Machine Learning in Motion

Engineering Self-Adaptive Systems

Firas Bayram

Machine Learning in Motion

Engineering Self-Adaptive Systems

Firas Bayram

Faculty of Health, Science and Technology

Computer Science

DOCTORAL THESIS | Karlstad University Studies | 2025:14

Machine Learning in Motion

Engineering Self-Adaptive Systems

Firas Bayram

DOCTORAL THESIS | Karlstad University Studies | 2025:14

Machine Learning in Motion - Engineering Self-Adaptive Systems

Firas Bayram

DOCTORAL THESIS

Karlstad University Studies | 2025:14

urn:nbn:se:kau:diva-103579

ISSN 1403-8099

ISBN 978-91-7867-560-9 (print)

ISBN 978-91-7867-561-6 (pdf)

https://doi.org/10.59217/rjrp8643

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Mathematics and Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2025

WWW.KAU.SE

To my mother, my everything.
To my brother, my backbone.

And to all freedom fighters who defend justice and dignity—
keep inspiring us.

Machine Learning in Motion: Engineering Self-Adaptive Systems

Firas Bayram
Department of Mathematics and Computer Science
Karlstad University

Abstract

In the rapidly evolving landscape of artificial intelligence (AI) and machine learning (ML), transitioning from theoretical models to robust, adaptive production systems remains a significant challenge in the digital age. As data-driven methodologies revolutionize problem solving across industries, a critical gap exists between research achievements and reliable real-world deployment. Key obstacles include concept drift and data quality issues that arise from unpredictable changes in data distributions and operational complexities that systematically affect model performance, reliability, and efficiency. This thesis addresses these challenges by introducing an overarching framework for adaptive ML systems that operate reliably in dynamic, real-world environments, incorporating innovative methodologies for dynamic drift detection and real-time data quality assessment bounded by robust Machine Learning Operations (MLOps) strategies. These integrated components enable the creation of production-grade ML systems that can efficiently adapt to shifts in data distributions and assess data quality in real-time, ensuring stable and reliable performance in dynamic environments. The proposed approaches are validated through real-world use cases, demonstrating significant improvements in predictive accuracy and operational efficiency. By deploying these adaptive systems in industrial contexts, the thesis highlights their potential to deliver reliable, high-performance ML solutions tailored to the demands of complex, time-sensitive applications. This work offers concrete solutions for translating theoretical advances into practical applications, contributing to developing robust and scalable ML systems for real-world deployment.

Keywords: Machine Learning, Concept Drift, Data Quality, Adaptive learning, MLOps, Data-Driven Development, Performance Robustness

Maskininlärning i rörelse: Utveckling av självanpassande system

Firas Bayram
Institutionen för matematik och datavetenskap
Karlstads universitet

Sammanfattning

I den snabbt föränderliga världen av artificiell intelligens (AI) och maskininlärning (ML) kvarstår utmaningen att överföra teoretiska modeller till robusta, adaptiva produktionssystem som en avgörande fråga i den digitala eran. Datadrivna metoder revolutionerar problemlösning inom en mängd olika industrier, men en kritisk klyfta existerar mellan forskningsframsteg och tillförlitlig implementering i verkliga tillämpningar. Centrala hinder inkluderar konceptdrift och problem med datakvalitet, som uppstår till följd av oförutsägbara förändringar i datadistributioner och operativa komplexiteter, vilket systematiskt påverkar modellernas prestanda, tillförlitlighet och effektivitet. Denna avhandling adresserar dessa utmaningar genom att presentera ett övergripande ramverk för adaptiva ML-system som kan fungera tillförlitligt i dynamiska, verkliga miljöer. Ramverket integrerar innovativa metoder för dynamisk driftdetektering och realtidsbedömning av datakvalitet, och bygger på robusta Machine Learning Operations (MLOps)-strategier. Dessa integrerade komponenter möjliggör skapandet av produktionsklara ML-system som effektivt kan anpassa sig till förändringar i datadistributioner och kontinuerligt utvärdera datakvalitet i realtid, vilket säkerställer stabil och tillförlitlig prestanda i dynamiska miljöer. De föreslagna metoderna valideras genom tillämpningar i verkliga användningsfall och demonstrerar betydande förbättringar i prediktionsnoggrannhet och operativ effektivitet. Genom att implementera dessa adaptiva system i industriella sammanhang belyser avhandlingen deras potential att leverera tillförlitliga, högpresterande ML-lösningar anpassade till kraven i komplexa, tidskritiska applikationer. Detta arbete erbjuder konkreta lösningar för att överbrygga klyftan mellan teoretiska framsteg och praktiska tillämpningar och bidrar till utvecklingen av robusta och skalbara ML-system för verklig användning.

Nyckelord: Maskininlärning, konceptglidning, datakvalitet, adaptiv inlärning, MLOps, datadriven utveckling, prestandarobusthet

Acknowledgements

To my mother: You are the heart of my existence, the one who taught me the true meaning of strength, compassion, and perseverance. Every sacrifice you made, every prayer you whispered into the quiet hours of night became the compass guiding me through both storms and sunlit paths, shaping everything I am today. Everything I have achieved, I owe to you. You are my inspiration, my foundation, my home. To my beloved late grandmother: Although you are no longer with us, your spirit continues to guide me. The lessons you imparted through your hard work, resilience, and love remain engraved in my heart. You were the first real-life superhero I knew, teaching me the value of sacrifice and dedication. Your words, your stories, and your unwavering commitment to our family are gifts I carry with me every day. To my brother: Watching you grow and succeed has been one of the greatest privileges of my life. Your example continues to inspire me in both my personal and professional pursuits. I owe much of my own success to your example, your guidance, and the advice you have always offered.

To my supervisor, Bestoun: Your mentorship has been a cornerstone of my academic journey. Our relationship extends far beyond academia. Your constant support, invaluable insights, and belief in my potential have been crucial in shaping my research and my personal development. Working with you has been one of the most rewarding experiences of my academic career. To my examiners, Andreas Kassler and Simone Fischer-Hübner: Your thoughtful feedback and suggestions have greatly enriched this work. I truly appreciate your contributions to refining my research.

To my uncle, Adeeb: Your wisdom and resilience have always been a guiding force in my life. The lessons you've shared, along with your enduring presence, have provided clarity during moments of uncertainty. Your strength has always been a source of guidance and inspiration. To my sisters: You have been a constant source of joy and inspiration. Your love and support have made this journey meaningful. The bond we share has been an unshakable foundation, and I am deeply grateful for your presence in my life. To Dr. Zakaria: Thank you for your mentorship and encouragement. Your wisdom has broadened my thinking and expanded my horizons in ways I never anticipated. Your presence has been a constant source of motivation throughout this journey.

To my friends and colleagues: Your shared time, laughter, and memories have been indispensable in this journey. Thank you for your companionship and for everything we have learned together along the way.

Karlstad, April, 2025 Firas Bayram

Contents

Introductory Summary

1 Introduction

Integrating artificial intelligence (AI) and machine learning (ML) into real-world applications has fundamentally transformed industrial decision-making processes, particularly in mission-critical systems. Modern production environments now rely on complex ML pipelines to process vast sensor data streams, optimize multi-objective control systems, and make real-time predictions across interconnected processes [1]. This shift from traditional rule-based systems to continuously evolving ML models has enabled unprecedented levels of automation and optimization, but it has also introduced new challenges to ensure system reliability and maintain performance under dynamic operating conditions.

The transition from experimental to production ML systems marks a fundamental shift in system requirements and operational constraints [2]. While controlled research environments primarily optimize for model accuracy and computational efficiency, production systems must operate under more complex constraints: they must process non-stationary data streams, assure data quality, adapt to evolving system dynamics, operate within strict resource and latency bounds, and maintain performance guarantees across diverse operating conditions. These requirements challenge fundamental assumptions in traditional ML frameworks, particularly in relation to data distribution stability, computational resource availability, and the trade-off between model complexity and operational reliability [3].

To ensure high-quality deployment, ML solutions must exhibit two fundamental characteristics: adaptability and quality-driven design [4]. Adaptability enables the system to maintain reliability when confronting unexpected changes in the operating environment, while a quality-driven approach ensures consistent performance standards and data integrity throughout operation. To effectively manage these characteristics, it is essential to address two primary challenges in production environments: concept drift, where relationships between input features and target variables evolve over time, and data quality issues, which manifest as missing values, sensor noise, and measurement inconsistencies.

