Bridging Process Mining and the Courtroom: A Data-Driven Framework for Fairness, Efficiency, and Transparency in Judicial Decision-Making

DINH THUY NHAT VY, University of Twente, The Netherlands

Judicial systems around the world continue to struggle with inefficiency, inconsistency, and limited transparency, which can undermine public trust and slow down access to justice. This paper presents a broadly applicable, modular framework that leverages process mining, statistical analysis, machine learning, and a novel activity role classification to assess and improve fairness, efficiency, and transparency in judicial decision-making. Applied to 247 German social court cases, the framework identified four process clusters: (1) fast-track cases resolved through early settlements or withdrawals; (2) complex, trial-driven cases with extended durations; (3) expert-heavy cases with frequent coordination bottlenecks; and (4) moderate-complexity resolutions. A multi-level fairness analysis (statistical, predictive, and causal) found only minor differences between court chambers, with no substantial effect on case duration. Activity role classification revealed that individual orders amplify procedural complexity, while repeated medical assessments are major bottlenecks. Transparency analysis showed that unpredictability and process opacity are concentrated in administrative and assessment-related transitions, pinpointing where targeted improvements could enhance clarity and predictability. Overall, these findings support more proactive and transparent workflow management, providing courts and other organizations with interpretable tools to address inefficiencies while preserving their autonomy.

Additional Key Words and Phrases: Process mining, judicial systems, trace clustering, fairness analysis, causal inference, workflow bottlenecks, transparency, activity role classification.

1 INTRODUCTION

1.1 Background and Motivation

The saying *"justice delayed is justice denied"* continues to hold significance as ever, yet courts across the globe still face persistent delays, inconsistencies, and a lack of transparency that threaten public confidence in the legal system [10, 13, 23]. While industries like healthcare and manufacturing have embraced data-driven approaches to optimize their processes, the judicial sector has been slower to adopt such innovations [6].

Process mining—a set of techniques that blends data science with process management—provides an advantageous way to uncover, monitor, and improve real-world processes by analyzing event logs [24]. Although these methods have already transformed other fields, their use in the legal domain is still in its early days. Moreover, the lack of established frameworks tailored to the judicial context—unlike the sector-specific taxonomies available in other domains—makes it difficult to interpret and apply process mining results effectively. This gap presents a real opportunity to use computational tools to tackle the long-standing challenges facing legal workflows.

1.2 Problem Statements and Research Question

Legal systems worldwide face three fundamental, interconnected challenges that impede judicial effectiveness:

(1) Procedural inefficiencies: Backlogs and delays in court proceedings create significant barriers to timely justice, with some cases taking months or even years to resolve [23].
(2) Inconsistency in process features: Similar cases may follow divergent procedural paths due to implicit bias, varying interpretations, or institutional practices, challenging the principle of equal treatment [14].
(3) Limited transparency: High variability and unpredictability in process flows can obscure the rationale behind procedural decisions, making it difficult for stakeholders to interpret case progress and reducing accountability [9].

These issues are compounded by the lack of data-driven tools specifically designed for judicial workflows—tools that can provide actionable insights while respecting the nuanced nature of legal decision-making. In response to this gap, and to support the shift from reactive to proactive judicial systems, this study addresses the following research question: ***How can process mining and data-driven methods help identify procedural risks, inefficiencies, and transparency gaps in judicial workflows, thereby promoting fairer, more efficient, and more transparent justice systems?***

1.3 Contributions

This paper tackles these challenges by developing, implementing, and testing a comprehensive process mining framework for judicial systems, focusing on German social court proceedings. The main contributions are:

(1) **A unified, modular framework** that brings together process mining, statistics, and machine learning, and is designed to be easily adapted to other event-driven processes.
(2) **Empirical identification of process clusters** that reveal actionable patterns in judicial workflows.
(3) **A multi-level fairness assessment** combining statistical, predictive, and causal inference approaches for bias-agnostic evaluation of disparities while respecting judicial independence.
(4) **Delay and transparency diagnostics** using survival analysis, queueing theory, and activity role classification to detect and quantify bottlenecks and sources of unpredictability.
(5) **A generalizable activity role classification tool** for identifying injectors, transmitters, amplifiers, and buffers in any

Author's address: Dinh Thuy Nhat Vy, dinhthuynhatvy@student.utwente.nl, University of Twente, P.O. Box 217, Enschede, The Netherlands, 7500AE.

TScIT 43, July 4, 2025, Enschede, The Netherlands
 The definitive Version of Record was published in *Proceedings of 43rd Twente Student Conference on IT (TScIT 43)*, https://doi.org/10.1145/nnnnnnn.nnnnnnn.

process-driven domain, supporting explainable analytics and targeted improvements.

(6) **Practical support for continuous improvement** through clear visualizations and monitoring tools that empower legal professionals to manage workflows proactively, without undermining institutional autonomy.

1.4 Paper Organization

The remainder of this paper reviews related work, outlines the methodology and framework, presents the main findings, discusses implications, limitations, and directions for future research, and concludes with a summary of key contributions. For clarity and ease of reference, all abbreviations used throughout this paper are compiled in Appendix A, Table 1.

2 RELATED WORK

Process mining has become a powerful tool for extracting actionable insights from event logs, supporting the discovery, monitoring, and improvement of real-world processes. However, the complexity and variability of event logs, especially in judicial systems, call for advanced analytical techniques to produce interpretable and precise models. To better understand and address these challenges, this section reviews key methodological advances in three areas: clustering and pattern discovery for managing process variability, statistical methods for robust process comparison, and recent applications of process analytics within judicial contexts. These topics collectively establish the foundation for the integrated analytical approach developed in this study.

2.1 Clustering and Pattern Discovery in Process Mining

Traditional process mining often struggles with highly variable or "spaghetti-like" event logs, leading to overly complex or generalized models. To address this, clustering-based approaches partition event logs into more homogeneous subsets, enabling the discovery of clearer and more precise process models. Alves de Medeiros et al. introduced a clustering methodology where event logs are iteratively grouped so each cluster represents a coherent set of cases, improving model precision and interpretability in complex domains [4].

A key component of pattern discovery in process mining is the identification of frequent behavioral patterns or variants. The Apriori algorithm, originally developed for association rule mining, has been adapted for use in process mining to efficiently identify frequent sequences and refine clusters [1]. Integrating Apriori-based pattern mining with clustering enables the extraction of meaningful process variants and supports the construction of more robust and representative process models, as discussed in the literature [4].

2.2 Statistical Analysis for Process Comparison

Comparing process features across groups, such as judicial chambers, requires robust statistical methods. The Mann–Whitney U test is a widely used non-parametric test for assessing whether two independent samples come from the same distribution, making it suitable for process mining data that may not be normally distributed [20]. In practice, process analytics often involve comparing distributions of durations, expert involvement, and other key features across data segments. To account for multiple comparisons, ***p*-values** from Mann–Whitney tests can be aggregated using Fisher's method [12], and the false discovery rate correction (Benjamini-Hochberg procedure [5]) is commonly applied to control for false positives. These techniques are standard in process analytics for evaluating fairness and consistency in complex datasets.

2.3 Process Analytics in Judicial Contexts

Recent work has demonstrated the applicability of process analytics to judicial processes. For example, Aleknonytė-Resch et al. conducted a case study in a German social court, combining data-driven analysis with expert knowledge to identify bottlenecks and factors influencing case duration. Their findings highlighted the impact of expert witness involvement, assessment documents, and reminders on case duration, and validated the practical relevance of process analytics through collaboration with domain experts [2].

In a related context, Caponecchia et al. applied process mining to Italian civil court proceedings and found a significant enhancement in process efficiency over time, characterized by a substantial reduction in median case processing times and an increased degree of process standardization [7]. Together, these studies illustrate how process analytics can both diagnose procedural bottlenecks and support measurable improvements in judicial workflows.

2.4 Bridging the Gap: Towards an Integrated Analytical Framework

While these works provide valuable methods for process analysis, they often lack a unified framework for simultaneously assessing fairness, delays, and transparency. Building upon the foundational analysis of Aleknonytė-Resch et al. and extending beyond, this study introduces a transferable analytical pipeline that synthesizes causal inference, survival analysis, queueing theory, and a novel activity role classification—alongside established process mining, clustering, and statistical analysis techniques—into a cohesive framework. The modular design enables adaptation to diverse event-driven domains, while the multi-tiered fairness assessment provides an empirically grounded, bias-agnostic evaluation. This integrated approach supports a holistic analysis of workflows and enables the systematic detection of inconsistencies, procedural delays, and transparency gaps.

In summary, the literature demonstrates the potential of process analytics in judicial contexts, but also highlights the need for a comprehensive, modular framework that can address multiple dimensions of judicial performance. The following section details the methodology and analytical framework developed in this study to bridge this gap.

