

A Systematic Literature Review of Machine Learning Approaches for Migrating Monolithic Systems to Microservices

Imen Trabelsi, Brahim Mahmoudi, Jean Baptiste Minani, Naouel Moha, and Yann-Gaël Guéhéneuc

Abstract—Scalability and maintainability challenges in monolithic systems have led to the adoption of microservices, which divide systems into smaller, independent services. However, migrating existing monolithic systems to microservices is a complex and resource-intensive task, which can benefit from machine learning (ML) to automate some of its phases. Choosing the right ML approach for migration remains challenging for practitioners. Previous works studied separately the objectives, artifacts, techniques, tools, and benefits and challenges, of migrating monolithic systems to microservices. No work has yet investigated systematically existing ML approaches for this migration to understand the migration phases automated, inputs used, ML techniques applied, evaluation processes followed, and challenges encountered. We present a systematic literature review (SLR) that aggregates, synthesizes, and discusses the approaches and results of 81 primary studies (PSs) published between 2015 and 2024. We followed the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement to report our findings and answer our research questions (RQs). We extract and analyse data from these PSs to answer our RQs. We synthesise the findings in the form of a classification that shows the usage of ML techniques in migrating monolithic systems to microservices. The findings reveal that some phases of the migration process, such as monitoring and service identification, are well-studied, while others, like packaging microservices, remain unexplored. Additionally, the findings highlight key challenges, including limited data availability, scalability and complexity constraints, insufficient tool support, and the absence of standardized benchmarking, emphasizing the need for more holistic solutions.

Index Terms—Systematic Literature Review, Microservices, Machine Learning, Migration

I. INTRODUCTION

As software systems evolve, they often grow in size and complexity due to the continuous addition of new functionalities. This growth frequently results in tightly coupled components with diminished cohesion, leading to architectural and operational challenges that impede scalability, maintainability,

and deployment. [1]. Traditionally, such systems have been built using monolithic architectures, where functionality is centralised within large, interconnected components [2]. While this architecture was widely adopted in the past, the increasing demand for flexibility and faster release cycles has revealed its inherent limitations [3].

In response to these limitations, microservice architecture has emerged as a solution to divide monolithic systems into or build new systems as smaller, independent services [4]. This division enables improved scalability, faster deployment, and more efficient maintenance, making microservices an attractive alternative for modern software development and evolution [5].

However, migrating existing monolithic systems to a microservices architecture remains a complex and resource-intensive task [6]. It involves identifying service boundaries and packaging them into self-contained microservices with defined APIs. Effective migration further requires robust strategies for deploying microservices and guaranteeing their scalability, security, and fault tolerance. Additionally, continuous monitoring is essential to ensure optimal functionality and maintain the system's operational health.

While many approaches have been proposed to address the various phases of migrating monolithic systems to microservices, such as heuristic-based methods, and domain-driven design, these techniques often require manual intervention and lack adaptability across different systems [7]. Recent advancements in machine learning (ML) offer a promising avenue to tackle these challenges, including identifying service boundaries, analysing interdependencies, and predicting failure. ML techniques excel at processing large and complex datasets, uncovering patterns, and supporting decision-making processes that are otherwise difficult to achieve with traditional approaches. For example, ML can assist in automating repetitive and error-prone tasks, such as clustering related components or optimising deployment strategies, potentially reducing migration time and improving accuracy and consistency [8]. [R3C3] Although business process (BP) analysis and modernisation are often crucial first steps in practice, most ML-based studies focus on migration tasks involving source code or system architecture. A few works leverage business process models as input, but typically assume these models are already available. The automation of BP discovery and refactoring prior to migration remains largely unexplored and presents a valuable opportunity for future research using NLP, process mining, or large language models.

This systematic literature review (SLR) aims to explore

Imen Trabelsi is with the Department of Computer Science, École de Technologie Supérieure (ÉTS) – Université du Québec, Canada
E-mail: iman.trabelssi@gmail.com

Brahim Mahmoudi is with the Department of Computer Science, École de Technologie Supérieure (ÉTS) – Université du Québec, Canada
E-mail: brahim.mahmoudi.1@ens.etsmtl.ca

Jean Baptiste Minani is with the Department of Computer Science and Software Engineering, Concordia University, Canada
E-mail: baptiste2k8@gmail.com

Naouel Moha is with the Department of Computer Science, École de Technologie Supérieure (ÉTS) – Université du Québec, Canada
E-mail: moha.naouel@etsmtl.ca

Yann-Gaël Guéhéneuc is with the Department of Computer Science and Software Engineering, Concordia University, Canada
E-mail: yann-gael.gueheneuc@concordia.ca

the role of ML in this domain by addressing key research questions, such as the phases of migration it automates, the characteristics of the input data it leverages, the types of ML approaches applied, how these approaches are evaluated, and the challenges faced by researchers. By answering these questions, this study seeks to provide a comprehensive understanding of the current state of ML-based migration techniques and identify opportunities for future research.

Overall, the findings suggest that while certain phases of the migration process from monolithic systems to microservices, such as monitoring and service identification, have been extensively studied, other phases remain in their early stages of exploration [9]. Notably, there is a significant gap in approaches specifically designed for the later stages of the migration process, including packaging microservices, generating the necessary code for microservice APIs, and implementing design patterns critical to their operation. Despite the progress made in some phases like identification, deployment and monitoring, other phases like pre-migration and packaging are not treated. We also found that the current approaches face several challenges that hinder their adoption in practical scenarios. These challenges include the insufficient availability of high-quality data, which is essential for training reliable machine learning models, as well as scalability and complexity concerns that restrict the applicability of these methods to large-scale and real-world systems [10]. Furthermore, many existing techniques lack tool support, making them difficult to implement effectively in the industry. The absence of standardised metrics, datasets, and baselines further complicates the evaluation and comparison of these approaches, limiting the ability to measure progress and identify best practices [11].

While several studies have explored specific aspects of microservice migration and the associated challenges, to the best of our knowledge, no prior work investigated systematically existing ML approaches for this migration to understand the automated phases, used inputs, applied ML techniques, followed evaluation processes, and encountered challenges.

[R4C25] The goal of this study is to understand how machine learning techniques are used in the migration of monolithic systems to microservices. From this goal, we derive the following research questions (RQs):

- ❶ **RQ1:** Which migration phases are automated by ML?
- ❷ **RQ2:[R1C1]** How are ML migration inputs characterised in terms of type, granularity, source, and preprocessing?
- ❸ **RQ3:[R1C1]** What ML techniques and learning paradigms are applied in migration approaches?
- ❹ **RQ4:[R1C1]** How are ML-based migration approaches evaluated with respect to metrics, benchmarks, tools, and success criteria?
- ❺ **RQ5:** What challenges arise in using ML approaches?

To answer these research questions, we followed the updated Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement for reporting systematic reviews [12]. We screened 2,301 potentially relevant studies from eight digital libraries, considering publications between 2015 and 2024. Using inclusion and exclusion criteria along with snowball sampling, we assessed the quality of the primary

studies based on their design, methodology, analysis, and conclusions. We retained a total of 81 PSs for analysis.

The primary contribution of this study is the development of a comprehensive understanding of how ML can be used to support the migration of monolithic systems to microservices. This is achieved through the following key contributions:

- ❶ A systematic analysis of the migration phases automated using ML techniques, identifying gaps in automation across the migration lifecycle (RQ1).
- ❷ An organised synthesis of the types of inputs used in ML-driven migration, including their granularity, sources and how it is preprocessing, forming a basis for understanding the input usage in migration (RQ2).
- ❸ A detailed analysis of machine learning techniques employed for migration, highlighting commonly used models, emerging techniques, and their features (RQ3).
- ❹ An exploration of evaluation practices for ML-based migration approaches, identifying common metrics, benchmarking, and success criteria (RQ4).
- ❺ A discussion of the challenges encountered when applying ML for migration, including scalability, data availability, and the interpretability of models, along with directions for future research (RQ5).
- ❻ [R4C22] A set of recommendations for practitioners and researchers, derived from the insights gathered, to help address the challenges of applying ML techniques during monolithic-to-microservices migration.

We organise the remainder of this paper as follows. Section II presents the background and related work on monolithic to microservices migration. Section III describes the SLR methodology used and provides an overview of the analysis process. Section V focuses on phases of migration that are currently automated using ML. Section VI examines the types of data collected and processed as inputs for ML during migration. Section VII identifies the ML techniques applied in this context. Section VIII discusses how ML-based approaches are evaluated for migration. Section IX investigates the challenges and limitations of applying ML to migration. Section X discusses the observations and recommendations. Section XI outlines the potential threats to the validity of our study. Finally, Section XII summarises the findings and suggests directions for future research.

II. RELATED WORK

The migration of monolithic applications to microservices has gained significant attention in recent years due to its potential to improve scalability, flexibility, and maintainability in software systems. Several studies have explored various aspects of this migration, ranging from decomposition frameworks to the challenges faced during the migration. This section presents an overview of the most relevant works, focusing on decomposition frameworks, migration strategies, and the challenges associated with microservice adoption.

[R4C20],[R1C5],[R2C3] Table I summarises the existing studies closely related to this SLR. Following established frameworks in the literature [16], [18], [22], the migration process considered in this study is structured into five phases: Pre-

TABLE I: [R4C27]Closely Related Works

Study Focus	Year	Method	RQs
[13] Use of AI techniques for improving any quality attribute of microservices during the DevOps phases	2025	SMS	<ul style="list-style-type: none"> • In which industry domains is AI used for microservices? • What challenges arise in applying AI to microservices design? • What benefits does AI offer over traditional microservices design? • What are the emerging AI trends in microservices design? • What input artifacts are used in AI-driven microservices design?
[14] AI for improving microservices design, decomposition, and validation	2025	SLR	<ul style="list-style-type: none"> • What AI methods support microservices design in new software systems? • What challenges arise in applying AI to microservices design? • What benefits does AI offer over traditional microservices design? • What are the emerging AI trends in microservices design? • What input artifacts are used in AI-driven microservices design?
[15] Microservice: identification techniques, tools, factors, issues, and migration benefits	2024	SMS	<ul style="list-style-type: none"> • How are monolithic systems to microservices migration done? • Which factors have caused the migration of monolithic systems to microservices? • What are the issues and benefits of migrating towards microservices?
[9] Microservice: Evolution	2024	SLR	<ul style="list-style-type: none"> • How has the area of study on microservices identification evolved? • What is the current state-of-the-art in microservices identification research? • What are the current and potential challenges associated with microservices identification?
[16] Microservice: Migration Techniques, Tools, Factors, and Benefits	2024	SMS	<ul style="list-style-type: none"> • How are monolithic systems to microservices migration done? • Which factors have caused the migration of monolithic systems to microservices? • What are the issues and benefits of migrating towards microservices?
[17] Migration Approaches, Challenges, Successful factors, and Potential for industrial adoption	2022	SMS	<ul style="list-style-type: none"> • How many articles in this field of research can be found each year? • What are the primary venues for the production and printing of the research? • What are the principal publication types in the research area?
[18] Microservices: Decomposition	2023	SLR	<ul style="list-style-type: none"> • What are the existing approaches, tools, and methods observed in the decomposition of monolith to microservices? • What are the metrics, datasets, and benchmarks used for evaluating monolith decomposition to microservices? • What research gaps can be identified in the current literature?
[19] Microservices: Quality Assurance	2021	SLR	<ul style="list-style-type: none"> • What problems and challenges are there for the migration process of monolithic applications to microservices?
[20] Microservice: Evolution	2021	SMS	<ul style="list-style-type: none"> • What methods and techniques are used in microservice analysis? • What are the problems or opportunities that are addressed using microservice analysis techniques? • Does microservice analysis overlap with other areas of software analysis, or are new methods or paradigms needed? • What potential future research directions are open in the area of microservice analysis?
[21] Microservice: Migration Techniques	2019	RR	<ul style="list-style-type: none"> • What are the migration techniques proposed in the literature? • In what types of systems have the proposed techniques been applied? • What type of validation do the authors of the techniques use? • Are there challenges associated with migration from monolith to microservices?
[22] Microservice: Refactoring Approaches	2019	SLR	<ul style="list-style-type: none"> • What are existing architectural refactoring approaches in the context of decomposing a monolithic application architecture into Microservices and how can they be classified with regards to the techniques and strategies used?
[23] Microservice: Migration Strategies	2019	SLR	<ul style="list-style-type: none"> • Which strategies have been reported in the literature to support the migration of legacy software systems to microservices-based architecture? • Which lessons learned have been reported in the literature regarding challenges and advantages perceived as a consequence of the aforementioned migration?
[10] Microservice: Migration Intentions, Strategies, and Challenges	2019	IS	<ul style="list-style-type: none"> • What are the intentions for migrating existing systems to Microservices? • Which Microservices migration strategies and decomposition approaches do companies apply? • What are the major technical and organisational challenges during a Microservices migration?
[24] Microservice: Migration Artefacts	2023	SMS	<ul style="list-style-type: none"> • During the process of re-architecting a legacy system, are data-driven artifacts like database schema and state of data used for identifying potential microservices? • How data-driven artifacts, like the database, are used in the process of software migration to services/microservices-based architecture?
[25] Microservice: Migration practices	2018	IS	<ul style="list-style-type: none"> • What are the activities carried out by practitioners when migrating towards a microservice-based architecture? • What are the challenges faced by practitioners when migrating towards a microservice architecture?
[26] Microservice: Migration Reasons and Challenges	2018	IS	<ul style="list-style-type: none"> • NP
[27] Microservice: Migration Problems and Challenges	2021	SMS	<ul style="list-style-type: none"> • NP
[28] Microservice: Migration Quality Attributes	2022	SLR	<ul style="list-style-type: none"> • Which studies implement a quality-driven approach to migrate to microservices? • Which are the quality attributes analyzed in the migration phases? • In which migration phase the quality-driven process is implemented?

IS: Industry Survey; **SLR:** Systematic Literature Review; **SMS:** Systematic Mapping Study; **RR:** Rapid Review; **NP:** Not Provided.

migration, Identification, Packaging, Deployment, and Monitoring. These phases result from a synthesis and refinement of prior works, tailored to the scope of ML-supported migration. This structuring aims to capture the full lifecycle of monolith-to-microservices transition while highlighting phases where automation through machine learning could provide concrete benefits.

Abgaz et al. [18] introduced the Monolith to Microservices Decomposition Framework (M2MDF), identifying major phases and key elements involved in the decomposition process.

They analyzed existing methods, tools, and metrics used for decomposition, proposing future directions for refining techniques. Fritzsche et al. [22] similarly classified refactoring approaches for decomposition. However, these studies do not consider the role of machine learning (ML) in supporting or automating migration tasks. Migration challenges and quality attributes have also been addressed. Velepucha et al. [19] identified quality improvement tactics, while Ponce et al. [21] categorized migration techniques by automation and validation. Razzaq et al. [17] examined organizational factors

influencing successful migration. Bushong et al. [20] explored overlaps between traditional and microservice-specific software analysis. While insightful, these works do not explore ML-driven migration strategies, inputs, or evaluation methods.

[R4C21] Saucedo et al. [16] proposed a structured migration process based on real-world industrial cases, offering a catalogue of tools and influencing factors. However, their work does not examine ML-based automation. [R3C2] Daniel et al. [14] discussed microservices design inputs and challenges, while Sergio [13] mapped AI techniques in DevOps but did not focus on migration-specific ML strategies.

Other studies reviewed migration strategies and lessons learned. Silva et al. [23] and Capuano et al. [28] highlighted strategic challenges, whereas Kalske et al. [26] focused on organizational motivations. Mpampoutis et al. [24] leveraged data artifacts for service identification, and Kazanavicius et al. [29] examined legacy migration drawbacks. None of these works, however, evaluated ML-based techniques, their automation potential, or supporting inputs and metrics. Di Francesco et al. [25] studied practitioner challenges during migration, and Fritzsch et al. [10] explored migration intentions and organizational factors. Yet, these studies do not consider ML or its evaluation in migration automation.

[R4C20] In summary, existing works provide valuable foundations for understanding microservice migration challenges, strategies, and decomposition methods. However, to our knowledge, no prior study has systematically investigated the use of ML techniques in this context. Our study addresses this gap by examining the phases of migration automated using ML, the input types, techniques applied, evaluation practices, and challenges faced.

III. RESEARCH METHOD

We followed the updated PRISMA guidelines [30], [31] and Kitchenham et al. [32] guidelines to review and report our findings. We used three main phases: planning, conducting, and reporting the review. During the planning phase, we defined the objective of SLR and reviewed the protocol. The objective of this SLR is defined in Section I. This section defines the review protocol for conducting the SLR. It consists of six steps: ① defining the research questions, ② formulating the search query, ③ selecting the studies, ④ snowballing, ⑤ assessing the quality of the studies, and ⑥ extracting and analysing the data. In the following subsections, we explain each step of our review protocol.