These challenges, however, are not isolated but are deeply interconnected, often arising simultaneously in real-world application environments. The presence of concept drift can affect data quality issues, and vice versa, creating cascading effects that significantly impact ML model performance in deployment. Traditional approaches that address these issues independently fail to account for their intertwined nature, leading to incomplete solutions. Consequently, the development of new frameworks is essential to integrate and mitigate these challenges in a comprehensive and holistic manner.

This thesis presents a comprehensive study of ML systems in production environments that addresses the unique challenges associated with the deployment and maintenance of reliable, adaptive systems in dynamic, real-world settings. It bridges the gap between theoretical research and practical implementation by developing methodologies and frameworks that ensure robustness, adaptivity, and operational efficiency. The research contributes to advancing ML in production by establishing formal foundations, introducing adaptive data-driven mechanisms tailored for real-world application requirements, and validating these solutions in complex real-world scenarios.

2 Problem Statement and Research Motivation

Industrial deployment environments introduce a distinct set of challenges that are not typically encountered during the development phase. In production environments, ML systems must meet stringent performance criteria and continue to operate effectively under unpredictable, dynamic conditions [5]. Unlike the controlled settings of development, where models can be fine-tuned and optimized for prediction accuracy, industrial systems must autonomously adapt to evolving data, resource constraints, real-time processing demands, and unforeseen operational changes. These systems are expected to maintain high reliability and performance over extended periods, necessitating continuous adaptation to the prevailing conditions in the operational environment. This shift demands a broader focus beyond model accuracy, prioritizing operational efficiency, robustness, and resilience in the face of real-world complexities [6].

A key factor influencing the reliability of ML systems in production is the nature of data itself [7]. Unlike development settings, where training data is static and carefully preprocessed, real-world data is dynamic, noisy, and often incomplete. The assumptions made during model development do not always hold in deployment, as data distributions evolve over time and unexpected variations emerge. These variations stem from multiple sources, including user behavior shifts, system wear, and changes in external conditions. As a result, models trained on historical data may struggle to maintain their initial level of performance when faced with new, unseen patterns [8].

In the ML literature, this disparity between development and production environments is commonly referred to as *dataset shift* [9]. A fundamental manifestation of this phenomenon is *concept drift*, where the statistical relationships between input features and target variables evolve over time [10]. Traditional ML models are typically designed under the assumption of statistical stationarity, expecting these relationships to remain fixed . However, in real-world environments, continuous changes such as evolving customer preferences, sensor degradation, or modifications in operational processes cause shifts in data distributions. Static models, unable to adapt to these ongoing changes, suffer from performance degradation, reducing their effectiveness in dynamic settings [11]. While drift detection mechanisms exist, many rely on fixed thresholds for triggering model updates, which introduces rigidity in dynamic systems. These rigid thresholds fail to account for the evolving nature of real-world applications, where adaptation should align with both the temporal context and the severity of drift [12]. Since the magnitude of change is inherently dynamic, drift detection mechanisms must adjust accordingly, ensuring timely and context-aware responses rather than relying on predefined, static criteria.

Beyond concept drift, *data quality* itself poses another major challenge in production environments. Unlike the well-curated datasets used in research, real-time production data is susceptible to noise, missing values, and inconsistencies resulting from sensor malfunctions, communication errors, or variations in data collection methods [13]. Traditional approaches to data quality assessment, which rely on offline analysis and manual intervention, become impractical in production settings where data flows continuously and decisions must be made in real time [14]. Furthermore, within the context of ML systems, data quality is not a static attribute but is instead shaped by the system's state, external conditions, and operational phase. For example, what is considered high-quality data in one operational context may be unsuitable in another, such as during system startup or seasonal transitions. This variability is closely linked to the broader challenge of drift management, as the system must adapt to changing conditions. As a result, static quality assessment methods fall short in addressing the evolving nature of data quality standards, emphasizing the importance of establishing adaptive frameworks within the context of ML systems that can dynamically adjust to the system's state and operational demands.

Although concept drift and data quality issues have been extensively studied in the ML research community, their management in real-world industrial applications remains challenging. The velocity and volume of production data create fundamental barriers to implementing many existing solutions [15]. Industrial systems often generate massive amounts of data at high speeds, requiring efficient processing mechanisms that can operate within strict time constraints [16]. Most state-of-the-art drift detection and data quality assessment techniques rely on batch processing, retrospective analysis, or human-in-the-loop interventions. These approaches, while theoretically sound, become impractical in real-time, autonomous ML systems operating under dynamic conditions.

Furthermore, the complexity of these challenges are particularly pronounced in industrial contexts where ML systems interface directly with physical processes [17]. In manufacturing environments, for instance, the interplay between equipment degradation, environmental variations, and process modifications creates a multifaceted problem space that exceeds the capabilities of traditional ML approaches. The critical interface between digital systems and physical processes demands solutions that can balance rapid adaptation with operational stability, a requirement that becomes particularly challenging when dealing with high, velocity data streams and real-time decision constraints.

The implications of these challenges extend far beyond theoretical concerns into significant practical and economic consequences. In critical industrial applications, ML system failures due to data-related issues can trigger costly production disruptions, quality issues, and safety risks. This scenario illustrates the essential need for next-generation frameworks that can effectively integrate adaptive mechanisms with quality-driven approaches while maintaining real-time performance. Such solutions must transcend current limitations, establishing new paradigms for robust, efficient, and autonomous ML systems in industrial settings. Developing such systems is essential for realizing the full potential of ML in industrial applications while minimizing risks and maximizing operational efficiency.

3 Overarching Themes

Machine learning systems in production environments require a systematic approach that integrates multiple specialized components to ensure reliable operation. These components must interact and function cohesively to address the fundamental challenges highlighted in Section 2, which extend beyond the capabilities of traditional ML approaches. MLOps has emerged as a key methodology for managing the monitoring, adaptation, and deployment of ML systems in real-world environments [18]. By providing systematic management, logging, and versioning, MLOps fosters more robust ML model deployment, ensuring that models remain effective and adaptable as they evolve in dynamic environments [19]. This makes MLOps a foundational principle upon which other components, such as drift detection and data quality assessment, are built.

Figure 1 illustrates our proposed framework, which establishes a structured flow from data generation to task execution, ensuring robust component interaction. The framework utilizes the MLOps ecosystem as its foundation, orchestrating the integration of various components while maintaining extensibility for additional services. Central to this approach is a comprehensive drift handling system that coordinates detection and adaptation mechanisms. This system acts as an intelligent monitoring driver, continuously analyzing data streams for significant changes and triggering alerts when deviations are detected. Upon identifying drift, the system appropriate responses across both the data quality assessment pipeline and ML inference components. This cohesive approach ensures that the system remains responsive and operationally relevant in dynamic environments. A key innovation lies in the framework's ability to balance real-time adaptability with operational stability, addressing a critical limitation of traditional approaches that often sacrifice one for the other. The upcoming subsections provide context for the discussed topics, aligning them with the broader scope of this thesis.