3 METHODOLOGY AND ANALYTICAL FRAMEWORK

The Cross-industry standard process for data mining (CRISP-DM) methodology is adopted for its proven structure and adaptability in data-driven research [8]. Its modular phases are particularly well-suited to address the complexities of judicial process mining, supporting systematic analysis, transparency, and stakeholder relevance. By tailoring each phase to the legal context, the approach ensures that data preparation, modeling, and evaluation are both rigorous and interpretable. The overall workflow, adapted to these requirements, is summarized in Figure 1, with each phase detailed in the following subsections.

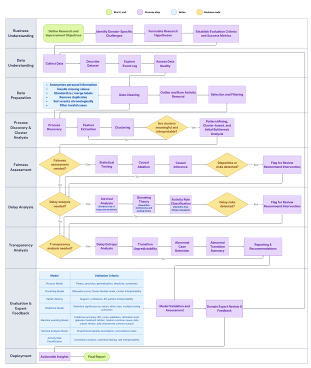

Fig. 1. Workflow of the data-driven process analysis framework.

3.1 Data Description

The dataset used in this study is identical to that described by Aleknonytė-Resch et al. [2]. It comprises 260 cases from three chambers within a single German social law court, extracted from redacted portable document format (PDF) documents using optical character recognition (OCR). All data and initial preprocessing were performed by Aleknonytė-Resch and colleagues, as detailed in their original publication. The final event log contains 19,948 events and 59 unique activities. Case durations range from 65 to 1,626 days (mean: 559 days, median: 548 days), and the number of events per case ranges from 20 to 194 (mean: 77 events per case). A complete list of activity labels is provided in Appendix A, Table 2.

3.2 Preprocessing and Event Log Preparation

Data preparation involved standard quality checks (see Figure 1), including the removal of cases with missing key fields or invalid outcomes (such as *Final ruling*, *Settlement declaration*, or *Withdrawal*), as well as the exclusion of outliers above the 95th percentile (1,062 days) and activities occurring in fewer than 2% of cases (Appendix A, Table 3). After these steps, the dataset included 247 cases, 18,399 events, and 51 activities, with analysis focused on the most common process variants covering 85% of cases. This resulted in a clean and representative event log for subsequent analysis.

3.3 Process Discovery and Clustering

To reveal the structure of judicial workflows, I applied process discovery to the filtered event log. The Heuristics Miner algorithm (PM4Py) was chosen for its ability to handle complex, variable event logs typical of judicial processes [15]. Unlike Inductive or Alpha Miner, which may yield overly complex or fragmented models in noisy data, Heuristics Miner filters out infrequent behavior and focuses on significant control-flow relations, making it well-suited for domains with high variability. Noise thresholds from 0.0 to 0.35 (in 0.05 increments) were systematically evaluated by adjusting **dependency** and **AND-measure** parameters, balancing model fitness, precision, generalization, simplicity, and soundness. The best-fitting global process model was selected based on a composite score of these metrics.

Given the diversity of cases, trace clustering was used to group similar cases. Features such as expert involvement, repeated activities, and average days between events were standardized. The optimal number of clusters was determined using **silhouette score** (ranging from -1 to 1), **Davies-Bouldin index** ($DBI \geq 0$), the elbow method, and a minimum cluster size of 25 to ensure reliable results. Several clustering algorithms were compared, including k-means, DBSCAN, agglomerative, spectral, and a hybrid approach combining feature-based and sequence-based distances—specifically, normalized Levenshtein (edit) distance [16]. Clusters were further refined using Apriori-based pattern mining to identify distinctive activity sets and subdivide larger, more varied groups. Each cluster was validated with descriptive statistics, including case volumes, event frequencies, characteristic activities, start and end activities, expert involvement rates, duration distributions, and chamber distributions. For each final cluster, a dedicated process model was discovered to support further analysis of process variants, bottlenecks, and activity sequences.

This approach provided a structured foundation for the subsequent analyses of fairness, delays, and transparency in judicial case processing.

3.4 Cluster-Based Pattern Analysis

To validate and contextualize the clustering results, pattern analysis was conducted within each cluster, focusing on deviations in activity frequencies (relative to global statistics), repeated activity loops, and average waiting times between activities to identify bottlenecks. Transitions were flagged as bottlenecks if delays exceeded 15% of the cluster's mean case duration. Direct transition frequencies were also compared to global patterns, with transitions classified as notably overused or underused if their frequency deviated by more than ±21.5% from the baseline[1]. These distinctive transitions were visualized using network diagrams and heatmaps, providing an intuitive overview of how process flows diverge across clusters.

Overall, this cluster-based pattern analysis complements the statistical validation by revealing the behavioral mechanisms that distinguish each process variant, offering actionable insights into case progression and potential sources of delay.

3.5 Assessing Fairness: Statistical and Causal Approaches

Fairness in this study refers to the absence of systematic disparities in process features and outcomes (such as case duration, number of expert activities, and temporal span) between chambers, rather than differences in case verdicts or parties' success (see Tables 5 and 4 for outcome and feature variables, respectively). To assess fairness across judicial chambers, a multi-stage analytical framework was implemented, integrating statistical testing, causal ablation, and causal inference, each with carefully selected parameters and evaluation metrics. First, a comprehensive set of case-level features—such

[1] This threshold was chosen empirically to balance sensitivity and specificity: lower thresholds introduced excessive noise, while higher ones risked overlooking meaningful process deviations.

as chamber load, expert involvement, activity diversity, settlement attempts, and temporal attributes—was extracted. Key outcomes were analyzed across all statistical and causal tests, with additional outcomes included as needed. These selections ensured control for confounders and comparability between chambers.

3.5.1 Statistical Testing. Fairness was first evaluated using the Mann–Whitney U test, a non-parametric method suitable for comparing outcome distributions between groups without assuming normality. Tests were conducted within strata defined by cluster and year, ensuring comparability across chambers. Primary metrics were **p-values** (with Benjamini-Hochberg correction) and **effect sizes** (**area under the curve, AUC**), quantifying both significance and magnitude of disparities. This approach reliably identifies statistical differences in outcomes such as duration and complexity, while controlling for confounders and reducing false positives.

3.5.2 Causal Ablation. To assess the sensitivity of outcome predictions to chamber assignment, causal ablation was performed using Random Forest regression. Random Forests were chosen for their robustness to non-linear relationships, ability to handle high-dimensional features, and resistance to overfitting [21]. Risks from irrelevant or noisy features were mitigated through prior feature selection and removal of uninformative attributes. Predictive models were trained on pre-chamber features with and without explicit chamber information, using 10-fold cross-validation. The key metric was the change in coefficient of determination (ΔR^2) between models, with 95% **confidence intervals (CI)** computed across folds. Paired t-tests assessed the statistical significance of ΔR^2. This step comes before causal inference to quickly check if chamber assignment adds any predictive value beyond the case features. If it does, a deeper causal analysis is needed.

3.5.3 Causal Inference. To estimate the direct effect of chamber assignment on key outcomes, causal inference techniques were applied. Double machine learning (DoubleML) with cross-validated Lasso and XGBoost models controlled for observed confounders, while propensity score matching provided robustness checks. Model hyperparameters (**number of estimators**, **learning rates**, **regularization strengths**) were selected empirically using cross-validation and standard tuning procedures. The specific values used in this study are available in the accompanying code repository[2]; however, these may require adjustment for other datasets or applications. The primary metrics included **average treatment effects (ATE)**, **CIs**, and results from a suite of refutation tests—namely, the **placebo treatment**, **random common cause**, **data subset**, and **add unobserved common cause refuters**—to assess the stability and credibility of causal estimates. These methods were chosen to move beyond mere association and estimate the genuine causal effects of chamber assignment.

This combined approach enables a thorough assessment of fairness. By using a range of analyses, the framework distinguishes between statistical differences, the predictive impact of chamber assignment, and genuine causal effects. This supports both the detection and practical improvement of fairness in judicial processes.

[2]https://gitlab.utwente.nl/s2803100/research-project.git

3.6 Delay Analysis: Survival Model, Queueing Theory, and Activity Roles

Delay in judicial case processing was analyzed using three complementary approaches: survival analysis, queueing theory, and activity role classification. Each method was chosen to provide a different perspective on where and why delays occur, and to support actionable recommendations for process improvement.

3.6.1 Survival Analysis. Survival analysis was used to predict which cases would take a long time to resolve, with "long duration" defined as cases above the 80th percentile in length. To enable early prediction, features were extracted from the first 45 days of each case. This window was chosen based on the dataset's characteristics: the median case duration is about 548 days, and even the shortest cases last over two months. Using a 45-day window allows for early detection of potential delays while still capturing enough process activity for meaningful analysis. The main features included how often events happened, the event rate (number of events per 45 days), whether an attorney or expert was involved early, and how many different activities took place.

