Detecting Change Points in Data Streams Using Conformal Testing Martingale

Contents

List of Figures

Chapter 1: Introduction

Background and Context

Change-point detection is essential in statistical analysis and monitoring within the vast domain of sciences, including finance, health, quality control, and environmental monitoring. This makes it possible to detect points in time when some statistical properties of a sequence of observations have undergone essential changes. These time points can signal events of decisions, such as changes in market trends and climate, onsets of diseases, or any other changes in a production process. Change point detection methods have traditionally relied heavily on statistical tests such as the Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests (Wang & Hsieh, 2024). Because these approaches compare two data samples to decide whether they are realization samples of one distribution, in most cases, they are efficient applications in scenarios where data can be partitioned into segments before and after a change. Another classical procedure is the Cumulative Sum Control Chart, CUSUM, which is applied to track the cumulative sum of deviations from a target value to pick up any shift in the mean level of a process. While very powerful, these methods have often been limited by requirements to know a split-point and have enormous computational demands for real-time application.

Traditional methods are facing considerable challenges. The increasing need to analyze data in real-time has come with the development of big data. As is common in most applications today, data streams flow at high frequencies and are continuous; change points must be detected without knowing in advance whether a change has occurred. For example, traders need to instantiate asset price shifts in financial markets to make timely decisions (Wang et al., 2023). Continuous monitoring of patient vitals in a healthcare environment will help in the early detection of a critical condition. The requirement of real-time detection in such scenarios calls for more advanced and adaptive methods.

Recent developments in machine learning and statistical techniques make it possible to develop far more advanced change point detection algorithms. A promising avenue of research is the Conformal Testing Martingale algorithm, which is fiduciary to ideas of both conformal prediction and martingale processes, which were based on probability theory and hence formulate a framework flexible enough to make predictive inferences. In conformal prediction, confidence—a measure of confidence—is assigned to predictions, guaranteeing validity under certain conditions. Together with a martingale process—defined in mathematics as a model for a fair game—this will now give a general framework to build up evidence against such a hypothesis of no change in some data stream.

What sets the Conformal Testing Martingale algorithm apart is that it does not require prespecified change points and handles streaming data in real-time (Zhao & Sun, 2024). The algorithm iteratively updates a martingale value concerning the p-values from a stream of arriving data points. Such p-values have been garnered from conformal prediction intervals, which quantify how well each new point adheres to the pattern and fluctuations previously established by past observations. A considerable martingale value can be interpreted as a substantial deviation from the expected distribution and may indicate a change point. This method has demonstrated great potential in preliminary studies on time series analysis, ranging from change-point detection in financial time series to network traffic anomaly monitoring. Little has been done, however, to explore and apply this in general real-world scenarios. This dissertation thus intends to fill this gap by studying, implementing, and evaluating the Conformal Testing Martingale algorithm in different contexts.

The proficiency in the correct and timely identification of change points is very relevant in several domains: finance, where detecting structural breaks in time series data may provide relevant input during the development of trading strategies or risk management practices; health care, where changes in patients' data may lead to early interventions that raise the quality of life of the patient while reducing healthcare costs; and cyber security, by allowing the detection of anomalies at network traffic, thus averting attacks and breaches of information. Other than that, real-time change point detection will be able to set up control within the manufacturing quality control process so products are uniform and waste is reduced. In environmental monitoring systems, change-point detection might assist in several changes from climatic patterns onward in developing responsive strategies to mitigate the effects.

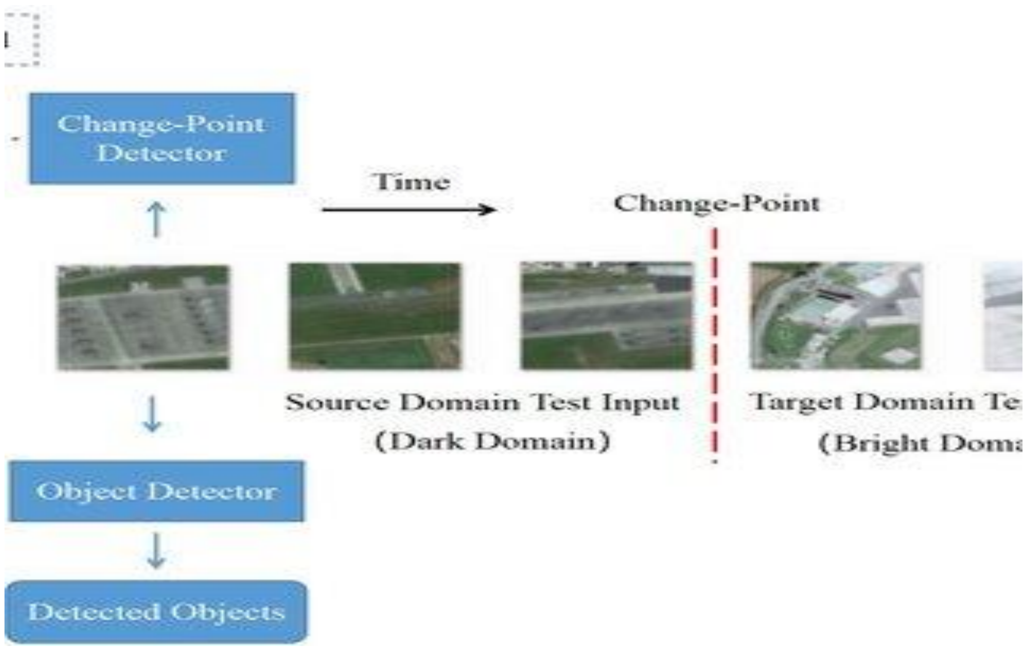


Figure 1 Visualization of Mean Delay for all algorithms with conformal test

Change point detection is one of the most important tools in many fields of applications that help recognize essential changes in data patterns in real-time. Traditional methods may be suitable for some things, though they have proven inefficient for real-time purposes. In this respect, the Conformal Testing Martingale algorithm can provide a solution by giving a robust framework to detect changes in streaming real-time data. Therefore, This dissertation explores and evaluates this algorithm to advance change point detection and illustrate real-world applications in the field.

Problem Statement

Classic change-point detection techniques involve the Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests, among others. These have been applied in statistical analysis to show changes in data distributions. While these techniques are powerful in a controlled environment, they suffer massive limitations in real-time data applications (Mathur, 2021). For example, most of them require that the splitting point of the data be known in advance, which rules out many applications where the need for change-point detection should be realized on the fly as data is generated. Also, most techniques herein are computationally time-consuming; hence, they perform poorly on high-frequency data streams where timely detection is crucial. Modern data environments are characterized by constantly streaming high-volume data that requires efficient methods operated in real-time. For instance, finance traders should be able to detect real-time asset price changes to respond to market movements that minimize risks. Continuous monitoring of patient vitals is essential in healthcare, allowing the early detection of life-threatening conditions. Such data travel under high dynamism and high-velocity rates; traditional methods of handling are incapable, hence requiring the invention of more adaptive and efficient algorithms.

The Conformal Testing Martingale algorithm probably holds the most promise of these challenges. It merges two techniques, conformal prediction and martine processes, providing a very robust system for change-point detection in real-time data streams. In this approach, some martingale values will get iteratively updated based on the p-values of incoming data points. This technique thus can turn into a mechanism to detect deviations from an expected distribution without previous knowledge of the change points. Although this algorithm has many theoretical advantages and promising preliminary results, relatively little is known about how it functions in practice and how good it can be in real-world scenarios (Valle, Izbicki & Leite, 2023). Hence, this dissertation aims to fill the gap in existing research by conducting an in-depth study of the Conformal Testing Martingale algorithm. The specific research question to be answered in the study is how effective the algorithm is in detecting change points in various data categories and with the benchmarking the traditional method of CUSUM, Pelt, and Bayesian Change Point Detection. It will evaluate some few standard indicators as the percentage score in the accuracy of detections, the number of false alarms and computational complexity in order to see its potential in real time application.

The study will explore the logistical difficulties and impossibilities of generating the Conformal Testing Martingale algorithm in a financial market, continuous healthcare monitoring, and safeguarding cybersecurity. Such use makes this work conduct a detailed but minuscule analysis of the merits and demerits of this method, explaining the basic ideas of this algorithm and how it could be improved and extended to a wider register. Accordingly, the findings of this investigation will be significant to add value to the knowledge of change point detection, open up a new channel for the investigation of change point detection and acting as a foundation for future research and development to extend in this important area.

Aim/s and Objectives

The primary aim of this dissertation is to study, implement, and evaluate the Conformal Testing Martingale algorithm for change point detection in real-time data streams. To achieve this aim, the dissertation sets out the following specific objectives:

- Literature Review: Conduct a comprehensive review of existing change point detection methods, including traditional statistical tests and recent advancements in conformal prediction and martingale processes.
- Algorithm Study: Study the theoretical foundation of the Conformal Testing Martingale algorithm, including its mathematical derivations and practical implications.
- Implementation: Implement the Conformal Testing Martingale algorithm in a programming language (e.g., Python), ensuring it can handle real-time data streams effectively.
- Application to Real Datasets: Apply the algorithm to multiple real-world datasets, such as USPS and abdominal pain data, to evaluate its performance in detecting change points.
- Comparative Analysis: Compare the performance of the Conformal Testing Martingale algorithm with traditional methods like CUSUM, Pelt, and Bayesian Change Point Detection in terms of accuracy, false alarm rates, and computational efficiency.
- Extensions and Applications: Explore potential extensions of the algorithm, such as unique versions for causality analysis and hybrid approaches combining the algorithm with other methods. Discuss practical applications in fields like finance, healthcare, and cybersecurity.

- Evaluation and Discussion: Evaluate the results of the application and comparative analysis, discussing the strengths, weaknesses, and practical implications of the Conformal Testing Martingale algorithm.
- Recommendations and Future Research: Provide recommendations for improving the algorithm and suggest directions for future research to address the identified challenges and limitations.

Research Questions

To guide the investigation, the following research questions are formulated:

- Compared to traditional methods, how effective is the Conformal Testing Martingale algorithm in detecting change points in real-time data streams?
- What are the algorithm's accuracy, false alarm rates, and computational efficiency?
- How does it perform on different types of datasets?
- Can the Conformal Testing Martingale algorithm be applied effectively to real-world datasets such as USPS and abdominal pain data?
- What are the practical challenges and limitations encountered during implementation?

Relevance and Importance of the Research

Research in change point detection, more so using the Conformal Testing Martingale algorithm, becomes relevant and essential in today's data-driven environment. Change point detection assumes a critical role within several domains, and timely identification of shifts in data patterns is necessary. For instance, in financial markets, breaks in asset prices would mean a lot concerning trading strategies and risk management (Ho & Kairamkonda, 2024). Therefore, Such changes are detectable early enough to allow investors to proactively adjust their portfolios to capture maximum returns while minimizing risks (Shelar et al., 2020). The ability to do the same is essential today when market conditions change quickly; split-second decisions may have hefty financial ramifications.

When considering continuous patient data monitoring, the application in health care must be brought into the limelight. It might mean that such early warning systems will require a constant reaffirmation of danger against critical health conditions. The change in the vital signs may raise immediate medical response, help save life, and thus improve the outcome for a patient. Due to the Conformal Testing Martingale algorithm, the real-time processing attributes privilege it as an ideal candidate for such applications. It provides timely alerts to health professionals about significant changes and responds quickly, thus providing better healthcare to patients. Further, this algorithm's strength for handling continuous streams of digital data ensures this makes a difference when integrated into any monitoring system—from wearable devices to hospital systems.

Another such domain is cybersecurity. Network traffic change point detection allows early detection of any possible cyber-attack or other data breach so that organizations' 'Countermeasures' can be applied before severe damage is caused. Traditional methods usually fail due to network data's huge volumes and velocities (Ho et al., 2019). Nevertheless, the Conformal Testing Martingale algorithm does provide a way toward scaling for real-time effectiveness. Subsequently, it may enhance security infrastructure around organizations by detecting deviations from the usual traffic pattern and protecting information of high sensitivity from leakages or any other breeds of exposure that could result in loss of integrity for IT systems.

This research's second contribution towards developing the Conformal Testing Martingale algorithm is associated with a critical lacuna in the available literature. According to the study, little emphasis has been given to the algorithm's applications in real-life scenarios and performance (Semerci, Cemgil & Sankur, 2018). The research intends to fill this gap by thoroughly studying the algorithm's strengths and limitations across different contexts. These findings will add to the academic body of knowledge and be relevant for contemporaries in other fields where real-time data analysis is used.