3.1 Scope and Boundaries

This research focuses on the challenges of deploying and maintaining supervised ML systems in industrial production environments characterized by non-stationary, high-velocity data streams. Although the proposed frameworks and methodologies have broad applicability, the work explicitly targets applications where ML predictions directly influence operational decisions in domains such as manufacturing and energy forecasting. The scope encompasses systems processing real-time structured time series data from sensors. These

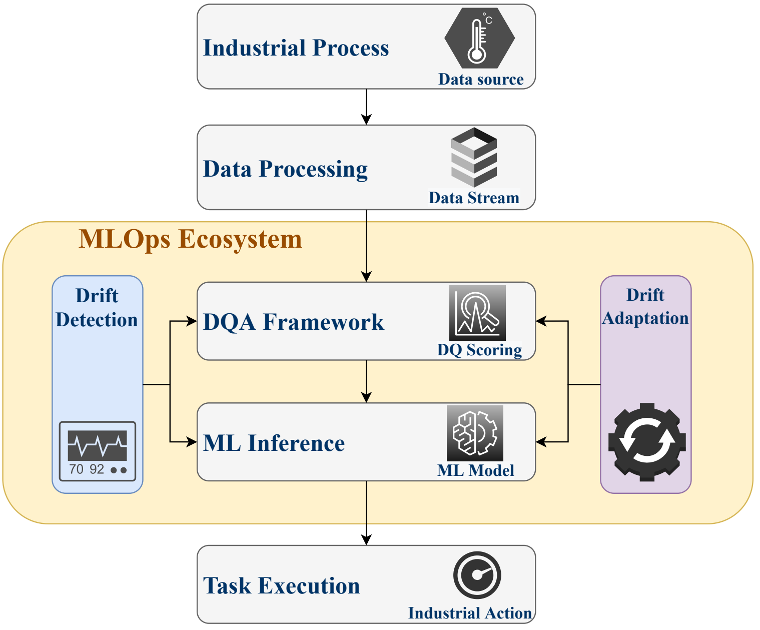

Figure 1: Key components of a robust ML operations ecosystem.

environments demand solutions that are both computationally efficient and capable of adapting in real-time to evolving data distributions. The research particularly addresses scenarios where concept drift manifests gradually due to equipment degradation, environmental shifts, or evolving operational patterns. The validation and implementation details prioritize these use cases while maintaining the broader applicability of the underlying principles.

3.2 Machine Learning Operations (MLOps) as the Foundational Framework

MLOps serves as the foundational framework for deploying and maintaining ML systems in production. Drawing upon DevOps principles while addressing the unique challenges of machine learning solutions, MLOps provides the operational backbone required to ensure robust, reproducible, and adaptive ML pipelines [20]. This foundation is indispensable, as it establishes the infrastructure for systematic management of ML artifacts, scalable deployment, and ongoing system monitoring.

The challenges of deploying ML systems stem from their dual dependency on both data and models, requiring frameworks that go beyond traditional software engineering practices [21]. Unlike conventional software, ML systems must continuously adapt to evolving data distributions, maintain performance under resource constraints, and meet stringent operational requirements. MLOps addresses these challenges through a structured lifecycle that integrates automated monitoring, version control, and adaptation mechanisms. This ensures that the models remain accurate, reliable, and efficient throughout their operational lifetime, even as the data and environmental conditions change [22].

In the context of this research, MLOps functions as a central orchestrator for industrial ML deployments, coordinating specialized components for data quality assessment and concept drift detection. As illustrated in Figure 1, the framework aligns these components with domain-specific requirements while maintaining consistent performance monitoring. The extensible architecture enables the integration of specialized components for different industrial applications, allowing the system to adapt to diverse operational requirements while preserving a unified approach to system management.

Our research contributes to this domain through a comprehensive analysis of MLOps robustness characteristics, an indispensable component of trustworthy ML systems. By establishing a systematic taxonomy that defines and operationalizes robustness in production ML systems, we provide practitioners with clear guidelines for implementing and evaluating robust MLOps pipelines. Through a systematic review of existing MLOps tools, we identify which aspects of robustness they currently support, highlighting critical gaps between available capabilities and the requirements for truly robust ML systems in production environments.

3.3 Drift Handling in Dynamic Production Environments

Managing concept drift is a critical aspect in maintaining the reliability and accuracy of ML systems in dynamic deployment environments. Unlike controlled experimental settings, production systems face constantly evolving operational conditions, where the statistical properties of input data and their relationships with target variables change over time, leading to degraded model performance and increased error rates if left unaddressed [23]. In deployment contexts, such changes can result from equipment wear, environmental fluctuations, or shifts in user behavior. These variations directly impact prediction and operational results, making effective drift management a cornerstone of robust ML deployment frameworks.

The real-world manifestations of concept drift defy simplistic categorization, with prior literature lacking consensus on taxonomy [24]. We resolve this ambiguity through a novel classification framework based on the underlying changes in the probability distribution, formally distinguishing between drift types based on their probabilistic sources and temporal transition patterns. This mathematical rigor enables targeted detection and adaptation strategies aligned with specific drift characteristics in real-world use cases.

Drift detection in production systems differs fundamentally from traditional approaches because of the need for real-time adaptability and scalability. Industrial data streams are often high-velocity and subject to frequent and unpredictable variability, making static thresholds and offline methods insufficient [25]. Our research proposes an innovative solution that employs dynamic drift thresholds that continuously monitor and analyze the distribution of observed drift magnitudes throughout the system. These adaptive threshold mechanisms dynamically adjust to evolving data patterns, enabling more effective detection of significant changes based on the prevalent conditions. When combined with automated drift-handling strategies such as real-time performance monitoring and model updates, these mechanisms ensure that systems remain aligned with operational requirements. This proactive approach minimizes disruptions and supports continuous reliability in production deployments.

Furthermore, the integration of drift handling mechanisms benefits from the MLOps principles throughout the lifecycle of the ML system [26]. Using continuous integration, deployment, and monitoring practices, drift detection mechanisms are designed to align with these principles, ensuring systematic and scalable drift management. This approach enables systems to meet strict performance requirements, which directly influence quality, safety, and efficiency in real-world applications. Through automated detection, diagnosis, and adaptation, ML systems can maintain high performance despite evolving operational challenges, ensuring reliability and robustness in dynamic environments.

3.4 Data Quality Assessment in Production Systems

Data quality assessment is a critical component of production ML systems, especially in industrial contexts where data directly impacts decision-making, operational efficiency, and model reliability. In these environments, data often originate from diverse, dynamic, and complex sources such as sensors, logs, and enterprise systems [27]. The continuous real-time influx of data requires rigorous quality control mechanisms to ensure that ML models remain accurate, robust, and aligned with operational goals. This is particularly vital in industries such as manufacturing, energy, and logistics, where ML-driven decisions can have significant environmental, financial, or safety consequences.

Real-world production systems require real-time mechanisms capable of identifying and addressing issues as they arise, moving beyond traditional offline quality checks [28]. Key challenges include detecting shifts in data distributions, identifying outliers, and managing schema changes in continuously evolving datasets. These issues, if left unaddressed, can affect ML model performance or lead to system failures. Additionally, production environments often face constraints on computational resources and latency, demanding solutions that balance thorough quality assessments with operational efficiency.

Quantitative data quality assessment methodologies represent a significant advance in addressing these challenges by providing measurable insights into the validity and integrity of data. These methods evaluate key quality dimensions, such as completeness, consistency, accuracy, and timeliness, assigning scores to quantify the health of incoming data streams [29]. This quantitative approach enables operators to make informed decisions about data suitability for downstream processes, enhancing visibility into data integrity and providing systematic quality monitoring.

Our research contributes to this domain through DQSOps (Data Quality

Scoring Operations), a novel framework that revolutionizes data quality assessment in production environments. DQSOps introduces an innovative ML prediction-based scoring approach that efficiently generates quality scores, serving as reliable indicators for industrial applications. This lightweight approach achieves significant computational speedup while maintaining high accuracy, making it particularly valuable for high-velocity industrial data streams. The framework is further extended with adaptive mechanisms that address the dynamic nature of quality dimensions, incorporating a dynamic change detector that continuously monitors and adapts to evolving data quality patterns.

At the system level, this quantitative approach ensures that ML systems consistently operate on high-quality data through automated monitoring systems that assess data characteristics in real time [30]. These systems enable proactive detection of quality deviations while maintaining scalability as data volumes grow. By embedding robust data quality evaluations into production infrastructure, industries can effectively safeguard the reliability and trustworthiness of their ML applications, mitigating data-related risks in mission-critical operations.

3.5 System-Level Synergies and Dependencies

Production ML systems require carefully integrating multiple specialized components to ensure reliable operation. Our research addresses this challenge by proposing a novel framework architecture that orchestrates the complex interactions between drift detection, data quality assessment, and model adaptation components. This architecture, illustrated in Figure 2, comprises an Integration Layer and a System Evolution Layer collaborating to achieve operational efficiency and system reliability.

The Integration Layer manages component synchronization and workflow orchestration within the MLOps ecosystem, while the System Evolution Layer handles adaptive responses and continuous optimization through the drift handling system. Together, these layers facilitate systematic monitoring and adaptation mechanisms, enabling robust performance in varying operational conditions.