A. Research Questions (RQs)

This study answers the following RQs:

RQ1: Which migration phases are automated by ML?

Rationale: By analysing the phases and the tasks supported by ML, we want to understand the role and impact of ML in the migration process. This investigation provides insights into how ML is integrated into current migration approaches.

RQ2:[R1C1] How are ML migration inputs characterised in terms of type, granularity, source, and preprocessing?

Rationale: By categorising inputs based on their types, granularities, sources, and preprocessing methods, we want to

create a structured classification of the inputs and understand how inputs are gathered, prepared, and utilised in machine learning models for microservices migration.

RQ3:[R1C1] What ML techniques and learning paradigms are applied in migration approaches?

Rationale: By investigating the ML approaches in the migration, we want to (1) identify and classify the ML models used, (2) examine their learning paradigms, and (3) analyse the features leveraged to enhance model performance. This RQ provides a comprehensive understanding of how ML contributes to different phases of the migration.

RQ4:[R1C1] How are ML-based migration approaches evaluated with respect to metrics, benchmarks, tools, and success criteria?

Rationale: By investigating how ML-based methods for microservices migration are evaluated, we want to (1) identify the evaluation metrics used to assess the quality and performance of these methods, (2) explore the benchmarks employed for comparative analyses, (3) understand the success criteria that define their effectiveness, and (4) analyse the tools available for implementing and evaluating these approaches.

RQ5: What challenges arise in using ML approaches?

Rationale: By examining the limitations, we aim to identify and analyze the key challenges associated with applying machine learning techniques in the migration from monolithic architectures to microservices.

B. Search Query

We formulated our search query by applying the PICO (Population, Intervention, Comparison, Outcome) framework [33]. We followed the following steps:

- ① Obtaining the main terms from our main research question as stated in the introduction.
- ② Identifying the possible synonyms of the main terms.
- ③ Applying the Boolean OR to combine possible synonyms of the main terms.
- ④ Applying the Boolean AND to combine expressions in the previous step.

As a result of PICO framework, we formulated the following search query:

```
(Monolith* OR Exist* OR Legac*) AND (Microservice* OR Micro-service* OR MSA) AND (Migrat* OR Identif* OR Decompos* OR Extract* OR Transform* OR Refactor* OR Transit* OR Creat* OR Genera*) AND (Machine learning OR ML OR Neural Network* OR *coder OR *supervised OR Reinforcement Learning OR Model*)
```

To obtain more comprehensive results, we used the asterisk (*) in search queries as a wildcard to match any sequence of characters.

C. Studies Selection

We applied the PRISMA steps to select the PSs. The main steps include database identification, removal of duplicates,

screening, eligibility assessment, multiple rounds of both backward and forward snowballing, and quality assessment of each study. Figure 1 summarises those steps.

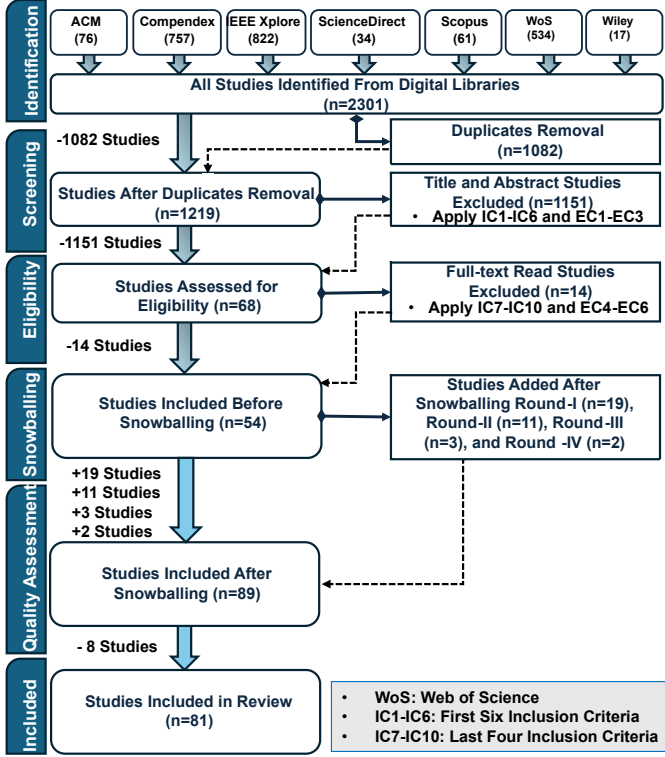


Fig. 1: PRISMA Flow for Primary Studies Selection

1) *Databases Identification*: We selected seven online digital libraries: ACM Digital Library, Compendex, IEEE Xplore, ScienceDirect, SpringerLink, Scopus, Web of Science, and Wiley. These libraries are widely used for literature reviews in software engineering, as recommended by Dyba et al. [34].

We applied our search query to each of these digital libraries. However, some libraries impose restrictions when performing queries. For instance, ScienceDirect limits queries to a maximum of eight connectors, while the ACM Digital Library does not allow wildcards. We adjusted the search query to meet the specific requirements of each library. Table II shows the search queries executed in each digital library and the number of studies retrieved. Our search was confined to

TABLE II: Database Search Results

Database	All Search Studies	Selected Studies (PSs)
ACM	76	10
Compendex	757	21
IEEE Xplore	822	37
ScienceDirect	34	3
Scopus	61	6
Web of Science	534	1
Wiley	17	3
Total	2301	81

English-language, peer-reviewed scholarly articles published in journals, conferences, and workshops between 2015 and 2024. This time frame was chosen due to the increase in

microservices-related literature starting in 2015 [35]. We initially retrieved a total of 2301 studies from seven libraries.

2) *Duplicates Removal*: [R4C19] Duplicates were identified based on an exact match of the study's title, first author, and venue (conference or journal). This reduced the total from 2301 studies to 1219.

3) *Screening*: We defined inclusion and exclusion criteria and applied them to select relevant primary studies while excluding irrelevant ones.

Inclusion Criteria: We considered the following inclusion criteria for PSs selection:

- **IC1**: The study is written in English.
- **IC2**: The study is published between 2015 and 2024.
- **IC3**: The study is published in journals, conferences, or workshops.
- **IC4**: The study has at least 4 pages.
- **IC5**: The study focuses on the migration of monolith applications into microservices.
- **IC6**: The study uses a machine/deep learning algorithm.
- **IC7**: The study provides enough information to answer at least three research questions (RQs).
- **IC8**: The study has its full text available online.
- **IC9**: The study provides sufficient migration details.
- **IC10**: The study uses an automated or semi-automated migration approach.

Exclusion Criteria: We considered the following exclusion criteria:

- **EC1**: [R4C18] The study is a secondary source (e.g., literature review, survey, opinion piece, or discussion).
- **EC2**: The study has not been peer-reviewed.
- **EC3**: The study is a graduate thesis or project report.
- **EC4**: The study's full text is not available online.
- **EC5**: The study does not provide enough details.
- **EC6**: The study does not provide an automated or semi-automated migration approach.

We applied our inclusion and exclusion criteria in two steps, first using the titles and abstracts, then the full texts. Criteria IC1-IC6 and EC1-EC3 were applied to the titles and abstracts, while IC7-IC10 and EC4-EC6 were applied to the full texts. [R1C2][R2C1] We considered that a study provided 'sufficient migration details' (IC9) if it described at least the system analyzed, the ML technique applied, and the evaluation procedure used. Conversely, studies lacking two of these three aspects were excluded under EC5. Additionally, EC5 encompassed studies that solely described theoretical approaches without providing experimental validation or application to real systems or datasets, in order to ensure the practical relevance of the selected primary studies. We acknowledge that some inclusion and exclusion criteria are logically complementary (e.g., IC8/EC4, IC9/EC5 and IC10/EC6) describe opposing conditions. Additionally, IC7 partially overlaps with EC5. We explicitly state both inclusion and exclusion criteria separately to comply with the PRISMA [12] guidelines, which recommend clearly articulating both positive and negative selection criteria to enhance transparency and reproducibility. [R4C17] To clarify the practical application of the criteria, we present in Table VII (in Appendix) examples of excluded studies for

IC7-IC10, EC5, and EC6, along with brief justifications for their exclusion. During the initial screening, we assessed the titles, abstracts, and page counts, and determined if the studies qualified as primary. Two authors independently conducted the screening using predefined inclusion and exclusion criteria. To ensure consistency, we compared the screening results for 50 randomly selected studies using Cohen's Kappa [36]

We calculated Cohen's Kappa, achieving near-perfect agreement ($k=0.85$), highlighting the consistency of our screening. For the remainder of the process, authors met regularly to review results and resolved disagreements through discussion and consensus. After completing the initial screening, we selected 68 studies for full-text review as potential PSs.

4) *Eligibility Assessment*: In the second round of screening, two authors independently applied IC7-IC10 to 68 studies by thoroughly reviewing them. To ensure a shared understanding of the inclusion and exclusion criteria, we compared the results of 15 randomly selected studies using Cohen's Kappa. We calculated a near-perfect agreement ($k=0.93$), demonstrating strong consistency between the two authors. This allowed us to confidently proceed with the eligibility assessment for the remaining studies. Ultimately, we identified 54 PSs that met the eligibility criteria.

D. Snowballing

[R1C3][R2C2] We conducted four rounds of both forward and backward snowballing to identify additional primary studies (PSs). In Round 1, we reviewed the references and citations of 54 initial PSs and identified 1,347 potential studies. After removing duplicates, we retained 1,172 studies and applied our inclusion and exclusion criteria, which resulted in 19 new PSs, bringing the total to 73. In Round 2, we used the 19 newly added PSs to identify 602 potential studies and selected 11 new PSs using the same criteria, raising the total to 84. In Round 3, we reviewed the references and citations of the 11 new PSs, found 190 potential studies, and selected 3 additional PSs, bringing the total to 87. In Round 4, we examined the references and citations of the 3 PSs from the previous round, identified 43 potential studies, and selected 2 more PSs. This brought the final total to 89 PSs. Table III presents the details from each round.

TABLE III: [R1C3, R2C2] Snowballing Results

Snowballing	Round	Retrieved	No duplicates	Included
Backward & Forward	1	1346	1172	19
Backward & Forward	2	602	375	11
Backward & Forward	3	190	135	3
Backward & Forward	4	43	43	2
Total			1725	35

E. Quality Assessment

[R4C16] Following the guidelines proposed by Li et al. [37] and Dyba et al. [38], we developed a quality assessment checklist comprising eight questions (or quality criteria) to evaluate the quality of our primary studies (PSs). Table IV shows the quality criterias we used in this study based on Li

et al.'s work [37]. Responses to each question were: "Yes" (1 point), "Partially" (0.5 points), or "No" (0 points). Three authors independently applied this checklist to each study, discussed any discrepancies, and achieved consensus on the scores. We calculated the quality score for each study by summing the individual scores and converting this total into a percentage. Studies achieving an 80% score or higher were retained. Using this criterion, we excluded eight studies falling below the threshold and included 81 PSs in our review.

F. Data Extraction and Analysis

We analyzed and extracted the necessary data from each primary study (PS) to address our research questions (RQ1-RQ5). Table V presents the key data items extracted, including the item name, a brief description, and the corresponding research question it addresses. The full details of the extracted data are available in our publicly accessible replication package.

[R4C14] To ensure accuracy and consistency, two researchers independently extracted data using a shared Excel template with dedicated columns for each data field. Initially, each researcher processed 50% of the PSs. This was followed by a full cross-validation phase, during which each extraction was reviewed by the other researcher. We achieved strong inter-rater reliability, with only 8% of the data points requiring arbitration. In those cases, a senior third researcher resolved discrepancies by re-examining the original source papers and documenting the final decision.

IV. OVERVIEW OF THE SELECTED LITERATURE

We analysed the primary studies (PSs) (presented in Table VIII in appendix A) to gain deeper insights into the selected set. In this section, we present our findings on publication trends, distribution of PSs by study type, databases used, and the publication venues associated with these PSs.

A. Publication Trends

We analysed the publication trends of the selected primary studies (PSs). Notably, 17.2% of the PSs were published in 2022, marking the year with the highest publication volume. An increasing trend in conference publications is observed from 2016 to 2022. In contrast, journal publications show a rise starting in 2021, with no journal publications recorded before this period except for two PSs in 2019. Only a few PSs were published in workshop proceedings, specifically in 2022 (1 PS), 2023 (2 PSs), and 2024 (1 PS). Figure 2 shows the publication trends over the years for the PSs.

B. Distribution of PSs by Type

We analysed the types of the selected primary studies (PSs). As shown in Figure 2 most of the PSs are from conferences, comprising 65% (53 PSs), followed by journals with 30% (24 PSs). The smallest group is from workshops, accounting for only 5% (4 PSs). The high conference publication rate may indicate rapid field growth, while journal articles add rigour. Limited workshop presence may reflect field maturity or reduced exploratory discussions.

TABLE IV: [R4C16] Quality Criteria extracted from [37]

No	Quality Criteria
Q1	Is there a clear statement of the aims of the research? Consider: Is there a rationale for why the study was undertaken?
Q2	Was the research design appropriate to address the aims of the research? Consider: Did the researcher justify the design of the research?
Q3	Was the research method implemented in a way that addressed the research issue? Consider: Has the researcher discussed the process or the details of the methods that were chosen/proposed?
Q4	Is there a clear statement of findings? Consider: Has an adequate discussion/evaluation of the evidence identified or method proposed, both for and against the researchers' arguments, been demonstrated?
Q5	Has the limitation or future work been considered adequately? Consider: Did the researcher examine the limitations and future work?
Q6	Is there an adequate description of the context in which the research was carried out? Consider: Did the researcher explain how the context influenced the study, and was this context relevant to the aim of the research?
Q7	Was the data collected in a way that addressed the research issue? Consider: Did the researcher collect data from relevant sources, and was the data collection process clearly described and ethically sound?
Q8	Was the data analysis sufficiently rigorous? Consider: Did the researcher provide a thorough and unbiased examination of the data, and were the limitations of the analysis acknowledged?

TABLE V: [R4C12] Data Extraction Template

Data Item	Short Description	Example	Value	RQs
Code	A unique identifier assigned to the paper	trabelsi2024_magnet	Varied	
Title	The title of the research paper	Magnet: Method-Based Approach Using Graph Neural Network for Microservices Identification	Varied	
Year	The year in which the paper was published	2024	Fixed	
Venue	The conference, journal, or workshop where the study was published	International Conference on Software Architecture	Varied	
Authors	The individuals who contributed to conducting the study	Trabelsi, I and Moha, N and Guéhéneuc, YG and ...	Varied	
Migration Phase	The specific phase of system migration from monolith to microservices	Identification	Fixed	RQ1
Automated Task	Exact task automated during the migration phase	Identification	Varied	RQ1
ML Integration	Specifies where machine learning (ML) is integrated	Clustering and semantic analysis	Varied	RQ1
Input type	Defines the types of input used, such as source code, logs, UML diagrams, etc.	Source code	Varied	RQ2
Granularity of Data	The level of detail or precision in the data, such as files, classes, lines of code, or logs	Methods	Fixed	RQ2
Source of Data	Indicates the data source: open source, industrial projects, user-generated, or collected	Open Source	Fixed	RQ2
Data preprocessing	Steps to prepare data, including collection, cleaning, normalization, and transformation	Static analysis with KDM and semantic analysis using word2vec	Varied	RQ2
ML Technique	Indicates the machine learning techniques used	Deep Modularity Networks (DMoN)	Varied	RQ3
Learning Approach	Indicates the learning approach: supervised, semi-supervised, unsupervised, etc.	Unsupervised	Fixed	RQ3
Feature Selection	Specifies the features used for model training	method calls, method bodies, and class structures	Varied	RQ3
Evaluation Metrics	Specifies the metrics used to evaluate the output	Precision, recall, f-measure, SMQ (Structural Modularity Quality), CMQ (Conceptual Modularity Quality), CHM (Cohesion at Message level), and CHD (Cohesion at Domain level).	Varied	RQ4
Benchmark	Indicates performance benchmarks or comparisons with other systems	Compared against Service-Cutter and MicroMiner	Varied	RQ4
Success Criteria	Specifies success criteria, such as improved recall, quality, or performance	Improved modularity, functional independence, and reduced coupling.	Varied	RQ4
Tool Availability	Indicates tool availability: open-source, commercially available, or a proof of concept (PoC)	Available, Open Source	Fixed	RQ4
Challenges	Identifies the challenges discussed in the paper	Generalizability across different types of monoliths and data availability	Varied	RQ5

C. Publication Venue

We identified 72 distinct venues for the PSs. Most venues have a single PS, except for ASPLOS (2 PSs), ICWS (2 PSs), ISSRE (2 PSs), ASE (3 PSs), ICSA (3 PSs), and IC SOC (3 PSs). [R4C11]The broad distribution of PSs across different publication venues suggests that the research topic attracts interest from various subfields within software engineering and related areas. However, most studies remain monodisciplinary in nature, focusing predominantly on aspects of software architecture, migration engineering, and software modernization.