To model the impact of early case features on duration, the Cox proportional hazards model was applied. This widely used method in survival analysis estimates how covariates influence the hazard rate—the instantaneous probability of case resolution at a given time—without requiring the specification of an underlying time-to-event distribution [11]. Before modeling, the Cox model's key assumption of proportional hazards was checked. This assumption means that the effect of each feature on the hazard rate is constant over time. It was assessed using both statistical tests (such as the Schoenfeld residuals test [19]) and graphical diagnostics. Model accuracy was measured using the concordance index. This approach helps identify early signs that a case might be delayed, allowing for timely intervention.

3.6.2 Queueing Theory. To better understand systemic bottlenecks, the expert witness process was modeled as a queueing system. This approach was motivated by findings in Aleknonytė-Resch et al., which revealed that cases requiring expert witness reports experience significant delays—averaging an additional 121 days. In this model, each case entering the expert witness phase was treated as a customer in a queue. As discussed in Chapter 5 on Queueing Theory [3, pp. 149–233], the system was first analyzed using LITTLE's LAW:

$$L = \lambda \cdot W \tag{5.11}$$

where L represents the average work-in-progress (cases active in the expert phase), λ denotes the average arrival rate, and W is the average time per case in this phase. This relationship quantifies backlog severity and resource requirements.

Building on this foundation, variability was assessed through the **coefficient of variation (CoV)**. A high **CoV** signals inconsistent processing times—a key indicator of unstable process segments and potential bottlenecks.

Finally, activity-level contributions to delays were quantified using the FORCED FLOW LAW, as described in Chapter 6 on Queueing Models of Computer Systems [3, pp. 234–270]:

$$X_k = V_k \cdot \lambda \tag{6.2}$$

Here, V_k denotes the average number of visits to expert witness activity k per case, while X_k represents its throughput. By identifying activities with high visit counts and substantial throughput, this approach helps pinpoint process stages that may disproportionately contribute to delays, thereby informing where targeted improvements may be most effective.

3.6.3 Activity Role Classification. To trace how delays spread through the workflow, activity role classification assigns each activity one of four roles based on its overall impact. Rather than focusing on step-by-step transitions, it highlights key points where intervention will have the greatest effect. Please refer to Table 6 in the Appendix B for a detailed summary of each role's definition, statistical indicators, and managerial implications.

3.7 Transparency Analysis

Transparency in this study is defined as the clarity and predictability of case progress, operationalized through measures of entropy and variability in process durations and transitions, focusing on process flow characteristics instead of final case outcomes. To assess transparency, this analysis quantified the predictability of case durations using **Shannon entropy**, both overall and within each cluster.

To further examine process irregularities, transitions between activities were analyzed using both the **coefficient of variation** and **entropy** of inter-step durations. This helped identify which transitions contribute most to inconsistent delays and timing uncertainty.

For cases exceeding the 90th percentile in duration within their respective clusters, transition paths were compared against typical cluster behavior. This highlighted unusually slow or unique steps, providing insight into why certain cases take longer than average.

Together, these analyses reveal where unpredictability and bottlenecks emerge, offering actionable guidance for improving procedural transparency and judicial workflow efficiency.

4 RESULTS

This section presents the main findings of the study, organized according to the analytical framework described previously. The results are structured around the key methodological components: process discovery and clustering, cluster-based pattern analysis, fairness assessment, delay and bottleneck analysis, and transparency analysis.

4.1 Process Discovery and Clustering Results

The global process model generated from the event log shows that judicial workflows are complex and variable, resulting in a highly intricate "spaghetti" process map. Quantitative evaluation of the best model—discovered using the Heuristics Miner algorithm with a noise threshold of 0.35—showed strong performance across key metrics: fitness (0.829), precision (0.902), generalization (0.715), simplicity (0.424), and soundness (all criteria met). The model was generated in 2.151 seconds.

However, given the complexity and limited interpretability of the full process map, clustering analysis was performed as described in the methodology to obtain more actionable insights. k-means clustering achieved the highest silhouette score (0.602) and lowest Davies–Bouldin index (0.772), initially identifying two clusters. Apriori-based pattern refinement further subdivided these into four distinct clusters, summarized in Appendix C, Table 7. The most significant rules driving this refinement included transitions such as `Court order` → `Request for medical findings and treatment report` and `Written statement EW` → `Medical findings and treatment report`, both with a confidence of 1.0. These rules enabled the identification of more homogeneous process variants within the data. Each cluster represents a distinct process variant: (1) **Cluster 0** – fast-track cases with simple paths and no expert involvement; (2) **Cluster 1** – court-driven, high-complexity cases with the longest durations; (3) **Cluster 2** – cases with frequent expert witness involvement and coordination delays; and (4) **Cluster 3** – moderate-complexity cases, often resolved after expert or medical reporting. These clusters provide a structured foundation for further analysis of process patterns, fairness, delays, and transparency, highlighting areas—such as expert coordination and court order issuance—where targeted improvements may be most effective.

4.2 Cluster-Based Pattern Analysis Results

Building on the clustering results, I analyzed each group's process model, activity frequencies, loops, bottlenecks, and transition patterns to validate and contextualize the earlier insights. For reference, all process models for each cluster are provided in Appendix C (Figures 4 to 7).

Cluster 0 (fast-track, simple cases) is characterized by a linear process model with minimal branching and few loops. Activity frequency analysis confirms the dominance of administrative steps, while expert-related activities are entirely absent. Bottlenecks are rare and isolated, with only occasional delays (e.g., "Attachment AP" to itself: 203 days). Overused transitions are limited, and the process is generally efficient. Given this simplicity (model simplicity score: 0.578), no main figures are included here; detailed significant transition patterns map and transition heatmap are provided in Appendix C Figures 8 and 9, respectively.

Cluster 1 (court-driven, high-complexity) stands out for its intricate process model, marked by frequent procedural loops involving activities such as "Medical findings and treatment report" (41 repetitions) and "Directive" (34 repetitions). Activity frequency analysis reveals a high occurrence of court orders and assessment-related steps, which are typical of trial-driven cases. The significant transition patterns map for Cluster 1 (Figure 2) provides a clear visualization of the cluster's procedural complexity and helps identify key transitions and major sources of delay.

Cluster 2 (expert-heavy) is defined by a dense network of transitions involving expert witness activities, as illustrated in the significant transition patterns map for this cluster (Figure 3). Overused transitions prominently connect "Written statement EW," "Official letter EW," and "Directive," emphasizing the central role of expert communication in these cases. Additionally, "Official letter PP" and "Official letter AP" serve as key connectors, linking expert-related steps with broader administrative actions. This pattern underscores the pivotal influence of both expert input and official correspondence in driving the progression of cases within this cluster.

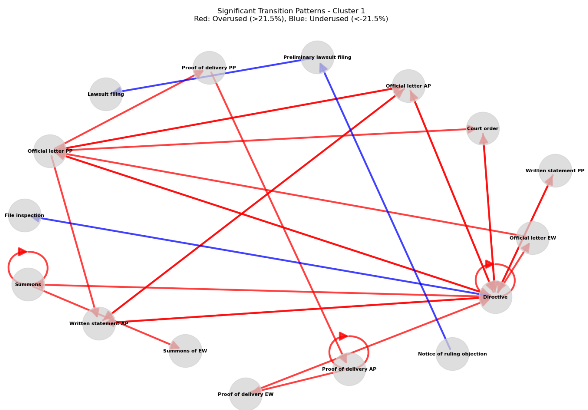

Fig. 2. Significant transition patterns map for Cluster 1.

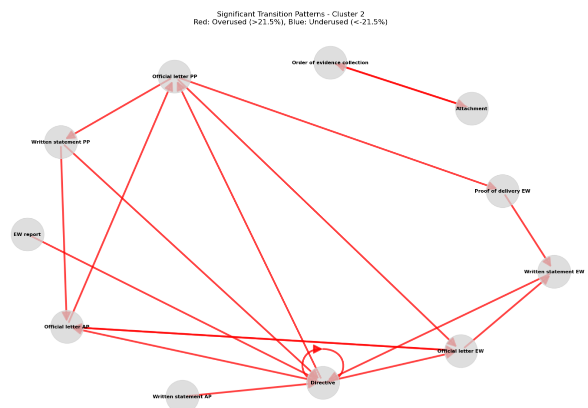

Fig. 3. Significant transition patterns map for Cluster 2.

Cluster 3 (moderate complexity) is defined by a diverse set of activities and frequent process loops, particularly in "Medical findings and treatment report" (186 repetitions), "Directive" (78), and "Summons" (57). The absence of activities such as "Court order" and "Written statement EW" distinguishes these cases from more formal or expert-driven proceedings. Many cases in this cluster progress through multiple cycles of medical assessment and administrative actions before reaching resolution. Bottleneck analysis reveals substantial delays, with transitions like "Assessment documents" to itself taking 420 days and "Attachment other" to "Notice of ruling objection" taking 357 days, indicating that document review and clarification are primary sources of inefficiency. Overall, while these cases involve some expert input, they tend to be resolved after medical findings are clarified, thus avoiding the extended complexity and delays seen in trial-driven clusters. Compared to the expert-heavy cluster, delays are also mitigated by the relative absence of expert reminders, which occur only 5 times in this group.