The synergy between the system components is evident in their interdependencies. Drift detection relies heavily on high-quality data to accurately identify distributional changes, while data quality assessment depends on drift detection to measure data skewness and monitor data quality changes. This inherent coupling is facilitated by the bidirectional connections shown in Figure 2, where both layers coordinate the flow of Data Quality Metrics and ML Performance Metrics between the Data and Model pipelines. Similarly, adaptive model updates are effective only when supported by a precise and timely identification of performance degradation. The framework orchestrates these dependencies through systematic workflows that ensure components enhance rather than impede each other's functionality. For instance, when drift is detected, the System Evolution Layer triggers necessary adaptations while the Integration Layer coordinates model updates with data quality re-evaluation

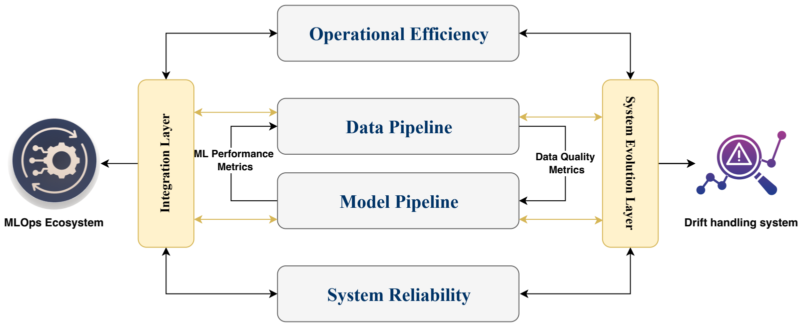

Figure 2: Synergistic framework for ML system component integration and evolution.

to maintain consistency across all components. This dual-layer coordination becomes particularly critical in time-sensitive industrial applications where component synchronization directly affects operational efficiency.

Our research validates this integrated approach through implementation in real-world industrial settings. By consolidating the integration of the data quality assessment and drift detection mechanisms, we achieved significant improvements in both the model performance and the computational efficiency. Implementation in a steel manufacturing process demonstrated a 12% improvement in model performance through optimal data quality thresholds while reducing prediction latency by a factor of four. These results confirm the framework's ability to balance prediction quality with operational efficiency in time-sensitive industrial environments.

These interdependencies highlight the importance of designing MLOps frameworks with a focus on system-level coherence. As demonstrated in the framework, the consistent transfer of data and insights between components through both layers facilitates collaborative functioning, ensuring that individual modules do not operate in isolation but contribute collectively to the goals of reliability and efficiency [31]. The bidirectional flows to both Operational Efficiency and System Reliability emphasize how maintaining reliable performance in dynamic production contexts necessitates the proactive management of these dependencies. Therefore, the achievement of the ML system's objectives is influenced by the robustness of its individual components and their effectiveness in functioning together as an integrated whole.

4 Synopsis

The primary aim of this thesis is to develop adaptive and quality-driven ML systems for industrial applications through systematic integration of drift adaptation and data quality assessment within MLOps frameworks. This thesis tackles these challenges by proposing innovative frameworks, methodologies, and adaptive techniques that enable ML systems to operate efficiently in non-

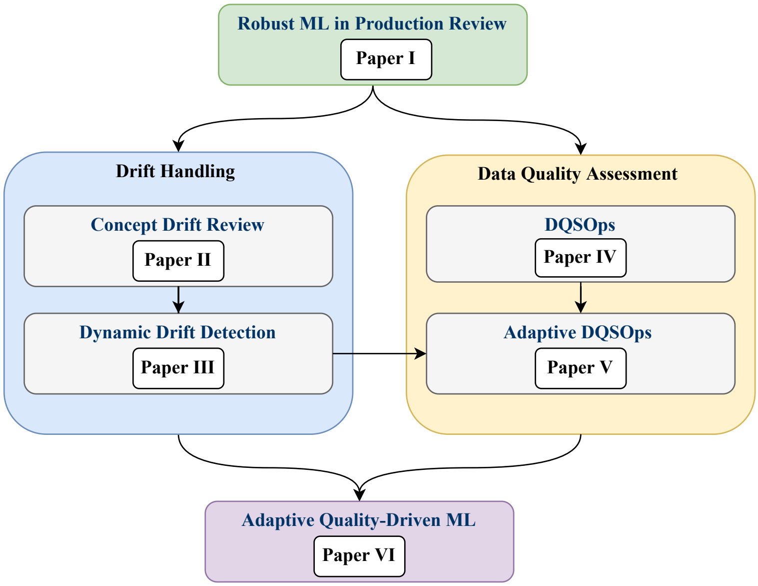

Figure 3: Alignment of the six papers with the overall framework for developing adaptive quality-driven ML systems in industrial applications.

stationary environments.

The dissertation comprises six interconnected papers that collectively establish a comprehensive framework for engineering self-adaptive ML systems. These papers address concept drift, data quality assessment, and robust MLOps practices in dynamic industrial settings. Their alignment with the phases of adaptive system design is illustrated in Figure 3.

- **Paper I** establishes the foundational principles of trustworthy and robust MLOps systems. It introduces a taxonomy of robustness dimensions and explores the integration of MLOps principles to support the reliable deployment and maintenance of ML systems in production environments. This paper aims to standardize robustness practices in ML deployment and establish clear specifications for trustworthy AI systems in production environments.

- **Paper II** presents a consolidated taxonomy of concept drift types and detection methods, addressing the ambiguity in drift terminology in the literature. The paper provides a systematic review of performance-based drift detection approaches and the types of concept drift that can be present in production environments.

- **Paper III** introduces DA-LSTM, a dynamic drift-adaptive learning framework that integrates both passive and active adaptation strategies. The framework includes a novel drift detection methodology that eliminates the need for fixed thresholds, along with adaptation mechanisms that balance prediction performance with computational costs.

- **Paper IV** proposes DQSOps, a framework for efficient data quality scoring in production environments. The framework introduces ML-based quality prediction to dramatically reduce computational overhead while maintaining assessment accuracy. The effectiveness is demonstrated through implementation in real industrial settings.

- **Paper V** extends the DQSOps framework with drift-aware mechanisms, enabling dynamic adaptation of quality assessment to evolving data characteristics. The framework integrates quality scoring with drift detection to provide a comprehensive solution for maintaining data quality assessment systems in non-stationary environments.

- **Paper VI** bridges the gap between adaptive ML frameworks and industrial applications by proposing a unified approach to handling concept drift, data quality, and operational efficiency. The paper validates this approach through an industrial use case, providing insights into the practical deployment and maintenance of robust ML systems.

The roadmap of this dissertation and its research attributes, which will be discussed in more detail in the subsequent sections, are summarized in Figure 4.

5 Research Challenges

The increasing adoption of MLOps principles has highlighted the need to ensure robustness and reliability in ML systems that operate in dynamic industrial environments. Despite the growing importance of MLOps, several challenges remain unaddressed, hindering the effective deployment and maintenance of ML solutions in production settings.

One of the primary challenges is the lack of a clear definition and operationalization of robustness within the context of ML system lifecycle management. This ambiguity leads to inconsistent practices across applications, making it difficult to build reliable pipelines capable of handling the complexities of real-world scenarios. Based on the aforementioned considerations, the first challenge is formulated to define and operationalize robustness in MLOps systems to ensure the reliable deployment and maintenance of ML solutions in production environments.

Challenge 1 (C1): Defining and operationalizing robustness in MLOps systems.