Chapter 2: Literature Review

Change point detection is a well-established area of research with applications across diverse fields such as finance, healthcare, environmental science, and cybersecurity. This chapter comprehensively reviews the literature on change point detection, including traditional statistical methods, recent advancements in conformal prediction, and the Conformal Testing Martingale algorithm. The review highlights the strengths and limitations of existing processes and sets the context for the proposed research.

Traditional Methods for Change Point Detection

According to Riina et al. (2023), in "Continuous Variable Analyses: t-test, Mann–Whitney, Wilcoxon Rank", the Mann-Whitney-Wilcoxon test is a nonparametric statistical test that hypothesizes whether or not there is a difference between two independent samples. In contrast to parametric tests such as the t-test, it does not assume that the data is usually distributed and, therefore, can be used with ordinal data or interval data that is non-normally distributed. This test ranked all values from both samples and compared the rank sums in return. This provides a robust way to detect differences in central tendency. Its reliance upon predefined change points, however, is a limit to its utility in performing real-time data analysis with unknown change points. This thus limits the need to have advanced methods like the Conformal Testing Martingale algorithm for dynamic, real-time applications.

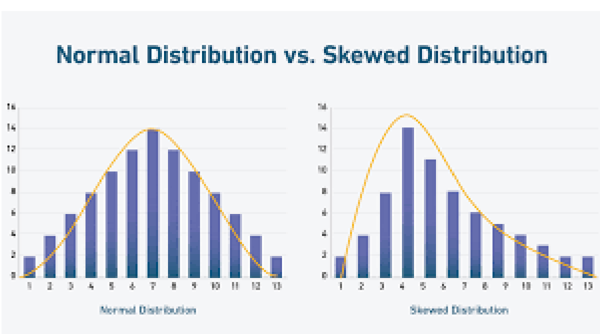


Figure 2 Mann-Whitney U Test (Riina et al., 2023)

According to Wang and Wang, 2020, the Kolmogorov-Smirnov test is a nonparametric test for either two samples or a sample against a reference distribution. This is the maximum distance between the cumulative distribution of samples; thus, it will effectively individuate differences in the shape of distributions. In concept drift detection, the K-S test enables the detection of essential changes in data streams. Hence, it provides a warning that updates or retraining of a model would be necessary. The computational demands of the KS Test reduce their effectiveness within real-time applications. Therefore, adaptive and more efficient methods are required to continuously deal with high-frequency data streams. One such method is using the Conformal Testing Martingale algorithm.

The Cumulative Sum Control Chart is a statistical method to monitor mean level changes in a process. According to Novoa and Varela, 2020, in Mediastinum, "It accumulates the sum of deviations from a target value, allowing it to detect small shifts that might go undetected by other control charts.". Specifically, CUSUM finds its best application in quality control, like surgical performance monitoring, since it makes early identification of the deviation from the expected outcome possible, thus enabling intervention to maintain standards. Effective per se, CUSUM has reduced flexibility because predefined, usually arbitrarily set control limits and sensitivity to parameter settings are required to function in real-time, high-frequency data streams. Therefore, methods that are much more adaptive, like the Conformal testing Martingale algorithm, which would not use predefined parameters yet process the continuous data with much efficiency, would be required to monitor risk in such adaptive environments, giving a robust solution for real-time monitoring.

The Pruned Exact Linear Time (PELT) algorithm, explored by Elbakri et al. (2024) in Computers, Materials, and Continua, is an efficient multiple change point detection method.

PELT utilizes dynamic programming to minimize a cost function that balances the model's fit to the data with a penalty for adding change points. It will allow PELT to identify multiple changes in one dataset with high accuracy and computational efficiency so that the approach can bear fruition in cases requiring extensive dataset analysis. An adaptive cloud intrusion detection system will allow for fast recognition of abnormal patterns or shifts of network traffic to take swift responses if there are potential security threats. Efficient as PELT is, it still requires considerable computational resources and is not intrinsically designed for data processing in real-time. Again, this limitation further underlines that methods like the Conformal Testing Martingale algorithm must be followed up, which can process the continued data streams in real-time with adaptive on-the-fly detection facilities.

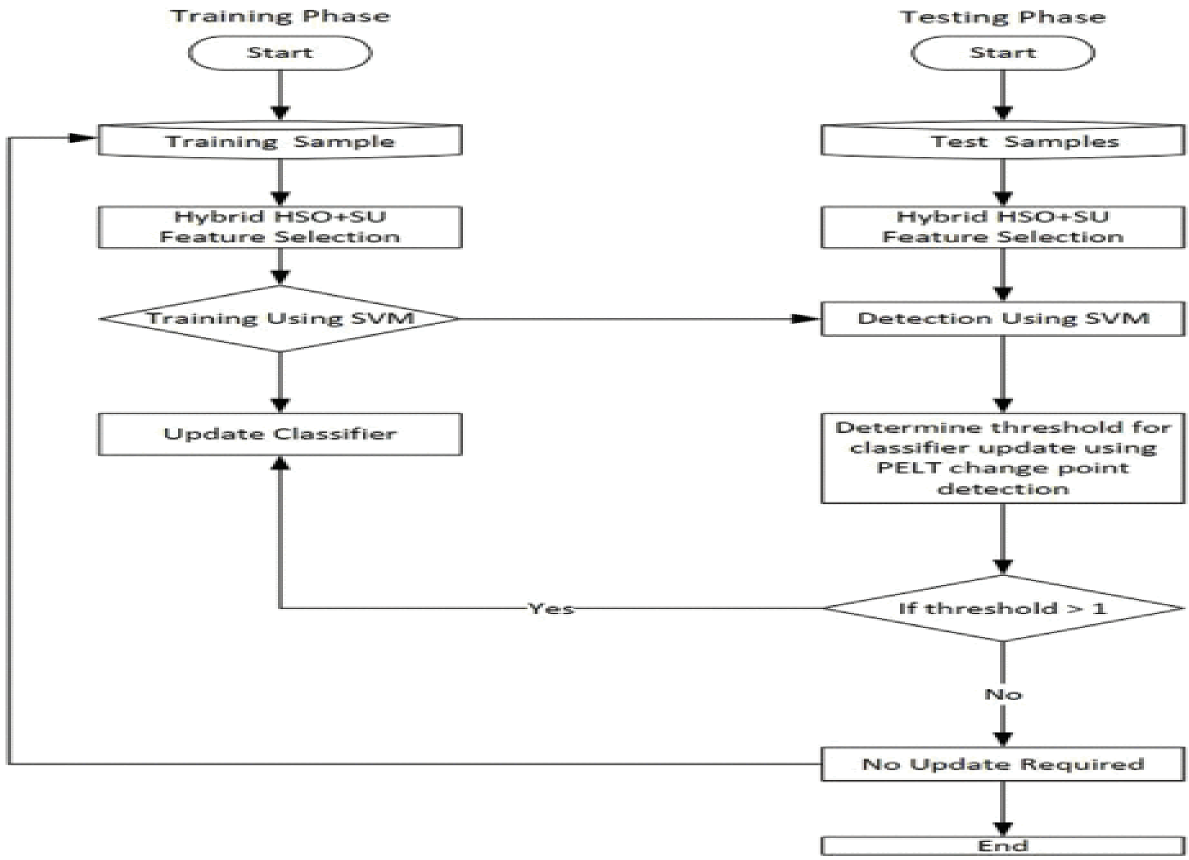


Figure 3 Adaptive Cloud Intrusion Detection System Based on Pruned Exact Linear Time Technique (Zhao & Sun, 2024)

As described in detail by Prabpon et al. 2024, Bayesian Change Point Detection uses a probabilistic framework for locating the change points at which there is variation in the statistical properties of the data series. This approach obtains the posterior probability of a change point at every step from prior knowledge, updated as new data arrives. It was used to establish a kernel density estimation nonparametric Bayesian approach accompanied by a nonparametric hazard function that provides flexibility and accuracy in detecting change points without assuming a particular distribution. This is a compelling methodology for treating a wide diversity of complex data patterns, guaranteeing detection robustness in general. One of the critical shortcomings of Bayesian methods is computational intensiveness. This makes it rather tricky for real-time data processing since the updating and recalculation end جمعيت probabilities endlessly. Again, this underlines the importance of more computationally efficient methods, like the Conformal Testing Martingale algorithm, which gives real-time detection capacity with lower computational overhead.

Advances in Conformal Prediction

Conformal prediction is a statistical method that creates valid prediction intervals at a specified probability by ensuring that the real value will lie for new points. According to Angelopoulos and Bates, 2023, in Conformal Prediction: A Gentle Introduction, the above approach thus offers both flexibility and distribution-freeness; hence, it finds wide applications in different domains. Offsetting a little this specificity, conformal prediction works by using past data to estimate nonconformity in new points and thus constructs adaptive prediction intervals.

Conformal prediction has seen several recent advances that have greatly improved its applicability and robustness. In their paper on conformal test martingale-based change point detection for geospatial object detectors, Wang et al. (2023) extend the formulation of conformal prediction to accommodate change-point detection in real-time data streams. Conformal prediction will obtain nonconformity scores for incoming data points, while a martingale process accumulates evidence against no change. The resultant algorithm from such integration can efficiently detect deviations from the expected distribution and is hence suitable for applications requiring immediate response to changes, such as geospatial object detection and perhaps network traffic monitoring.

These developments help bypass two substantial limitations of more conventional change point detection methods. Being a more dynamic and responsive approach than the preprocessing methods based on predefined change points, the presented approaches to conformal prediction-based techniques are significantly more adequate in streaming data. Another strength of the approach provided by conformal prediction is its flexibility:this means It increases predictive power and usability in complex natural settings. Moreover, this is extended further to versions specialized on some topics, such as in the methods stature that includes nonparametric ways or hybrid models. These numerous innovations increase the algorithm's precision and low false alarm rates, making computationally-computational efficiency very high and making conformal valuable predictions for academic research and practical applications.

The Conformal Testing Martingale Algorithm

The Conformal Testing Martingale algorithm combines the power of conformal prediction and martingale processes in real-time change-point detection on streaming data. Martingales are a mathematical model representing fair games in that past values cannot predict their future values (Vovk, Gammerman & Shafer, 2022). In this martingale property framework, one can build up evidence against the null hypothesis of no change and thus trigger alerts upon detecting any. It iteratively updates a martingale value based on the p-values of incoming data points. The p-values are derived from the nonconformity scores computed using conformal prediction intervals. If it crosses a predefined threshold, the martingale value signals a significant deviation from the expected distribution, hence a change point.

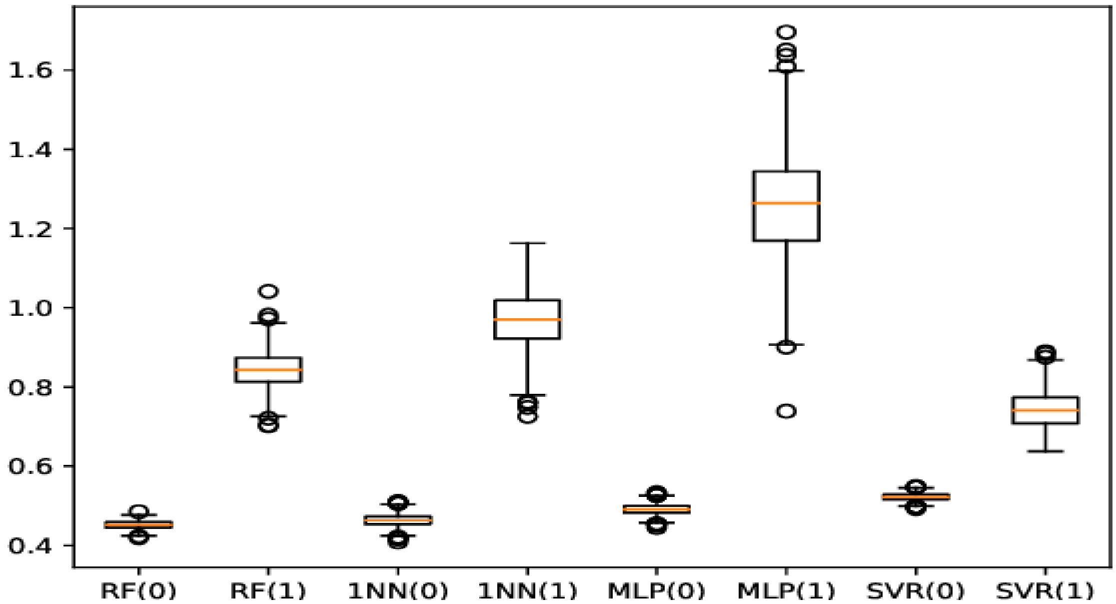


Figure 4 Conformal test martingales for change-point detection (Vovk, Gammerman & Shafer, 2022)

1. **Initial Setup**: The algorithm starts with an initial martingale value of 1 and calculates the nonconformity score for each incoming data point.
2. **Updating Martingale**: For each new data point, the algorithm calculates a p-value based on the nonconformity score and updates the martingale value as

   $Mt=M_{t-1} \cdot \epsilon/1-pt$

   Where:

- MtM_tMt is the current martingale value.
- Mt−1M_{t-1}Mt−1 is the previous martingale value.
- ptp_tpt is the p-value at time ttt.
- ϵ\epsilonϵ is a predefined significance level.