These findings further validate the initial clustering, revealing how specific procedural mechanisms and transition patterns shape the efficiency and complexity of each group.

4.3 Fairness Assessment Results

Non-parametric statistical testing (Mann–Whitney U) found no significant disparities between chambers for most metrics, including duration, number of court activities, number of events, and expert involvement (all p-values > 0.177). The only exception was a moderate disparity in temporal span for Chamber 1 ($p = 0.036$), where cases tended to have a longer temporal spread, but this effect was not observed for other outcome measures. The distribution of case durations by cluster and chamber (see Appendix C, Figure 10) reveals some variability, but no chamber consistently exhibits longer or shorter durations across all clusters, visually further supporting the statistical findings.

Regression-based sensitivity analysis was additionally consistent with these results. Adding chamber information to predictive models for duration, expert activities, loops, reminders, and temporal span did not significantly improve model performance ($\forall \Delta R^2 \in \mathbb{R},\ \Delta R^2 \approx 0;\ \forall p \in (0, 1],\ p > 0.226$), indicating that chamber assignment does not provide additional explanatory power beyond case features.

Causal inference analyses, including DoubleML, propensity score matching, and S-Learner, were used to estimate the average treatment effect of chamber assignment on key outcomes. Across all outcomes tested, the estimated causal effects were negligible or not statistically significant, and refutation tests consistently supported the robustness of these results.

Taken together, the lack of significant differences in statistical tests, the minimal predictive value of chamber assignment, and the absence of causal effects provide strong, converging evidence that chamber assignment does not predict or cause meaningful disparities in case outcomes. This suggests that chambers do not systematically favor or neglect particular types of cases, even when cases are grouped by distinct process patterns, which further supports the overall fairness of the judicial process in this setting.

4.4 Delay and Bottleneck Analysis Results

Survival analysis of 247 cases highlighted key factors linked to prolonged case durations. The Cox proportional hazards model performed well (concordance: 0.690), identifying the average time between early events ($\beta = -0.470$, $p = 0.010$) and event rate ($\beta = 0.420$, $p = 0.060$) as the most influential risk factors. Notably, shorter intervals between early events were significantly associated with a lower risk of long delays, emphasizing the value of early procedural momentum. In contrast, simply receiving medical reports or requests early did not have a significant effect (high p-values), suggesting that active assessment, rather than early receipt, is more important for timely case resolution. This finding aligns with Aleknonytė-Resch et al. [2], who observed that the presence of an assessment document is linked to longer case durations—highlighting assessment as a key driver of delay. The relationship between assessment documents and average time between events further supports this view. Scaled Schoenfeld residual plots (see Appendix C Figures 11 to 14) confirm that the proportional hazards assumption holds for all main covariates.

Expert involvement emerged as the second most significant contributor to delays, warranting closer examination. Queueing analysis of the expert witness phase revealed a substantial bottleneck: cases spent an average of 323 days in this stage, with about 42 cases in progress at any time. Since the arrival rate of new expert witness (EW) cases ($\lambda = 0.13$ cases/day) is largely fixed, reducing the average time spent in this phase (W) is the most effective way to decrease overall work-in-progress (L), as described by LITTLE's LAW. The most significant delays stemmed from repeated cycles of "Medical findings and treatment report," which occurred 510 times and showed high variability (CoV = 2.621). This pattern may reflect frequent rework, incomplete initial assessments, or a mismatch between report content and the information judges need. Additionally, the average transition time from "Proof of delivery EW" to "EW report" was 121 days (CoV = 0.629), and from "Reminder EW" to the next step averaged 155 days (CoV = 0.674), both indicating substantial and unpredictable waiting periods. These findings point to the need for clearer judicial expectations for expert reports, standardized initial submissions, and more proactive follow-up to reduce both the length and variability of delays in the expert witness phase.

Beyond phase-level bottlenecks, an activity-wise perspective—captured in the activity percolation roles heatmap (Appendix C, Figure 15)—reveals how specific steps contribute to the propagation of delays within the process.

Among injector activities, "Court order" stands out as an expected result, since it naturally initiates new procedural requirements, deadlines, or actions that introduce delays. More surprisingly, "Receipt" also emerged as a strong injector. While it may seem like a routine administrative step, in practice it often triggers follow-up tasks or new procedural timelines, thereby injecting fresh delays into the workflow. This shows how even minor administrative actions can significantly affect overall process efficiency.

Turning to transmitter activities, "Medical findings and treatment report" is a particularly notable example, as it often carries existing delays forward to subsequent steps—a finding that is consistent with the survival analysis and prior research emphasizing the critical role of assessment, rather than mere receipt, in driving case duration.

In judicial practice, an "Individual order" functions as an amplifier because issuing supplementary or case-specific orders often initiates additional procedural steps, follow-up actions, or clarifications. These orders can extend timelines, particularly when they require responses from multiple parties or generate further documentation, thereby compounding existing delays and intensifying the procedural load. To mitigate this effect, streamlining the drafting and communication of individual orders, and ensuring that their scope is as clear and comprehensive as possible from the outset, can help prevent unnecessary iterations and reduce cumulative delays.

Meeting-at-courtroom actions—such as official invitations, summons, and court rulings—typically serve as buffers in the process. These activities often absorb accumulated delays by consolidating procedural steps and bringing cases to key decision points, thereby helping to stabilize timelines and limit the further spread of delays.

Taken together, the combined insights from survival analysis, queueing theory, and activity role classification provide a comprehensive understanding of why, where, and how delays occur—laying the groundwork for data-driven improvements in judicial efficiency.

4.5 Transparency Analysis Results

Transparency was assessed by measuring the unpredictability of case durations and process transitions. Overall, case duration entropy was moderate (1.609), with Cluster 2 and Cluster 3 exhibiting the highest variability. Analysis of transitions revealed that certain handoffs—such as `Official letter AP` → `Official letter PP`—were especially unpredictable, with a high coefficient of variation (CoV = 14.313) and frequent occurrence across clusters. Transitions with the highest entropy often involved administrative or assessment-related steps, indicating that procedural uncertainty is concentrated in these areas. Abnormal cases, defined as those exceeding the 90th percentile in duration within their cluster, were most prevalent in Cluster 3. Examination of these cases showed that excess delays frequently accumulated in transitions involving directives, official letters, and written statements, highlighting these as key contributors to timeline opacity. These findings suggest that while overall process transparency is moderate, specific administrative and assessment-related transitions are major sources of unpredictability, and targeted improvements in these areas could enhance the predictability and clarity of judicial workflows.

5 DISCUSSION

5.1 Key Findings and Practical Implications

The results show that applying process mining and data-driven analytics can move beyond descriptive statistics to identify specific procedural risks, inefficiencies, and transparency gaps in judicial workflows. By applying the analytical framework described in Section 3 to German social court data, the analysis yields actionable insights for judicial administration.

Finding these distinct groups—ranging from fast-track cases to those involving lengthy trials or heavy expert involvement—shows just how varied court processes can be, and points to the need for management approaches that fit each type. This finding aligns with previous research [4], highlighting the value of clustering and pattern mining for targeted process improvement over generic reforms.

Beyond mapping out the different types of workflows, the study also looked closely at whether the process was fair. The fairness analysis did not uncover any consistent differences between chambers, which suggests that cases are handled fairly across the court. This finding is particularly significant given ongoing concerns about bias in judicial systems, as noted by Kleinberg et al.[14]. By combining statistical, predictive, and causal approaches, it becomes possible to create a practical framework that others can adapt when evaluating fairness in different organizations.

Beyond fairness, the analysis of delays and bottlenecks reveals inefficiencies in key phases, particularly expert witness involvement and cycles of medical assessment. These findings resonate with earlier work [2] and, through queueing theory and survival analysis, precisely quantify where and why delays arise. By identifying "injector" and "amplifier" activities—such as receipt and individual orders steps—the study points to concrete targets for process redesign, including standardizing expert reports and streamlining administrative handoffs. At the same time, it is important to recognize that *"the role of a judge is a complex one.*[3] *"* [22]. This complexity means that, although analytical methods can reveal important patterns, they may overlook the human and contextual aspects of judicial work. By combining these tools with expert insight, we can make meaningful improvements while respecting judicial independence.

When looking at transparency, the results indicate that most unpredictability happens during administrative and assessment steps, which matches ongoing worries about how clear court procedures really are. While much of the literature on judicial transparency has focused on the challenges of understanding or predicting final decisions [18], the present analysis—limited to process data rather than case outcomes—suggests that significant sources of unpredictability may instead lie in the supporting administrative and assessment steps. Looking at the problem this way points to practical steps—like improving communication or using digital tracking—in parts of the process that are often overlooked, which could make case progress more predictable and build trust.