Based on the review, a fundamental characteristic of robustness is maintaining performance under changing conditions, which inherently involves adapting to concept drift [32]. Concept drift refers to the phenomenon of evolving data distributions over time. However, the existing literature reveals inconsistencies in the taxonomy of concept drift, with ambiguities in the

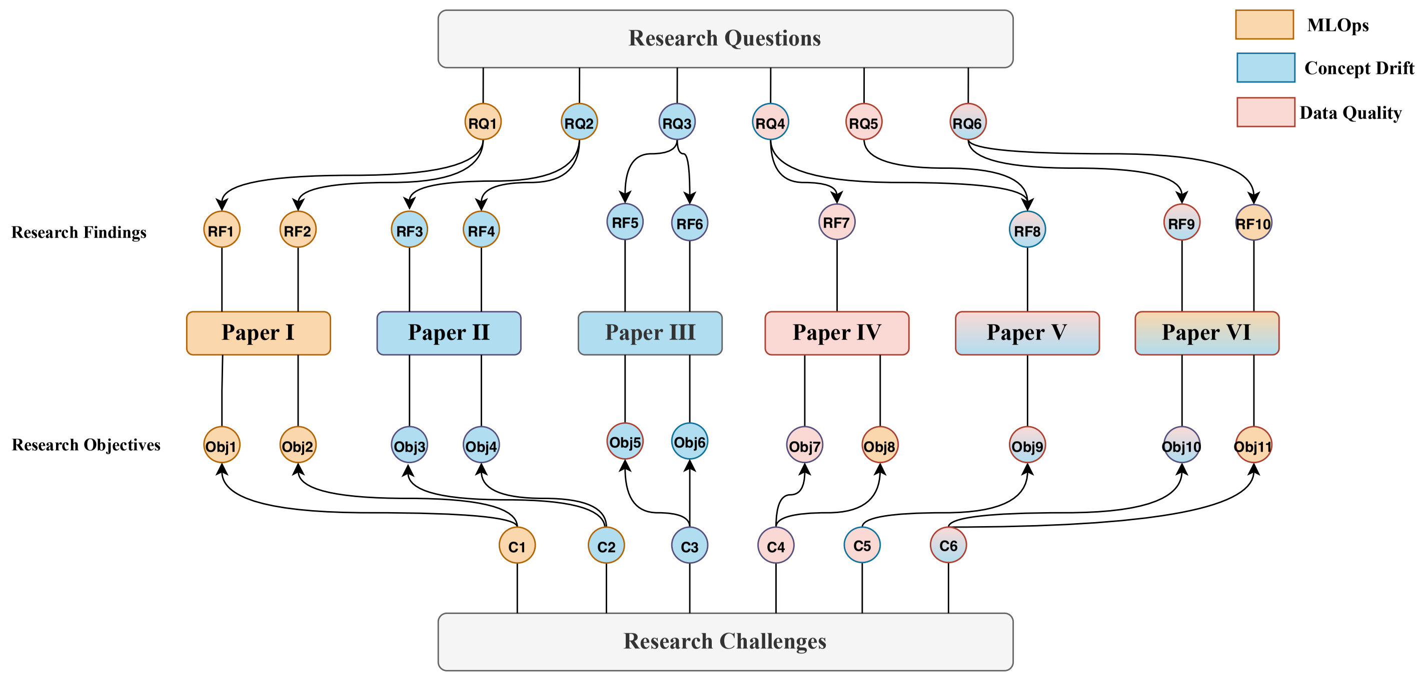

Figure 4: Connecting the dots: Mapping research questions to key findings and research challenges to the objectives of this dissertation.

categorization and definition of different drift types. These inconsistencies pose significant obstacles to the development of effective drift detection and adaptation frameworks. Consequently, the second challenge is formulated to identify and address the ambiguities in concept drift taxonomy to enable the development of robust drift detection and adaptation frameworks.

Challenge 2 (C2): Identifying and addressing ambiguities in concept drift taxonomy.

A relevant issue in the literature on drift detection is specifying a drift threshold to detect drift [25]. Drift detection mechanisms for high-velocity data streams often rely on static thresholds or pre-defined parameters, which limit their ability to respond effectively to dynamic changes. A key challenge is determining appropriate adaptive thresholds for identifying change points in the data. These thresholds must evolve with the system to maintain detection accuracy, as their configuration directly affects the performance of ML models in production. Based on these considerations, the third challenge is formulated to enable scalable and dynamic drift detection mechanisms that can effectively operate in non-stationary environments in real-world applications.

Challenge 3 (C3): Enabling scalable and dynamic drift detection mechanisms.

Data quality assessment is another crucial aspect of MLOps that poses significant challenges. Existing data quality assessment techniques often suffer from high computational complexity, making them impractical for deployment in high-volume production pipelines. These techniques may require extensive feature engineering, data preprocessing, or manual labeling, which can be time-consuming and resource-intensive [33]. Moreover, the lack of quantitative metrics for assessing data quality makes it difficult to objectively assess the suitability of data for ML tasks and to track the evolution of data quality over time. Therefore, the fourth challenge is formulated, which is developing quantitative data quality assessment methods that can operate efficiently in production environments without introducing significant computational overhead.

Challenge 4 (C4): Developing methods to assess the quality of quantitative data without added complexity

In addition, traditional data quality assessment approaches tend to focus on static quality dimensions such as accuracy, completeness, and consistency. However, in real-world scenarios, the relevance and importance of these dimensions can vary over time. For example, what is considered high-quality data with respect to an ML system may differ in various phases due to factors such as seasonality, changes in business requirements, or shifts in the underlying data distribution. This requires the development of flexible and adaptive quality assessment frameworks that can dynamically adjust to these evolving requirements [34]. These frameworks should be able to incorporate both static and dynamic quality dimensions while efficiently adapting to the changing characteristics of the data. Consequently, the fifth challenge is formulated, which is designing adaptive frameworks for data quality assessment that can handle both static and dynamic quality dimensions while adapting to evolving

data characteristics in real-time industrial applications.
Challenge 5 (C5): Designing adaptive frameworks for data quality assessment.

Lastly, integrating adaptive methodologies, such as drift adaptation techniques and quantitative data quality assessments with ML pipelines presents an opportunity to enhance the performance, reliability, and robustness of ML systems in production environments. By incorporating drift adaptation methodologies and data quality metrics directly into the ML workflow, it becomes possible to identify and address quality issues at the source while simultaneously adapting to evolving data characteristics. This approach can lead to an improved overall quality of the training data and the resulting models. However, aligning these assessments and adaptive methodologies with ML frameworks while balancing computational efficiency and operational constraints in real-world industrial settings remains a significant challenge. Based on these considerations, the sixth challenge is formulated, which is leveraging adaptive methodologies and quantitative data quality assessment to improve the performance, reliability, and robustness of ML systems in production environments.
Challenge 6 (C6): Leveraging adaptive methodologies and quantitative data quality assessment to improve ML model performance.

6 Research Questions

To address the challenges described in Section 5, a set of research questions (RQ) has been formulated to guide the exploration and development of innovative solutions to improve the robustness, adaptability, and reliability of ML systems in dynamic, real-world environments.

Robustness in MLOps encompasses the ability of ML systems to maintain performance and reliability despite evolving data distributions, unexpected inputs, and changing environmental conditions. Establishing clear guidelines and best practices for robust ML solutions requires a deeper understanding of the factors influencing robustness and the development of effective strategies to optimize it. Building on the challenge of defining and operationalizing robustness in MLOps systems (C1), the first research question is formulated as follows:

RQ1: How can robustness be defined and operationalized in MLOps systems to ensure reliable deployment and maintenance of ML solutions in production environments?

Concept drift, characterized by changes in the statistical properties of the target variable over time, often leads to performance degradation in ML models. The lack of a unified taxonomy for concept drift creates ambiguities that hinder the development of effective drift detection and adaptation strategies. To address this, a clearer understanding of drift types and their real-world implications is needed. Inspired by the challenge of addressing ambiguities in the concept drift taxonomy (C2), the second research question is formulated as follows:

RQ2: What are the fundamental taxonomic ambiguities in the characterization of concept drift, and how do different types of drift emerge and evolve in real-world ML systems, influencing model performance?

Current drift detection methods often rely on static thresholds or predefined parameters, which are not suitable for high-velocity data streams and dynamic environments. Developing adaptive mechanisms that can adapt to evolving data characteristics is critical to maintaining detection accuracy. Guided by the challenge of enabling scalable and dynamic drift detection mechanisms (C3), the third research question is formulated as follows:

RQ3: What novel methodologies can be devised for dynamic drift detection in environments with evolving data distributions and high variability, particularly in non-stationary settings?

Data quality assessment is a cornerstone of reliable ML systems, yet existing techniques often suffer from high computational complexity, making them impractical for production pipelines. Developing efficient and scalable methods for assessing data quality is essential for effective integration into real-world workflows. Motivated by the challenge of developing quantitative data quality assessment methods without added complexity (C4), the fourth research question is formulated as follows:

RQ4: How can a computationally efficient and quantitatively data quality assessment framework be constructed, capable of operating within production environments without introducing prohibitive computational overhead?