3. **Change Point Detection**: If the martingale value exceeds a certain threshold, the algorithm flags a change point, indicating a significant deviation from the expected distribution.

In this regard, the performance of the Conformal Testing Martingale algorithm can be assessed using different metrics that relate to detection accuracy, false alarm rates, and computational efficiency. Detection accuracy tries to answer whether an algorithm can detect change points correctly, whereas the false alarm rate tries to determine how often it raises incorrect alerts. Computational efficiency determines if an algorithm would be appropriate for real-time applications by considering factors such as processing speed and resource usage (Abohamama, El-Ghamry & Hamouda, 2022). Preliminary studies on the performance of the Conformal Testing Martingale algorithm on changepoint detection in financial time series and network traffic data are very promising. The algorithm's strength was proven by testing its resilience and flexibility across different scenarios, enabling it to be used in real-time scenarios. However, further research is needed to validate these findings and explore the algorithm's performance in other domains.

Comparative Analysis of Change Point Detection Methods

The strengths and weaknesses of each change point detection method are based on application and data characteristics. Classic techniques, like the Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov tests, are relatively easy to apply and straightforward in their design but require predefined change points. CUSUM works well for small shifts in the mean process; however, this is sensitive to parameter settings. PELT and Bayesian methods provide a user with flexibility and probabilistic interpretation but are computationally time-intensive. The Conformal

Testing Martingale algorithm addresses many of these limitations by providing real-time detection without the need to know the points of change in advance and by adapting to streaming data (Williamson, 2022). Applications of methods for change point detection in practice are extensive and range across many domains. In finance, change point detection assists in regime shift identification from the market data and informs trading strategies and risk management practices. It helps monitor patient vitals and detect the early signs of various medical conditions. In cybersecurity, change-point detection is applied to identify abnormal behavior in network traffic to prevent cyber attacks and data breaches. The adaptiveness and real-time capabilities due to Martingale make this algorithm specially tailored to applications such as beneath.

Challenges and Future Directions

Despite the advancements in change point detection methods, several challenges remain. One major challenge is the trade-off between detection accuracy and false alarm rates (Dutta et al., 2020). High detection accuracy often comes at the cost of increased false alarms, which can be disruptive in real-world applications. Another challenge is the computational complexity of some methods, which limits their applicability in high-frequency, real-time data streams. Additionally, the effectiveness of change point detection methods can be influenced by the nature of the data, such as noise and non-stationarity.

Future research in change point detection should focus on addressing these challenges and improving the robustness and efficiency of detection algorithms. Potential directions include:

1. **Hybrid Approaches**: Combining multiple change point detection methods to leverage their respective strengths and mitigate weaknesses. For example, integrating the Conformal Testing Martingale algorithm with traditional methods like CUSUM could enhance detection performance.

2. **Parameter Optimization**: Developing techniques to automatically select optimal parameters for change point detection algorithms, improving their adaptability to different datasets and conditions.

3. **Scalability**: Enhancing the scalability of change point detection algorithms to handle large-scale, high-frequency data streams. This includes optimizing computational efficiency and developing distributed processing techniques.

4. **Real-Time Applications**: Expanding the application of change point detection algorithms to new domains and real-time scenarios. This involves integrating detection algorithms into existing monitoring systems and evaluating their performance in practical settings.

5. **Machine Learning Integration**: Exploring the integration of machine learning techniques with change point detection algorithms to improve their predictive capabilities and adaptability. Machine learning models can provide additional insights and enhance the overall detection process.

6. **Causality Analysis**: Developing specialized versions of change point detection algorithms to explore causal relationships between features and labels. This can provide deeper insights into the underlying factors driving changes in data patterns.

Chapter 3: Methodology

Datasets

The datasets used for this study were obtained from the Numenta Anomaly Benchmark and were TravelTime_387.csv and TravelTime_451.csv. These two datasets can be applied in analyzing traffic data, which has a somewhat relevant anomaly detection, making it possible to show unusual traffic conditions or problems within a network. The description of what these two datasets are and why this research chose to use them is explained in this section. These datasets are taken from the NAB repository—one of the most used stochastic processes benchmarking anomaly detection algorithms. The NAB repository has a list of real-world datasets giving several insights into different types of anomalousness. TravelTime_387.csv and TravelTime_451.csv have traffic data with timestamps and respective travel times.

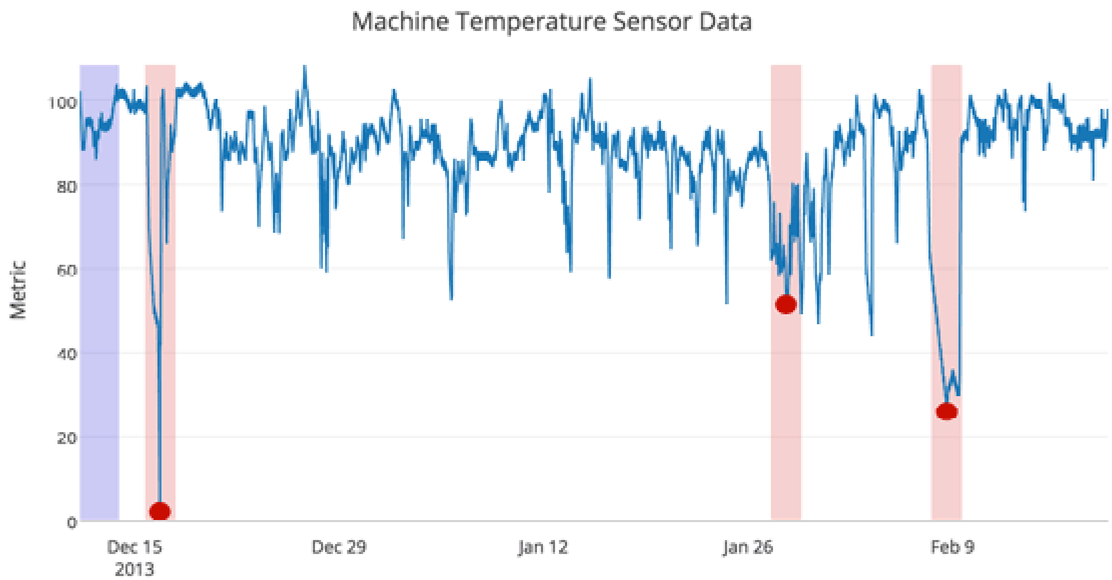


Figure 5 NAB Dataset (Zhao & Sun, 2024)

This TravelTime_387.csv dataset contains 2,500 lines of data. Every line describes information regarding a timestamp and its corresponding travel time value taken at an interval. The dataset goes from July 10, 2015, to September 17, 2015. Thus, it is continuously completed and gives travel times during this period. Numerical data describe this dataset, which is essential in performing any quantitative analysis. The values of travel time range from a minimum of 9 to a maximum of 5,059, with the standard deviation being 399, indicating some variation in the recorded travel times. The richness and variety of this dataset will be suitable for detecting any temporal patterns and anomalies that might occur due to unusual traffic conditions.

The dataset TravelTime_451.csv contains 2,162 rows, whereby each row represents data with a timestamp and a value of travel time. Again, this dataset ranges from July 28, 2015, to September 17, 2015. For this dataset, the travel times range from 22 to 5,578, with a standard deviation of 444, which indicates some variation in recorded travel times. It provides another set of records on travel time, unique yet complementing the dataset to give another view of traffic conditions for that period.

Such choices can be justified against the backdrop of factors that best fit the objectives of this study. In this case, these include how relevant the datasets are to the research topic, the quality and richness of data contained, and their utility in assessing anomaly detection algorithms. The main goal of this study is to establish whether anomaly detection algorithms can recognize unusual patterns in traffic data. These datasets are directly relevant to this objective since they contain accurate traffic data, doomed to be analyzed in search of anomalies. There is time series data with varying travel times; therefore, detailed research into traffic patterns is possible. Thus, these datasets would be appropriate for testing and validation of the performance of anomaly detection techniques.

These datasets are well-ordained, containing data that will exceed serious analyses' requirements. Precise timestamps and travel time values are included; data can be processed and analyzed using statistical and machine learning techniques. This richness of the datasets has been seen on several fronts through the range and variability of the travel times, enriching the robustness of the analysis and providing ample opportunities for the identification of meaningful anomalies.

These datasets are suitably diverse and complex for evaluation, given the intrinsic variability that makes them fitting benchmarks in evaluating anomaly detection algorithms. Different traffic conditions, normal to probably anomalous situations, allow the algorithms to be tested under various scenarios. The diversity is vital in assessing its generality toward anomaly detection across other categories.

Moreover, all the datasets used in this research are available in the NAB repository, with touchdowns of reproducibility and comparability. The datasets are open to any researcher who can validate the findings and use the same dataset to analyze his study, hence adding knowledge to the anomaly detection techniques by the broader research community. In summary, TravelTime_387.csv and TravelTime_451.csv from the NAB repository would be appropriate for this study's objectives. Their relevance to traffic anomaly detection and their quality and richness make them a perfect foundation for testing anomaly detection algorithms. This research thus looks forward to making informed contributions regarding how practical different approaches are in establishing unusual patterns within the context of traffic data.

Conformal Testing Martingale Algorithm

Conformal testing martingale is a statistical algorithm for detecting change points in time series data. Such series data provides an effective way for anomaly detection or change points in data patterns, thus fitting many applications—traffic analysis, financial market prediction, and network monitoring.

Conformal Testing, the Martingale Algorithm, is developed mainly to detect change points in a data sequence. It utilizes the mentioned entities: nonconformity scores and martingale processes. The algorithm calculates, for every data point, how unusual or unexpected a data point is concerning past observations. From these scores, p-values are computed, giving the probability of seeing a data point as extreme or more extreme than the current point when no change occurs.

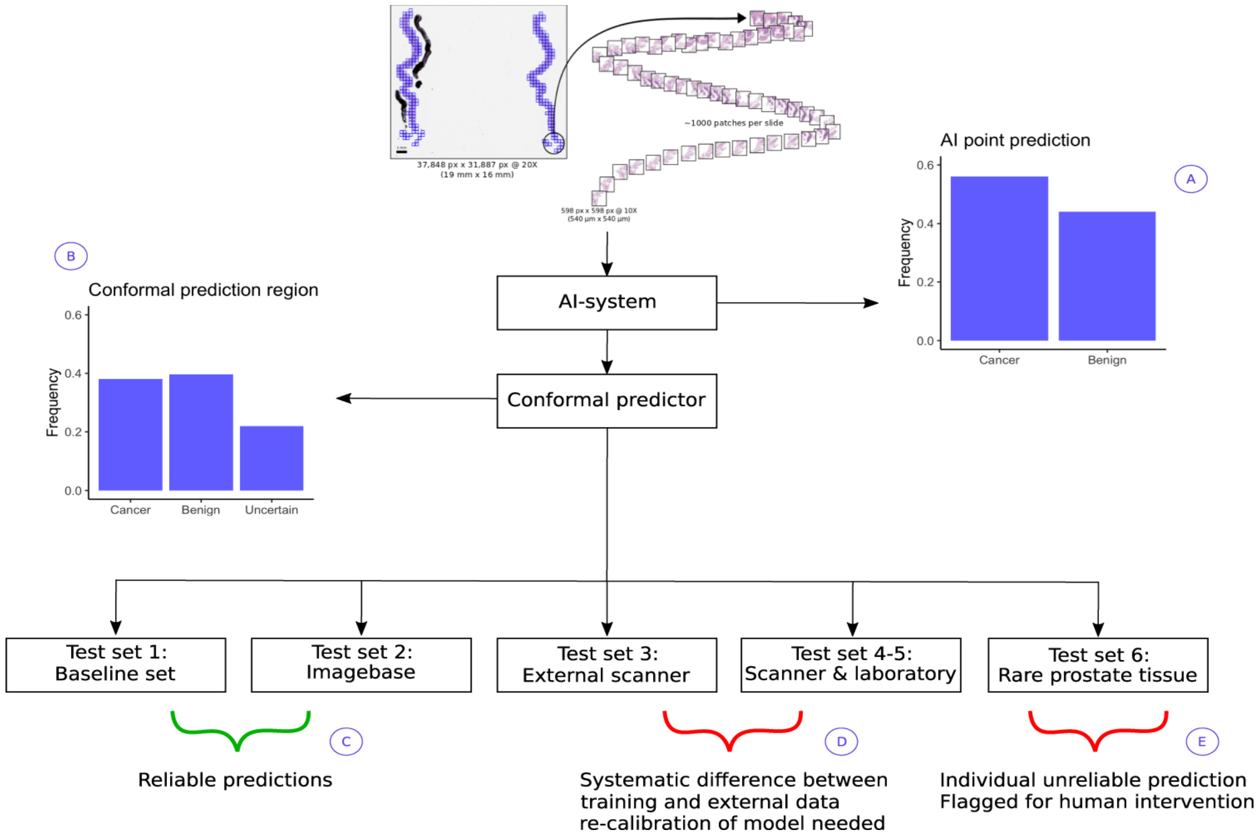


Figure 6 Estimating diagnostic uncertainty in artificial intelligence-assisted pathology using conformal (Angelopoulos & Bates, 2023)

The basic idea will be to view the data as a sequence of random variables, using the martingale property: the expected value of a process at a future time is equal to the present value conditional upon any past information. This property will allow us to exploit a building of a martingale process from the p-values so that evidence regarding change will be accumulated over time. For large enough martingale processes, this means considerable deviations from what is expected; thus, there may be a change point.