Taken together, these findings illustrate how the proposed methodology and results directly address the central research question. This work also shows that bringing these methods together in a modular framework makes it possible to pinpoint and understand where inconsistencies, inefficiencies, and lack of clarity arise.

5.2 Scientific Contributions

Beyond the immediate practical implications, this study makes several contributions that extend both to judicial workflows and to broader process analysis. By bringing together process mining, advanced statistical analysis, and machine learning within a modular framework, this work moves past isolated case studies and offers a coherent, multi-level approach to examining fairness, efficiency, and transparency in legal processes. A central element of the framework is its multi-level fairness assessment, which enables transparent evaluation of disparities without presuming bias and supports more nuanced, evidence-based insights. Furthermore, the activity role classification introduced here offers a practical means to trace where procedural delays originate and how they develop over time, enriching the analytical toolkit available for process analysis and providing clear guidance for workflow optimization. Applying the framework to real court data demonstrates both its methodological robustness and its practical relevance, while the recommendations are designed to improve procedures without undermining institutional autonomy or established decision-making. Collectively, these contributions advance the development of more rigorous, interpretable, and adaptable process analytics—within the judicial domain and in other complex, event-driven environments.

5.3 Evaluation and Outlook

There are, however, some limitations. The dataset includes very few cases that both started and ended within the same year (for example, only 19 such cases in the year-cluster-chamber breakdown shown in Appendix C, Figure 16), which makes it hard to assess the effects of year-specific policies or events and limits the strength of subgroup comparisons. Some useful details are also missing from the data, such as key legal terms (e.g., injunction, jurisdictional dispute) or indicators of case complexity, which could help refine the analysis and control for confounding factors. While the framework is designed to be general, applying it to other domains may require further adjustments, like customizing features or standardizing activity labels. As with any process mining study, the results depend on the quality and completeness of the event log data; missing or inconsistent events can affect the findings [17].

Overall, this framework shows strong potential for improving fairness, efficiency, and transparency in judicial and other organizational processes. Future work should address these limitations by using more detailed and diverse datasets, adding more contextual variables, and testing the approach in other settings. Other directions include developing real-time monitoring tools, early warning systems for bottlenecks or fairness issues, and making the framework more accessible for non-technical users. Continued methodological improvements could further increase its usefulness and broader impact.

6 CONCLUSION

This study set out to address persistent challenges in judicial systems—inefficiency, inconsistency, and limited transparency—by developing a modular, data-driven framework integrating process mining, statistical analysis, machine learning, and a novel activity role classification. Applied to 247 German social court cases, the framework identified four distinct process clusters, each reflecting different procedural patterns and sources of delay. Fairness analysis found only minor differences between court chambers, with no substantial effect on case duration, supporting the overall fairness of proceedings. Delay analysis showed that repeated cycles of "Medical findings and treatment report" are a major bottleneck, with assessment activities as the key driver of prolonged durations, while activity role classification revealed that individual orders amplify procedural complexity. These findings further indicate that delays and complexity often coincide with reduced transparency, particularly in administrative and assessment-related transitions. By synthesizing these results, this research demonstrates that data-driven process analytics can equip courts and similar organizations with interpretable, actionable tools to proactively manage workflows, address inefficiencies, and enhance transparency—while preserving institutional autonomy. The modular framework developed here is broadly applicable to other event-driven domains, offering a practical pathway for continuous improvement.

Because public trust in legal institutions is essential, improving the fairness, efficiency, and transparency of judicial processes is not merely a technical exercise but a commitment to the core values of justice. Continued exploration of data-driven methods will help ensure that judicial systems can adapt to new challenges while upholding these values. In this way, process analytics can be a meaningful tool for supporting a justice system that is both principled and practical.

[3] It can incorporate activism, complex interactions with people, dispute settlement, case management, public and specific education activities, social commentary as well as adjudicatory functions that might be conducted with other judges or less commonly in some jurisdictions with lay people (juries)

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisors, Faiza Allah Bukhsh and Milda Aleknonytė-Resch, for their invaluable guidance, insights, and support throughout this research. I would also like to thank the members of the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), specifically within Department of Computer Science (EEMCS-CS) and Datamanagement & Biometrics group (EEMCS-CS-DMB) at the University of Twente. Their expertise and dedication have greatly contributed to my understanding and application of data-driven methods in process analytics, which was essential to my work.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*. Morgan Kaufmann, 487–499. https://www.vldb.org/conf/1994/P487.PDF

[2] Milda Aleknonytė-Resch, Anna-Katharina Dhungel, Fabian Elsaeßer, and Arvid Lepsien. 2025. Making the Case for Process Analytics: A Use Case in Court Proceedings. https://doi.org/10.13140/RG.2.2.34664.17924 Preprint.

[3] Arnold O. Allen. 1978. *Probability, Statistics, and Queueing Theory: With Computer Science Applications.* Academic Press. https://doi.org/10.1016/B978-0-12-051050-4.50011-3

[4] Ana Karla Alves de Medeiros, Antonella Guzzo, Gianluigi Greco, Wil M. P. van der Aalst, A. J. M. M. Weijters, Boudewijn F. van Dongen, and Domenico Saccà. 2007. Process Mining Based on Clustering: A Quest for Precision. In *Business Process Management Workshops (BPM 2007) (Lecture Notes in Computer Science, Vol. 4928)*. Springer, 17–29. https://doi.org/10.1007/978-3-540-78238-4_4

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 1 (1995), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[6] Devis Bianchini, Carlo Bono, Alessandro Campi, Cinzia Cappiello, Stefano Ceri, Francesca De Luzi, Massimo Mecella, Barbara Pernici, and Pierluigi Plebani. 2024. Challenges in AI-supported Process Analysis in the Italian Judicial System: What After Digitalization? *Digital Government: Research and Practice* 5, 1 (2024), 10:1–10:10. https://doi.org/10.1145/3630025

[7] V. Caponecchia, B. D'Agostino, G. Comandè, D. Licari, and A. Vandin. 2024. Towards visualizing and analysing legal proceedings with process mining. In *1st International Workshop on Processes, Laws and Compliance - PLC 2024 (CEUR Workshop Proceedings, Vol. 3850)*. CEUR-WS, 46–57. https://ceur-ws.org/Vol-3850/paper5.pdf

[8] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth. 2000. CRISP-DM 1.0: Step-by-step data mining guide. https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf

[9] Ignacio N. Cofone. 2019. Algorithmic Discrimination Is an Information Problem. *Hastings Law Journal* 70, 6 (2019), 1389–1444. https://repository.uclawsf.edu/hastings_law_journal/vol70/iss6/1

[10] Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences* 108, 17 (2011), 6889–6892. https://doi.org/10.1073/pnas.1018033108

[11] Samar Abd ElHafeez, Graziella D'Arrigo, Daniela Leonardis, Maria Fusaro, Giovanni Tripepi, and Stefanos Roumeliotis. 2021. Methods to Analyze Time-to-Event Data: The Cox Regression Analysis. *Oxidative Medicine and Cellular Longevity* 2021, 1 (2021), 1302811. https://doi.org/10.1155/2021/1302811 Part of Special Issue: Statistical Issues and Methodology Concerns in Clinical Studies of Oxidative Stress.

[12] Ronald A. Fisher. 1992. Statistical Methods for Research Workers. In *Breakthroughs in Statistics*, Samuel Kotz and Norman L. Johnson (Eds.). Springer, 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6

[13] JUSTICE. 2023. The State We're In: Addressing Threats & Challenges to the Rule of Law. https://justice.org.uk/our-work/system-wide-reform/rule-of-law/

[14] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. Human Decisions and Machine Predictions. *The Quarterly Journal of Economics* 133, 1 (2018), 237–293. https://doi.org/10.1093/qje/qjx032

[15] Radosław Klimek and Julia Witek. 2025. Logic Mining from Process Logs: Towards Automated Specification and Verification. arXiv:2506.08628 [cs.SE] https://arxiv.org/pdf/2506.08628v1

[16] Vladimir I. Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* 10, 8 (1966), 707–710. https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf Originally published in Russian in 1965 in *Doklady Akademii Nauk SSSR.*

[17] Heidy M. Marin-Castro and Edgar Tello-Leal. 2021. Event Log Preprocessing for Process Mining: A Review. *Applied Sciences* 11, 22 (2021), 10556. https://doi.org/10.3390/app112210556

[18] Masha Medvedeva, Martijn Wieling, and Michel Vols. 2023. Rethinking the field of automatic prediction of court decisions. *Artificial Intelligence and Law* 31, 1 (2023), 195–212. https://doi.org/10.1007/s10506-021-09306-3

[19] Statistical Modeling and Forecasting. [n. d.]. Schoenfeld Residuals: What are Schoenfeld residuals and how to use them to test the proportional hazards assumption of the Cox model. https://timeseriesreasoning.com/contents/schoenfeld-residuals/ Accessed 2025-06-11.