Traditional data quality frameworks often focus on static dimensions, such as accuracy and completeness, which may not account for the dynamic nature of real-world data. In non-stationary environments, the relevance of these dimensions can shift over time, necessitating adaptive frameworks that accommodate evolving requirements. Building on the challenges of developing quantitative data quality assessment methods (C4) and designing adaptive frameworks (C5), the fifth research question is formulated as follows:

RQ5: How can data quality scoring frameworks be designed to efficiently handle static and dynamic quality dimensions while adapting to evolving data characteristics in real-time industrial applications?

Integrating adaptive data quality assessment and drift adaptation techniques into ML pipelines offers significant potential to improve system performance, reliability, and robustness. However, aligning these methodologies with ML workflows while maintaining computational efficiency and operational feasibility remains a complex challenge. Drawing from the challenges of enabling dynamic drift detection (C3), designing adaptive frameworks (C5), and leveraging data quality assessment to improve ML performance (C6), the sixth research question is formulated as follows:

RQ6: How can the fusion of dynamic drift adaptation with adaptive data quality assessment techniques improve the overall performance, robustness, and reliability of ML systems deployed in production environments while maintaining computational efficiency and scalability?

By addressing these research questions, this work aims to advance ML research and contribute to the development of robust, adaptive, and reliable ML systems capable of thriving in dynamic, real-world environments. The insights gained will provide valuable guidance to practitioners and researchers, ultimately leading to improved system performance, enhanced reliability, and better decision-making capabilities in real-world industrial applications.

7 Research Methodology

This thesis employs a multi-faceted methodological approach—integrating theoretical analysis, framework development, and empirical validation—to comprehensively address the research questions outlined in Section 6. The methodological design balances academic rigor with practical applicability, ensuring that the proposed solutions are both theoretically sound and industrially relevant. By combining literature review, taxonomy development, computational experiments, and industrial case studies, the research bridges critical gaps between macML theoretical foundations and real-world deployment challenges.

7.1 Literature Review

To address RQ1 and RQ2, a comprehensive literature review was conducted, analyzing 107 peer-reviewed papers—41 on MLOps robustness and 66 on concept drift. The objective was to investigate the key factors that define robustness aspects within MLOps frameworks (RQ1) and to resolve taxonomic ambiguities in concept drift (RQ2). The review included a comparative analysis of definitions and methodologies across multiple domains to establish a structured understanding of these challenges.

7.2 Taxonomic Exploration

Mathematical frameworks were developed to characterize concept drift (RQ2) and data quality dimensions (RQ4). Probabilistic models of distributional changes ($P(X, y)$) were employed to formalize drift terminology, while data quality metrics were operationalized through measurable indicators aligned with industrial standards. This theoretical foundation enabled precise definitions of drift types and quality attributes, ensuring a systematic approach to framework design.

7.3 Algorithmic Framework Design

Addressing RQ3, RQ4, and RQ5, the research developed innovative algorithmic frameworks for adaptive machine learning deployment. The DA-LSTM framework was designed for dynamic drift adaptation to maintain model accuracy in non-stationary environments. Additionally, the DQSOps framework and its extension were introduced as scalable data quality scoring mechanisms, enabling efficient real-time quality assessment in production pipelines.

7.4 Industrial Case Studies

To comprehensively validate the proposed methodologies (RQ3, RQ4, RQ5, RQ6), the research involved thorough evaluations through industrial case studies. Real-world datasets from electricity load forecasting and steel manufacturing were utilized to rigorously assess the frameworks' predictive performance, computational efficiency, and adaptive capabilities. The case studies provided empirical evidence of the frameworks' robustness, usability, and impact on decision-making processes, demonstrating their potential for large-scale industrial adoption.

8 Research Contributions

This thesis contributes to the development of ML systems in dynamic industrial environments, ensuring their long-term operation in complex real-world settings. It establishes a comprehensive framework that integrates concept drift detection, adaptive learning mechanisms, and real-time data quality assessment into cohesive MLOps pipelines, components that are typically addressed separately, therefore ensuring that ML models remain effective as production systems evolve. The primary contribution is an end-to-end methodology that enables ML systems to continuously detect and respond to changing data characteristics while managing data quality without introducing computational overhead over time.

The research encompasses six papers that address different aspects of ML robustness in production environments. Table 1 presents the key quantitative achievements in these articles, demonstrating significant improvements in areas such as computational efficiency, prediction accuracy, and system scalability. Building upon these quantitative results, the following sections detail the research objectives that guided this work and the key findings that emerged from each paper.

8.1 Research Objectives

The primary research objectives of this study, referred to as Research Objectives (**Objs**), are:

Table 1: Summary of research papers and their corresponding key quantitative contributions.

Paper	Key Quantitative Results
Paper I	• Identification of **12 robustness dimensions** • **41 papers** reviewed on robustness in ML production • **5 major MLOps tools** analyzed
Paper II	• **66 papers** analyzed (2011–2021) • **40 drift-related terms** identified in literature • **14 drift types** reviewed and defined
Paper III	• Up to **64.03% accuracy improvement** • **44.48% RMSE reduction** • **53% computational cost reduction**
Paper IV	• **819x processing speedup** achieved • **R^2 maintained above 0.89** • Processing time reduced from **0.94s to 0.001s**
Paper V	• **52.64% accuracy improvement** ($\tau = 0.15$) • **88% reduction in computational overhead** • **R^2 up to 0.978** achieved ($\tau = 0.09$)
Paper VI	• **12% model performance improvement** • **$R^2 = 0.94$** in production settings • **4x reduction in prediction latency**

Paper I

Obj1: To provide a comprehensive overview of robustness in MLOps systems, defining its key dimensions and operationalizing it for reliable continuous deployment of ML solutions in production environments.

Obj2: To survey existing approaches and tools that address robustness in MLOps, identifying gaps and proposing future directions for research and practice.

Paper II

Obj3: To consolidate the taxonomy of concept drift types, addressing ambiguities in terminology and providing a unified classification based on mathematical definitions.

Obj4: To survey the types of concept drift that may appear in real-world applications.

Paper IV

Obj5: To develop a novel drift detection methodology that eliminates the need for fixed thresholds, enabling more flexible and accurate detection of concept drift.

Obj6: To propose a dynamic drift-adaptive learning framework (DA-LSTM) for interval load forecasting, integrating both passive and active adaptation strategies.

Paper IV

Obj7: To propose a scalable and efficient data quality scoring framework (DQSOps) for data-driven applications in industrial settings.

Obj8: To introduce ML-based quality prediction mechanisms that reduce computational overhead while maintaining high accuracy in data quality assessment.

Paper V

Obj9: To extend the DQSOps framework with drift-aware mechanisms, enabling dynamic adaptation of data quality assessment to evolving data characteristics.

Paper VI

Obj10: To propose a unified approach for handling concept drift, data quality, and operational scalability in industrial ML systems.

Obj11: To validate the proposed approach through industrial use cases, demonstrating its ability to improve ML performance in dynamic environments.

8.2 Overview of Research Findings

The following points summarize the key Research Findings (**RFs**) from the papers:

Paper I

RF1: A formalized understanding of robustness in MLOps systems was established, highlighting its importance in production-grade ML pipelines.

RF2: Existing MLOps tools were found to lack sufficient support for adaptive and scalable robustness practices, indicating areas for improvement.

Paper II

RF3: A unified taxonomy of concept drift types was established, clarifying ambiguities in terminology and providing a systematic classification based on probabilistic sources and transition patterns.

RF4: The review of performance-based drift detection methods highlighted the importance of adaptive and hybrid approaches for improving detection accuracy and applicability in real-world ML systems.

Paper III

RF5: The proposed drift detection methodology, which eliminates the need for fixed thresholds, proved to be more flexible and accurate in identifying concept drift in high-velocity data streams.

RF6: A trade-off analysis revealed that the DA-LSTM framework offers a balanced approach between prediction performance and computational cost, making it suitable for real-world industrial applications.

Paper IV

RF7: The DQSOps framework significantly reduced computational overhead while maintaining high accuracy in data quality assessment, making it suitable for scalable industrial processes.

Paper V

RF8: The drift-aware extension of the DQSOps framework demonstrated its ability to dynamically adapt to evolving data characteristics, ensuring consistent data quality assessment in non-stationary environments.