D. Databases for Selected PSs

We observe that the majority of PSs were sourced from IEEE Xplore (37 PSs or 45.6%), followed by Compendex (21 PSs or 25.9%) and ACM Digital Library (10 PSs or 12.3%). We observed the presence of PSs in other databases, such as Scopus (6 PSs or 7.4%), ScienceDirect (3 PSs or 3.7%), and Wiley (3 PSs or 3.7%). The Web of Science database provided the least contribution with 1 PS (1.2%). This diversity in sources enhances the robustness of the literature review by ensuring a comprehensive collection of relevant studies.

For more visuals and tables, see Appendix A.

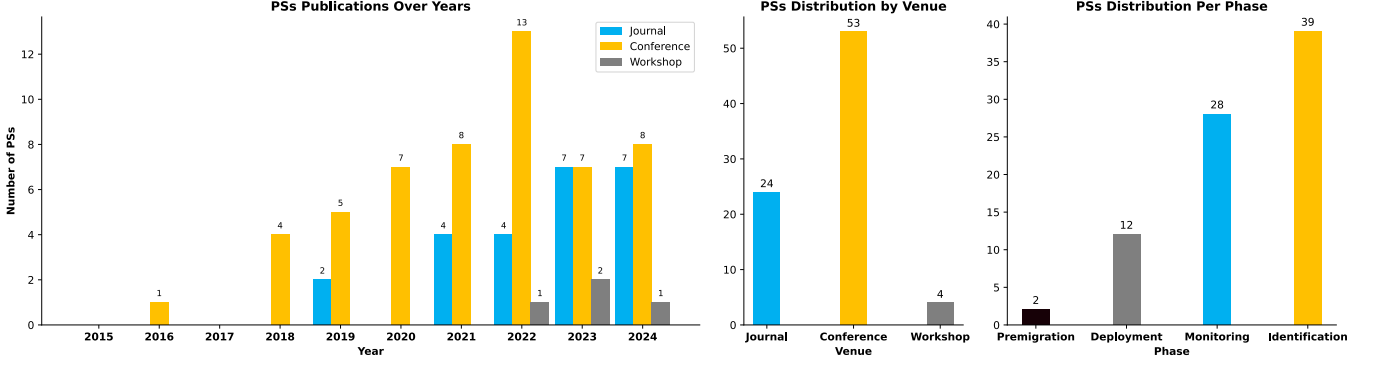


Fig. 2: Distribution of PSs

V. MIGRATION PHASES AUTOMATED BY ML (RQ1)

The migration of legacy systems to microservices involves a sequence of complex phases, many of which are now supported or automated by machine learning. In this section, we analyze the phases of the migration process and identify the specific tasks that ML automates in this process. [R1C5][R2C3][R4C10] Following prior works on microservices migration: Abgaz et al. [5], Fritzsch et al. [22], and Saucedo et al. [16], we structured the migration process into five phases: Pre-migration, Identification, Packaging, Deployment, and Monitoring. The identification phase is the most frequently covered in our primary studies (PSs), with 48% (n=39), followed by the monitoring phase at 35% (n=28). The deployment phase is also notable, covered by 15% of PSs (n=12), while the pre-migration phase is the least discussed, with only 2.24% (n=2). Figure 2 shows the number of PSs that target each phase. We provide the definitions for all key terms related to RQ1, including migration phases and associated automated tasks, in Appendix A for clarity and consistency.

A. Pre-migration

The pre-migration phase focuses on evaluating the current approaches, methods, and tools used in the legacy system. This phase also involves planning the migration by identifying suitable strategies, defining objectives, and outlining the steps needed to ensure a smooth transition. Researchers have leveraged ML to automate tasks such as designing microservices, which assists in planning microservice architecture by analyzing legacy systems [P1], and predicting success rates of migration, which uses predictive models to evaluate potential migration outcomes [P2].

B. Identification

The identification phase defines the boundaries of prospective microservices. This involves detecting functional modules, mapping dependencies, and clustering components with business requirements. ML techniques have been applied to automate tasks such as boundary identification, which detects functional boundaries within monolithic systems [P3]; clustering, which groups related components or services using ML-based clustering algorithms [P4]; and microservices identi-

cation, which automates the process of determining candidate microservices [P5]–[P13].

C. Packaging

The packaging phase encapsulates the identified components into functional microservices. This phase includes defining service interfaces, managing dependencies, and generating missing components. However, no studies in our review proposed the use of ML for this phase, highlighting a significant gap in the research landscape.

D. Deployment

The deployment phase focuses on deploying the created microservices into a target environment. This involves setting up orchestration tools, configuring infrastructure, and integrating with existing systems. ML techniques are increasingly used to automate tasks such as automated deployment, which automates the deployment of microservices into production environments [P13], [P14]; resource management, which ensures efficient use of infrastructure resources during deployment [P15]–[P18]; resource allocation, which dynamically allocates resources to microservices based on workload demands [P15], [P17], [P19]–[P23]; microservice autoscaling, which adjusts the scale of microservices to match performance requirements [P24]; microservice orchestration, which manages dependencies and interactions between microservices [P25]; microservice placement, which decides optimal placements for microservices within the infrastructure [P16], [P26]; and resource estimation, which estimates the resource needs of microservices to optimize deployment strategies [P27].

E. Monitoring

The monitoring phase ensures the reliability and efficiency of the microservices-based system post-migration. This phase involves continuous performance tracking, anomaly detection, and dynamic resource management to meet changing workload demands. ML models enhance the monitoring process by enabling proactive detection of issues and optimizing resource usage. Tasks automated in this phase include anomaly detection, which identifies abnormal behaviors or performance issues in microservices [P28]–[P30]; performance analysis,

which evaluates the performance metrics of deployed microservices [P31]; detection of failure types, which classifies failure types to facilitate troubleshooting [P32], [P33]; fault diagnosis, which diagnoses root causes of failures to enable faster resolution [P31]; root cause analysis, which pinpoints the underlying issues causing anomalies or failures [P21], [P34]; privacy risks detection, which identifies potential privacy concerns in microservice interactions [P35]; and sanity checks, which verifies the correctness of system states and responses [P27].

VI. INPUTS USED BY ML MIGRATION APPROACHES(RQ2)

[R2C4] It is essential to understand the types of inputs collected and processed in the migration process to identify how machine learning techniques are applied. To this end, we categorized inputs along four dimensions: type, granularity, source, and preprocessing method. These categories were not predefined, but emerged inductively from our analysis of the primary studies. [R1C4, R4C3] Figure 3 summarizes data we extracted in our PSs. The resulting taxonomy allowed

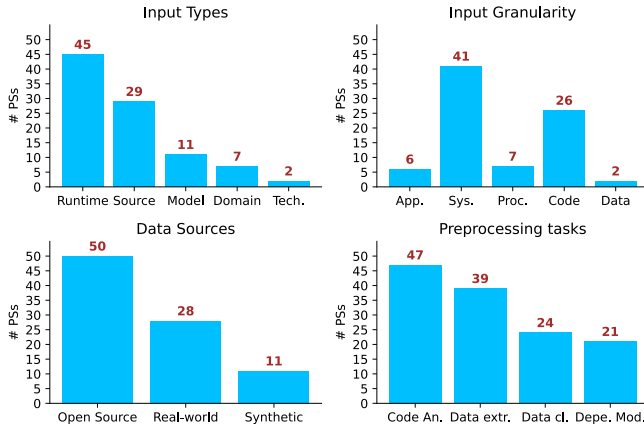


Fig. 3: [R1C4, R4C3] Inputs and Preprocessing Tasks in PSs

us to systematically examine the role and characteristics of inputs used in ML-based migration approaches. We provide the definitions for all key terms, including input types, granularity levels, data sources, and preprocessing categories, in Appendix A for clarity and consistency.

A. Input Types and Their Role in Migration

In this section, we classify the input types identified in the primary studies (PSs) into five distinct groups which illustrates their use in the migration phases. Our analysis reveals that runtime artifacts are the most commonly used inputs, appearing in 55% of the PSs ($n = 45$), emphasizing their critical role in understanding system performance during migration. [R4C4]Source Artifacts, such as source code and configuration files, are the second most frequently used, accounting for 35% of PSs ($n = 29$). In contrast, domain artifacts ($n = 7$) and model artifacts ($n = 11$), provide valuable high-level system descriptions, while technical artifacts are the least common, found in only 2% of PSs ($n = 2$).

1) *Domain Artifacts*: Domain artifacts capture high-level information about the system’s business and functional requirements, aiding in the understanding and identification of microservices. These include API documentation, which represents details about the available APIs and their interactions. For example, [P10] highlighted the use of the OpenAPI specification of the legacy system’s RESTful APIs to identify microservices. Quality of Service (QoS) constraints represent parameters that must be maintained during migration. The authors of [P35] used service communication traces to meet predefined QoS requirements, such as availability and reliability. Architecture recommendations, as provided by [P36], offer expert suggestions for migration strategies and component separation. Functional descriptions, such as those provided by [P37], represent high-level descriptions of system functionalities, capturing technological attributes and relationships between services.

2) *Runtime Artifacts*: Runtime artifacts are derived from the system’s runtime behavior and are crucial for understanding performance and operational characteristics. These include resource metrics, which focus on information about resource consumption, such as memory and CPU usage. The study by [P27] explored the use of API traffic logs to analyze the runtime behavior of systems. Similarly, [P32] leveraged system metric data, such as CPU, memory, and disk usage, along with failure propagation information, to identify resource bottlenecks. [P38] used memory utilisation, network latency, and packet loss to detect anomalies in containerized microservices, while [P19] used CPU usage metrics to analyse performance across different stages of system operation. Monitoring metrics are another key runtime artifact. [P39] introduced SpanGraph, a tool that leverages monitoring metrics, trace logs, and configuration files to construct directed graphs representing microservice interactions. Performance metrics, such as response time, throughput, and latency, are also critical. [P40] used time series data of performance metrics to predict system performance. Trace logs provide detailed insights into the execution flow of the system, as demonstrated by [P41], who used trace logs and performance metrics for anomaly detection. Workloads, which describe system usage scenarios and patterns, are used by [P42] to predict response time of microservices.

3) *Model Artifacts*: Model artifacts describe the system from an architectural or design perspective, often providing an abstract representation. [R3C3] Business process models (BPMN) are used to capture the workflows supported by the system and align microservices with organizational functions. Several studies (e.g., [P6]–[P8], [P43]) leverage BPMN to support microservice identification, demonstrating that business process understanding can guide the decomposition of monolithic systems into well-bounded services. UML diagrams, which represent details of the system design, are also commonly used. For instance, [P29] combined code, logs, and UML diagrams to provide unique insights for anomaly analysis, leveraging the visual representation of system design to identify issues. User stories are another key model artifacts, used by [P44] to guide the organisation of microservice development tasks.

4) *[R4C4]Source Artifacts*: Source artifacts pertain to the actual software and its configuration, offering insights into the system’s implementation. For example, [P13] used source code to identify microservices, while [P1] analyzed source code files from a monolithic application to generate a calling-context tree for better system understanding and transformation. Configuration files are used by [P39] alongside trace logs and monitoring metrics to construct SpanGraph, a directed graph representing microservice interactions for fault localization. Data files are used by [P20] to support latency-aware provisioning.

5) *Technical Artifacts*: They provide supporting information necessary for system operation. These include server information, which details the infrastructure supporting the system. For example, [P26] used service radius and capacity, as well as runtime traces to predict microservice placement.

B. Input Granularity

[R2C6] The granularity of input data, ranging from high-level abstractions like system architecture diagrams to fine-grained elements such as lines of code or individual function calls, plays a critical role in determining the precision and scope of machine learning tasks. We organize the granularity levels in descending order of abstraction: application-level, system-level, process-level, code-level, and data-level. Application-level granularity (7%, $n=6$) examines higher-level abstractions, with [P45] using use case-level data for workflow modeling and [P29] analyzing application files for anomaly detection. URI and API-level granularity, as in [P46] and [P47], support system decomposition and service identification. System-level granularity is the most frequently used (51%, $n=41$), as seen in [P14] and [P48], which monitor and analyze system components and interactions. Process-level granularity (9%, $n=7$) examines workflows, with [P47] using API operation-level data for decomposition and [P8] using activity-level granularity for workflow modeling. Code-level granularity (32%, $n = 26$) targets classes and methods for tasks like refactoring and microservice identification, as seen in [P49] and [P12]. Finally, data-level granularity (2%, $n = 2$) is the least explored and focuses on stored information, such as table-level data in [P50] and entity-level representations in [P3].

C. Data Sources

Our analysis shows that open-source datasets are the most widely used, representing 62% of PSs ($n=50$), showcasing the reliance on publicly available repositories for experimentation and validation. For example, [P51] validated their approach on open-source applications like JPetStore. Real-world data, which provides insights from operational systems, is used in 35% of PSs ($n=28$). For instance, [P52] used the AIOps Challenge dataset to detect anomalies and faults in microservice systems. Meanwhile, synthetic data, generated to simulate real-world conditions, is utilized in 14% of PSs ($n=11$). For example, [P34] combined synthetic data from Apache Thrift with open-source datasets like DeathStarBench to debug performance in microservice environments, while [P2] collected data using a survey and expert interviews to optimize migration strategies.

D. Preprocessing Tasks

[R4C7]Preprocessing ensures the quality and usability of input data for machine learning models. Code analysis is the most frequently employed technique (58%, $n = 47$), underscoring the importance of understanding system structure and semantics. Typical approaches rely on static analysis tools such as JavaParser, which parses source code to extract classes, methods, and dependencies, providing structured inputs for machine learning models. Data extraction, used in 48% of PSs ($n = 39$), captures essential operational and dependency information, such as parsing execution traces to build graph representations, as in [P53]. Data cleaning, employed in 30% of PSs ($n = 24$), improves data quality through noise reduction and normalization. For example, [P41] applied density-based clustering to filter noisy data points, while [P27] normalized trace data for compatibility with further analysis. Dependency modeling, found in 26% of PSs ($n = 21$), structures relationships for better analysis. For instance, [P54] used vectorization and embedding for clustering, and [P55] built dependency graphs using normalized matrices for analysis with Variational Autoencoders (VAE). Modeling and task structuring, as in [P48], represented microservice dependencies using adjacency matrices, while [P17] captured features like CPU usage and request counts for workload analysis.

VII. ML APPROACHES APPLIED BY RESEARCHERS (RQ3)

In this section, we discuss the key ML models used in microservices migration, their learning paradigms, and the features that enhance model performance. [R1C4, R4C3] Figure 4 summarizes data we extracted in our PSs.

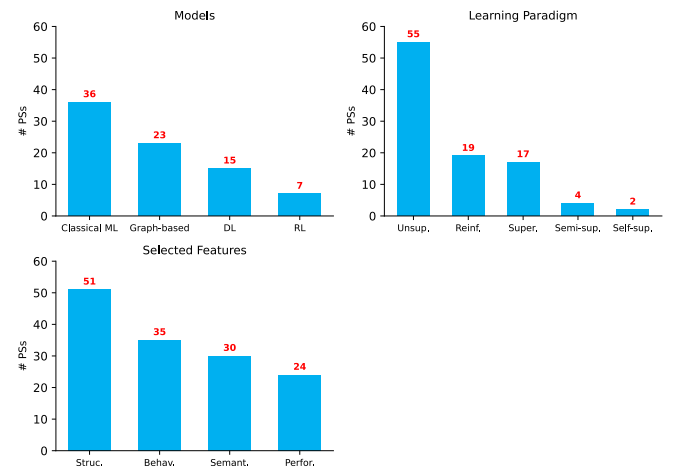


Fig. 4: [R1C4, R4C3] ML Models, Learning Paradigms, and Feature in PSs

We provide the definitions for all key terms related to RQ3 in Appendix A for clarity and consistency.

A. Models

Different machine learning models have been applied to address various phases of the microservices migration process.

[R2C5] Based on our systematic analysis of the literature and aligned with established machine learning taxonomies [39], [40], we categorise these models into four distinct classes: Classical machine learning, Deep learning, Graph-based models and Reinforcement learning. Classical ML models were the most frequently used (44.40%, $n = 36$), followed by graph-based models (28.40%, $n = 23$), deep learning (18.50%, $n = 15$), and reinforcement learning (8.6%, $n = 7$).

1) *Classical Machine Learning*: [R1C7] Classical machine learning techniques are predominantly applied to structured datasets. Nevertheless, several models, such as Naive Bayes or Support Vector Machines, also perform well on unstructured data (e.g., text) once transformed into feature vectors through methods like TF-IDF or word embeddings. These techniques are widely applied to automate tasks such as service identification, dependency analysis, and performance prediction. Classification techniques, such as Support Vector Machines [P11], Random Forest [P6], and Naive Bayes [P56], are used to categorise components and analyse dependencies. Regression techniques, including Linear Regression [P56] and Support Vector Regression [P5], predict performance metrics and resource usage. Clustering methods, such as K-means [P56] and Density-Based Spatial Clustering of Applications with Noise [P57], group components for service decomposition. Search-based techniques, like Genetic Algorithms [P2], optimise migration strategies.

