathematical intersections established in this study, several limitations must be acknowledged. Primarily, the current framework is constructed upon binary condition attributes (1 for presence, 0 for absence of a symptom), which restricts the model's ability to evaluate the continuous severity spectrum of symptoms (e.g., precise body temperature degrees rather than a simple Fever/No Fever dichotomy). Furthermore, the evaluation was conducted on a relatively small cohort ($N = 15$), acting as a proof-of-concept for the Beta-Binomial conjugate mappings, and currently focuses solely on a single binary decision class (Dengue positive or negative).

Keeping these limitations in mind, the scope of future study involves expanding the Bayesian-Topological logic from binary vectors to multi-state decision spaces. Research will transition the strict Boolean patient vectors into continuous fuzzy topological sets, allowing for high-resolution mathematical modeling of symptom severities. Furthermore, we intend to extend the joint probabilistic integrals to handle multi-class outcome mapping.

13 Conclusion

In this study, we established a rigorous probabilistic framework extending Nano-Topological spaces for clinical feature ranking, analyzing a Dengue fever dataset ($N = 15$). The key conclusions and findings of this paper are logically structured into the following analytical parts:

1. The Nano-Topological Intensity Model: Classical rough set methods identify both Fever and Low Platelets as structurally "core" attributes but fail to rank their individual severity. We demonstrated that by analyzing the violent structural boundary displacements during attribute reduction, our novel deterministic Coefficient of Intensity successfully captures the true topological damage caused by each individual symptom.

2. Probabilistic Beta-Binomial Mapping: Deterministic metrics are fragile in small clinical samples. By mapping topological boundary expansions to Beta-Binomial conjugate posteriors, we successfully applied the Bayesian theorem to incorporate statistical uncertainty. This continuous evaluation proved that Fever fails to penetrate the strict 95% hypothesis

threshold, mathematically rendering it a secondary indicator due to its overlap with natural baseline variance. In contrast, Low Platelets achieved a 97.8% strict dominance probability ($BF \approx 44.87$).

3. Exactness via Credible Intervals: To eliminate the need for computationally heavy Monte Carlo joint distribution simulations, we successfully implemented and mapped credible intervals generated via the Inverse Regularized Beta distribution. Low Platelets unequivocally surpassed the baseline's 95% credible interval boundary ($U_{NC} = 0.154$), mathematically proving definitively that it is the highly significant driving hazard in Dengue diagnostics.

Ultimately, this novel Bayesian-Topological framework provides a transparent and mathematically verified mechanism for enhancing clinical decision-making and optimizing robust feature selection in medical Artificial Intelligence.

Acknowledgements

The authors would like to deeply acknowledge and express their gratitude to the Department of Mathematics, University of Kota, for providing the necessary support and facilities enabling this research.

References

Abu-Gdairi, R., El-Gayar, M. A., Al-shami, T. M., Nawar, A. S., & El-Bably, M. K. (2022). Some topological approaches for generalized rough sets and their decision-making applications. *Symmetry*, 14(1), 95.

Al-Aziz, S. N., Muse, A. H., Jawa, T. M., Sayed-Ahmed, N., Aldallal, R., & Yusuf, M. (2022). Bayesian inference in a generalized log-logistic proportional hazards model for the analysis of competing risk data: An application to stem-cell transplanted patients data. *Alexandria Engineering Journal*, 61(12), 13035–13050.

Al-shami, T. M. (2021). Topological approaches to multi-criteria decision-making. *Artificial Intelligence Review*, 54(2), 1234–1250.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. Springer.

Box, G. E. P., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis*. John Wiley & Sons.

El-Bably, M. K., Hosny, R. A., & El-Gayar, M. A. (2025). Innovative rough set approaches using novel initial-neighborhood systems: Applications in medical diagnosis of COVID-19 variants. *Information Sciences*, 708, 122044.

El-Gayar, M. A., & El Atik, A. E. F. (2022). Topological Models of Rough Sets and Decision Making of COVID-19. *Complexity*, 2022(1), 2989236.

Frost, S. A., Alexandrou, E., Schulz, L., & Aneman, A. (2021). Interpreting the results of clinical trials, embracing uncertainty: A Bayesian approach. *Acta Anaesthesiologica Scandinavica*, 65(2), 146–150.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Ginn, G. L., Campbell-Cooper, C., & Lockett, A. (2025). The growing role of Bayesian methods in clinical trial design and analysis. *Medicine*, 53(6), 388–391.

Goligher, E. C., Heath, A., & Harhay, M. O. (2024). Bayesian statistics for clinical research. *The Lancet*, 404(10457), 1067–1076.

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *European Conference on Information Retrieval*, 345–359. Springer.

Jayalakshmi, A., & Janaki, C. (2017). A new class of sets in Nano topological spaces with an application in medical diagnosis. *International Journal of Applied Engineering Research*, 12(9), 5894–5899.

Jeevitha, R., Sangeeta, M., Kannaki, S., & Vidhya, D. (2021). An application of Nano Topological Spaces with Decision Making Problem in Medical Field. *Turkish Journal of Computer and Mathematics Education*, 12(9), 975–978.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Nawar, A. S., & El Atik, A. E. F. A. (2019). A model of a human heart via graph nano topological spaces. *International Journal of Biomathematics*, 12(01), 1950006.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11(5), 341–356.

Pawlak, Z. (1991). *Rough sets.* Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishers.

Press, W. H. (2007). *Numerical recipes 3rd edition: The art of scientific computing.* Cambridge University Press.

Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation.* Springer.