Paper VI

RF9: The unified approach to handling concept drift, data quality, and operational scalability was successfully validated in various industrial use cases, demonstrating its effectiveness in real-world applications.

RF10: The proposed approach provides a comprehensive solution to maintaining robust ML systems in dynamic environments, addressing key challenges such as drift detection, data quality assessment, and system scalability.

9 Limitations

While this research makes significant contributions to the field of ML systems in production environments, several limitations should be acknowledged to provide context for the findings and guide future research directions:

9.1 Model-Specific Constraints

The frameworks and methodologies developed in this research focus mainly on supervised learning scenarios with structured time series data. Although the proposed approaches are model-agnostic and designed to be applicable across various ML algorithms, they have been primarily validated using traditional ML models, such as ensemble-based methods and basic neural networks. This focus was intentional, prioritizing computational efficiency and real-time processing capabilities critical for industrial applications. However, this focus means that the frameworks' effectiveness with more complex architectures, such as deep learning models or transformer-based architectures, requires further investigation.

9.2 Data Characteristics

The research predominantly explores data streams with relatively stable feature spaces and well-defined quality dimensions, focusing primarily on univariate data for drift detection. However, the same methodologies can be applied to multivariate data streams, provided that appropriate data distributions are constructed to account for the relationships between variables. Although the principles of the proposed methods remain unchanged, adapting them to multivariate scenarios requires extending the underlying frameworks to capture more complex data interactions. Similarly, additional data quality dimensions can be defined based on specific use case requirements without altering the core approach.

9.3 Resource Optimization

The framework currently focuses on balancing computational efficiency with performance accuracy. Although this approach is effective for validated use cases, there is potential to further optimize resource utilization, particularly in scenarios involving multiple concurrent data streams or distributed processing requirements. These optimizations could enhance the framework's scalability without compromising its real-time processing capabilities.

10 Conclusions and Outlook

Machine learning systems in production environments must exhibit high robustness and reliability to operate effectively in dynamic real-world conditions. While traditional offline machine learning focuses on solving isolated problems in static environments, deploying ML systems in production requires managing multiple interacting components while maintaining consistent performance. This thesis addresses the fundamental challenges of implementing ML in production environments, where concept drift, data quality issues, and operational constraints can systematically affect model reliability.

The research bridges theoretical advances with practical implementation by establishing a comprehensive framework for engineering self-adaptive ML systems. Key contributions include novel methodologies for dynamic concept drift detection without fixed thresholds, efficient real-time data quality assessment integrated within MLOps workflows, and robust adaptation strategies that balance performance with computational costs. The framework demonstrated significant improvements in both prediction accuracy and processing efficiency when evaluated in real-world industrial use cases. Specifically, the case studies showed stable behavior and consistent performance in handling dynamic changes, validating its suitability for time-sensitive industrial applications. This work provides valuable insights for researchers and practitioners looking to deploy reliable ML solutions in production environments while maintaining high performance under dynamic conditions.

Looking ahead, future research should extend this work to several critical dimensions based on the limitations presented in Section 9. Enhancing the scalability of the framework to handle various types of data, such as multimodal or unstructured data, and a wider array of industrial processes would significantly broaden its applicability. Ethical considerations, including bias detection and fairness-aware adaptation, must be deeply integrated into MLOps pipelines to ensure trustworthy and transparent deployments. Additionally, exploring synergies with emerging technologies, such as federated learning for decentralized data quality assessment, could further enhance resilience in non-stationary environments. Sustainability should also be a central focus of future advancements, prioritizing energy-efficient algorithms and resource optimization to align machine learning deployments with global environmental goals. By harmonizing scalability, ethics, technological agility, and environmental stewardship, the next generation of adaptive ML systems will thrive in the face of real-world unpredictability while evolving responsibly, ensuring long-term reliability, ethical integrity, and scalability across industries.

References

[1] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, "Machine learning for industrial applications: A comprehensive literature review," Expert Systems with Applications, vol. 175, p. 114820, 2021.

[2] R. Rai, M. K. Tiwari, D. Ivanov, and A. Dolgui, "Machine learning in manufacturing and industry 4.0 applications," 2021.

[3] H. ElMaraghy, L. Monostori, G. Schuh, and W. ElMaraghy, "Evolution and future of manufacturing systems," CIRP Annals, vol. 70, no. 2, pp. 635–658, 2021.

[4] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," World Information Technology and Engineering Journal, vol. 10, no. 07, pp. 3897–3904, 2023.

[5] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," SN computer science, vol. 2, no. 3, p. 160, 2021.

[6] S. Sahoo, S. Kumar, M. Z. Abedin, W. M. Lim, and S. K. Jakhar, "Deep learning applications in manufacturing operations: a review of trends and ways forward," Journal of Enterprise Information Management, vol. 36, no. 1, pp. 221–251, 2023.

[7] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," Expert Systems with Applications, vol. 166, p. 114060, 2021.

[8] A. Paleyes, R.-G. Urma, and N. D. Lawrence, "Challenges in deploying machine learning: a survey of case studies," ACM computing surveys, vol. 55, no. 6, pp. 1–29, 2022.

[9] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, Dataset shift in machine learning. Mit Press, 2022.

[10] I. Žliobaitė, M. Pechenizkiy, and J. Gama, "An overview of concept drift applications," Big data analysis: new algorithms for a new society, pp. 91–114, 2016.

[11] H. Yu, W. Liu, J. Lu, Y. Wen, X. Luo, and G. Zhang, "Detecting group concept drift from multiple data streams," Pattern Recognition, vol. 134, p. 109113, 2023.

[12] D. M. V. Sato, S. C. De Freitas, J. P. Barddal, and E. E. Scalabrin, "A survey on concept drift in process mining," ACM Computing Surveys (CSUR), vol. 54, no. 9, pp. 1–38, 2021.

[13] C. A. Escobar, M. E. McGovern, and R. Morales-Menendez, "Quality 4.0: a review of big data challenges in manufacturing," Journal of Intelligent Manufacturing, vol. 32, no. 8, pp. 2319–2334, 2021.

[14] J. Wang, Y. Liu, P. Li, Z. Lin, S. Sindakis, and S. Aggarwal, "Overview of data quality: Examining the dimensions, antecedents, and impacts of data quality," Journal of the Knowledge Economy, vol. 15, no. 1, pp. 1159–1178, 2024.

[15] A. L. Suárez-Cetrulo, D. Quintana, and A. Cervantes, "A survey on machine learning for recurring concept drifting data streams," Expert Systems with Applications, vol. 213, p. 118934, 2023.

[16] J. Parmar, S. Chouhan, V. Raychoudhury, and S. Rathore, "Open-world machine learning: applications, challenges, and opportunities," ACM Computing Surveys, vol. 55, no. 10, pp. 1–37, 2023.

[17] R. A. Khalil, N. Saeed, M. Masood, Y. M. Fard, M.-S. Alouini, and T. Y. Al-Naffouri, "Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications," IEEE Internet of Things Journal, vol. 8, no. 14, pp. 11016–11040, 2021.

[18] L. Faubel, K. Schmid, and H. Eichelberger, "Mlops challenges in industry 4.0," SN Computer Science, vol. 4, no. 6, p. 828, 2023.

[19] M. M. John, H. H. Olsson, and J. Bosch, "Towards mlops: A framework and maturity model," in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 1–8, IEEE, 2021.

[20] J. Diaz-De-Arcaya, A. I. Torre-Bastida, G. Zárate, R. Miñón, and A. Almeida, "A joint study of the challenges, opportunities, and roadmap of mlops and aiops: A systematic survey," ACM Computing Surveys, vol. 56, no. 4, pp. 1–30, 2023.

[21] C. Amrit and A. K. Narayanappa, "An analysis of the challenges in the adoption of mlops," Journal of Innovation & Knowledge, vol. 10, no. 1, p. 100637, 2025.

[22] J. Antony, D. Jalušić, S. Bergweiler, Á. Hajnal, V. Žlabravec, M. Emődi, D. Strbad, T. Legler, and A. C. Marosi, "Adapting to changes: A novel framework for continual machine learning in industrial applications," Journal of Grid Computing, vol. 22, no. 4, p. 71, 2024.

[23] G. A. Lewis, S. Echeverría, L. Pons, and J. Chrabaszcz, "Augur: A step towards realistic drift detection in production ml systems," in Proceedings of the 1st Workshop on Software Engineering for Responsible AI, pp. 37–44, 2022.