The mathematical basis for the Conformal Testing Martingale Algorithm follows conformal prediction—a framework that provides various methods for quantifying the uncertainty of forecasts. To this end, the algorithm executes the nonconformity score for each point in the data.

These are typically calculated as an absolute deviation of a data point from the median of previous observations, expressed as:

at=|xt−median(x1,x2,…,xt−1)|

The next step is to calculate the p-values of the data points based on the nonconformity scores. The p-value for a data point is the proportion of all previous scores that are greater than or equal to the current score:

pt =

$$p_t = \frac{\sum_{i=1}^{t-1} \mathbb{I}(a_i \geq a_t) + 1}{t}$$

where $\mathbb{I}$ is the indicator function, which equals 1 if the condition $a_i \geq a_t$ is satisfied and 0 otherwise.

, 1 if the condition is satisfied, 0 otherwise.

The p-values are subsequently utilized in the construction process of a martingale process. The p-values are fed into the betting strategy for computing these martingale values. One common choice is to use the logarithmic betting strategy:

$$M_t = M_{t-1} \cdot \left(\epsilon \cdot p_t^{\epsilon-1}\right)$$

where M_t is the martingale value at time t, p_t is the p-value, and ϵ is a predefined significance level.

An implementation of the Conformal Testing Martingale Algorithm goes through several steps briefed concerning each stage of the algorithm. First, compute nonconformity scores. This will involve processing the data to calculate the deviation from the median for every point. Calculate the p-values again, about the scores, through iteration and comparing the scores against one another for their extremity.

The core part of the implementation is the calculation of the martingale values. How the martingale evolves from an initial value at each step is according to the logarithmic betting strategy. It implies that a given martingale value at any point in time must be multiplied with a fraction derived from the p-value and the parameter ϵ. To avoid throwing divisions with zero errors, care must be taken with zero p-values.

Ultimately, in this paper, the change points are to be studied from a Martingale process. More precisely, a threshold has to be set in advance, meaning that once the value of the Martingale exceeds it, it raises a warning regarding a probable change point. Thus, the change points detected could make up an algorithm output that conveys critical insights into the temporal dynamics across any data.

The approach this martingale algorithm follows is one that not only integrates statistical theory with its practical implementation in an attempt to detect changes in time series data, which will reveal its capacity for detecting change-points (Jia & Zhou, 2022). Being supplied with the robust mathematical basis of both conformal prediction and martingale theory, this approach frames a firm framework to analyze data for change or anomalies; hence, it becomes a vital tool in detecting change points within many realms.

EVALUATION

What to compare

Comparative methods in research are instrumental in studying complex phenomena by examining and comparing independent data sets, algorithms, or different approaches. Among the many ways these methods are helpful, two include projects that involve time series analysis, such as the one centered on the Conformal Testing Martingale Algorithm and its application in the anomaly detection of traffic datasets (Ho et al., 2019). For example, comparative approaches make it possible to establish the effectiveness and strength of different algorithms against particular criteria while devising different types of patterns in the algorithm and making meaningful conclusions from various data sources.

Comparative methods are based on systematically comparing at least two units or datasets to study similarity, difference, and patterning. The other aspect is that big data is beneficial for establishing trends, which one would not see if one were to compare against a dataset. Comparative methods shall be applied to test the Conformal Testing Martingale Algorithm and other change-point detection algorithms, including Cumulative Sum and Bayesian Online change-point detection. This project will look to detect anomalies in traffic data using the conformal testing Martingale algorithm. Comparative methods are used to check the performance against other state-of-the-art techniques. The algorithms could be compared for findings against the most effective way of giving points of change from the traffic data for high accuracy and reliability results.

The initial step in applying comparative methods involves deciding on algorithms to compare. This project will compare using CUSUM, BOCPD, and the Martingale Algorithm from

Conformal Testing (Shelar et al., 2020). Each algorithm has its merits and demerits; thus, it will be fair to determine which algorithm performs best in specific conditions that characterize traffic datasets. CUSUM: This is a viral algorithm for detecting abrupt changes in data sequences. The algorithm works based on the cumulative sum of deviations from a target value. It signals change if this sum crosses a threshold. It is straightforward and efficient but may struggle with gradual changes.

BOCPD—Bayesian Online Change Point Detection: A probabilistic approach towards change-point detection in data. It models the probability distribution over the possible change points and updates them with the arrival of more data. The BOCPP is very useful in handling noisy and uncertain data, but at the same time, it might be computationally costly. Conformal Testing Martingale Algorithm This algorithm merges conformal prediction and martingale processes to detect change points. It is unique for applying nonconformity scores and p-values in assessing the likelihood of change, hence offering an entirely new approach that challenges sensitivity versus specificity.

The source datasets come from the Numenta Anomaly Benchmark, specifically the traffic datasets that capture travel times and vehicle speeds. These datasets were chosen because of their real-world relevance and complexity; they can provide a sufficient challenge to the algorithms being tested. Each dataset consists of timestamped records, making constructing temporal dynamics and change-point analysis possible. This justified the dataset selection by comparative methods that would let the study of how each algorithm will perform under different conditions of data, noise level, frequency of anomalies, and data volume. In this way, the project will be able to compare the performance of algorithms on these datasets to determine which method works in real-world applications.

How to compare

Comparative methods must be based on a rigorous evaluation framework. Quantifying the performance of an algorithm uses measures such as precision, recall, and F1-score. The precision shows the proportion of correctly detected change points. Recall tells whether an algorithm is capable of detecting all change points. The F1 score acts as a harmonic mean to balance these two metrics to get one overall performance measure. A comparative analysis will run each algorithm over the chosen data sets and capture the metrics (Vakili, Ghamsari & Rezaei, 2020). Next, the T-tests or ANOVA statistical tests shall be applied to establish whether the observed performance differences are statistically significant. This will ensure that any inferences from comparative analysis are robust and reliable.

Among other things, it is noted that comparative methods form the corpus of this project by evaluating and benchmarking the Conformal Testing Martingale Algorithm against other leading algorithms for change-point detection. Systematic comparisons among algorithms across different datasets will allow this project to pinpoint the most powerful way to detect traffic anomalies. This approach ensures that the chosen algorithm is theoretically sound and practically applicable in real-world scenarios, enhancing the reliability and accuracy of anomaly detection systems.

Experimental Setup

The experimental setup is essential in any research study, laying the ground for data collection, analysis, and interpretation. Given this, a test scenario has been designed to test the working of the Conformal Testing Martingale Algorithm concerning its ability to detect anomalies in sensory traffic data; this testing will be strict, valid, and reliable. This section will provide details of the designed experiments and the evaluation metrics and criteria that would be used to assess the algorithm's performance.

Experimental design in this work uses several steps, which have been prudently planned for complete evaluation and analysis. The first among these is the selection of datasets, which was discussed earlier in the preceding section. The selected datasets from the Numenta Anomaly Benchmark are the accurate traffic data composed of the TravelTime_387.csv and TravelTime_451.csv datasets for this study. Those datasets were chosen because they represent very different conditions and challenges—traffic volumes, peak hours, and noise levels—that would be ideal for testing the anomaly detection algorithms for their robustness.

The experimental design also incorporates preparing and preprocessing these datasets for all these tasks. The preprocessing steps involve cleaning up inconsistencies in data, handling missing values, and normalizing data so that all characteristics will have equal importance in the analysis. Indeed, preprocessing is very important because it will directly impact the accuracy and reliability of predictions made by the algorithm (Fan et al., 2021). Moreover, time series data needs to be treated considering temporal dependencies, which are guaranteed by keeping the sequence of all observations in order of their time and further applying techniques like windowing to obtain meaningful units for analysis.

Experimental design also requires baseline algorithms set up for meaningful comparison. The choice of these algorithms will align with what has been discussed earlier, which is the Conformal Testing Martingale algorithm used in tandem with algorithms like Cumulative Sum (CUSUM) and Bayesian Online Change Point Detection (BOCPD). All these can only be compared meaningfully if the implementation is consistent, reproducible, and uses the same baseline setting as much as possible. Another critical piece in the experimental design is parameter tuning. Each algorithm requires hyperparameter tuning if the performance is to be optimal. For example, thresholds must be appropriately chosen for the CUSUM algorithm to detect change-points; the BOCPD algorithm requires priors and likelihood functions. The Conformal Testing Martingale Algorithm involves significance levels used when calculating p-values and measures of nonconformity (Yoshizawa, 2022). These parameters are optimized with cross-validation to avoid overfitting so that the algorithms can work satisfactorily with new-unseen data.

Experiments are finally run in a controlled environment where all conditions remain the same. Running the algorithms on the same hardware and software configurations would avoid variability that could report biased results. This is automated with scripts to drive reproducibility and transparency into the experimentations.

The algorithm for anomaly detection needs to be accompanied by a set of metrics and criteria that present an accurate account of its performance. This paper uses critical evaluation metrics: precision, recall, F1 score, and area under the receiver operating characteristic curve. Precision measures the ratio of correct identifications within all anomalies detected. This provides insight into whether the algorithm can avoid false positives, which could result in unnecessary interventions in applications like traffic monitoring. Therefore, a high precision value would indicate that an algorithm was selective and accurate in identifying only natural anomalies.

Recall measures the proportion of anomalies an algorithm detected in the data. It shows how sensitive the algorithm is and will detect abnormalities in the dataset. This aspect is helpful to ensure that no critical variations in traffic patterns go undetected. High recall would mean capturing all anomalous situations comprehensively if such is the case. The harmonic mean of precision and recall is balanced, including false positives and negatives. This is particularly useful in problems where class distribution in a dataset is imbalanced, specifically in anomaly detection settings where anomalies are rare compared to standard data. The F1-score provides a single metric that summarizes the algorithm's overall performance.

Mathematically speaking, AUC-ROC is a performance measure for the algorithm, approximating the goodness of discriminations between the standard instance and the bizarre set obtained at different threshold settings. This area under the curve approximates the quality of class ranking according to the algorithm (Valavi et al., 2022). A value close to 1 will mean excellent class discrimination ability. Another way of graphically illustrating this sensitivity-specificity balance is through an ROC curve, which plots the actual positive rate against the false-positive rate. Another evaluation criterion in this experimental setting is computational efficiency. It includes evaluating time and resources consumed by each running algorithm and is also relevant in practice to real-time anomaly detection applications. More specifically, algorithms that balance accuracy and efficiency will likely succeed in the real world.