[20] Nadim Nachar. 2008. The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology* 4, 1 (2008), 13–20. https://www.tqmp.org/RegularArticles/vol04-1/p013/p013.pdf

[21] George Ekow Quaye. 2024. *Random Forest For High-Dimensional Data.* Ph. D. Dissertation. University of Texas at El Paso. https://scholarworks.utep.edu/cgi/viewcontent.cgi?article=5200&context=open_etd Doctoral Program in Data Science. Open Access Theses & Dissertations, 4201.

[22] Tania Sourdin. 2018. Judge v Robot?: Artificial intelligence and judicial decision-making. *University of New South Wales law journal* 41, 4 (2018), 1114–1133. https://www.unswlawjournal.unsw.edu.au/wp-content/uploads/2018/12/Sourdin.pdf

[23] Tania Sourdin and Naomi Burstyner. 2014. Justice Delayed is Justice Denied. *Victoria University Law and Justice Journal* 4, 1 (2014), 49–62.

[24] Wil van der Aalst. 2016. *Process Mining: Data Science in Action* (2 ed.). Springer. https://link.springer.com/book/10.1007/978-3-662-49851-4

A ABBREVIATIONS, ACTIVITY LABELS, AND EXCLUSIONS

Table 1: List of abbreviations.

Abbreviation	**Full Term**
AP	Applicant Party
ATE	Average Treatment Effect
AUC	Area Under the Curve
CI	Confidence Interval
CoV	Coefficient of Variation
CRISP-DM	Cross-Industry Standard Process for Data Mining
DBI	Davies-Bouldin Index
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DoubleML	Double Machine Learning
EW	Expert Witness
OCR	Optical Character Recognition
PDF	Portable Document Format
PP	Party Proceeding

Table 2: List of activity labels.

Activity Label	**Description**
Assessment documents	Evaluations (e.g., financial, health) submitted

Continued on next page

Table 2: List of activity labels. (Continued)

Activity Label	Description
Notice of ruling objection	Formal objection to a ruling
Power of attorney	Authority given to lawyers or reps
Lawsuit filing	Plaintiff (AP) formally starts the case. Preliminary version may be incomplete
Directive	An administrative or procedural instruction from the court or other authorized body. Is a soft control point, shapes flow but with procedural flexibility
Official letter AP	Generic court communication with applicant party
Official letter PP	Generic court communication with party proceeding
Written statement PP	Legal arguments and case facts submitted
Declaration of settlement AP	Agreement reached without formal judgement. Emphasizes the act of declaring the settlement. It might imply a specific form or procedure that must be followed
Receipt	Court logs receipt of submitted documents
Proof of delivery AP	Confirmation that documents reached their recipient
Written statement AP	Legal arguments and case facts submitted
Individual order	Specific action order unique to this case
Request for medical findings and treatment report	Court asks for medical proof (key in social law cases)
Medical findings and treatment report	Hospital or doctor submits evidence
Attachment other	Supplementary documents added to the file. More specific event that identifies the source or origin of the attachment
Order of evidence collection	Court commands more evidence gathering
Official letter EW	Generic court communication with experts

Continued on next page

Table 2: List of activity labels. (Continued)

Activity Label	Description
Proof of delivery EW	Confirmation that documents reached their recipient
Reminder EW	A notification or prompt issued by the court or authorized body, either in response to a missed deadline or as a proactive measure to guide parties and prevent future delays
EW report	Expert witness submits their opinion
Summons	Orders to appear in court
Summons of EW	Orders to appear in court
Proof of delivery PP	Confirmation that documents reached their recipient
Transcript	Official written record of proceedings
Pronouncement of judgement	Judge announces the decision in court
Court ruling	Decisions made by the court. Refers to any formal decision made by the court during the course of a case. It can address a wide range of issues and court rulings can be issued at any point in the case
Final ruling	Decisions made by the court. The ultimate decision in the case
Attachment AP	Supplementary documents added to the file. More specific event that identifies the source or origin of the attachment
Preliminary lawsuit filing	Plaintiff (AP) formally starts the case. Preliminary version may be incomplete
Declaration of acknowledgement PP	Party proceeding agrees to certain facts
Declaration of acknowledgement AP	Applicant party agress to certain facts

Continued on next page

Table 2: List of activity labels. (Continued)

Activity Label	Description
Reminder	A notification or prompt issued by the court or authorized body, either in response to a missed deadline or as a proactive measure to guide parties and prevent future delays
Settlement declaration PP	Agreement reached without formal judgement. Emphasizes the document or statement itself. The focus is on the content of the declaration
Settlement declaration AP	Agreement reached without formal judgement. Emphasizes the document or statement itself. The focus is on the content of the declaration
Attachment PP	Supplementary documents added to the file. More specific event that identifies the source or origin of the attachment
Official letter	Generic court communication with other parties
Declaration of case closing	Formal closure
Attachment	Supplementary documents added to the file. Does not specify who added it or their role in the case
Attachment EW	Supplementary documents added to the file. More specific event that identifies the source or origin of the attachment
Official invitation to court AP	Scheduling notices for hearings
Official invitation to court PP	Scheduling notices for hearings
Official invitation to court EW	Scheduling notices for hearings
Withdrawal	Case is ended early by the claimant
File inspection	Parties review the case file for accuracy and completeness
Written statement EW	Legal arguments and case facts submitted

Continued on next page

Table 2: List of activity labels. (Continued)

Activity Label	Description
Court order	Instructions guiding the case timeline. A legally enforceable directive issued by the judge that compels a specific action or outcome. Is a hard control point, forces an action or change in process flow
Written statement other	Legal arguments and case facts submitted
Attachment of evidence order	Court commands more evidence gathering
Preliminary written statement PP	Early defense from the opposing party
Official invitation to court other	Scheduling notices for hearings
Preliminary withdrawal	Case is ended early by the claimant
Preliminary power of attorney	Authority given to lawyers or reps
Request for continuation	Party asks to postpone or extend proceedings
Written statement AP via fax	Legal arguments and case facts submitted. Possible early indicator of time pressure / procedural risk due to urgency implied by using fax instead of standard electronic submission
Attachment of expert report	Expert witness submits their opinion
Direct court ruling	Decisions made by the court. A ruling made without extensive preliminary procedures or a full trial to resolve simpler cases or issues quickly and efficiently
Legal brief	Structured written argument from a lawyer
Order of referral	Case is sent to a different court or chamber. This could be due to subject matter expertise, caseload management, or conflict of interest

Table 3: List of excluded activities.

Activity Label	Case Coverage (%)
Attachment of evidence order	0.769
Attachment of expert report	0.385
Direct court ruling	1.154
Legal brief	0.385
Order of referral	0.385
Preliminary power of attorney	0.385
Preliminary written statement PP	0.385
Written statement AP via fax	1.923

B ANALYTICAL FRAMEWORK SPECIFICATIONS

Table 4: Case Features.

Variable Name	Description	Used In
`chamber`	Judicial chamber assigned to the case	Mann–Whitney *U*
`similarity_group`	Combined cluster and year (e.g., "2_-2020")	Mann–Whitney *U*
`chamber_load`	Number of cases assigned to a chamber at the time of case initiation	Causal ablation and causal inference
`cluster`	Cluster label indicating the process pattern group for the case	Causal ablation and causal inference
`count_attachments`	Total number of case attachments	Causal ablation and causal inference
`count_settlement_events`	Number of settlement-related events recorded	Causal ablation and causal inference
`day`	Day of the month when the case was initiated	Causal ablation and causal inference
`early_expert_days`	Days elapsed from case start to the first expert involvement	Causal ablation and causal inference

Continued on next page

Table 4: Case Features. (Continued)

Variable Name	Description	Used In
`expert_complexity`	Number of unique activities if an expert was involved, otherwise zero	Causal ablation and causal inference
`has_expert`	1 if an expert was involved, 0 otherwise	Causal ablation and causal inference
`has_lawsuit_filing`	1 if the case has a lawsuit filing, 0 otherwise	Causal ablation and causal inference
`has_medical_report_received`	1 if a medical report was received, 0 otherwise	Causal ablation and causal inference
`has_medical_request`	1 if a medical report was requested, 0 otherwise	Causal ablation and causal inference
`has_power_of_attorney`	1 if a power of attorney was filed, 0 otherwise	Causal ablation and causal inference
`has_settlement_attempt`	1 if a settlement attempt was made, 0 otherwise	Causal ablation and causal inference
`month`	Month when the case was initiated	Causal ablation and causal inference
`month_sin`	Sine-transformed month value for cyclical encoding	Causal ablation and causal inference
`month_cos`	Cosine-transformed month value for cyclical encoding	Causal ablation and causal inference
`unique_activities`	Number of distinct activities recorded	Causal ablation and causal inference
`year`	Year when the case was initiated	Causal ablation and causal inference
`year_sin`	Sine-transformed year value for cyclical encoding	Causal ablation and causal inference
`year_cos`	Cosine-transformed year value for cyclical encoding	Causal ablation and causal inference

Table 5: Outcome Variables.