2) *Graph-Based Models*: Graph-based methods model component relationships as graphs, enabling service interaction and dependency analysis. Graph Convolutional Networks [P58] and Graph Attention Networks [P12] aggregate features from neighbouring nodes to improve graph representations. Variational Graph Autoencoders [P55] combine graph convolutional networks with variational inference for tasks like link prediction.

3) *Deep Learning*: Deep learning techniques leverage neural networks to analyse complex data structures. Autoencoders and Variational Autoencoders [P55] are used for dimensionality reduction and feature learning. Recurrent networks, such as Long Short-Term Memory [P52] and Gated Recurrent Units [P29], process sequential data for anomaly detection. Transformers, including Bidirectional Encoder Representations from Transformers [P30] and CodeBERT [P9], excel in natural language and code analysis tasks.

4) *Reinforcement Learning*: Reinforcement learning learns optimal strategies through interactions with the environment. Fuzzy Q-Learning [P16] used to handles uncertainty in resource allocation, while Deep Q-Learning [P42] and Deep Deterministic Policy Gradient [P48] used to optimise autoscaling and resource usage. Multi-Agent Deep Deterministic Policy Gradient [P26] improves service placement in edge computing environments.

B. Learning paradigm

The learning paradigms applied in the studies include unsupervised learning, which is the most popular (67.9%, $n = 55$), used for tasks like microservice extraction, clustering, and anomaly detection [P11], [P59]; reinforcement learning

(23.46%, $n = 19$), primarily for resource provisioning and deployment optimisation [P20], [P48]; supervised learning (20.99%, $n = 17$), applied in anomaly detection and resource optimisation [P42], [P60]; semi-supervised learning (4.94%, $n = 4$), which combines labelled and unlabeled data to refine service boundaries [P56]; and self-supervised learning (2.47%, $n = 2$), which generates pseudo-labels for tasks like failure localisation and resource provisioning [P23], [P33].

C. Selected Features

Machine learning models in these PSs rely on diverse feature types to analyze, predict, and optimize different aspects of the migration process. Structural features are the most popular (62.96%, $n = 51$), including class dependencies [P61], method calls [P4], data dependencies [P50], transactional dependencies [P6], and call graph dependencies [P62]. Behavioral features (43.21%, $n = 35$) capture runtime interactions, such as invocation paths [P4], log events [P59], contextual log entries [P59], and temporal patterns [P59]. Semantic features (37.04%, $n = 30$) focus on the meaning of system components, leveraging semantic embeddings [P11], function names [P63], and API descriptions [P4]. Performance features (29.63%, $n = 24$) measure system efficiency, including CPU usage [P45], memory usage [P55], network traffic [P64], response times [P45], and availability [P65]. These features collectively enable tasks such as microservice identification, dependency analysis, anomaly detection, and performance optimisation.

VIII. EVALUATION OF ML APPROACHES (RQ4)

In this section, we begin by discussing the various evaluation metrics commonly employed to assess the performance of machine learning models in migration scenarios. Next, we explore the conducted comparative analyses. Following this, we outline the success criteria adopted in the literature. Finally, we will discuss the tools availability of those works. [R4C6] Although the *Data Sources* category is discussed in Section VI as part of input characterization, it also plays an important role in RQ4. The same data are frequently used not only as inputs to the ML-based migration techniques but also as a foundation for their evaluation and validation. [R1C4, R4C3] Figure 5 summarizes data we extracted in our PSs.

We provide the definitions for all key terms related to RQ4 in Appendix 8.

A. Evaluation Metrics

Evaluation metrics are categorized into five key areas: classification and prediction metrics (46%, $n=38$), software design metrics (19%, $n=16$), system behavior metrics (27%, $n=22$), clustering metrics (17%, $n=14$), and developer-centric metrics (5%, $n=4$). Each category addresses specific aspects of microservices migration.

1) *Classification and Prediction Metrics*: These metrics assess model performance in labeling and prediction tasks, primarily during the identification phase. Precision measures the accuracy of positive predictions, with variations like Precision@K (PR@K) and Mean Average Precision (MAP) used

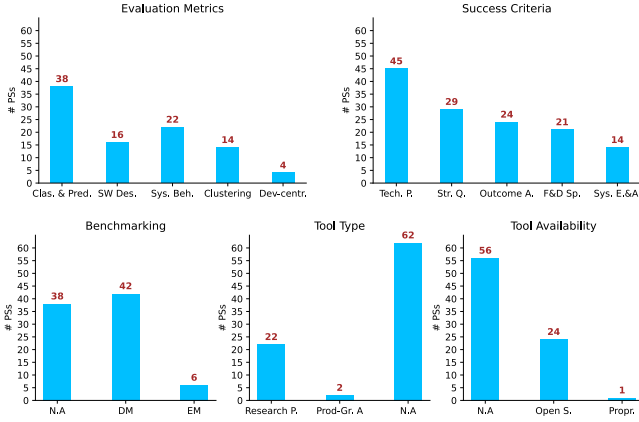


Fig. 5: [R1C4, R4C3] Evaluation of ML Approaches in PSs

for ranked predictions or multiple queries. It is commonly applied in monitoring for anomaly detection, as seen in [P28], [P30], [P38], [P59]. Recall evaluates the model's ability to identify all relevant instances, particularly in anomaly detection and fault identification tasks, as demonstrated in [P38], [P56]. The F1-Score balances precision and recall, making it critical for monitoring and service identification tasks, as shown in [P6], [P38], [P53], [P59], [P66]. Accuracy represents the overall correctness of predictions, with variations like Top-K Accuracy used in multi-class classification tasks, as seen in [P1], [P5], [P24], [P60]. The Area Under the ROC Curve (AUC) evaluates binary classification models, particularly in anomaly detection, as highlighted in [P43], [P67]. Finally, the Matthews Correlation Coefficient (MCC) provides a balanced measure of classification quality, especially in imbalanced datasets, as used in [P59].

2) *Clustering Metrics*: These metrics evaluate the quality of service clustering during the identification phase. The Dunn Index measures the separation between clusters, as demonstrated in [P5], [P53], [P57], [P66]. The Silhouette Score assesses the similarity of objects within their cluster compared to others, as seen in [P24], [P35], [P43], [P57]. Newman-Girvan Modularity evaluates the strength of network division into clusters, as highlighted in [P5], [P53], [P60], [P63]. Non-Extreme Distribution ensures balanced cluster sizes [P49], [P55], [P58], [P62]. Maximum Cluster Size limits the largest cluster size to avoid granularity issues [P65]. The Number of Singleton Clusters counts clusters with a single element to prevent excessive fragmentation [P65].

3) *System Behavior Metrics*: These metrics assess operational performance, resource utilization, and scalability during and after migration. Response Time measures the time taken for the system to respond to a request, as highlighted in [P38], [P56], [P60], [P66]. Resource Utilization evaluates the effectiveness of resource use, as seen in [P24], [P35], [P57], [P63]. Energy Consumption assesses the total energy required for operations, as demonstrated in [P1], [P6], [P60], [P66]. The SLA Violation Rate measures the percentage of time a system fails to meet its Service Level Agreement obligations, as shown in [P24], [P43], [P57], [P63]. Scalability evaluates the system's capacity to maintain or improve performance

under increased workload, as in [P5], [P6], [P35], [P53].

4) *Software Design Metrics*: These metrics evaluate software quality during the evaluation phase, focusing on maintainability and modularity. Cohesion measures the relatedness of functionalities within a service, as demonstrated in [P6], [P35], [P57], [P66]. Coupling assesses dependencies between services, as seen in [P5], [P56], [P60], [P63]. The Granularity Metric evaluates service size and scope, as highlighted in [P1], [P24], [P53], [P57]. Structural Modularity measures the degree of system decomposition into independent components, as shown in [P5], [P38], [P43], [P66]. Cognitive Complexity assesses code understandability for developers, as demonstrated in [P1], [P57], [P60], [P63].

5) *Developer-Centric Metrics*: These metrics evaluate alignment with business requirements and developer feedback during the validation phase. Developer Validation assesses qualitative insights from developers, as seen in [P1], [P6], [P57], [P66]. Closeness to Manual Expert Analysis compares automated classifications to expert evaluations, as highlighted in [P24], [P43], [P53], [P57].

B. Benchmarking

The reviewed PSs demonstrate a strong emphasis on benchmarking to evaluate and compare the effectiveness of their approaches. Direct method comparisons are the most commonly conducted, appearing in 51.85% of the PSs (n=42), while evolutionary method comparisons are less frequent, accounting for only 7.41% of the PSs (n=6).

1) *Direct Method Comparisons*: This category includes methodologies that assess the performance of ML models against existing approaches and techniques. Most PSs utilized comparisons with the State of the Art to benchmark current techniques against the latest advancements in the field, as well as against traditional rule-based and heuristic methods. For example, [P11], [P12], [P42], [P45], [P53], [P56], [P64] highlight comparisons across ML-based approaches and traditional techniques. Additionally, some PSs focused on the effectiveness of widely used algorithms and techniques, such as clustering algorithms like K-Means, DBSCAN, and their variations. These were explored in studies like [P68]–[P70].

2) *Evolutionary Method Comparisons*: This category emphasizes the development of new methodologies built upon previous research. A few PSs aimed at enhancing existing Research by improving established techniques through the integration of novel features or adaptations for specific microservices migration scenarios. For instance, [P8], [P58] demonstrated enhancements to clustering and service identification methodologies. Other PSs explored combinatory approaches, combining different methodologies to leverage their strengths and improve migration outcomes. Examples include hybrid methods that integrate clustering with deep learning models for better service decomposition, as seen in [P27], [P53].

C. Success Criteria

The success of ML-based migration approaches is determined by various criteria. Technical performance is the most emphasized, appearing in 55.56% of the PSs (n=45), with

key indicators including improvements in precision, recall, and resource optimization. Structural quality is considered in 35.80% of the PSs (n=29), focusing on modularity and cohesion. Outcome alignment with standards and baselines is assessed in 29.63% of the PSs (n=24). Functional and domain-specific success is addressed in 25.93% of the PSs (n=21), while system effectiveness and adaptability are evaluated in only 17.28% of the PSs (n=14).

1) *Technical Performance Metrics*: These metrics focus on the operational effectiveness of the ML models and their impact on system performance. Key indicators include increased precision, recall, and F1-Score, which demonstrate the effectiveness of the proposed ML-based approach in performing specific tasks, as shown in [P28], [P31], [P45], [P47]. Reduced response times and latency indicate improved system performance, leading to better user experiences and satisfaction, as highlighted in [P14], [P25], [P61]. Enhanced efficiency, reflecting the ability of the migration process to operate with limited resources while achieving desired outcomes, is another critical indicator, as seen in [P16], [P51]. Increased throughput, which measures the system's capacity to handle requests, reflects the effectiveness of the ML approach in optimizing service delivery, as demonstrated in [P5], [P52].

2) *Outcome Alignment with Standards and Baselines*: This criterion evaluates how well the outcomes of the ML-driven migration align with established standards and benchmarks. Achieving a defined baseline of performance metrics is crucial for evaluating the success of the migration, as improvements over this baseline indicate the effective use of ML methodologies. For example, [P3], [P4], [P10] demonstrated surpassing baseline metrics as a success criterion. Additionally, developer architectural alignment ensures that the architecture of the migrated services aligns with developer expectations, with positive feedback from developers serving as a key indicator of success, as seen in [P37], [P65].

3) *Structural Quality*: This category focuses on the integrity and quality of the microservices architecture. Increased modularity ensures that microservices are independently deployable and maintainable, allowing for easier updates and scaling, as noted by [P6], [P71]. Increased cohesion assesses the relatedness of functionalities within a service, as highlighted in [P44], [P58]. Decreased coupling measures the reduction in dependencies between services [P12], [P51].

4) *System Effectiveness and Adaptability*: This criterion evaluates how well the system performs and its ability to adapt to changing requirements. Increased scalability measures the system's ability to scale resources up or down as needed to accommodate varying workloads, as emphasized in [P40], [P48]. Increased resource utilization aims to maximize the effectiveness of resource use during migration, as shown in [P29], [P60]. Increased automation evaluates the extent to which processes are automated, reducing manual intervention and enhancing efficiency, as highlighted in [P13].

5) *Functional and Domain-Specific Success*: This criterion evaluates how well ML models meet domain-specific requirements and objectives. Increased anomaly detection quality enhances the reliability of the migration process and operational stability, as demonstrated in [P30], [P59]. Reduced QoS viola-

tions indicate that the migrated services are performing within acceptable parameters, as seen in [P21], [P42]. Increased cost efficiency reflects the ability to achieve desired outcomes with minimal financial expenditure, as highlighted in [P19], [P46].

D. Tool Types and Availability

Research prototypes are the most commonly reported tool type, appearing in 27.18% of the PSs (n=22), while only 2.47% (n=2) describe production-grade applications. Regarding availability, 69.14% of the PSs (n=56) did not provide information on tool availability, whereas 30% (n=24) described open-source tools. Proprietary tools are mentioned in one PS.

1) *Tool Types*: Research prototypes, developed within research projects for validation purposes, are often partial implementations (20 PSs). Examples include tools proposed in [P18], [P61], [P66], [P71], [P72]. In contrast, production-grade applications are fully developed, robust, and scalable tools ready for real-world deployment, as highlighted in [P3].

2) *Tool Availability*: Open-source tools, made publicly available through platforms like GitHub, enable reuse and collaboration, as seen in [P12], [P51], [P61], [P71]. Proprietary tools, such as those discussed in [P27], have restricted access and are not publicly available. However, the majority of PSs (74%) provide no information on tool availability, limiting reproducibility and reuse.

IX. CHALLENGES IN USING ML APPROACHES (RQ5)

Machine learning has shown potential in supporting various phases of microservices migration. However, several challenges limit its effectiveness and broader adoption. [R4C15] These challenges were derived through manual analysis of the primary studies (PSs), specifically by reviewing: i) the Discussion section, ii) the Threats to Validity section, iii) the Future Work section, and iv) the Conclusion section of each study. Only challenges explicitly reported by the study authors, rather than inferred, were considered to ensure objectivity. The extracted challenges range from technical complexities to data quality issues, scalability concerns, integration difficulties, and specialized skill requirements. This section discusses these challenges and explores possible solutions.

A. Technical Complexities

One of the primary challenges in applying ML to software migration is handling the variability and complexity of legacy systems. Identifying service boundaries within monolithic architectures remains a difficult task, often leading to suboptimal service decomposition that affects system performance, scalability, and maintainability [P9], [P49], [P65]. Nunes *et al.* [P65] emphasize the need to focus on transactional contexts rather than structural dependencies when decomposing monolithic systems. Their approach aims to preserve business logic while enabling effective service extraction. Similarly, Rathod *et al.* [P49] propose using Relational Topic Modeling (RTM) to combine structural dependencies with semantic analysis, ensuring that refactoring operations improve design quality.

Potential Solution: A hybrid approach combining ML techniques with domain-driven design and expert knowledge

can refine results and ensure practical applicability. Increasing model explainability can also help software architects validate and refine service decomposition outcomes.

B. Data Quality and Availability

ML models rely on diverse datasets, including source code repositories, execution traces, logs, and performance metrics. However, data unavailability, inconsistency, incompleteness, and noise significantly impact the accuracy and reliability of ML-based approaches [P7], [P8], [P11]–[P13], [P18], [P22], [P28], [P29], [P31], [P32], [P38], [P41], [P44], [P45], [P47], [P49], [P50], [P53]–[P55], [P58], [P59], [P62], [P65], [P67], [P73]–[P76]. Mathai *et al.* [P58] highlight the dependency of decomposition techniques on external artifacts, such as runtime traces and commit histories, which are often incomplete or unavailable. Similarly, Bajaj *et al.* [P45] stress that inconsistent SDLC artifacts, including use case models and functional requirements, pose challenges in greenfield developments. [R1C10] Beyond data quality issues, the lack of data also plays a critical role in shaping the type of machine learning techniques. The scarcity of large, labeled datasets likely limited the use of deep learning techniques, favoring classical ML and graph-based models instead.

Potential Solution: Robust data preprocessing pipelines, data augmentation techniques, and validation mechanisms can mitigate inconsistencies and improve reliability. Establishing standardized datasets for benchmarking ML models can also facilitate the development of more robust approaches.

C. Scalability Concerns

Scalability becomes a critical issue when ML models are applied to large and complex monolithic systems. These models often experience increased computational overhead, leading to performance bottlenecks and delays in the migration process [P4], [P6], [P17]–[P20], [P23]–[P26], [P29], [P31], [P33], [P35], [P40], [P44], [P48], [P49], [P51], [P54], [P55], [P66], [P69], [P72], [P75]. Tong *et al.* [P24] discuss the inefficiencies of centralized ML strategies in edge-cloud environments, where distributed architectures complicate data collection and synchronization. These challenges highlight the need for efficient processing strategies to handle large-scale migration tasks.