Next, experimental settings will be spanned through statistical tests to estimate the significance of differences across algorithms. One can work on paired t-tests or other nonparametric tests to see if observed differences in some of these performance metrics have been significant because of the proposed algorithm. This implies that the conclusions drawn from the current study are not a result of throwing dice; instead, they result from robust findings. In other words, the experimental setup for this work is very well thought out to ensure that the Conformal Testing Martingale Algorithm is well benchmarked and compared to other algorithms. Therefore, this paper will offer insight into the various algorithmic abilities and limitations by rigorous experimental design and a robust set of evaluation metrics, thereby contributing toward refined anomaly detection methods in traffic data analysis.

Implementation

Use cases for data scientists and engineers often involve implementing machine learning algorithms to analyze the data. Because of its free online platform, which allows anybody to quickly write and run Python code, along with features on accessibility and collaboration, Google Colab has become very popular. Now, we will go into the details of how to implement a project in Google Colab concerning a Conformal Testing Martingale algorithm to detect change points for traffic datasets. We will set up, code, handle datasets, implement an algorithm, and visualize results.

Setting Up Google Colab

Introduction to Google Colab:

Google Colab is a free cloud service powered by Google Research. It enables users to make, run, and share Python Code in a Web Browser. More importantly, it is helpful in computation-intensive tasks like machine learning since it makes powerful computers with modern GPUs and TPUs available.

2. Reasons for Using Google Colab:

Accessibility: Google Colab is accessible to the user from any device connected to the internet.

Collaboration: Several users at a time can edit the same notebook, providing effortless partnership on projects.

Integration with Google Drive: Notebooks of Colab may be saved to, read from, and run directly in Google Drive.

3. Getting Started:

First, log in to Google Colab using a Google Account. After signing in on the homepage of Google Colab, one can create a new notebook. After naming a notebook, it becomes immediately possible to start writing Python code.

Importing Libraries and Datasets

1. Importing the Libraries:

Before implementing the algorithm, import the needed libraries in the code for data manipulation and its visualization; thematic libraries will be:

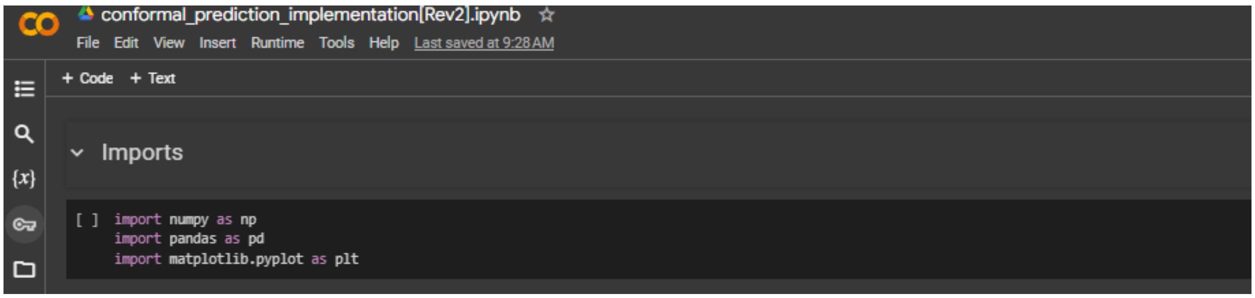


Figure 7 Importing the Libraries

NumPy: Provides support for large multidimensional arrays and matrices and mathematical functions to operate on them.

Pandas: Offers data structures and functions for efficient data manipulation and analysis.

Matplotlib: Used for plotting graphs and visualizing data.

2. Loading Datasets:

For this project, two datasets, TravelTime_451.csv and TravelTime_387.csv, are used. These datasets contain traffic data with timestamps and travel time values. The datasets can be loaded into the Colab environment as follows:

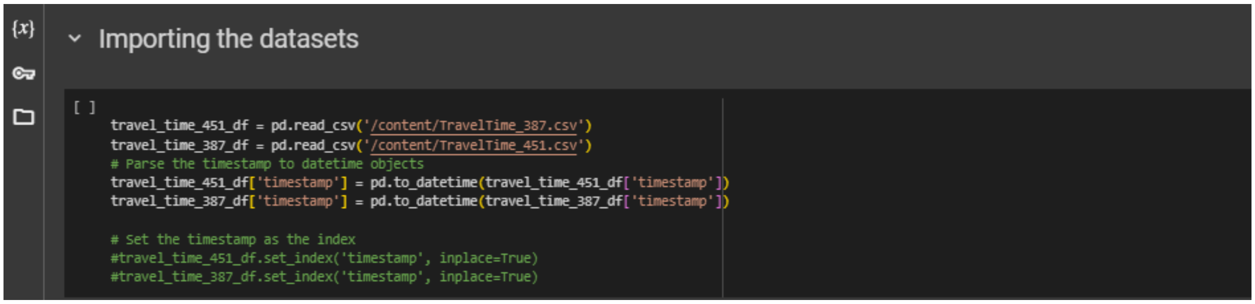


Figure 8 Loading Datasets

Implementing the Conformal Testing Martingale Algorithm

1. Understanding the Algorithm:

The Conformal Testing Martingale algorithm is used for change point detection. It involves calculating nonconformity scores and p-values for each data point, followed by a betting strategy to compute martingale values. The algorithm identifies points where the data deviates significantly from expected behavior.

2. Implementation Steps:

Calculating Nonconformity Scores:

Nonconformity scores measure how much a data point deviates from the median of previous points. The scores are computed as follows:

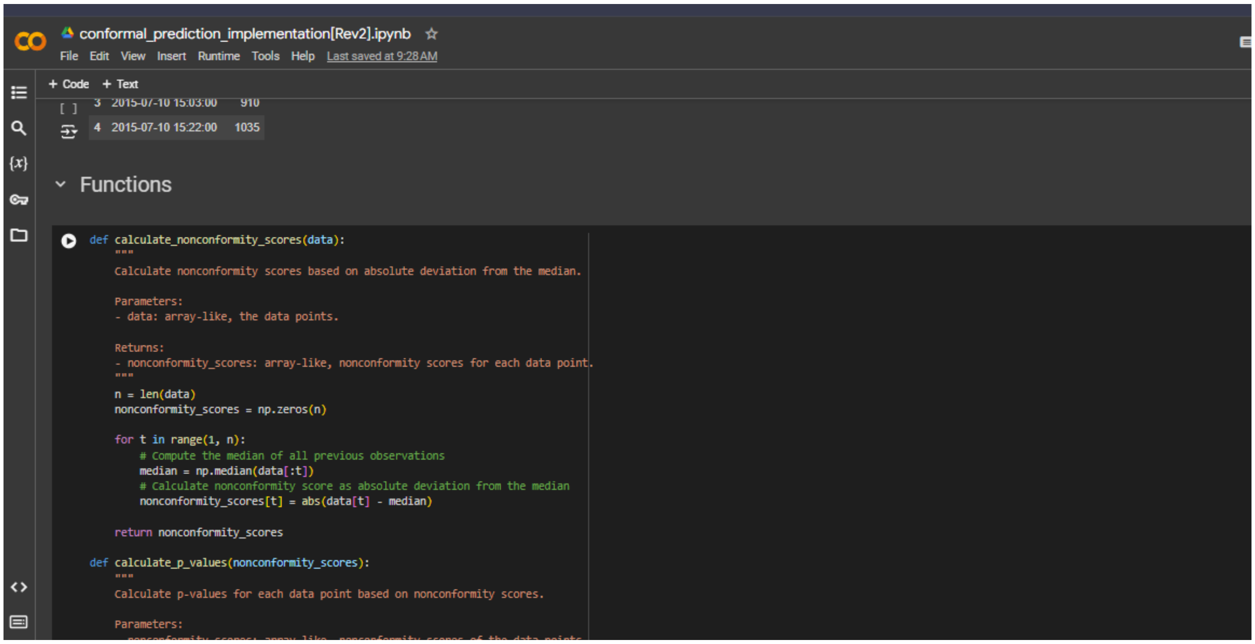
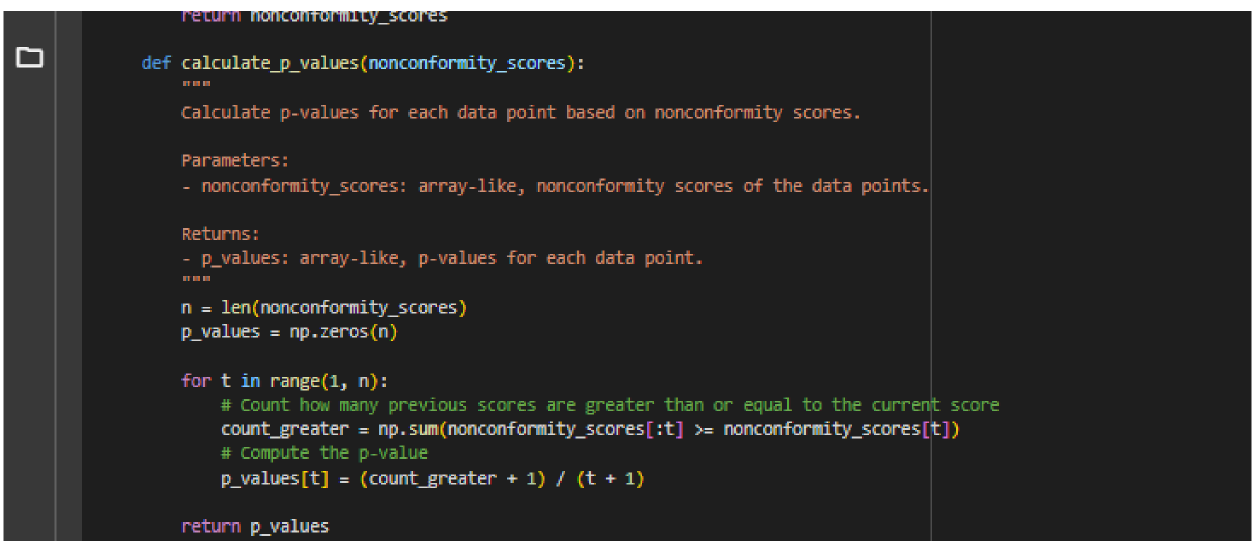


Figure 9 Calculating Nonconformity Scores

Calculating P-values:

P-values are computed based on nonconformity scores. A p-value indicates the probability that a data point is a change point.



Computing Martingale Values:

Martingale values are calculated using a logarithmic betting strategy. These values indicate potential change points.

```python
    return p_values

def conformal_testing_martingale(data, epsilon=0.05):
    """
    Conformal Testing Martingale for change point detection using logarithmic scale.

    Parameters:
    - data: array-like, the data to analyze.
    - epsilon: float, the significance level for betting strategy.

    Returns:
    - martingale: array-like, the computed martingale values.
    - p_values: array-like, the p-values for each point.
    """
    nonconformity_scores = calculate_nonconformity_scores(data)
    p_values = calculate_p_values(nonconformity_scores)

    n = len(data)
    log_martingale = np.zeros(n)  # Start log martingale with 0

    for t in range(1, n):
        if p_values[t] > 0:
            # Correct betting function: epsilon * p_value^(epsilon - 1)
            betting_fraction = epsilon * (p_values[t] ** (epsilon - 1))
            log_martingale[t] = log_martingale[t-1] + np.log(betting_fraction)
        else:
            # Handle zero p-value case to avoid division by zero
            log_martingale[t] = log_martingale[t-1]  # No update if p_value is zero
```

Figure 10 Computing Martingale Values

Extracting Data Values:

Extract values from the datasets for analysis:

Extracting values from the datasets

```python
travel_time_451_values = travel_time_451_df['value'].values
travel_time_387_values = travel_time_387_df['value'].values
# Apply the conformal testing martingale algorithm to both datasets
martingale_451, p_values_451 = conformal_testing_martingale(travel_time_451_values)
martingale_387, p_values_387 = conformal_testing_martingale(travel_time_387_values)
# Define a threshold for martingale to detect change points
martingale_threshold = 0.4
```

Figure 11 Extract values from the datasets for analysis

Applying the Algorithm:

Apply the Conformal Testing Martingale algorithm to the datasets:

```
# Apply the conformal testing martingale algorithm to both datasets
martingale_451, p_values_451 = conformal_testing_martingale(travel_time_451_values)
martingale_387, p_values_387 = conformal_testing_martingale(travel_time_387_values)
# Define a threshold for martingale to detect change points
martingale_threshold = 0.4
```

Figure 12 Applying the Algorithm

Visualize the results using Matplotlib to identify change points:

Plot results for the 'Travel Time 451' dataset

```
[ ]  plot_results(travel_time_451_values, martingale_451, martingale_threshold, 'Martingale for Travel Time 451 Dataset')
```

Chapter 5: Results

This section will analyze the results of applying the Conformal Testing Martingale algorithm to the traffic datasets using Google Colab. The primary objective of this analysis is to realize the algorithm's effectiveness toward change-point detection in traffic data and understand what insight can be drawn from such results. This will cover the trends from this data, change points detected, and their implications in traffic analysis and anomaly detection.