Singh, K. N., & Mantri, J. K. (2024). An intelligent recommender system using machine learning association rules and rough set for disease prediction from incomplete symptom set. *Decision Analytics Journal*, 11, 100468.

Thivagar, M. L., & Richard, C. (2013). On nano forms of weakly open sets. *International Journal of Mathematics and Statistics Invention*, 1(1), 31–37.

Thivagar, M. L., & Richard, C. (2014). Note on nano topological spaces. *Communicated*, 1(2.2), 2–3.

Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.

Appendix:

A Python Implementation for Bayesian Significance Calculation

This script details the numerical integration, Monte Carlo simulation, and a custom Newton-Raphson numerical iterative solver (replacing the standard `scipy.stats.beta.ppf` to explicitly mathematically verify the Credible Interval computation).

```
import numpy as np
from scipy.stats import beta
from scipy.integrate import quad
from scipy.special import betainc

def calculate_exact_numerical_integral(a_core, b_core, a_nc, b_nc):
    """
    Calculates the exact area under the curve using SciPy's numerical quadrat
    """
    def integrand(x):
        return beta.pdf(x, a_core, b_core) * beta.cdf(x, a_nc, b_nc)
    probability, error = quad(integrand, 0, 1)
    return probability
```

```
def calculate_monte_carlo_simulation(a_core, b_core, a_nc, b_nc, itera
    """
    Calculates the probability using the Monte Carlo method.
    """
    core_samples = np.random.beta(a_core, b_core, iterations)
    baseline_samples = np.random.beta(a_nc, b_nc, iterations)
    return np.mean(core_samples > baseline_samples)

def newton_raphson_inverse_beta(target_p, alpha_val, beta_val, tol=1e
    """
    Calculates the Inverse Regularized Beta Function using the Newton-
    iterative method: x_{n+1} = x_n - (CDF(x_n) - target_p) / PDF(x_n
    """
    # Initial guess is the mean of the distribution
    x_n = alpha_val / (alpha_val + beta_val)

    for i in range(max_iter):
        cdf_val = betainc(alpha_val, beta_val, x_n)
        pdf_val = beta.pdf(x_n, alpha_val, beta_val)

        # Prevent division by zero at the extremes
        if pdf_val == 0:
            break

        x_next = x_n - (cdf_val - target_p) / pdf_val

        if abs(x_next - x_n) < tol:
            return x_next
        x_n = x_next

    return x_n

def calculate_credible_intervals(alpha_val, beta_val, confidence=0.95
    """
    Calculates the Credible Intervals explicitly using the Newton-Rap
    """
    lower_bound = (1.0 - confidence) / 2.0
    upper_bound = 1.0 - lower_bound
    ci_lower = newton_raphson_inverse_beta(lower_bound, alpha_val, be
    ci_upper = newton_raphson_inverse_beta(upper_bound, alpha_val, be
    return ci_lower, ci_upper

if __name__ == "__main__":
```

```
# Baseline Parameters (Non-Core) -> Be(0.5, 15.5)
alpha_nc, beta_nc = 0.5, 15.5

print("=============================================================")
print(" BAYESIAN SIGNIFICANCE INTEGRAL CALCULATOR")
print("=============================================================")

# 1. Evaluate Fever (F) -> Be(1.5, 14.5)
alpha_F, beta_F = 1.5, 14.5
prob_F_exact = calculate_exact_numerical_integral(alpha_F, beta_F, alpha_
prob_F_mc = calculate_monte_carlo_simulation(alpha_F, beta_F, alpha_nc, b
ci_F = calculate_credible_intervals(alpha_F, beta_F)

print("-- Evaluating FEVER (F) --")
print(f"Exact Numerical Integral: {prob_F_exact:.4f} ({prob_F_exact*100:.
print(f"Monte Carlo Simulation   : {prob_F_mc:.4f} ({prob_F_mc*100:.2f}%)"
print(f"95% Credible Interval    : [{ci_F[0]:.3f}, {ci_F[1]:.3f}]")
print("Result: FAILS 95% Threshold\n")

# 2. Evaluate Low Platelets (LP) -> Be (3.5, 12.5)
alpha_LP, beta_LP = 3.5, 12.5
prob_LP_exact = calculate_exact_numerical_integral(alpha_LP, beta_LP, alp
prob_LP_mc = calculate_monte_carlo_simulation(alpha_LP, beta_LP, alpha_nc
ci_LP = calculate_credible_intervals(alpha_LP, beta_LP)

print("-- Evaluating LOW PLATELETS (LP) --")
print(f"Exact Numerical Integral: {prob_LP_exact:.4f} ({prob_LP_exact*100
print(f"Monte Carlo Simulation   : {prob_LP_mc:.4f} ({prob_LP_mc*100:.2f}%
print(f"95% Credible Interval    : [{ci_LP[0]:.3f}, {ci_LP[1]:.3f}]")
print("Result: PASSES 95% Threshold")
```

Simulation Console Output

```
=============================================================
 BAYESIAN SIGNIFICANCE INTEGRAL CALCULATOR
=============================================================
-- Evaluating FEVER (F) --
Exact Numerical Integral: 0.8357 (83.57%)
Monte Carlo Simulation   : 0.8354 (83.54%)
95% Credible Interval    : [0.011, 0.268]
Result: FAILS 95% Threshold

-- Evaluating LOW PLATELETS (LP) --
```

```
Exact Numerical Integral: 0.9782 (97.82%)
Monte Carlo Simulation  : 0.9780 (97.80%)
95% Credible Interval   : [0.064, 0.438]
Result: PASSES 95% Threshold
```