Potential Solution: Scalable ML frameworks, distributed computing, and incremental learning approaches can enhance efficiency. Optimized data vectorization and clustering techniques can also reduce computational costs.

D. Integration Complexities

Integrating ML into migration workflows is complex due to the diversity of tools, platforms, and architectures involved [P22], [P26], [P48], [P65]. Lv *et al.* [P48] introduce a Graph Convolutional Network (GCN) combined with deep reinforcement learning (DRL) to optimize microservice deployment. While this approach enhances decision-making, its integration into existing migration frameworks remains challenging. Ray

et al. [P26] explore ML-driven microservice placement strategies in edge computing, emphasizing the need for adaptive, automated decision-making processes.

Potential Solution: Developing modular ML components with standardized APIs can improve interoperability. Using containerization and microservices for ML components simplifies deployment and integration with existing architectures.

E. Specialized Skills and Resource Requirements

ML-based migration requires expertise in both machine learning and software engineering, which many development teams lack [P13], [P15], [P18], [P20]–[P22], [P40], [P42], [P48], [P52], [P58], [P68], [P70], [P74], [P75]. Santos *et al.* [P40] highlight the complexity of implementing ensemble learning models for performance forecasting, while Dehghani *et al.* [P70] emphasize the difficulty of applying AI-based migration techniques due to their reliance on reinforcement learning and neural networks.

Potential Solution: [R4C13] Combining specialized training for developers with thoughtfully designed black-box tools that abstract technical complexity while preserving essential controls, supported by cross-disciplinary collaboration between software engineers and ML practitioners.

X. DISCUSSION

Figure 6 presents a high-level classification derived from our answers to the research questions, summarising the key findings discussed in the previous sections. Building on this classification, we further analyse the four core dimensions addressed by our study, including Phases, Input Types, ML Approaches, and Evaluation Methods. In addition to examining each dimension independently, we conduct a cross-dimension analysis to reveal how these aspects co-occur across the selected primary studies. Finally, we synthesise these insights into concrete recommendations to support practitioners and researchers in effectively applying ML techniques during migration.

A. Phases

Our findings highlight that Identification is the phase in which ML approaches have been the most applied by researchers, with 39 PSs addressing automating this phase. It is followed by Monitoring (28 PSs) and Deployment (12 PSs), reflecting the research community’s emphasis on defining microservices, monitoring their functionality, and ensuring their deployment into production environments. The pre-migration is perceived as straightforward or reliant on domain-specific knowledge, which limits its appeal to academic exploration. Also, the variability in business logic and technical landscapes across domains makes it challenging to generalise findings, further discouraging research. Additionally, we found no PSs proposing an ML-based approach to assist the Packaging phase. This phase remains largely overlooked, despite relying on extensive code generation and refactoring—both of which are still predominantly manual and time-consuming. Earlier machine learning approaches struggled to automate

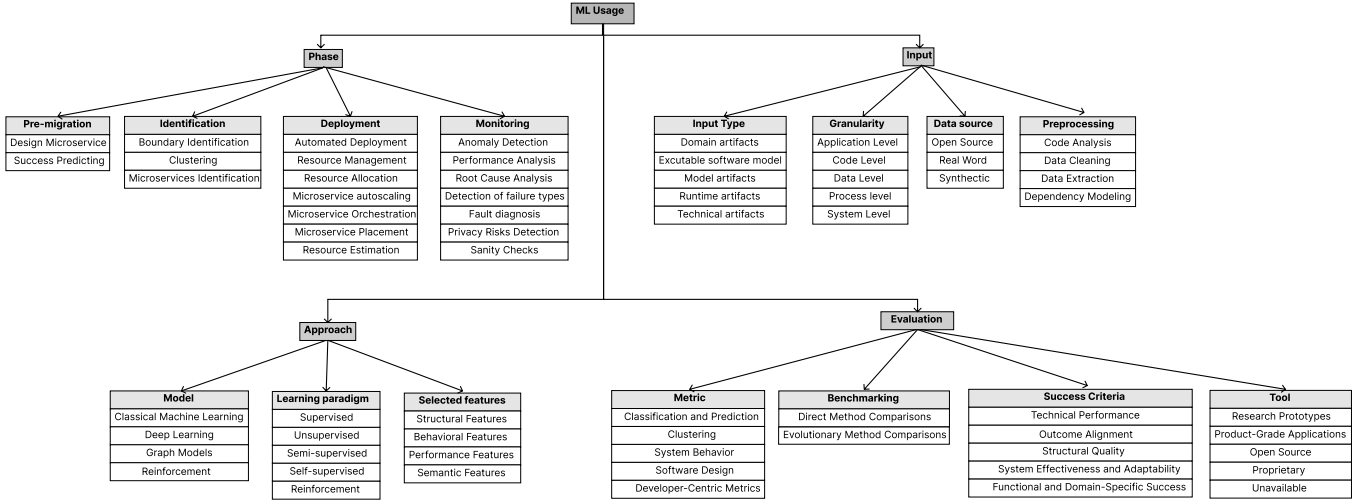


Fig. 6: ML Usage in Migration: A Classification Overview

this phase effectively. However, the advent of Large Language Models (LLMs) presents new opportunities to address these challenges. LLMs have the potential to automate substantial aspects of the code generation process, such as API generation, offering solutions that are not only automated but also more consistent and accurate [41]. [R4C24] This shift highlights an opportunity for future research to explore the potential integration of LLMs into the packaging phase to assist automation efforts. While their actual effectiveness in this context remains to be empirically validated, recent successes of LLMs in tasks such as code generation, bug fixing, and software refactoring suggest that they may offer valuable capabilities for addressing migration-related challenges [42], [43]

B. Input

The effectiveness of any machine learning-based approach, including those for microservices migration, depends on the quality, diversity, and preparation of its input data. The classification categorises inputs into five main types: domain artefacts (e.g., API documentation), Executable software model (e.g., source code), model artifacts (e.g., use cases), runtime artefacts (e.g., resource metrics, logs), and technical artifacts (e.g., Servers information). Among these, runtime artifacts are the most frequently utilised, appearing in 49 PSs, because they are crucial for capturing dynamic system behaviours, making them particularly valuable in monitoring and deployment phases. In contrast, domain and model artifacts are predominantly used during the identification and pre-migration phases.

Data sources vary, with most PSs relying on open-source data (50 PSs), due to their accessibility. Real-world data sources (28 PSs), and synthetic data sources (11 PSs) provide additional data, though their use is limited by accessibility and confidentiality challenges. The limited use of real-world and industrial systems justifies the gap in using academic findings on practical, large-scale applications.

Preprocessing steps, including data extraction (39 PSs), data cleaning (24 PSs), code analysis (47 PSs), and dependency modelling (21 PSs), are used for preparing data for machine learning models. These steps ensure that raw data is transformed into a structured and meaningful format, enabling models to effectively analyze system components, dependencies, and relationships. While these preprocessing techniques have been explored in various studies, challenges such as handling incomplete or noisy data, and automating complex code analysis remain critical areas for further research.

C. Approaches

The PSs employ a diverse range of ML models that we categorised into classical machine learning, deep learning, graph-based methods, and reinforcement learning. Among these, classical machine learning is the most prominent, appearing in 30 PSs, with clustering, classification, and regression techniques commonly used in the PSs. [R1C8] However, a cross-analysis of ML technique usage over time (see Figure 7) reveals a temporal trend: while classical ML dominated earlier works (before 022), recent PSs show increasing adoption of deep learning and graph-based approaches. Deep learning is used particularly in the monitoring phase to model complex behaviors and predict root causes of failures, leveraging architectures such as autoencoders, transformers, and recurrent neural networks. Graph-based methods, used in 23 PSs, are particularly effective in capturing structural relationships between code components or log traces. Reinforcement learning appears in 19 PSs, mainly in deployment and monitoring phases, where it is applied for adaptive resource management and anomaly detection. Supervised learning is predominantly employed in monitoring, where labeled data is available. Unsupervised learning techniques are reported in 55 PSs, especially in the identification phase, where labeled datasets are typically unavailable. For instance, clustering algorithms like DBSCAN

and K-Means are used in 6 PSs to identify microservices in monolithic architectures, with evaluation criteria focusing on cohesion and coupling.

ML models in these PSs leverage a variety of feature types to enhance different phases of the migration process. Performance features (24 PSs), such as CPU usage, memory consumption, and response time, are commonly used for performance prediction, anomaly detection, and deployment optimization. Structural Features (51 PSs) enable ML models to learn structural and relational patterns, facilitating service decomposition, impact analysis, and dependency resolution. Semantic features (30 PSs), such as function names, API descriptions, and business logic embeddings, support domain-aware clustering and automated service identification by capturing functional similarities. Behavioural features (35 PSs), including invocation paths, transactional dependencies, and execution traces, provide insights into runtime behaviour, helping ML models understand component interactions and operational patterns. While these feature types enhance model effectiveness, challenges remain in automating feature extraction and handling noisy dependencies.

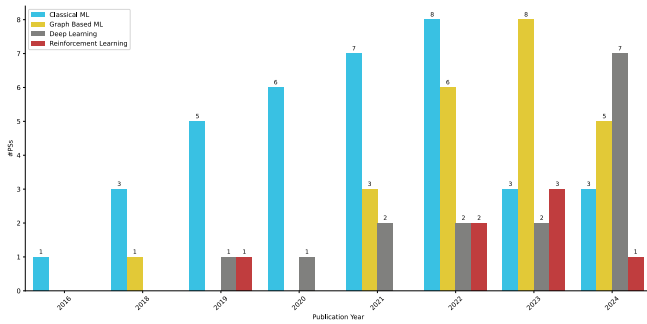


Fig. 7: [R1C8] Trends in ML Models usage across publication years

D. Evaluation

The evaluation of microservices migration approaches relies on a combination of metrics, benchmarking, and success criteria. The analysis of evaluation metrics in the reviewed PSs highlights a strong emphasis on classification and prediction metrics (45 PSs), which are predominantly used to assess the accuracy and effectiveness of ML models in microservices identification and anomaly detection. Software design metrics (23 PSs) follow, focusing on cohesion, coupling, and modularity to ensure maintainability and architectural quality exclusively used in the identification phase. System behaviour metrics (18 PSs), such as response time and resource utilization, are widely used to evaluate the performance and scalability of migrated microservices. Clustering metrics (12 PSs) provide insights into the quality of service decomposition, ensuring well-formed and meaningful microservices groupings. In contrast, developer-centric metrics (3 PSs) remain underexplored, despite their importance in validating the practicality and usability of automated migration approaches.

Benchmarking is a key validation method, with direct method comparisons being the most common (42 PSs), where

approaches are evaluated against state-of-the-art techniques. Evolutionary method comparisons, which assess how approaches improve with new features, are significantly less frequent (6 PSs). Expanding evolutionary benchmarking could offer deeper insights into model robustness, scalability, and long-term adaptability.

[R3C4] Tool availability remains a major concern across both academic and industrial contexts. Among the 81 primary studies reviewed, only 21 provide open-source implementations, while 60 do not disclose any accessible tool, severely limiting reproducibility, evaluation, and practical adoption. To complement our literature analysis, we conducted a supplementary web search using the query "microservices migration" (tool OR platform OR product) (commercial OR enterprise OR SaaS). This search revealed that three commercial tools exist, notably IBM Mono2Micro, vFunction, and Commercetools Composable Commerce, which support aspects of migration such as service boundary detection, observability-based decomposition, and incremental modernisation. However, these tools are commercially licensed, not open source, and do not explicitly leverage machine learning techniques. This reinforces our observation that ML-driven support for automated migration remains limited, non-reproducible, and poorly standardised. Future work should prioritise the development of publicly accessible tools and transparent ML-based solutions, along with standardised evaluation frameworks, to ensure both industrial relevance and scientific rigour.

E. [R1C9, R3C5, R4C1] Cross-Dimensions Analysis

While each research question focused on a specific dimension, we extended our analysis with a cross-dimensions examination to explore how migration phases, input types, ML models, and evaluation metrics co-occur across the selected primary studies. To support this analysis, we constructed a *Sankey Diagram* (Figure 8) that visualises complete migration paths derived from the selected primary studies. Each stream in the diagram corresponds to a unique combination of four key dimensions: Migration Phase → Input Type → ML Model → Evaluation Metric. The width of each stream is proportional to the number of PSs following that path, while the colour indicates the originating migration phase. An interactive version of this diagram is available online at this link¹, where hover tooltips display the full path along with the corresponding number of supporting PSs.

To structure our cross-dimensional analysis, we first identified the most frequent migration paths across all reviewed studies. These represent combinations of migration phase, input type, ML technique, and evaluation metric that occur most often. Table VI presents the prominent paths reported in five or more primary studies.

This analysis highlights some dominant paths in how machine learning is applied across microservices migration phases. The most frequent migration paths emerge in the monitoring phase. These studies typically rely on runtime artifacts,

¹An interactive version of this figure is available at: <https://imen-trabelsi.github.io/SLR-MS-Migration/Cross-RQs-Interactive-Diagram.html>. Hovering over a stream reveals the complete path and the associated number of PSs.

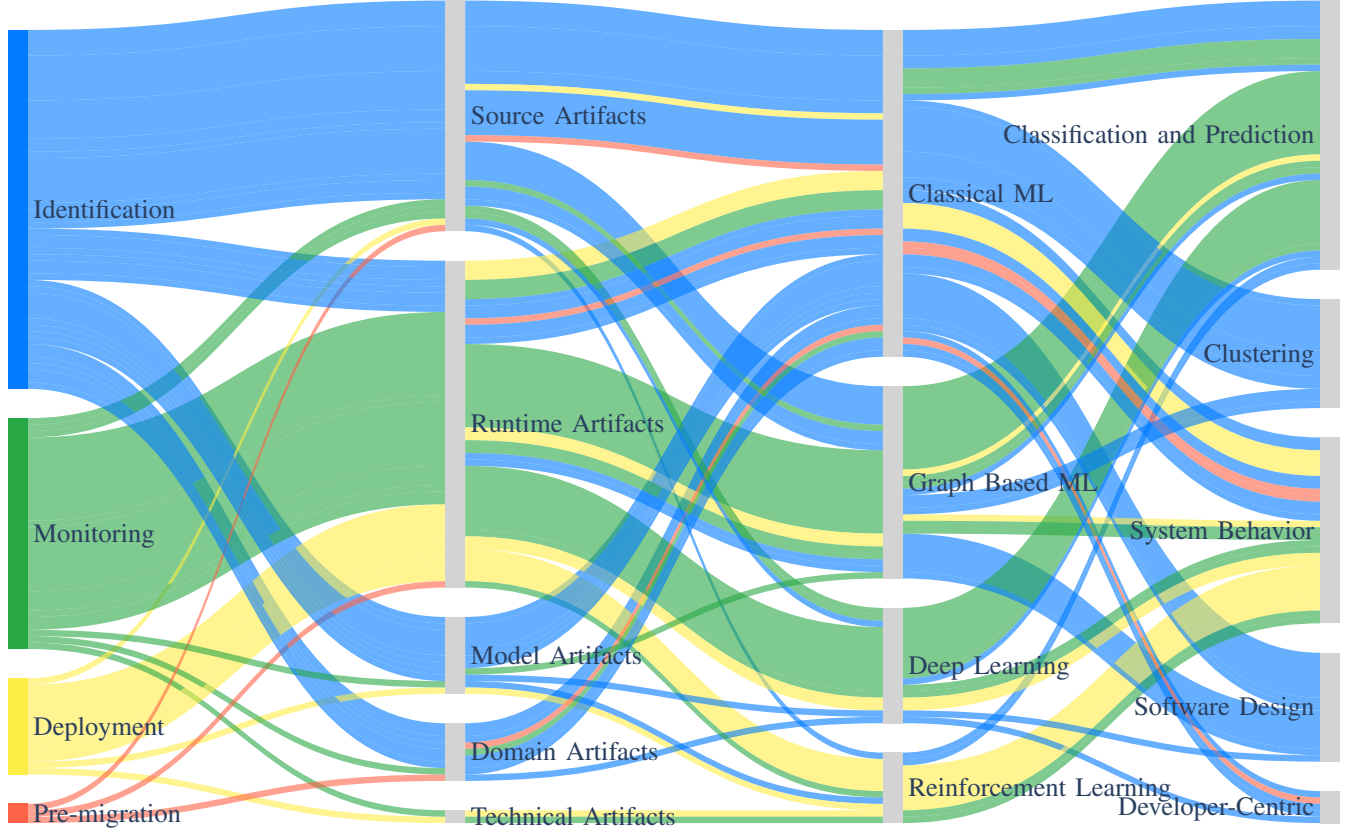


Fig. 8: [R1C9, R3C5, R4C1] Cross-dimensions analysis of ML-based migration approaches. Each stream represents a path linking Migration Phase, Input Type, ML Technique, and Evaluation Metric.