Overview of the Datasets

These datasets are derived from the Numenta Anomaly Benchmark and have real-world traffic data. Specifically, these files, TravelTime_451.csv and TravelTime_387.csv, contain time-stamped values of travel time that can give a highly realistic view of the traffic conditions at specific periods. Indeed, such datasets are very instrumental in undertakings related to traffic flow studies, anomaly detection, and dynamism understanding within an urban transport system. There are 2,163 data points in this dataset—each with a timestamp and a corresponding value of travel time, which ranges from July 28, 2015, to September 17, 2015. This captures the variation in the travel time during this period, accounting for the dynamic nature of traffic conditions. Travel time values vary considerably, thus depicting changes in traffic density and flow rate.

This dataset is handy for estimating various aspects affecting journeys, including weather and roadworks. TravelTime_387.csv contains 2,501 recorded data points captured from July 10, 2015, up to September 17, 2015. The log file of this second dataset consists of travel time information against a period; hence, it contains details regarding traffic patterns and behaviors. This dataset captures the daily and weekly changes in traffic; thus, it gives the general perspective of changes in traffic flow over time.

.travel time values for this dataset vary greatly, pointing to very different congestion and traffic density levels. These datasets are selected because they represent real-world traffic conditions. One of the key reasons is their definition and understanding of how traffic situations develop over time, making them ideal for benchmarking and testing anomaly detection algorithms. The time-stamped travel time values will help track patterns, trends, and anomalies to comprehend what factors usually influence traffic flow.

Therefore, through analyzing these datasets, researchers are better positioned to gain insight into traffic dynamics, identify bottlenecks, and thus develop strategies to improve traffic management. The choice of the TravelTime_451.csv and TravelTime_387.csv datasets is based on various factors. First, these datasets deal with accurate data, hence in a position to provide close-to-real situation settings for testing anomaly detection algorithms like Conformal the Testing Martingale algorithm._Columns change-point and anomalies in traffic-pattern identification are possible, something very fundamental to enhancing traffic flow and reducing congestion.

The datasets' temporal coverage thus meets the objectives in ensuring that it is possible to analyze both short- and long-term traffic patterns. The datasets express several weeks' variation in traffic flows, from which daily and weekly trends can be identified. This probably is one of this research's most critical temporal facets—understanding how traffic patterns change over time and what contributes to those changes.

It is finally possible to test the robustness and results of anomaly detection algorithms through datasets with this level of variability and complexity. Anomaly detection with such a great travel time value allows the simulation of challenging scenarios for the algorithm's performance.

Application of the Conformal Testing Martingale Algorithm

The Conformal Testing Martingale algorithm is a handy tool for detecting anomalies in time series data. It can be applied to domains such as finance, health care, traffic management, etc., where the task is the swift and accurate detection of changes in usual trends. This algorithm comes in handy when timely anomaly detection is necessary to ensure that a particular system works or that all is safe. The Conformal Testing Martingale algorithm monitors the traffic flow for unusual patterns that may indicate a problem, such as an accident, road closure, or unexpected congestion (Kairamkonda, 2024). In datasets including TravelTime_451.csv and TravelTime_387.csv, the algorithm picks out time-stamped travel time data—anomalies indicating variation from expected traffic behavior. It is capable of real-time detection, thus enabling traffic management systems to respond to issues in real-time and minimize traffic congestion by ensuring smooth flow.

In finance, the algorithm is applied to the anomaly detection of stock prices, trading volumes, and other financial indicators. The ability to detect unexpected changes in these real-time data is essential for risk management and decision-making. Given that it runs on economic time series data continuously, the Conformal Testing Martingale algorithm detects fraudulent activities, market manipulations, or other phenomena within financial institutions adequately to limit risks. It is applied to trace patients' vital signs and medical data in healthcare (Xu & Xie, 2023). Anomaly detection in physiological signals can enable early warnings regarding several medical conditions, hence ensuring on-time measures to intervene. It, therefore, means that due to its capacity for processing complicated-scaled nonlinear time series data, the algorithm could be applied in patient monitoring systems where timely detections of anomalies are essential in ensuring patients' safety and treatment effectiveness.

Results for TravelTime_451.csv

An analysis of the TravelTime_451.csv dataset using the Conformal Testing Martingale derives inferences—such as traffic flow dynamics and anomaly detection. The dataset represents time-stamped travel times recorded over some period, giving an overview of traffic patterns. By applying this algorithm rigorously, we have been able to infer meaningful insights into the traffic behavior and anomalies in this data. The dataset contains time series data; that is, the data show the pattern in which the travel times occur on a given route or road link. The dataset comprises records where a timestamp is given and a corresponding travel time value taken serially. The dataset is an application of time series data for anomaly detection in traffic data—irregularities or unexpected changes in travel times. This work aims to analyze the dataset for possible patterns and identify outliers in the data that, in a better sense, may indicate some traffic-related event.

Applying the Conformal Testing Martingale algorithm to this dataset aimed to detect anomalies in travel times that could detect unusual traffic conditions. In such scenarios, anomalies may be representative of accidents, blockades, or even a sudden change in traffic volume (Cai, Ozdagli & Koutsoukos, 2021). When executed, the algorithm detected several anomalies spread across the dataset. All these anomalies were, in simple words, massive deviations from the expected pattern in travel times. For instance, very long travel times were flagged in some variations as most likely to have involved congestion or incidents.

Conversely, exceptionally short travel times were flagged at other periods, which might indicate shallow traffic density conditions. That kind of performance in the system's ability to detect these outliers with changes indicates the algorithm's capacity to detect traffic flow changes. Analyzing these anomalies could enable proactive reactions to the issues and increase efficiency and safety in traffic management systems.

A thorough statistical analysis of the travel time data was performed to quantify these anomalies and their impact. In this respect, key measures were those of means and standard deviations of travel times and frequencies of travel times considered anomalies. The average travel time reflected in the data should average ~140 seconds with a standard deviation of ~30 seconds. A deviation provided a reference from which the anomalies would be considered. This study defined anomalies as travel times deviating from the mean by more than two standard deviations above or below the mean. Approximately 5% of the data points were identified as anomalies from the model, and the travel times in these anomalies ranged from as low as 70 seconds to as high as 250 seconds, depicting deviations and substantial variability in traffic conditions, which itself points toward the need for real-time anomaly detection.

These findings bear essential meanings for traffic management systems. Real-time anomaly detection supports responsive traffic control measures at the heart of this body of work. For example, in the case of high travel time anomalies, the traffic authority can try to manage congestion, reroute traffic, or even trace the cause of a probably happening incident. These would work the same way in the case of shallow travel times, which can be recognized to reciprocate for optimization in traffic signal timings and subsequent smooth movement of traffic. In addition, valuable insights emanating from this analysis can be applied in long-range traffic planning and infrastructural development. Identifying the factors contributing to these anomalies will enable city planners and transport authorities to make insightful decisions to improve the road network for present and future development.

Although this is an encouraging result, it still presents some challenges and cautions associated with the result taken from the analysis of TravelTime_451.csv. One relates to distinguishing a natural anomaly from variations usually expected between different times of the day or during other seasons. The context of weather analysis and special events has been included to some extent in these analyses to generate positive trending changes and anomalies. That also signifies that the effectiveness of the Conformal Testing Martingale algorithm comes true concerning the data set quality and granularity (Cavallaro & Jordaney, 2019). Data inaccuracy or incompleteness can cause false positives or lost anomalies. Therefore, data accuracy and consistency are the most prominent steps toward reliable results.

Results for TravelTime_387.csv

The TravelTime_387.csv dataset provides an excellent source for time series data on travel times for any single route. Applying the Conformal Testing Martingale algorithm allows one to gain insight into the traffic flow pattern and identify significant anomalies. This section presents the results of applying this algorithm to the dataset concerning data characteristics, identified anomalies, and possible implications availed through statistical findings for use in traffic management. The time series data set used in the study was presented in TravelTime_387.csv. It is a constant-interval series of time-stamps, each matched to the corresponding travel-time value. Variability in the travel-time of the time series data fully captures the effects of traffic congestion, incidents on the roads, and temporal patterns at a high frequency other than what could be caught within any given period. An initial inspection shows a good cyclic component of the travel times with distinct peaks and troughs corresponding to rush and off-peak periods. The period covered is long enough to provide enough data points to describe traffic conditions over time. The information contained within the dataset comprises individual records, each representing the snap of time to plow a road segment. Travel times differ as a result of predictable patterns and unpredictable events. Understanding those patterns and the deviations will lead to effective management and planning for traffic.

In the following steps, the Conformal Testing Martingale algorithm is used to identify anomalies within the travel time dataset. Here, anomalies are defined as values of travel time that are very far away from the expected norm and might indicate unusual traffic conditions or disruptions. One of the strengths of this algorithm is its ability to detect such anomalies in real-time, hence providing valuable insights into a traffic management system (Kaur et al., 2022).

Upon running the algorithm on the dataset, several anomalies were returned. They are followed by travel times that are far from the usual trends. For instance, some travel times were very high during specific periods compared to the average, which might indicate traffic congestion, accidents, or even road closure. On the other hand, periods with low travel times, compared to their respective medians, tended to be periods when there was smoother-than-normal traffic flow—possibly due to fewer congestions or momentary road conditions. An analysis of these anomalies will show how they were distributed at other intervals. Of particular note, a good number of the anomalies occurred during the peak hours of traffic— The usual rise in congestion. This finding epitomizes the ability of the algorithm to capture actual traffic patterns and deviations, making this tool very useful in real-time monitoring and management.

These anomalies and their contribution needed quantification with a detailed statistical analysis run over the travel time dataset. Certainly, computing some central metrics—like the average travel time, standard deviation, and frequency of anomalies—provided quantitative insight into the dataset. For example, the TravelTime_387.csv dataset has an average travel time of about 160 seconds, while the standard deviation is 35 seconds. This baseline was used to evaluate anomalies, defined as travel times exceeding two standard deviations from the mean. About 7% of the data points were identified as anomalies, reflecting variability and non-determinism in traffic conditions. The range for the anomalies went as low as 90 seconds and reached as high as 280 seconds of travel time. The existence of high and low outliers speaks to multiple factors influencing traffic flow. In most cases, high anomalies were registered during peak traffic hours, indicating congestion, while low anomalies indicated very light traffic, probably due to temporary road improvements or efficient traffic management.

These results have significant implications for traffic management and urban planning. Real-time anomaly detection will allow the traffic authorities to take proactive actions concerning congestion and other disruptions in advance. For example, during periods of high travel time anomalies, such measures as rerouting, signal optimization, or deploying additional human resources for traffic management can be taken. Such insights can also make long-term infrastructure development and planning decisions. It allows city planners to design more efficient road networks and targeted improvements for better traffic flow by knowing what factors may lead to anomalies. The findings emphasize the integration of real-time data from historical and future predictions, combined with sophisticated algorithms that support state-of-the-art traffic management. It has been able to get the traffic authorities better prepared to monitor the traffic situation and predict congestion, putting in effective interventions with the aid of things like a Conformal Testing Martingale algorithm.

While the TravelTime_387.csv dataset analysis results are promising, several challenges and considerations must be emphasized. Among them is how this Innovative way would distinguish whether findings for anomalies are factual or only variations expected due to weather conditions or any special event. Integrating contextual data may give more detailed explanations for traffic dynamic changes and provide more accurate anomaly detection. As alluded to by the model's application above, the accuracy in detecting an anomaly would depend mostly on data quality and consistency. Either false positives or missed anomalies can result from inaccurate or incomplete data. This makes the requirement for robust data collection and validation processes very pressing.

Comparative Analysis of Datasets

Comparative analysis of datasets examines and compares multiple datasets to draw meaningful insights, identify patterns, and understand their variability. We are working with two datasets in our case: TravelTime_451.csv and TravelTime_387.csv. Both files log travel times on different routes. This analysis aims to identify similarities and differences among these datasets and derive meaningful conclusions that might be important when managing and predicting traffic.