[24] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," Pattern recognition, vol. 45, no. 1, pp. 521–530, 2012.

[25] A. Liu, J. Lu, Y. Song, J. Xuan, and G. Zhang, "Concept drift detection delay index," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 5, pp. 4585–4597, 2022.

[26] L. Piano, F. Garcea, V. Gatteschi, F. Lamberti, and L. Morra, “Detecting drift in deep learning: A methodology primer,” IT Professional, vol. 24, no. 5, pp. 53–60, 2022.

[27] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, “Sensor data quality: A systematic review,” Journal of Big Data, vol. 7, no. 1, p. 11, 2020.

[28] G. Aad, B. Abbott, D. C. Abbott, A. A. Abud, K. Abeling, D. K. Abhayasinghe, S. H. Abidi, O. AbouZeid, N. L. Abraham, H. Abramowicz, et al., “Atlas data quality operations and performance for 2015–2018 data-taking,” Journal of instrumentation, vol. 15, no. 04, pp. p04003–p04003, 2020.

[29] L. Cai and Y. Zhu, “The challenges of data quality and data quality assessment in the big data era,” Data science journal, vol. 14, pp. 2–2, 2015.

[30] S. Schelter, D. Lange, P. Schmidt, M. Celikel, F. Biessmann, and A. Grafberger, “Automating large-scale data quality verification,” Proceedings of the VLDB Endowment, vol. 11, no. 12, pp. 1781–1794, 2018.

[31] C. K. Tantithamthavorn, F. Palomba, F. Khomh, and J. J. Chua, “Mlops, llmops, fmops, and beyond,” IEEE Software, vol. 42, no. 01, pp. 26–32, 2025.

[32] A. Manure and S. Bengani, “Robustness and reliability,” in Introduction to Responsible AI: Implement Ethical AI Using Python, pp. 133–158, Springer, 2023.

[33] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer, “Quality assessment for linked data: A survey,” Semantic Web, vol. 7, no. 1, pp. 63–93, 2016.

[34] A. Karkouch, H. Mousannif, H. Al Moatassime, and T. Noel, “Data quality in internet of things: A state-of-the-art survey,” Journal of Network and Computer Applications, vol. 73, pp. 57–81, 2016.

List of Appended Papers

1. **Firas Bayram**, Bestoun S. Ahmed. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. *ACM Computing Surveys, 57(5), Article 121, May 2025.*

2. **Firas Bayram**, Bestoun S. Ahmed, Andreas Kassler. From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors. *Knowledge-Based Systems, 245, 108632, 2022.*

3. **Firas Bayram**, Phil Aupke, Bestoun S. Ahmed, Andreas Kassler, Andreas Theocharis, Jonas Forsman. DA-LSTM: A Dynamic Drift-Adaptive Learning Framework for Interval Load Forecasting with LSTM Networks. *Engineering Applications of Artificial Intelligence, 123, 106480, 2023.*

4. **Firas Bayram**, Bestoun S. Ahmed, Erik Hallin. DQSOps: Data Quality Scoring Operations Framework for Data-Driven Applications. *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, 2023.*

5. **Firas Bayram**, Bestoun S. Ahmed, Erik Hallin. Adaptive Data Quality Scoring Operations Framework using Drift-Aware Mechanism for Industrial Applications. *Journal of Systems and Software, 217, 112184, 2024.*

6. **Firas Bayram**, Bestoun S. Ahmed, Erik Hallin. End-to-End Data Quality-Driven Framework for Machine Learning in Production Environment. *Submitted the first revision to Heliyon.*

Comments on my Participation

Paper I I designed and conducted the systematic survey methodology, including the identification, extraction, and analysis of MLOps methods and tools. I performed the investigation and wrote the original manuscript. Bestoun contributed to the conceptualization of the study, and reviewed and edited the manuscript

Paper II I designed the research methodology, conducted the comprehensive literature review of drift detection methods, developed the mathematical taxonomy and hierarchical classification of concept drift types, and wrote the original manuscript. Bestoun provided feedback on the methodology and reviewed and edited the manuscript, while Andreas reviewed and edited the manuscript.

Paper III I developed the drift-adaptive methodology including both active and passive adaptation strategies, performed the conceptualization of the DA-LSTM framework, conducted the analysis and experimentation, and wrote the original manuscript. Phil developed the LSTM model and conducted the experimentation, while Bestoun, Andreas K, and Andreas T contributed to the conceptualization, editing, and provided comments on the results analysis. Jonas provided feedback on the results.

Paper IV I developed the DQSOps methodology, conducted the analysis and experimental evaluation of the framework performance, and wrote the original manuscript. Bestoun contributed to the design of the methodology and reviewed and edited the manuscript. Erik provided the industry use case and experiments, contributing to the real-world validation of the framework.

Paper V I developed the adaptive data quality scoring methodology integrating drift-aware mechanisms, conducted the analysis and experimental evaluation of the framework, validated and visualized the results, and wrote the original manuscript. Bestoun contributed to the design of the methodology and reviewed and edited the manuscript. Erik provided the industry use case and experiments, contributing to the real-world validation of the framework.

Paper VI I designed and implemented the framework architecture, developed the adaptive scoring mechanism, conducted the industrial case study, and wrote the manuscript. I also led the experimental validation and analysis of results. Bestoun contributed to the design of the methodology and reviewed and edited the manuscript. Erik provided the industry use case and experiments, contributing to the real-world validation of the framework.

Other Publications

- **Firas Bayram**, and Bestoun S. Ahmed. *"A domain-region based evaluation of ML performance robustness to covariate shift." Neural Computing and Applications*, 35(24) (2023): 17555-17577.
- Chahed, Hamza, Muhammad Usman, Ayan Chatterjee, **Firas Bayram**, Rajat Chaudhary, Anna Brunstrom, Javid Taheri, Bestoun S. Ahmed, and Andreas Kassler. *"AIDA—A holistic AI-driven networking and processing framework for industrial IoT applications." Internet of Things*, 22 (2023): 100805.
- **Firas Bayram**, Bestoun S. Ahmed, Erik Hallin, and Anton Engman. *"A Drift Handling Approach for Self-Adaptive ML Software in Scalable Industrial Processes."* In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, pp. 1-5. 2022.

Machine Learning in Motion

In the era of AI-driven innovation, ensuring the reliability and adaptability of machine learning (ML) systems in dynamic environments remains a fundamental challenge. This thesis addresses key obstacles such as concept drift and data quality, which affect model performance in real-world applications. By integrating novel drift detection mechanisms, real-time data quality assessment, and robust MLOps strategies, this work proposes a comprehensive framework for self-adaptive ML systems. These solutions enable industrial ML deployments to maintain predictive accuracy, efficiency, and resilience against evolving data conditions. Through real-world case studies, this research demonstrates significant improvements in operational stability and decision-making, bridging the gap between theoretical advancements and practical implementation. By tackling these challenges, this thesis contributes to the development of scalable, trustworthy, and high-performance ML solutions tailored for industrial and mission-critical applications.

ISBN 978-91-7867-560-9 (print)

ISBN 978-91-7867-561-6 (pdf)

ISSN 1403-8099

DOCTORAL THESIS | Karlstad University Studies | 2025:14

Machine Learning in Motion

In the era of AI-driven innovation, ensuring the reliability and adaptabilityof machine learning (ML) systems in dynamic environments remains a fundamental challenge. This thesis addresses key obstacles such as concept drift and data quality, which affect model performance in real-world applications. By integrating novel drift detection mechanisms, real-time data quality assessment, and robust MLOps strategies, this work proposes a comprehensive framework for self-adaptive ML systems. These solutions enable industrial ML deployments to maintain predictive accuracy, efficiency, and resilience against evolving data conditions. Through real-world case studies, this research demonstrates significant improvements in operational stability and decision-making, bridging the gap between theoretical advancements and practical implementation. By tackling these challenges, this thesis contributes to the development of scalable, trustworthy, and high-performance ML solutions tailored for industrial and mission-critical applications.

DOCTORAL THESIS
Karlstad University Studies, 2025:14
ISBN 978-91-7867-560-9 (print)
ISBN 978-91-7867-561-6 (pdf)