TABLE VI: [R1C9, R3C5, R4C1] Most Frequent Migration Paths across Phase, Input, ML Technique, and Evaluation

Phase	Input Type	ML Technique	Evaluation Metric	#PSs
Monitoring	Runtime Artifacts	Graph Based ML	Classification and Prediction	13
Monitoring	Runtime Artifacts	Deep Learning	Classification and Prediction	10
Identification	Source Artifacts	Classical ML	Software Design	7
Identification	Source Artifacts	Classical ML	Clustering	7
Identification	Source Artifacts	Graph Based ML	Software Design	6
Deployment	Runtime Artifacts	Reinforcement Learning	System Behavior	5

such as logs, execution traces, and system-level metrics, and apply either graph-based ML or deep learning. Graph-based ML is used in 13 primary studies ([P21], [P28]–[P30], [P32]–[P34], [P39], [P41], [P69], [P73], [P74], [P77]). Deep learning is used in 10 PSs ([P17], [P18], [P31], [P52], [P59], [P60], [P68], [P76], [P78], [P79]). Both configurations are primarily evaluated using classification and prediction metrics, such as precision, recall, F1-score, and AUC. These studies address anomaly detection, fault prediction, or behavioral monitoring in production environments.

In the identification phase, the dominant input remains source artifacts, including source code, call graphs, and code embeddings. Classical machine learning techniques such as K-Means, DBSCAN, and Agglomerative Clustering are applied in conjunction with either software design metrics 7 PSs ([P9],

[P11], [P51], [P56], [P57], [P61], [P63]) apply classification models to source code, or clustering to model-level inputs such as use-case diagrams and class diagrams.

In the deployment phase, most approaches rely on runtime artifacts and leverage reinforcement learning for dynamic optimization, resource allocation, or autoscaling. These appear in 5 studies ([P14], [P20], [P22], [P26], [P48]), where the evaluation relies on system behavior metrics such as response time, memory consumption, or CPU utilization. A few additional studies (e.g., [P16]) use classical ML or deep learning, but they are less common in this phase.

Several underexplored paths also emerged. These include configurations where domain-level artifacts (e.g., API documentation) are combined with classical ML and evaluated with system behaviours metrics (e.g., [P3]) or cases where runtime input is introduced in the identification phase to guide service boundary prediction through behavioral insight (e.g., [P37], [P46]).

Beyond full migration paths, we examined pairwise relationships between individual dimensions to surface finer-grained patterns. For instance, input types vary significantly across phases. Source artifacts dominate the identification phase in 26 PSs, where they support architectural decomposition and

service extraction (e.g., [P5], [P11], [P49], [P61], [P71]). Runtime artifacts are prevalent in both monitoring and deployment, used to inform operational decisions or detect abnormal behaviors [P35], [P75], [P78]. Domain and model artifacts are primarily used in pre-migration or identification, reflecting their importance in early planning [P2], [P43]. Configuration artifacts appear almost exclusively in deployment studies that target environmental tuning or resource scheduling [P22], [P26].

Similarly, ML technique choices differ markedly across phases. Classical ML is most prominent in the identification phase, used for clustering and classification (e.g., [P11], [P37], [P46], [P66]). Graph-based models also used for the identification phase, particularly when structural relations are central to the decomposition logic [P12], [P65]. Approaches for monitoring, on the other hand, favor deep learning techniques that can model temporal or sequential patterns (e.g., [P34], [P41], [P60], [P74]). Reinforcement learning appears predominantly in the deployment phase, applied in contexts requiring adaptive policy learning [P14], [P20], [P22], [P26], [P48].

Evaluation practices reflect these distinctions. Studies in the identification phase rely mostly on software design metrics, such as modularity, cohesion, and coupling, which evaluate the structural quality of identified microservices candidates [P5], [P49], [P53], [P56], [P62], [P64], [P71]. Most monitoring studies (17 out of 18 PSs) rely on classification and prediction metrics such as precision, recall, F1-score, and AUC. However [P24] validated their approach using only system behavior metrics, including Average Waiting Time (AWT), SLA violation rate, latency, and the standard deviation of autoscaling performance. Other studies [P17], [P21], [P34] combined classification metrics with system behavior metrics. All deployment studies apply system behavior metrics such as latency or throughput to assess performance improvements and adaptability. Developer-centric or human-in-the-loop metrics are rare, particularly in pre-migration [P2], and during the identification phase [P3], [P37], [P80].

We also examined input–technique pairings. Source artifacts are mostly combined with classical ML and, to a lesser extent, graph-based ML [P12], [P39], [P81]. Runtime artifacts drive the use of deep learning and reinforcement learning, given their rich behavioral data and need for temporal modeling [P14], [P15], [P20], [P48], [P69], [P78]. Domain artifacts are used with supervised techniques aimed at classification or prediction tasks during specially during the identification phase [P10], [P47].

F. Recommendations for Researchers and Practitioners

Based on the insights gathered, we provide recommendations to address the challenges of using machine-learning during the migration from monolithic systems to microservices.

a) Recommendations for Researchers: We present researchers with current challenges and future research directions to address key obstacles and enhance the effectiveness of ML-based migration approaches.

Explore hybrid ML models that offer a promising approach for improving microservice identification, deployment optimization, and anomaly detection by leveraging the

strengths of multiple learning paradigms.

Rationale: Findings from RQ4 emphasise the need for hybrid models to tackle the challenges of microservices migration. For example, graph-based ML models (e.g., GNNs) combined with clustering algorithms (e.g., DBSCAN, K-Means) can enhance microservice identification by capturing both structural dependencies and functional clustering. For deployment optimization, integrating reinforcement learning (RL) with predictive models (e.g., LSTMs) can enable dynamic resource allocation and adaptive scaling based on workload patterns. Future research should focus on implementing hybrid ML techniques that enhance automation, adaptability, and interpretability in microservices migration, ensuring they effectively handle evolving architectures, dynamic workloads, and complex service interactions while maintaining high efficiency.

Address data accessibility issues by collaborating with industry stakeholders to create anonymised datasets.

Rationale: Insights from RQ2 (Input) and RQ5 (Data Availability) underscore the importance of data accessibility and quality to ensure the effectiveness and reliability of ML approaches. Collaboration with industry stakeholders facilitates the creation of datasets that reflect real-world complexities while safeguarding privacy. Researchers should develop standardised anonymisation techniques, propose privacy-preserving data-sharing frameworks, and advocate for open-access benchmark datasets to improve reproducibility. Additionally, they should engage with industry partners to establish domain-specific data collection methodologies that ensure ML models are trained on representative and diverse datasets.

Standardise evaluation metrics by developing universal benchmarking standards and engaging with academic and industrial communities to define success criteria for migration.

Rationale: Standardised metrics ensure fairness, reproducibility, and comparability across research efforts, addressing inconsistencies in assessing ML-based migration approaches. Insights from RQ4 (Evaluation) and RQ5 (Evaluation Challenges) emphasize the necessity of well-defined benchmarks to advance the field effectively.

[R4C2, R3C3] Explore the potential of ML in underexplored migration phases, including both pre-migration and packaging. While our findings (RQ1) show that most ML-based approaches focus on identification, deployment, and monitoring, we observed significantly fewer efforts targeting the earlier phases, particularly business process analysis and modernisation, or the intermediate steps like code refactoring and service packaging.

Rational: Targeted research in these areas could provide tangible benefits, especially where current approaches rely heavily on manual expertise. For instance, the use of ML and AI techniques (e.g., NLP, process mining, or large language models) to support automated business process understanding and transformation could bridge the gap between legacy processes and microservice-oriented design. Similarly, leveraging ML to automate packaging and infrastructure preparation (e.g., configuration, API scaffolding) could reduce migration cost and risk. We encourage future work to investigate these directions where ML’s strengths (pattern discovery, learning from examples, code synthesis) align with clearly defined and

repetitive tasks.

b) Recommendations for Practitioners: From the insights we collected, we provide practical guidance to help practitioners overcome key challenges and optimise the application of ML techniques during migration.

Adopt hybrid approaches that combine machine-learning techniques with business elements to address critical challenges in microservice migration.

Rationale: Drawing from our findings in RQ4 and RQ5, hybrid approaches are essential for resolving challenges such as defining microservice boundaries, determining appropriate service sizes, and selecting optimal deployment resources. These approaches integrate the strengths of machine-learning techniques with business elements, enabling a balanced resolution of technical and business challenges. By aligning ML techniques usage with specific business objectives through iterative workflows and cross-disciplinary discussions, organisations can improve the accuracy and relevance of the proposed approach. Adaptive frameworks, incorporating user feedback and domain knowledge, further ensure deployment optimisation, cost-efficiency, and alignment with organisational goals.

Invest in data quality by prioritising high-quality datasets and establishing workflows for preprocessing and validation.

Rationale: Findings from RQ2 and RQ5 underscore that high-quality data is critical for the reliability and effectiveness of ML models during migration phases. Practical workflows for validation and robust preprocessing in datasets. Utilising effective tools for cleaning and structuring data ensures its usability, while collaboration with domain experts aligns datasets with business objectives. This alignment improves the reliability of machine-learning techniques and ensures that the outcomes meet organisational needs.

Adapt scalable ML approaches to address the complexities of microservices migration.

Rationale: Findings from RQ4 and RQ5 emphasise that scalable ML approaches are essential for managing the increasing complexities of microservices migration as software systems are getting larger. Distributed computing techniques enable efficient processing of large-scale systems, reducing processing time and improving the scalability of the proposed approach. Monitoring tools can track resource usage, predict workload trends, and dynamically optimise resource allocation, ensuring efficient resource management during deployment. By aligning these approaches with microservices architecture constraints, organisations can minimise downtime, reduce operational bottlenecks, and ensure scalability to more complex systems.

Upskill teams through continuous learning initiatives tailored to both machine-learning and software engineers

Rationale: Findings from RQ5 highlight that successful microservices migration requires teams to possess a blend of specialised skills and collaborative expertise. Comprehensive learning initiatives, such as hands-on workshops, scenario-based case studies, and real-world projects, bridge the gap between theoretical knowledge and practical application. Fostering cross-functional collaboration among domain experts, ML engineers, and software architects ensures seamless teamwork and alignment. By keeping teams updated with the latest technological advancements, organisations can empower

them to tackle both technical and collaborative challenges effectively, driving successful migration outcomes.

XI. THREATS TO VALIDITY

In this section, we identify potential threats associated with construct validity, internal validity, external validity, and conclusion validity.

a) Construct Validity: This aspect focuses on the sources used and the approach adopted for data collection. It includes the selection process of primary studies and the methodology employed to extract data in relation to the research questions.

Exclusion of relevant studies: The possibility of overlooking relevant studies poses a threat to the study. Since our search was based on titles and keywords, there is a chance that some relevant studies were unintentionally excluded. The effectiveness of our search depends largely on how well digital libraries index research papers. To address this issue, we used seven widely recognized and comprehensive databases for literature reviews [34]. Additionally, we conducted four rounds of snowballing to identify further potentially relevant studies.

Selection bias in PSs: The process of selecting PSs may have led to the exclusion of relevant studies. As the selection was conducted manually, subjective judgment could have influenced the final set of chosen studies. To mitigate this risk, we clearly defined the study's objectives and research questions beforehand, adhering to the PRISMA guidelines. We also established well-defined inclusion and exclusion criteria.

Bias in data extraction: Since data extraction was performed manually, there is a risk of personal bias affecting the results. To reduce this risk, we designed a structured data collection form. Two researchers independently carried out the extraction process following the predefined form. We ensured consistency through interrater agreement and held multiple discussions involving all authors to reach a consensus. Furthermore, inconsistencies in terminology across PSs posed another challenge. We addressed this by systematically discussing and agreeing upon a standardized vocabulary.

[R3C1] Scope Boundary Definitions: Our study exclusively examined ML-based approaches. Consequently, we excluded approaches that do not explicitly involve ML techniques—such as those based solely on process mining or architectural heuristics. This restricted scope enabled a focused analysis of the specific characteristics and limitations of ML techniques. However, we acknowledge that non-ML approaches, including process mining methods (e.g., [44]), can play a complementary role. In particular, process mining can generate valuable artifacts such as event logs or execution traces, which may serve as input to ML models. Future research could beneficially explore such hybrid combinations, capitalizing on the complementary strengths of process-oriented and learning-based techniques.

b) Internal Validity: Internal validity pertains to the methods employed in this study and the accuracy of the conclusions derived from them.

Review completeness: This study primarily investigates the use of ML techniques in the migration process: phases, input, approach, evaluation, and challenges. Some PSs may not

provide exhaustive details on all these aspects. To minimize this limitation, we included only those studies that can answer at least three RQs.

Research methodology: The methodology used in this study may introduce certain biases, potentially impacting the reliability and validity of the findings. To mitigate this risk, we followed the updated PRISMA guidelines and carefully curated our study selection process to ensure relevance. The inclusion and exclusion criteria were rigorously defined and reviewed by all authors to enhance objectivity.

c) *External Validity:* This concerns the extent to which our findings can be generalized across all migration approaches that use ML. Our review focuses exclusively on academic literature, meaning that industry practices may not be fully represented if they were not documented in research papers. Additionally, we only considered studies published within a specific timeframe, which could limit the generalizability of our conclusions. Nonetheless, we plan to conduct an industry-focused study to supplement these findings.

d) *Conclusion Validity:* This aspect pertains to the soundness of the conclusions drawn from the extracted data. To ensure validity, we conducted a thorough analysis of the extracted data and engaged in multiple discussion sessions to cross-verify our conclusions. We strictly based our findings on the data derived from the selected PSs, ensuring that our conclusions remain well-supported and justified.

XII. CONCLUSION

This systematic literature review examines the role of machine learning in automating the migration from monolithic systems to microservices. Using a PRISMA-based methodology, we analyzed 81 primary studies to understand the automated phases, the types of inputs used, the ML techniques applied, the evaluation methods used, and the challenges encountered. Our findings indicate that ML-driven migration studies have primarily focused on service identification, monitoring, and deployment, while pre-migration analysis and microservice packaging remain largely unexplored. In particular, automating code generation and packaging tasks has received limited attention, despite the potential of emerging technologies such as Large Language Models. Additionally, runtime artifacts serve as the dominant data source, yet the scarcity of real-world datasets raises concerns about the practical applicability of ML techniques. Unsupervised learning remains the most common approach, particularly for service identification and anomaly detection, but evaluation practices vary widely, with system adaptability and real-world validation often overlooked. Despite ML's potential to automate migration tasks, key challenges persist, including data availability, scalability, tool support, and the absence of standardized benchmarking. Addressing these challenges requires further research on underexplored migration phases, particularly pre-migration planning and packaging. Enhancing data accessibility through industry collaboration, privacy-preserving data-sharing frameworks, and the development of benchmark datasets will be crucial. Furthermore, exploring hybrid ML approaches that integrate multiple learning paradigms could improve accuracy

and adaptability in migration processes. By addressing these gaps, future research can contribute to more effective and scalable ML-driven migration solutions.

REPLICATION PACKAGE

The replication package can be accessed using the following DOI: 10.5281/zenodo.15723658

ACKNOWLEDGMENTS

The Canada Research Chair program partly funded this work.