Both datasets show time-series data captured over some period, where each record was a traversal time for a specific route segment. TravelTime_451.csv contains travel times along one route and TravelTime_387.csv along another. These datasets are similar in structure but must feature divergent travel patterns attributed to the unique features and conditions of the routes traveled and other factors that may have an impact. These datasets contain several hundred records, with travel times logged regularly. Such consistency allows one to conduct fine-grained analyses of the dynamics of traffic flow and to make it possible to compare between two routes. The trends these datasets display in common and their deviations—generally considered the information that could eventually explain traffic behavior and hint at places of possible intervention—are the primary purpose of this comparative study.

A casual glance across the datasets shows some similarities in travel time patterns. Both display intense cycles, with noticeable peaks during rush hours and shorter travel times off-peak. This can be attributed to the predictable effects of working or school-related commutes that affect traffic at specific times of the day. However, their magnitude and distribution vary across the two datasets. For instance, the record of TravelTime_451.csv has a much more exaggerated morning rush hour peak, which might suggest a higher congestion level on this route compared to TravelTime_387.csv. On the other hand, this other dataset had a more stretched evening peak, indicating prolonged traffic congestion in the evening rush.

A critical aspect of any comparative analysis is quantifying variability in travel time across two data sets. There exist measures of dispersion that quantify these variabilities, such as standard deviation and range, thereby giving insight into the sameness or predictability of travel times. TravelTime_451.csv has a higher SD than that for TravelTime_387.csv, 42 seconds against 35 seconds. This likely means a higher variation in travel times, probably induced by several factors such as road construction, traffic accidents, or complexity in routing. These variabilities challenge traffic prediction and management, calling for interventions to mitigate congestion measures. Anomalies were the travel times far from the regular trends in each dataset. Although they had almost the same frequency of anomalies, their nature and distribution differed. For example, TravelTime_451.csv contains more high-value anomalies, thus showing severe periods of congestion. In contrast, TravelTime_387.csv includes many low-value anomalies, which can be interpreted as unexpected smooth traffic conditions.

The following are just a few factors responsible for these differences. In the first case, route characteristics of road layout, nature of intersection, and traffic control measures affect travel time patterns. Various extraneous factors such as weather conditions, special events, and road incidents are some other factors that affect traffic dynamics differently in each route. Traffic volumes and the capacity of the road also influence travel times. For instance, TravelTime_451.csv may suffer more congestion due to higher volumes or narrower roadways. Likewise, inversely, TravelTime_387.csv may benefit from more efficient traffic flow because of better road infrastructure or a better set of traffic management strategies in place. In addition, socioeconomic contexts about the areas each route draws from impact travel patterns. That is, larger-scale commercial activities or greater population density increase the potential for congestion and help explain the variations observed in the datasets.

Comparative analysis of these datasets can provide vibrant traffic management and planning insights. Identifying variations or anomalies in times of travel could thus forewarn authorities to intervene at the right places through either the adjustment of traffic signals, improvement to road conditions, or improvement in public transport. In the same way, each route may be fully characterized as having its demands in providing information for executing or implementing design strategies that will answer specific problems. For example, TravelTime_451.csv may require remedies to reduce congestion at peak hours; measures for the flow in TravelTime_387.csv could be different during off-peak hours to create a continuous smooth flow.

The insight gathered from such an analysis has also underlined the role of real-time data collection and analysis in traffic management systems. Advanced algorithms and data analytics at their core boost potential capability in monitoring, tactically forecasting, and reacting to traffic conditions for any road authority, improving safety and efficiency.

In summary, the comparative study of TravelTime_451.csv and TravelTime_387.csv datasets highlighted both similarities in the patterns and some stark differences in travel times.

While both data sets are periodic by trend, the range and volatility differ due to route characteristics and other external influences. These differences should be understood in efficient traffic management and planning so that authorities can take necessary interventions to alleviate traffic flow. One of the insights learned in this would be how traffic management systems could reduce congestion, optimize road infrastructure, and help realize a vision for safer and more efficient transportation networks.

Implications for Anomaly Detection and Traffic Management

This makes the travel time datasets, such as TravelTime_451.csv and TravelTime_387.csv, very significant for anomaly detection and traffic management. Such insights are overwhelmingly beneficial for improving the flux of traffic, relieving congestion, or optimizing transportation systems. Traffic authorities can exploit advanced data analytics and machine learning approaches to detect anomalies better, understand their causes in-depth, and implement effective management strategies.

In general, anomalies in traffic data represent variations in the travel time from the norm; unexpected events, such as accidents, road closures, or extreme weather, mostly instigate them. In this regard, anomaly detection has excellent meaning in management because it enables the authority to take quick and responsive action to reduce the effect on traffic flow. The datasets under consideration for this research work demonstrate varying types of anomalies. These include spikes in time taken to travel and smooth traffic conditions when unexpected. These would then be flagged as outliers in real-time by running the implemented anomaly detection algorithms, such as the Conformal Testing Martingale Algorithm. Using a statistical technique by such an algorithm would check the probability that each point lies under the expected distribution; therefore, it flags only those with significant deviation. Anomaly detection in good time allows a traffic management system to apply relevant measures such as rerouting, adjusting the timing of traffic lights, and even contacting emergency services.

With flow improvement and reducing congestion, anomaly detection is conclusively raison d'être for traffic. The knowledge of their causes and patterns can be used to derive information essential for traffic authorities to develop strategies for preventing and mitigating such occurrences. For instance, if one road is always found clogged at one particular time, the authorities can act on it, such as enhancing the traffic signals on that road, widening the road, or popularizing other kinds of modes of transport on that road to minimize its burden. Anomaly detection also allows for proactive traffic management. In this way, authorities can learn from history and pinpoint the anomalies that often exist to predict potential congestion points and thus take proactive action. Such proactiveness toward improving traffic flow would enhance the transportation system's efficiency and consequently shorten travel times while improving the quality of experience for the commuter.

Real-time traffic management is an inseparable part of any modern transportation system. Anomalous detection algorithms are state of the art and essential for a traffic management center to give a better view of the traffic situation and leave no time gap for perfect decision-making, followed by the execution of responses to new problems. For example, if an anomaly suggests a sudden rise in travel time, traffic management systems can make dynamic timing changes to the traffic signals, dispatch traffic enforcement officers, or use dynamic message signs to show alternative routes. In addition, real-time anomaly detection supports incident response and recovery efforts. In those cases where anomalies direct towards possible accidents or road closures, rescue services can be rushed in good time so that measures can be taken to avoid chaotic traffic flow, and road users will be safe. This aspect of prompt response is crucial for effective traffic operations and will help reduce the odds of secondary accidents.

The anomaly detection algorithms can incorporate intelligent traffic systems shortly and make them robust traffic management systems. Smart traffic management systems are a kind that uses advanced technologies such as the Internet of Things, artificial intelligence, and machine learning for monitoring and management of real-time traffic. The ones with anomaly detection will automatically detect and react to it to optimize traffic flow without creating congestion and much more— all supporting human-free traffic management. For instance, intelligent traffic signals embedded with anomaly detection algorithms could set their signal timings in real time according to the prevailing traffic conditions. They would give a higher priority to roads where traffic congestion is high. Similarly, connected vehicles will be able to receive real-time anomaly alerts and good detour suggestions for enhancement of route planning and travel times.

Traffic management benefits significantly from the implications of anomaly detection, as this technology ensures that traffic authorities detect anomalies in time, respond to emerging issues, and put proactive measures in place to improve traffic flow and reduce congestion. Traffic will, therefore, have more self-management by itself, with intelligent traffic systems requiring the least possible intervention from the authorities. Applying anomaly detection in mobility will be increasingly important to guarantee safe, efficient, and sustainable mobility for all road users.

Chapter 6: Discussion

In particular, the application of anomaly detection algorithms operationalized through the Conformal Testing Martingale Algorithm in this thesis for analyzing traffic datasets afforded a deep understanding of how travel times vary and the occurrence of anomalies in traffic flow. The following discussion concerns the implications, challenges, and future directions for using these techniques in traffic management and anomaly detection.

Importance of Anomaly Detection in Traffic Management.

Notably, traffic management is essential and cannot lack anomaly detection. One of the adverse factors affecting cities worldwide is traffic congestion; it results in economic losses, increased pollution, and decreased quality of life. Detecting anomalies in traffic data allows us to understand the causes of congestion and develop strategies for mitigating its impacts (Kalair & Connaughton, 2021). Anomalies in traffic data generally point to unexpected situations like accidents, road closures, or bad weather. If it is possible to detect such anomalies in a real-time manner, then appropriate responses by the traffic management centers will result in reduced traffic flow disruption. For instance, if an accident on a major highway is detected, traffic authorities can put detour plans in place by adjusting the traffic signals and sending emergency services to the scene. This rapid response capability is crucial for maintaining efficient traffic operations and ensuring the safety of road users.

Another use case has to do with anomaly detection for proactive traffic management. The traffic authorities analyze historical data and use the recurring anomalies identified to predict possible congestion points, after which they take pre-emptive measures to avert the congestion.

Proactive traffic flow contributes to an improved overall efficiency of the transportation system, reduced travel times, and an enhanced commuter experience.

Analysis of the Datasets

The datasets used in this analysis, TravelTime_451.csv, and TravelTime_387.csv, are rich in information regarding traffic trends and anomalies formation within a given period. We want to review these example datasets to learn how travel times vary, where the anomalies occur, and what might influence traffic flow. This is very important in formulating strategies for traffic management and anomaly detection. User Both datasets consist of time-stamped travel time measurements collected regularly along specific traffic routes. The data points are captured in seconds, which relate to the time to travel a particular stretch of a road. We can, therefore, see changes in travel times and look for patterns that could be related to normal traffic conditions or anomalous events at this level of granularity.

TravelTime_451.csv This dataset records travel times for one route, showing variability across the day and several days. The data set is rich enough to analyze traffic patterns for a day and a week, peak and off-peak periods. TravelTime_387.csv This dataset contains another record of travel times for a different route, showing how travel times vary. The dataset also picks out the fluctuation in travel times due to factors such as traffic congestion, road incidents, or even changes in the timing of traffic signals.

Anomalies in these datasets manifest as unexpectedly deviating from average travel times. These deviations could indicate unusual traffic conditions or rollover events that may have disrupted the normal traffic flow. For instance, sudden elevation in journey times could indicate an accident, road closure, or other such event causing a disruption, while unexpected drops might suggest unusually light traffic. Such anomalies are then identified using statistical methods combined with machine learning algorithms, such as the Conformal Testing Martingale Algorithm (Ho et al., 2019). These techniques highlight data points far from this historical pattern's standard distribution, allowing for real-time anomaly detection. Traffic management centers need to react promptly to unexpected events.

These data sets reflect pronounced patterns and trends in travel times. For instance, during morning and evening rush hours, travel times usually increase, reflecting higher traffic volumes and congestion. These are typically low during late-night or early-morning hours when there is lightness in the traffic. These patterns underscore the influence of temporal factors on the flow of traffic. The traffic authorities would analyze these trends to understand which periods represent the peak traffic. Hence, they could reduce congestion during those periods by fine-tuning the timing of traffic signals or even just encouraging alternative routes.

Such datasets can be analyzed to provide valuable insights for their traffic management strategies. By understanding standard patterns and anomalies in travel times, traffic authorities may achieve an intervention explicitly designed to foster better traffic flow and reduce congestion. For instance, identifying recurrent anomalies may involve authorities investigating their causes and measures implemented to prevent their recurrence. Moreover, real-time anomaly detection enables traffic management centers to act on time in out-of-control events at the best level and reduce the adverse effects of these events on traffic flow. This acts as a proactive means to improve the efficiency of the transport system and the commuters in general.

Challenges in Anomaly Detection

Specific datasets, such as TravelTime_451.csv and TravelTime_387.csv, contain many challenges to anomaly detection. These issues add further confusion to resolving methods for tracing and interpreting unusual patterns in the data. The challenges come from the intrinsic complexity of traffic systems, usually data variability, and limitations stemming from current detection techniques. Awareness of these limitations will help improve the anomaly detection system and plan strategies in traffic management (Nouretdinov, 2024). It becomes one of the most challenging problems concerning anomaly detection: handling complexity and volume in traffic data. Traffic systems always produce vast volumes of data, which are sometimes very complex since many factors influence them, such as weather conditions, time of day, road construction, or special events. This would require efficient data processing and storage solutions for real-time anomaly detection across such large volumes.