Variable Name	**Description**	**Used In**
`duration`	Total case duration (days)	Mann–Whitney *U*, causal ablation, and causal inference
`num_court_activities`	Number of court activities in the case	Mann–Whitney *U*
`num_events`	Total number of events in the case	Mann–Whitney *U*
`num_expert_activities`	Number of expert-related activities in the case[4]	Mann–Whitney *U*, causal ablation, and causal inference
`temporal_span_days`	Mean time between consecutive events within a case	Mann–Whitney *U*, causal ablation, and causal inference
`is_judgement_case`	1 if outcome is judgement, 0 otherwise	Mann–Whitney *U*
`is_settlement_case`	1 if outcome is settlement, 0 otherwise	Mann–Whitney *U*
`num_loops`	Number of consecutive repeated activities in the case progress	Causal ablation and causal inference
`num_reminders`	Number of reminder events issued during the case	Causal ablation and causal inference

[4]Only activities reflecting true expert involvement, not just procedural intent, are included; therefore, requests alone are excluded from this feature.

Table 6: Activity roles classification in delay percolation.

Activity Role	**Key Scoring Metrics**	**Managerial Implications**
Delay Injector (an activity that originates new and often unpredictable delays within the process)	High average outgoing transition time. High **CoV** in outgoing transition times. High total delay contributed by all outgoing transitions (for activities that frequently trigger delays). Penalized for having high average incoming transition times.	**Identifies root causes of inefficiency.** Targeting injectors is crucial for preventing new delays from entering the process, leading to systemic improvements in timeliness and predictability
Delay Transmitter (an activity that passes along existing delays without significantly changing them)	Strong, statistically significant positive **Pearson correlation** between its incoming and outgoing transition delays (the correlation's statistical significance is enforced by de-weighting the score using its p-**value**). Score is weighted by the magnitude of the average transition times it handles, but penalized (reduced to 0) if the correlation is not positive	**Highlights pathways of delay propagation.** While not the source of delays, Transmitters reveal how they travel through the process. Improving handoffs at these points can break the chain

Continued on next page

Table 6: Activity roles classification in delay percolation. (Continued)

Activity Role	Key Scoring Metrics	Managerial Implications
Delay Amplifier (an activity that magnifies the duration or variability of existing delays)	Frequent involvement in process loops (both direct and indirect). A significantly higher median outgoing transition time compared to the median incoming time. High number of occurrences combined with high variability in outgoing transitions.	**Pinpoints critical bottlenecks and rework cycles.** Amplifiers are prime candidates for process re-engineering, automation, or resource allocation to reduce costly escalations and loops
Delay Buffer (an activity that absorbs incoming delays, reducing their downstream impact)	A high ratio of incoming-to-outgoing average delay, combined with low outgoing variability (**CoV**). Strongly rewarded for a statistically significant reduction in variance from incoming to outgoing delays (validated with **Levene's test**). Penalized if it frequently acts as a process start event.	**Shows what works well.** These activities are good at absorbing delays, and they can be studied to find best practices that can be applied to other parts of the workflow.

C ADDITIONAL RESULTS AND FIGURES

Cluster	Cases	Mean Days	Avg Events	Expert Cases	Distinctive Activities (x = times more frequent)	Potential Actions
0	25	288	34	0	Request for continuation (25.967x), Withdrawal (4.269x), Assessment documents (3.847x)	No urgent problem, but track why these cases exit early - are they legitimate settlements, or procedural dropouts? Consider recommender propmpts for parties early in the process to settle if similar patterns are detected.
1	32	713	92	32	Court order (94.916x), Assessment documents (1.476x), Official invitation to court AP (1.399x)	Investigate the specific legal paths (e.g, trial, appeals?) contributing to delay. Process optimization: can court orders or invitations be digitized or made asynchronous?
2	55	587	88	55	Written statement EW (11.379x), Reminder EW (2.175x), Official letter EW (1.748x)	Consider dashboards to flag expert involvement early and track report deadlines.
3	135	496	72	135	Written statement other (1.599x), Settlement declaration AP (1.529x), Medical findings and treatment report (1.524x)	Suggest implementing early medical/expert scheduling tools to reduce waiting periods.

Table 7. Cluster characteristics and insights.

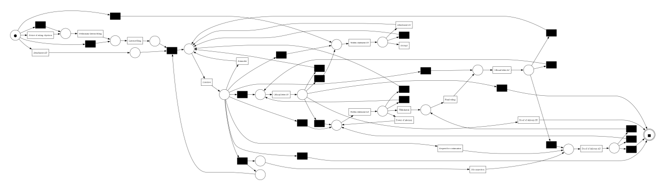

Fig. 4. Petri net for cluster 0.

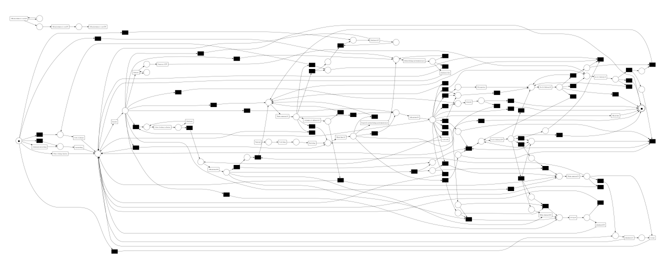

Fig. 5. Petri net for cluster 1.

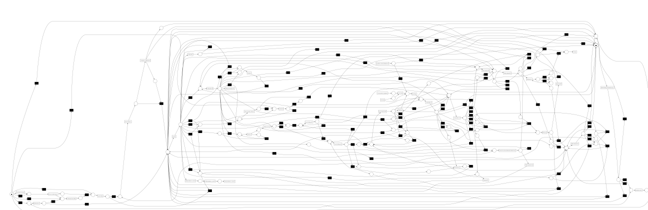

Fig. 6. Petri net for cluster 2.

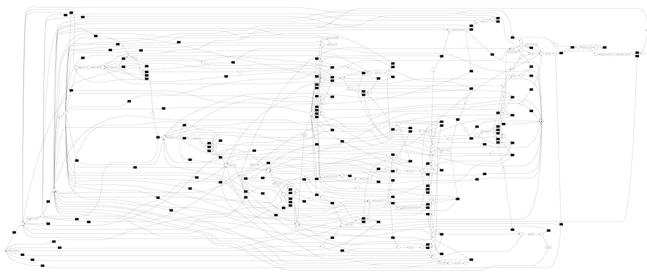

Fig. 7. Petri net for cluster 3.

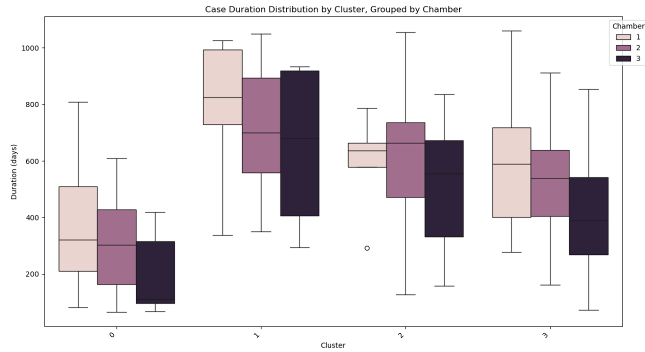

Fig. 10. Case duration distribution by cluster, grouped by chamber.

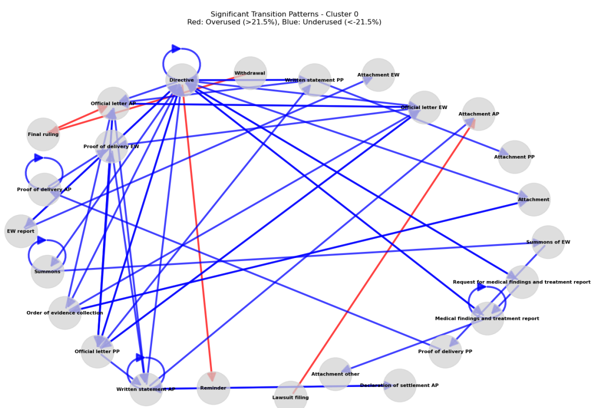

Fig. 8. Significant transition patterns map for Cluster 0.

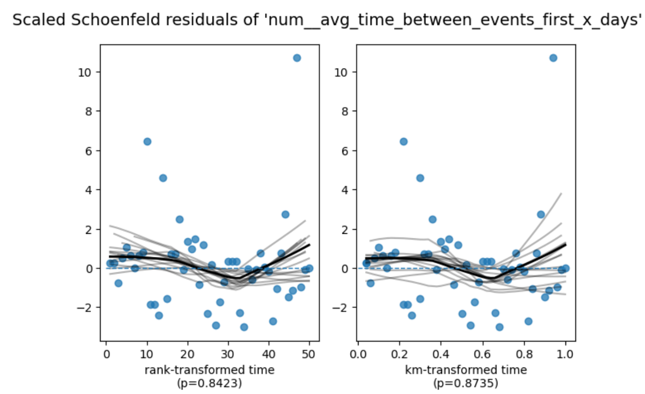

Fig. 11. Scaled Schoenfeld residuals of avg_time_between_events_-first_x_days.