REFERENCES

- [1] N. Ali, S. Baker, R. O’Crowley, S. Herold, and J. Buckley, “Architecture consistency: State of the practice, challenges and requirements,” *Empirical Software Engineering*, vol. 23, pp. 224–258, 2018.
- [2] N. Dragoni, I. Lanese, S. T. Larsen, M. Mazzara, R. Mustafin, and L. Safina, “Microservices: How to make your application scale,” *IEEE Software*, vol. 34, no. 5, pp. 81–85, 2017.
- [3] J. Fritzsch, J. Bogner, S. Wagner, and A. Zimmermann, “Microservices migration in industry: intentions, strategies, and challenges,” in *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2019, pp. 481–490.
- [4] F. Tapia, M. Á. Mora, W. Fuertes, H. Aules, E. Flores, and T. Toulkeridis, “From monolithic systems to microservices: A comparative study of performance,” *Applied sciences*, vol. 10, no. 17, p. 5797, 2020.
- [5] Y. Abgaz, A. McCarren, P. Elger, D. Solan, N. Lapuz, M. Bivol, G. Jackson, M. Yilmaz, J. Buckley, and P. Clarke, “Decomposition of monolith applications into microservices architectures: A systematic review,” *IEEE Transactions on Software Engineering*, vol. 49, no. 8, pp. 4213–4242, 2023.
- [6] S. Sarkar, G. Vashi, and P. Abdulla, “Towards transforming an industrial automation system from monolithic to microservices,” in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1. IEEE, 2018, pp. 1256–1259.
- [7] M. Abdellatif, A. Shatnawi, H. Mili, N. Moha, G. El Boussaidi, G. Hecht, J. Privat, and Y.-G. Guéhéneuc, “A taxonomy of service identification approaches for legacy software systems modernization,” *Journal of Systems and Software*, vol. 173, p. 110868, 2021.
- [8] N. Toumi, M. Bagaa, and A. Ksentini, “Machine learning for service migration: a survey,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 3, pp. 1991–2020, 2023.
- [9] I. Oumoussa and R. Saidi, “Evolution of Microservices Identification in Monolith Decomposition: A Systematic Review,” *IEEE Access*, 2024.
- [10] J. Fritzsch, J. Bogner, S. Wagner, and A. Zimmermann, “Microservices migration in industry: intentions, strategies, and challenges,” in *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2019, pp. 481–490.
- [11] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiaq, N. Rafa, M. Mofijur, A. Shawkat Ali, and A. H. Gandomi, “Deep learning modelling techniques: current progress, applications, advantages, and challenges,” *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13 521–13 617, 2023.
- [12] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, *et al.*, “The PRISMA 2020 Statement: An Updated Guideline For Reporting Systematic Reviews,” *International journal of surgery*, vol. 88, p. 105906, 2021.
- [13] S. Moreschini, S. Pour, I. Lanese, D. Balouek, J. Bogner, X. Li, F. Pecorelli, J. Soldani, E. Truyen, and D. Taibi, “AI Techniques in the Microservices Life-Cycle: a Systematic Mapping Study,” *Computing*, vol. 107, no. 4, p. 100, 2025.
- [14] D. Narváez, N. Battaglia, A. Fernández, and G. Rossi, “Designing Microservices Using AI: A Systematic Literature Review,” *Software*, vol. 4, no. 1, p. 6, 2025.
- [15] A. M. Saucedo, G. Rodríguez, F. G. Rocha, and R. P. dos Santos, “Migration of monolithic systems to microservices: A systematic mapping study,” *Information and Software Technology*, p. 107590, 2024.
- [16] —, “Migration of monolithic systems to microservices: A systematic mapping study,” *Information and Software Technology*, p. 107590, 2024.

- [17] A. Razzaq and S. A. Ghayyur, "A systematic mapping study: The new age of software architecture from monolithic to microservice architecture—awareness and challenges," *Computer Applications in Engineering Education*, vol. 31, no. 2, pp. 421–451, 2023.
- [18] Y. Abgaz, A. McCarren, P. Elger, D. Solan, N. Lapuz, M. Bivol, G. Jackson, M. Yilmaz, J. Buckley, and P. Clarke, "Decomposition of monolith applications into microservices architectures: A systematic review," *IEEE Transactions on Software Engineering*, vol. 49, no. 8, pp. 4213–4242, 2023.
- [19] V. Velepucha and P. Flores, "Monoliths to microservices-migration problems and challenges: A SMS," in *2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)*. IEEE, 2021, pp. 135–142.
- [20] V. Bushong, A. S. Abdelfattah, A. A. Maruf, D. Das, A. Lehman, E. Jaroszewski, M. Coffey, T. Cerny, K. Frajtak, P. Tisnovsky, et al., "On microservice analysis and architecture evolution: A systematic mapping study," *Applied Sciences*, vol. 11, no. 17, p. 7856, 2021.
- [21] F. Ponce, G. Márquez, and H. Astudillo, "Migrating from monolithic architecture to microservices: A Rapid Review," in *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*. IEEE, 2019, pp. 1–7.
- [22] J. Fritzsche, J. Bogner, A. Zimmermann, and S. Wagner, "From monolith to microservices: A classification of refactoring approaches," in *Software Engineering Aspects of Continuous Development and New Paradigms of Software Production and Deployment: First International Workshop, DEVOPS 2018, Chateau de Villebrumier, France, March 5-6, 2018, Revised Selected Papers I*. Springer, 2019, pp. 128–141.
- [23] H. C. d. Silva Filho and G. d. Figueiredo Carneiro, "Strategies reported in the literature to migrate to microservices based architecture," in *16th International Conference on Information Technology-New Generations (ITNG 2019)*. Springer, 2019, pp. 575–580.
- [24] A. Mpampoutis and G. Kakarontzas, "Using Database Schemas of Legacy Applications for Microservices Identification: A Mapping Study," in *Proceedings of the 6th International Conference on Algorithms, Computing and Systems*, 2022, pp. 1–7.
- [25] P. Di Francesco, P. Lago, and I. Malavolta, "Migrating towards microservice architectures: an industrial survey," in *2018 IEEE international conference on software architecture (ICSA)*. IEEE, 2018, pp. 29–2909.
- [26] M. Kalske, N. Mäkitalo, and T. Mikkonen, "Challenges when moving from monolith to microservice architecture," in *Current Trends in Web Engineering: ICWE 2017 International Workshops, Liquid Multi-Device Software and EnWoT, practi-O-web, NLPIT, SoWeMine, Rome, Italy, June 5-8, 2017, Revised Selected Papers 17*. Springer, 2018, pp. 32–47.
- [27] V. Velepucha and P. Flores, "Monoliths to microservices-migration problems and challenges: A SMS," in *2021 Second International Conference on Information Systems and Software Technologies (ICI2ST)*. IEEE, 2021, pp. 135–142.
- [28] R. Capuano and H. Muccini, "A systematic literature review on migration to microservices: a quality attributes perspective," in *2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C)*. IEEE, 2022, pp. 120–123.
- [29] J. Kazanavičius and D. Mažeika, "Migrating legacy software to microservices architecture," in *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*. IEEE, 2019, pp. 1–5.
- [30] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., "The PRISMA 2020 Statement: An Updated Guideline For Reporting Systematic Reviews," *International journal of surgery*, vol. 88, p. 105906, 2021.
- [31] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, et al., "PRISMA 2020 Explanation And Elaboration: Updated Guidance And Exemplars For Reporting Systematic Reviews," *bmj*, vol. 372, 2021.
- [32] B. A. Kitchenham, L. Madeyski, and D. Budgen, "SEGRESS: Software Engineering Guidelines For Reporting Secondary Studies," *IEEE Transactions on Software Engineering*, 2022.
- [33] A. Cooke, D. Smith, and A. Booth, "Beyond PICO: the SPIDER Tool For Qualitative Evidence Synthesis," *Qualitative health research*, vol. 22, no. 10, pp. 1435–1443, 2012.
- [34] T. Dyba, T. Dingsøyr, and G. K. Hanssen, "Applying Systematic Reviews To Diverse Study Types: An Experience Report," in *First international symposium on empirical software engineering and measurement (ESEM 2007)*. IEEE, 2007, pp. 225–234.
- [35] H. Vural, M. Koyuncu, and S. Guney, "A systematic literature review on microservices," in *Computational Science and Its Applications-ICCSA 2017: 17th International Conference, Trieste, Italy, July 3-6, 2017, Proceedings, Part VI 17*. Springer, 2017, pp. 203–217.
- [36] J. Cohen, "Weighted Kappa: Nominal Scale Agreement Provision For Scaled Disagreement Or Partial Credit," *Psychological bulletin*, vol. 70, no. 4, p. 213, 1968.
- [37] S. Li, H. Zhang, Z. Jia, C. Zhong, C. Zhang, Z. Shan, J. Shen, and M. A. Babar, "Understanding and addressing quality attributes of microservices architecture: A Systematic literature review," *Information and software technology*, vol. 131, p. 106449, 2021.
- [38] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and Software Technology*, vol. 50, no. 9–10, pp. 833–859, 2008.
- [39] K. P. Murphy, *Probabilistic machine learning: an introduction*. MIT press, 2022.
- [40] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.
- [41] X. Hou, Y. Zhao, Y. Liu, Z. Yang, K. Wang, L. Li, X. Luo, D. Lo, J. Grundy, and H. Wang, "Large language models for software engineering: A systematic literature review," *ACM Transactions on Software Engineering and Methodology*, 2023.
- [42] M. Chen, H. Touvron, T. Lavril, et al., "Code llama: Open foundation models for code," *arXiv preprint arXiv:2308.12950*, 2023.
- [43] W. U. Ahmad, S. Chakraborty, S. Yuan, and G. Neubig, "A context-aware natural language interface for code transformation tasks," in *Proceedings of the 45th International Conference on Software Engineering (ICSE)*, 2023.
- [44] L. Baresi, M. Garriga, and A. De Renzis, "Microservices identification through interface analysis," in *Service-Oriented and Cloud Computing: 6th IFIP WG 2.14 European Conference, ESOC 2017, Oslo, Norway, September 27-29, 2017, Proceedings 6*. Springer, 2017, pp. 19–33.
- [45] E. Djogic, S. Ribic, and D. Donko, "Monolithic to microservices redesign of event driven integration platform," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, pp. 1411–1414.
- [46] I. Oumoussa, S. Faieq, R. Saidi, and M. Daoud, "An nlp-based approach for identifying microservices from a set of business processes," *SSRN Electronic Journal*, 01 2022.
- [47] C.-Y. Fan and S.-P. Ma, "Migrating monolithic mobile application to microservice architecture: An experiment report," in *2017 IEEE international conference on ai & mobile services (aims)*. IEEE, 2017, pp. 109–112.
- [48] D. Faustino, N. Gonçalves, M. Portela, and A. R. Silva, "Stepwise migration of a monolith to a microservice architecture: Performance and migration effort evaluation," *Performance Evaluation*, vol. 164, p. 102411, 2024.

SELECTED PRIMARY STUDIES (PSSs)

- [P1] R. Nakazawa, T. Ueda, M. Enoki, and H. Horii, "Visualization Tool for Designing Microservices with the Monolith-First Approach," *2018 IEEE Working Conference on Software Visualization (VISsOFT)*, pp. 32–42, 2018.
- [P2] A. Alshammari, A. Almadhor, S. Qasem, J. Alkhateeb, and K. Amjad, "High-performance computing-enabled probabilistic framework for migration from monolithic to microservices architecture using genetic algorithms," *Soft Computing*, 2023.
- [P3] M. Gysel, L. Kölbener, W. Giersche, and O. Zimmermann, "Service Cutter: A Systematic Approach to Service Decomposition," *Lecture Notes in Computer Science*, pp. 185–200, 2016.
- [P4] M. Kamimura, K. Yano, T. Hatano, and A. Matsuo, "Extracting Candidates of Microservices from Monolithic Application Code," *2018 25th Asia-Pacific Software Engineering Conference (APSEC)*, pp. 571–580, 2018.
- [P5] T. Zhong, Y. Teng, S. Ma, J. Chen, and S. Yu, "A Microservices Identification Method Based on Spectral Clustering for Industrial Legacy Systems," *2023 IEEE Globecom Workshops (GC Wkshps)*, pp. 1331–1337, 2023.
- [P6] M. Daoud, A. El Mezouari, N. Faci, D. Benslimane, Z. Maamar, and A. El Fazziki, "A multi-model based microservices identification approach," *Journal of Systems Architecture*, vol. 118, p. 102200, 2021.
- [P7] M. Saidi, A. Tissaoui, D. Benslimane, and S. Faiz, "Automatic Microservices Identification Across Structural Dependency," *Lecture Notes in Networks and Systems*, pp. 386–395, 2022.
- [P8] M. Daoud, A. E. Mezouari, N. Faci, D. Benslimane, Z. Maamar, and A. E. Fazziki, "Automatic Microservices Identification from a Set of Business Processes," *Smart Applications and Data Analysis*, pp. 299–315, 2020.
- [P9] V. Faria and A. R. Silva, "Code vectorization and sequence of accesses strategies for monolith microservices identification," pp. 19–33, 2023.

- [P10] X. Sun, S. Boranbaev, S. Han, H. Wang, and D. Yu, "Expert system for automatic microservices identification using API similarity graph," *Journal of Software: Evolution and Process (JSEP)* 2022), 2022.
- [P11] I. Trabelsi, M. Abdellatif, A. Abubaker, N. Moha, S. Mosser, S. Ebrahimi-Kahou, and Y.-G. Guéhéneuc, "From legacy to microservices: A type-based approach for microservices identification using machine learning and semantic analysis," *Journal of Software: Evolution and Process*, vol. 35, p. e2503, 2023.
- [P12] I. Trabelsi, N. Moha, Y.-G. Gueheneuc, and L. Geffard, "Magnet: Method-Based Approach Using Graph Neural Network for Microservices Identification," *2024 IEEE 21st International Conference on Software Architecture (ICSA)*, pp. 1–11, 2024.
- [P13] I. Trabelsi, B. Popa, J. Pereyrol, P.-O. Beaulieu, and N. Moha, "Micro-Matic: Fully Automated Microservices Identification Approach From Monolithic Systems," *Proceedings of the ACM/IEEE 6th International Workshop on Software Engineering Research & Practices for the Internet of Things*, 2024.
- [P14] W. Lv, Q. Wang, P. Yang, Y. Ding, B. Yi, Z. Wang, and C. Lin, "Microservice Deployment in Edge Computing Based on Deep Q Learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, pp. 2968–2978, 2022.
- [P15] S. A. Khan, M. Abdullah, W. Iqbal, M. A. Butt, F. Bukhari, and S.-U. Hassan, "Automatic Migration-Enabled Dynamic Resource Management for Containerized Workload," *IEEE Systems Journal*, vol. 17, pp. 2378–2389, 2023.
- [P16] C. Joseph, J. Martin, K. Chandrasekaran, and A. Kandasamy, "Fuzzy Reinforcement Learning based Microservice Allocation in Cloud Computing Environments," *TENCON 2019 - IEEE Region 10 Conference (TENCON)*, pp. 1559–1563, 2019.
- [P17] S. Luan and H. Shen, "Minimize Resource Cost for Containerized Microservices Under SLO via ML-Enhanced Layered Queueing Network Optimization," *2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 631–637, 2024.
- [P18] Y. Gan, Y. Zhang, K. Hu, D. Cheng, Y. He, M. Pancholi, and C. Delimitrou, "Seer: Leveraging Big Data to Navigate the Complexity of Performance Debugging in Cloud Microservices," *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 19–33, 2019.
- [P19] N. Shafi, M. Abdullah, W. Iqbal, A. Erradi, and F. Bukhari, "Cds-caler: a cost-effective dynamic autoscaling approach for containerized microservices," *Cluster Computing - The Journal of Networks Software Tools and Applications*, 2024.
- [P20] C. Song, M. Xu, K. Ye, H. Wu, S. Gill, R. Buyya, and C. Xu, "Chains-Former: A Chain Latency-Aware Resource Provisioning Approach for Microservices Cluster," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14419, pp. 197–211, 2023.
- [P21] Y. Gan, M. Liang, S. Dev, D. Lo, and C. Delimitrou, "Enabling Practical Cloud Performance Debugging with Unsupervised Learning," *Operating Systems Review (ACM)*, vol. 56, pp. 34–41, 2022.
- [P22] Z. Yang, P. Nguyen, H. Jin, and K. Nahrstedt, "MIRAS: Model-based reinforcement learning for microservice resource allocation over scientific workflows," *Proceedings - International Conference on Distributed Computing Systems*, pp. 122–132, 2019.
- [P23] H. Zeng, T. Wang, A. Li, Y. Wu, and W. Zhang, "Topology-Aware Self-Adaptive Resource Provisioning for Microservices," *2023 IEEE International Conference on Web Services (ICWS)*, pp. 28–35, 2023.
- [P24] G. Tong, C. Meng, S. Song, M. Pan, and Y. Yu, "GMA: Graph Multi-agent Microservice Autoscaling Algorithm in Edge-Cloud Environment," *2023 IEEE International Conference on Web Services (ICWS)*, pp. 393–404, 2023.
- [P25] N. Li, Y. Tan, X. Wang, B. Li, and J. Luo, "SCORE: A Resource-Efficient Microservice Orchestration Model Based on Spectral Clustering in Edge Computing," *Service-oriented Computing (ICSOC 2022)*, vol. 13740, pp. 186–202, 2022.
- [P26] K. Ray, A. Banerjee, and N. Narendra, "Learning-Based Microservice Placement and Migration for Multi-Access Edge Computing," *IEEE Transactions on Network and Service Management*, vol. 17, pp. 1–10, 2023.
- [P27] K.-H. Chow, U. Deshpande, S. Seshadri, and L. Liu, "DeepRest: Deep Resource Estimation for Interactive Microservices," *EuroSys 2022 - Proceedings of the 17th European Conference on Computer Systems*, pp. 181–198, 2022.
- [P28] K. Shi, J. Li, Y. Liu, Y. Chang, and X. Li, "BSDG: Anomaly Detection of Microservice Trace Based on Dual Graph Convolutional Neural Network," *Service-oriented Computing (ICSOC 2022)*, vol. 13740, pp. 171–185, 2022.
- [P29] C. Zhang, X. Peng, C. Sha, K. Zhang, Z. Fu, X. Wu, Q. Lin, and D. Zhang, "DeepTraLog: Trace-Log Combined Microservice Anomaly Detection through Graph-based Deep Learning," *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*, pp. 623–634, 2022.
- [P30] L. Chen, Q. Dang, M. Chen, B. Sun, C. Du, and Z. Lu, "BERTHTLG: Graph-Based Microservice Anomaly Detection Through Sentence-Bert Enhancement," in *Web Information Systems and Applications*. Springer Nature Singapore, 2023, pp. 427–439.
- [P31] Y. Song, R. Xin, P. Chen, R. Zhang, J. Chen, and Z. Zhao, "Autonomous selection of the fault classification models for diagnosing microservice applications," *Future Generation Computer Systems*, vol. 153, pp. 326–339, 2024.
- [P32] Y. Xu, Z. Qiu, H. Gao, X. Zhao, L. Wang, and R. Li, "Heterogeneous Data-Driven Failure Diagnosis for Microservice-Based Industrial Clouds Towards Consumer Digital Ecosystems," *IEEE Transactions on Consumer Electronics*, pp. 1–1, 2023.
- [P33] Y. Sun, Z. Lin, B. Shi, S. Zhang, S. Ma, P. Jin, Z. Zhong, L. Pan, Y. Guo, and D. Pei, "Interpretable Failure Localization for Microservice Systems Based on Graph Autoencoder," *ACM Transactions on Software Engineering and Methodology*, vol. 33, pp. 326–339, September 2024.
- [P34] Y. Gan, M. Liang, S. Dev, D. Lo, and C. Delimitrou, "Sage: Practical and scalable ML-driven performance debugging in microservices," *International Conference on Architectural Support for Programming Languages and Operating Systems - ASPLOS*, pp. 135–151, 2021.
- [P35] J. Chou, E. Al-Masri, S. Kanzehev, and H. Fattah, "Detecting Security and Privacy Risks in Microservices End-to-End Communication Using Neural Networks," *2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII)*, pp. 105–110, 2021.
- [P36] A. Selmadji, A.-D. Seriai, H. L. Bouziane, R. Oumarou Mahamane, P. Zaragoza, and C. Dony, "From Monolithic Architecture Style to Microservice one Based on a Semi-Automatic Approach," *2020 IEEE International Conference on Software Architecture (ICSA)*, pp. 157–168, 2020.
- [P37] G. Morais, D. Bork, and M. Adda, "Towards an Ontology-driven Approach to Model and Analyze Microservices Architectures," *Proceedings of the 13th International Conference on Management of Digital EcoSystems (MEDES)*, pp. 79–86, 2021.
- [P38] Q. Du, T. Xie, and Y. He, "Anomaly Detection and Diagnosis for Container-Based Microservices with Performance Monitoring," *Lecture Notes in Computer Science*, pp. 560–572, 2018.
- [P39] H. Kong, T. Li, J. Ge, L. Zhang, and L. Li, "Enhancing fault localization in microservices systems through span-level using graph convolutional networks," *Automated Software Engineering*, 2024.
- [P40] W. Santos, A. J. Sampaio, N. Rosa, and G. Cavalcanti, "Microservices performance forecast using dynamic Multiple Predictor Systems," *Engineering Applications of Artificial Intelligence*, vol. 243, 2024.
- [P41] J. Chen, F. Liu, J. Jiang, G. Zhong, D. Xu, Z. Tan, and S. Shi, "TraceGra: A trace-based anomaly detection for microservice using graph deep learning," *Comput. Commun.*, vol. 204, pp. 109–117, 2023.
- [P42] M. Abdullah, W. Iqbal, A. Erradi, and F. Bukhari, "Learning Predictive Autoscaling Policies for Cloud-Hosted Microservices Using Trace-Driven Modeling," *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 119–126, 2019.
- [P43] M. Daoud, A. El Mezouari, N. Fati, D. Benslimane, Z. Maamar, and A. El Fazziki, "Towards an Automatic Identification of Microservices from Business Processes," *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, pp. 42–47, 2020.
- [P44] F. Vera-Rivera, E. Puerto, H. Astudillo, and C. Gaona, "Microservices Backlog-A Genetic Programming Technique for Identification and Evaluation of Microservices From User Stories," *IEEE Access*, vol. 9, pp. 117 178–117 203, 2021.
- [P45] D. Bajaj, U. Bharti, I. Gupta, P. Gupta, and A. Yadav, "GTMicro—microservice identification approach based on deep NLP transformer model for greenfield developments," *International Journal of Information Technology*, 2024.
- [P46] M. Abdullah, W. Iqbal, and A. Erradi, "Unsupervised learning approach for web application auto-decomposition into microservices," *Journal of Systems and Software*, vol. 151, pp. 243–257, 2019.
- [P47] O. Al-Debagy and P. Martinek, "A new decomposition method for designing microservices," *Periodica Polytechnica Electrical Engineering and Computer Science*, vol. 63, pp. 274–281, 2019.
- [P48] W. Lv, P. Yang, T. Zheng, C. Lin, Z. Wang, M. Deng, and Q. Wang, "Graph-Reinforcement-Learning-Based Dependency-Aware Microser-