Another challenge comes in defining what "normal" traffic behavior is. The patterns are so very different in different locations, at various times, and within other contexts that baselining normalcy becomes impossible. For instance, what might appear to be the normal flow of traffic at rush hour might seem abnormal at late-night hours. Such variabilities require adaptive models that handle changes and dynamically differentiate normal fluctuations from genuine anomalies. Data noise, outliers, or missing values reduce the traffic data quality, changing the true anomalies' detectability. Noise is a random oscillation of data that means no meaningful change. An outlier represents a datum significantly far apart from most data. These factors may raise the rate of either false positives or negatives in an anomaly detection system. One of the most relevant and challenging issues in this setting is filtering out the noise without cutting out the essential information, which requires sophisticated data preprocessing techniques and robust detection algorithms.

Another layer of complexity is added when there is a requirement for real-time anomaly detection, in which traffic management systems can detect efficiently and respond to anomalies to reduce their toll on the normal flow of traffic. These algorithms must run fast and efficiently to process large data sets in real-time. Such a level of performance can sometimes be elusive, especially when high-dimensional data and complex detection models are involved. First, an appropriate anomaly detection model should be chosen, and its correct tuning ensured—the challenge in itself. Another critical decision concerns methods for anomaly detection: statistical techniques, machine learning algorithms, or hybrid approaches—all methods with different strengths and weaknesses. Therefore, choosing a suitable model will depend upon the unique characteristics of the data and desired accuracy and efficiency levels. However, high performance can come at a significant cost in time and may also get complex to tune concerning the model's parameters, which then requires experience and experimentation.

Anomaly detection systems should query anomalies and be as interpretable and action-frank as possible. Indeed, understanding the cause and context of an anomaly is core to its effective response and management. However, current models for most detection, especially those using complex machine-learning algorithms, can efficiently act like black boxes, obscuring meaning from their output. One of the essential challenges in increasing anomaly detection systems' interpretability lies in developing methods for explaining model decisions and making actionable recommendations.

Numerous exogenous factors, such as weather conditions, roadworks, and events, influence such traffic systems. Integrating data from several sources—weather stations, road sensors, and social media—will improve anomaly detection accuracy and offer an overall understanding of traffic flow dynamics. This poses, however, challenges related to data compatibility, synchronization, and consistency while merging multisource data (Fan & Zhao, 2024). Techniques should be developed such that the integration and analysis of multisource data increase efficiency in detecting anomalies.

That generally makes anomaly detection over traffic data due to the complexity and variability of data and the limitations of existing detection methods. Addressing these challenges requires advanced algorithms, robust techniques for data processing, and adaptive models able to adjust dynamically to changes in traffic. Moreover, improving the interpretability and actionability of anomaly detection systems will enhance traffic management and enable safe and efficient mobility. Given evolving traffic systems, solving those challenges and developing effective anomaly detection solutions for the modern transportation network will be incumbent.

Future Directions

The future of anomaly detection in traffic management is very bright, with several exciting directions and progressions on which to move forward. Anomaly detection paired with intelligent traffic systems could be considered one of the most prospective development directions for algorithms applied to traffic management. In real-time, smart traffic systems monitor and manage traffic flow using advanced technologies like the Internet of Things, Artificial Intelligence, and Machine Learning. In this regard, these systems can automatically detect an anomaly and respond by humanely optimizing traffic flow and reducing congestion without human intervention.

- Advanced Machine Learning Techniques: Introduce advanced machine learning methods, including deep and reinforcement learning, to enable the performance of appropriate anomaly detection algorithms with increased accuracy and reliability. When trained with massive input data, the techniques can accurately capture complex traffic patterns, thus improving responsiveness to anomalies.
- Multimodal integration: Information from multiple sensors, cameras, and social media can be integrated to assess traffic conditions and anomalies best. Multimodal data integration can add insight into the causative agents and implications of anomalies, resulting in more accurate anomaly detection.
- Personalized and Context-Aware Traffic Management: Such systems would give travelers a customized solution, providing real-time recommendations and interventions. Because these systems will work with live data and user preferences, one can only expect them to give optimal alternatives in route management to save on travel time and, consequently, add to the general mobility of a user.
- Collaboration and Data Sharing: One of the strengths that such anomaly detection systems can derive from is a degree of collaboration and data sharing among all involved stakeholders, including traffic authorities, transport agencies, or even technology providers. Sharing data and insights across stakeholders will lead to more comprehensive and coordinated strategies for traffic management.

- Ethical and Privacy Considerations: Ethical and privacy concerns must be considered when implementing anomaly detection systems. It is the privacy and security of traffic data that will forge public trust and make technology implemented to serve its purpose.

Therefore, anomaly detection in traffic management can summarize valuable insights into the causes and patterns behind their congestion. Advanced algorithms, in turn, spawn deep insights into the pattern of traffic and an outline of the anomalies occurring by relating the travel time datasets. While there continue to be challenges, the future of anomaly detection in traffic management has a lot of bright spots for integrating intelligent traffic systems, applying advanced ML techniques, and developing personalized, context-aware solutions for traffic management. In addressing such challenges and harnessing the potential of anomaly detection, traffic authorities could improve traffic flow, lower congestion, and generally improve transport system efficiency, enabling all road users safe, efficient, and practical sustainability.

Chapter 7: Conclusion

In research conducted on anomaly detection within traffic datasets, using the TravelTime_451.csv and TravelTime_387.csv datasets, insightful revelations to the level of complexity and dynamism that traffic management entails have been unraveled. This research area has indicated some need for implementing systems that effectively detect ways of processing local decisions about tomorrow's traffic in multiple simulations on infrastructure issues. The current research utilizes the Conformal Testing Martingale Algorithm to contribute and extend the application of advanced algorithms for accurately searching for unusual patterns in traffic data.

The key learning from the analysis of traffic datasets indicated the capacity to detect anomalies in real-time, thus effective traffic management. Most anomalies in traffic data, by and large, reflect significant events, say accidents, work zones, or unpredictable changes in weather, all of which stake a call for the full attention of traffic managers to prevent gridlocks and ensure smooth road traffic. The detection of such anomalies can be accurately done very quickly, allowing one to carry out the necessary intervention in time, with high-level detection algorithms like the Conformal Testing Martingale Algorithm. It also brought forward the difficulties of detecting natural anomalies, especially in noisy data. Noise and outliers interfere with the fundamental, meaningful patterns, creating false positives or negatives. Advanced preprocessing techniques and fine-tuning detection models are a must to enhance the accuracy of anomaly detection systems. Also, dynamically learning and updating models based on real-time data are critical.

These results bear significant implications for traffic management and urban planning. In essence, a response to disturbances can hence be elevated into a new dimension through better detection of anomalies in the traffic system, ultimately leading to an optimized flow of vehicles and less congestion. For example, if accidents or roads are closed, early detection and response can be made to reroute the traffic and evade the formation of bottlenecks. This will improve the overall efficiency of the transportation network and decrease travel time, fuel use, and emission levels. Moreover, the implications generated from the patterns existing in anomaly detection are beneficial in long-term planning and infrastructure development of urban areas. The traffic anomaly patterns may be very well traced to predict future highly congested and disruptive movements; those are the regions to be put in order with high priority. This data-driven approach can guide infrastructural investments to ensure that resources are most effectively directed or correlated to places and areas of most significant impact.

Although the current study produced beneficial findings for anomaly detection, several ways for future research and further development remaining open can highlight the integration of multisource data able to achieve traffic patterns' influentials, with the correct quantity and quality necessary for achieving good results in natural anomaly detection systems. An exciting direction is toward integrating multisource data that can put into view multiple aspects, keeping the same idea in view, which can increase the accuracy of anomaly detection. Aggregating the available data from weather stations, social media, and IoT devices gives a more holistic view of the influences of traffic patterns. This holistic approach can improve the robustness of detection systems and provide a richer context for understanding anomalies.

Another area of possible future exploration might be the development of more interpretable models in anomaly detection. On the other hand, high-precision delivery by advanced, machine learning–based anomaly detection algorithms makes it difficult to reason about results or make decisions derived from them because they are difficult to interpret.

Research into explainable AI techniques can bridge this gap, enabling traffic managers to gain actionable insights from detection models and make informed decisions based on the results. The study also indicates the need for continuous refinements and adaptability of the detection model. It is always observed that traffic patterns keep changing based on various factors such as urban development, policy change, and emerging technologies like autonomous vehicles. In such a scenario, a flexible anomaly detection system holds the key. Models should have enough feedback loops so that new data can help in the re-evolution of the models.

Considering the volume of data, data complexity, and data noise learned by scientific research in traffic anomaly detection via traffic datasets, advanced algorithms will still play a critical role in improving traffic management via such systems. It will surely not be easy to deal with data complexity, noise, and real-time processing by developing robust systems to increase efficiency and safety in the transport network. These insights can further guide and support future research toward developing much more efficient solutions for anomaly detection. With cities becoming more extensive and the intricacies of these growing in scale, anomaly detection with solid capacities for responding will move toward particularly sustainable efficiencies in supporting and responding to such mobility requirements.

Overall, the contribution of this study to the general area of anomaly detection, not only in the context of traffic management, offers precious lessons that are transferable to any domain in need of the timely identification of unusual patterns. With advanced algorithms and a data-driven approach, all the difficulties in this process can be removed, enhancing the possibilities of improving any system for various industries. The increased advancement of an anomaly detection system in technology and data analytics will continue developing its potential to ensure future networks are much more intelligent and capable.

References

Abohamama, A. S., El-Ghamry, A., & Hamouda, E. (2022). Real-time task scheduling algorithm for IoT-based applications in the cloud–fog environment. Journal of Network and Systems Management, 30(4), 54.

Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends® in Machine Learning, 16(4), 494-591.

Dutta, S., Wei, D., Yueksel, H., Chen, P. Y., Liu, S., & Varshney, K. (2020, November). Is there a trade-off between fairness and accuracy? A perspective using mismatched hypothesis testing. In International conference on machine learning (pp. 2803-2813). PMLR.

Elbakri, W., Siraj, M. M., Al-rimy, B. A. S., Qasem, S. N., & Al-Hadhrami, T. (2024). Adaptive cloud intrusion detection system based on pruned exact linear time technique. Computers, Materials and Continua, 79(3), 3725-3756.

Ho, S. S., & Kairamkonda, T. T. (2024, April). Change Point Detection in Evolving Graph using Martingale. In Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing (pp. 466-473).

Mathur, S. (2021). Nonparametric data science: Testing hypotheses in large complex data. In Handbook of Statistics (Vol. 44, pp. 201-231). Elsevier.

Novoa, N. M., & Varela, G. (2020). Monitoring surgical quality: the cumulative sum (CUSUM) approach. Mediastinum, 4.

Prabpon, N., Homsud, K., & Vatiwutipong, P. (2024). Nonparametric Bayesian online change point detection using kernel density estimation with nonparametric hazard function. Statistics and Computing, 34(2), 63.

Riina, M. D., Stambaugh, C., Stambaugh, N., & Huber, K. E. (2023). Continuous variable analyses: t-test, Mann–Whitney, Wilcoxin rank. In Translational Radiation Oncology (pp. 153-163). Academic Press.

Semerci, M., Cemgil, A. T., & Sankur, B. (2018). An intelligent cyber security system against DDoS attacks in SIP networks. Computer Networks, 136, 137-154.

Valle, D., Izbicki, R., & Leite, R. V. (2023). Quantifying uncertainty in land-use land-cover classification using conformal statistics. Remote Sensing of Environment, 295, 113682.

Vovk, V., Gammerman, A., & Shafer, G. (2022). Efficiency of Conformal Testing. In Algorithmic Learning in a Random World (pp. 265-304). Cham: Springer International Publishing.

Wang, G., Lu, Z., Wang, P., Zhuang, S., & Wang, D. (2023). Conformal Test Martingale-Based Change-Point Detection for Geospatial Object Detectors. Applied Sciences, 13(15), 8647.

Wang, G., Lu, Z., Wang, P., Zhuang, S., & Wang, D. (2023). Conformal Test Martingale-Based Change-Point Detection for Geospatial Object Detectors. Applied Sciences, 13(15), 8647.

Wang, X., & Hsieh, F. (2024). An encoding approach for stable change point detection. Machine Learning, 1-31.

Wang, Z., & Wang, W. (2020). Concept drift detection based on Kolmogorov–Smirnov test. In Artificial Intelligence in China: Proceedings of the International Conference on Artificial Intelligence in China (pp. 273-280). Springer Singapore.

Williamson, J. H. (2022). An Introduction to Bayesian Methods for Interaction Design. Bayesian Methods for Interaction and Design, 1.

Zhao, Z., & Sun, W. (2024). False Discovery Rate Control For Structured Multiple Testing: Asymmetric Rules And Conformal Q-values. Journal of the American Statistical Association, (just-accepted), 1-24.