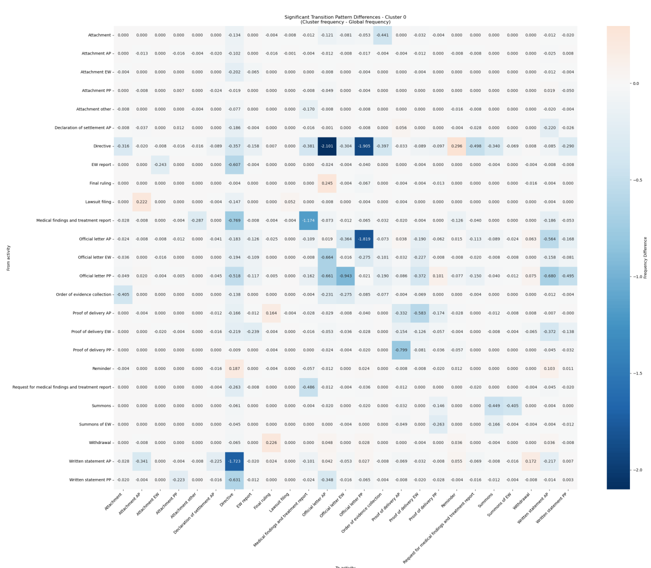

Fig. 9. Significant transition pattern differences for Cluster 0.

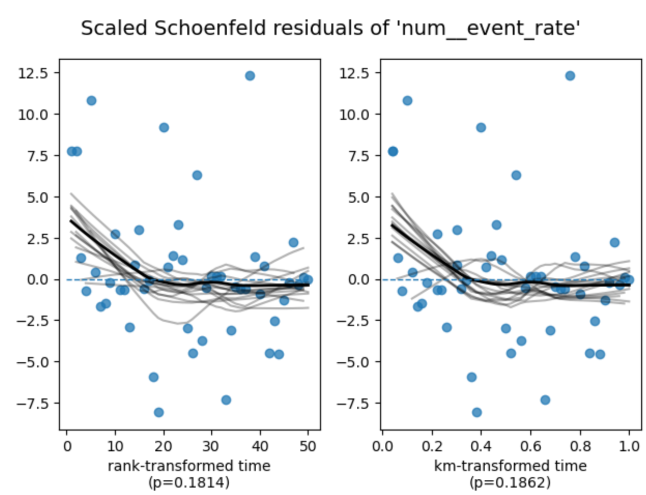

Fig. 12. Scaled Schoenfeld residuals of event_rate.

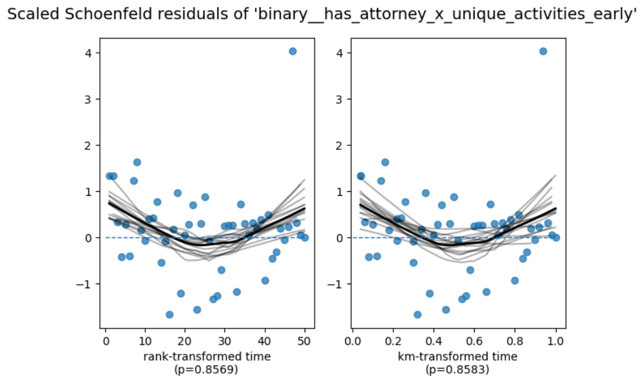

Fig. 13. Scaled Schoenfeld residuals of has_attorney_x_unique_-activities_early.

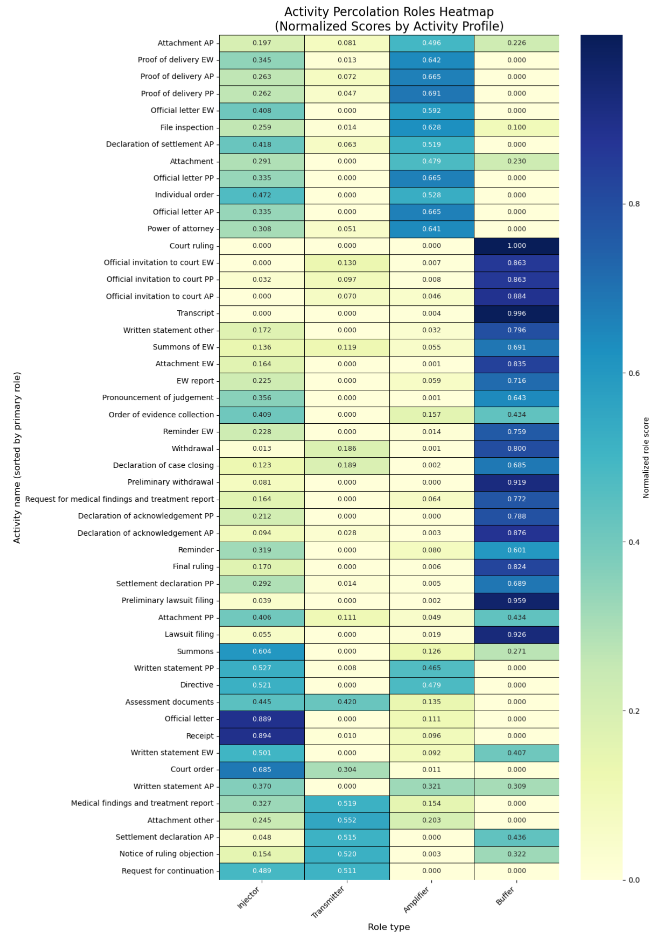

Fig. 15. Activity percolation roles heatmap.

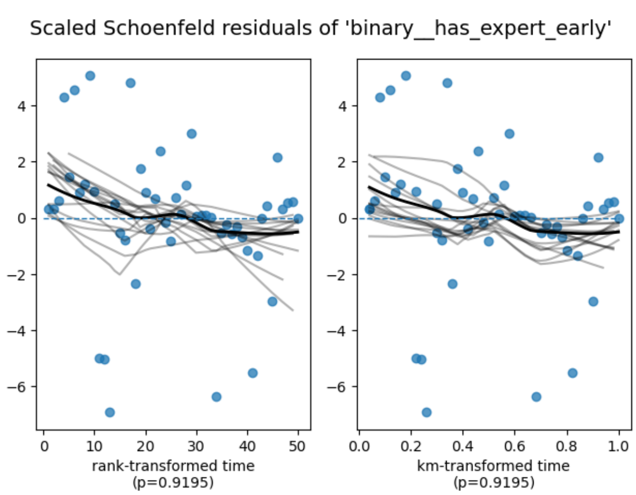

Fig. 14. Scaled Schoenfeld residuals of has_expert_early.

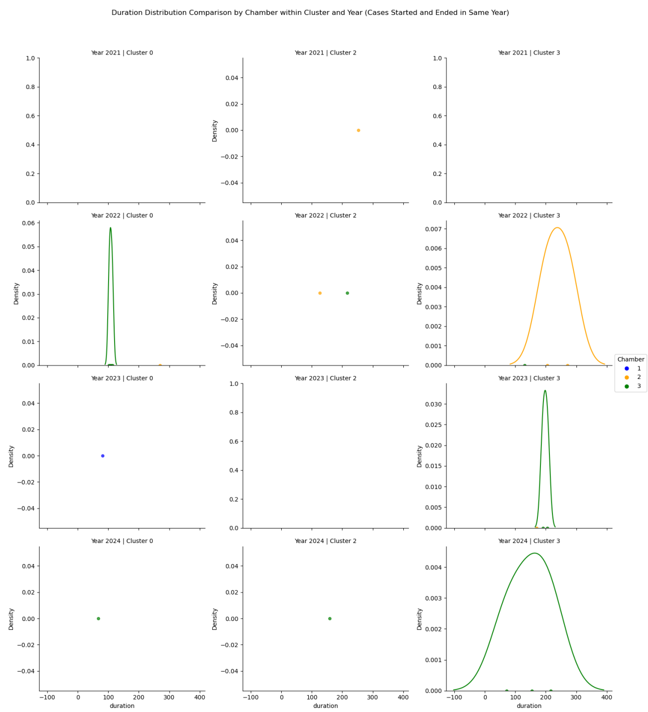

Fig. 16. Duration distribution comparison by chamber within cluster and year (cases started and ended in same year).

D STATEMENT ON THE USE OF AI TOOLS

During my research, I used AI tools such as ChatGPT and Cursor to assist with IEEE citations, identify relevant literature, suggest improvements, and review grammar and academic tone. However, all analysis, interpretation, writing, and reviewing were carried out solely by me, reflecting my own understanding and original contributions. I take full responsibility for the content and conclusions presented in this work.