- vice Deployment in Edge Computing,” *IEEE Internet of Things Journal*, vol. 11, pp. 1604–1615, 2024.
- [P49] T. Rathod, C. Joseph, and J. Martin, “Improving Industry 4.0 Readiness: Monolith Application Refactoring using Graph Attention Networks,” *Proceedings - 23rd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing Workshops, CCGridW 2023*, pp. 223–230, 2023.
- [P50] Y. Romani, O. Tibermacine, and C. Tibermacine, “Towards Migrating Legacy Software Systems to Microservice-based Architectures: a Data-Centric Process for Microservice Identification,” *2022 IEEE 19th International Conference on Software Architecture Companion (ICSAC)*, pp. 15–19, 2022.
- [P51] L. Cao and C. Zhang, “Implementation of Domain-oriented Microservices Decomposition based on Node-attributed Network,” *Proceedings of the 2022 11th International Conference on Software and Computer Applications*, pp. 136–142, 2022.
- [P52] R. Li, M. Du, H. Chang, S. Mukherjee, and E. Eide, “Deepstitch: Deep Learning for Cross-Layer Stitching in Microservices,” *Proceedings of the 2020 6th International Workshop on Container Technologies and Container Clouds*, pp. 25–30, 2021.
- [P53] L. Chen, M. Guang, J. Wang, and C. Yan, “Dynamic and static feature-aware microservices decomposition via graph neural networks,” *2023 Knowledge Science, Engineering and Management (KSEM)*, 2023.
- [P54] B. Liu, J. Lu, F. Zhang, W. Zhang, and M. Wang, “Method of Microservices Division for Complex Business Management System Based on Dual Clustering,” *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pp. 2259–2268, 2020.
- [P55] K. Sooksatra, R. Maharjan, and T. Cerny, “Monolith to Microservices: VAE-Based GNN Approach with Duplication Consideration,” *2022 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, pp. 1–10, 2022.
- [P56] O. Al-Debagy and P. Martinek, “A microservice decomposition method through using distributed representation of source code,” *Scalable Computing: Practice and Experience*, vol. 22, no. 1, pp. 39–52, 2021.
- [P57] O. A. Sellami K, Saied MA, “A hierarchical dbscan method for extracting microservices from monolithic applications,” *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*, pp. 201–210, 2022.
- [P58] A. Mathai, S. Bandyopadhyay, U. Desai, and S. Tamilselvam, “Monolith to Microservices: Representing Application Software through Heterogeneous Graph Neural Network,” *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, pp. 3905–3911, 2022.
- [P59] S. Shahini and H. Momeni, “Autoencoder-based Anomaly Detection in Microservices using Distributed Tracing,” *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISIP)*, 2024.
- [P60] R. Tan and Z. Li, “MAAD: A Distributed Anomaly Detection Architecture for Microservices Systems,” *38th IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, 2024.
- [P61] V. Nitin, S. Asthana, B. Ray, and R. Krishna, “CARGO: AI-Guided Dependency Analysis for Migrating Monolithic Applications to Microservices Architecture,” *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, ASE 2022*, 2022.
- [P62] L. Qian, J. Li, X. He, R. Gu, J. Shao, and Y. Lu, “Microservice extraction using graph deep clustering based on dual view fusion,” *Information and Software Technology*, vol. 158, 2023.
- [P63] M. Saidi, A. Tissaoui, and S. Faiz, “A ddd approach towards automatic migration to microservices,” in *2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, 2023, pp. 01–06.
- [P64] J. Liu and C. Zhang, “Migrating Monolith System to Microservices with Directed Graph Attention Neural Network,” *Third International Conference on High Performance Computing and Communication Engineering (HPCCE 2023)*, vol. 13073, 2024.
- [P65] L. Nunes, N. Santos, and A. Rito Silva, “From a Monolith to a Microservices Architecture: An Approach Based on Transactional Contexts,” *Lecture Notes in Computer Science*, vol. 11377, pp. 37–52, 2019.
- [P66] D. Bajaj, U. Bharti, A. Goel, and S. C. Gupta, “Partial migration for re-architecting a cloud native monolithic application into microservices and FaaS,” pp. 111–124, 2020.
- [P67] A. K. Kalia, J. Xiao, C. Lin, S. Sinha, J. Rofrano, M. Vukovic, and D. Banerjee, “Mono2Micro: an AI-based toolchain for evolving monolithic enterprise applications to a microservice architecture,” *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2020.
- [P68] Y. Cai, B. Han, J. Su, and X. Wang, “TraceModel: An Automatic Anomaly Detection and Root Cause Localization Framework for Microservice Systems,” *2021 17th International Conference on Mobility, Sensing and Networking (MSN)*, pp. 512–519, 2021.
- [P69] R. Chen, J. Ren, L. Wang, Y. Pu, K. Yang, and W. Wu, “MicroEGRCL: An Edge-Attention-Based Graph Neural Network Approach for Root Cause Localization in Microservice Systems,” *Service-Oriented Computing: 20th International Conference, ICSC 2022*, vol. 13747, pp. 264–272, 2022.
- [P70] M. Dehghani, S. Kolahdouz-Rahimi, M. Tisi, and D. Tamzalit, “Facilitating the migration to the microservice architecture via model-driven reverse engineering and reinforcement learning,” *Software and Systems Modeling*, vol. 21, pp. 1115–1133, 2022.
- [P71] U. Desai, S. Bandyopadhyay, and S. Tamilselvam, “Graph Neural Network to Dilute Outliers for Refactoring Monolith Application,” *Thirty-fifth AAAI Conference on Artificial Intelligence, Thirty-third Conference on Innovative Applications of Artificial Intelligence and the Eleventh Symposium on Educational Advances in Artificial Intelligence*, vol. 35, pp. 72–80, 2021.
- [P72] O. Al-Debagy and P. Martinek, “Dependencies-based microservices decomposition method,” *Scalable Computing: Practice and Experience*, vol. 22, pp. 39–52, 2021.
- [P73] X. Liang, L. Li, and H. Peng, “Unsupervised Microservice Log Anomaly Detection Method Based on Graph Neural Network,” *International Conference on Swarm Intelligence*, 2024.
- [P74] K. Zhang, C. Zhang, X. Peng, and C. Sha, “PUTraceAD: Trace Anomaly Detection with Partial Labels based on GNN and PU Learning,” *2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE)*, pp. 239–250, 2022.
- [P75] M. Li, D. Tang, Z. Wen, et al., “Microservice anomaly detection based on tracing data using semi-supervised learning,” *International Conference on Artificial Intelligence and Big Data*, pp. 38–44, 2021.
- [P76] J. Wang, Y. Li, Q. Qi, Y. Lu, and B. Wu, “Multilayered Fault Detection and Localization With Transformer for Microservice Systems,” *IEEE Transactions on Reliability*, 2024.
- [P77] J. Huang, Y. Yang, H. Yu, J. Li, and X. Zheng, “Twin Graph-Based Anomaly Detection via Attentive Multi-Modal Learning for Microservice System,” *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pp. 66–78, 2023.
- [P78] S. Ding, Y. E, J. Zhang, L. Li, L. Zhang, and J. Ge, “Trace Anomaly Detection for Microservice Systems via Graph-based Semi-supervised Learning,” *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2024.
- [P79] P. Liu, H. Xu, Q. Ouyang, R. Jiao, Z. Chen, S. Zhang, J. Yang, L. Mo, J. Zeng, W. Xue, and D. Pei, “Unsupervised Detection of Microservice Trace Anomalies through Service-Level Deep Bayesian Networks,” *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*, pp. 48–58, 2020.
- [P80] T. Stojanovic and S. D. Lazarević, “The application of ChatGPT for identification of microservices,” *E-business technologies conference proceedings*, vol. 3, no. 1, pp. 99–105, 2023.
- [P81] S. Eski and F. Buzluca, “An automatic extraction approach: transition to microservices architecture from monolithic application,” *Proceedings of the 19th International Conference on Agile Software Development: Companion*, 2018.

APPENDIX

METHODOLOGY DETAILS

Inclusion Criteria and Exclusion Criteria

Primary Studies (PSs) list

Table VIII show the list of the selected primary studies.

Databases for Selected PSs

Figure 9 shows the databases for selected PSs.

TABLE VII: [R1C2, R2C1, R4C17] Examples of Excluded Studies Based on Specific Criteria

Criterion	Excluded Study	Justification
IC7: The study provides enough information to answer at least 3 RQs	[45]	The study provided only a high-level summary of its approach, without sufficient technical or evaluation detail. Only two research (RQ1 and RQ2) questions could be addressed based on the available content.
IC8: The study has its full text available online	[46]	The publication metadata was available, but access to the full article text was missing, preventing proper assessment.
IC9: The study provides sufficient migration details	[47]	The paper outlines decomposition patterns and incremental migration steps but omits how evaluated and ML techniques used
IC10: The study uses an automated or semi-automated migration approach	[48]	The study proposed architectural guidelines but did not implement or evaluate any tool or algorithm to automate the migration process.
EC5: The study does not provide enough details	[47]	The paper briefly mentioned the use of ML for migration but did not describe the technique, inputs, or outputs, making it unsuitable for analysis.
EC6: The study does not provide an automated or semi-automated migration approach	[48]	The migration approach was based entirely on manual analysis and refactoring decisions by software architects, with no automation.

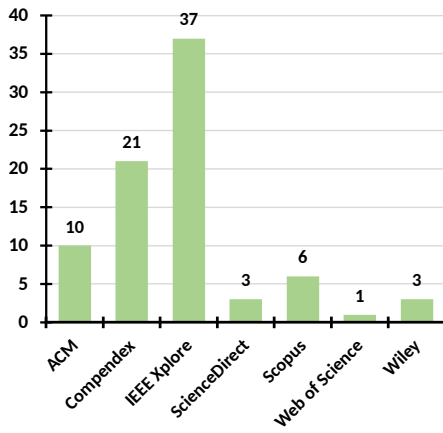


Fig. 9: Databases for Selected PSs

DEFINITIONS OF KEY TERMS FOR RQ1

Migration Phases

- **Pre-migration:** Focuses on evaluating the current approaches, methods, and tools used in the legacy system. This phase also involves planning the migration by identifying suitable strategies, defining objectives, and outlining the steps needed to ensure a smooth transition.
- **Identification:** Defines the boundaries of prospective microservices. This involves detecting functional modules, mapping dependencies, and clustering components with business requirements.
- **Packaging:** Encapsulates the identified components into functional microservices. This phase includes defining service interfaces, managing dependencies, and generating missing components.
- **Deployment:** [R4C9] Implements and configures the necessary runtime infrastructure to host microservices, including containerization, service orchestration, network configuration, and integration with legacy or third-party systems.
- **Monitoring:** Ensures the reliability and efficiency of the microservices-based system post-migration. This phase involves continuous performance tracking, anomaly detection,

and dynamic resource management to meet changing workload demands.

Automated Tasks• **Pre-migration Tasks:**

- Designing microservices: Assists in planning microservice architecture by analyzing legacy systems.
- Predicting success rates of migration: Uses predictive models to evaluate potential migration outcomes.

• **Identification Tasks:**

- Boundary identification: Detects functional boundaries within monolithic systems.
- Clustering: [R4C9] Groups related components or services based on structural, semantic, or performance similarity using unsupervised learning.
- Microservices identification: Automates the process of determining candidate microservices.

• **Deployment Tasks:**

- Automated deployment: Automates the deployment of microservices into production environments.
- Resource management: [R4C6] Allocates and monitors computing resources such as CPU, memory, and I/O bandwidth across microservices during runtime to optimize cost-efficiency and performance.
- Resource allocation: Dynamically allocates resources to microservices based on workload demands.
- Microservice autoscaling: [R4C9] Dynamically adjusts the number of microservice instances or resource allocations based on performance indicators such as CPU load, request latency, or throughput, in order to maintain service-level objectives and system responsiveness.
- Microservice orchestration: Manages dependencies and interactions between microservices.
- Microservice placement: Decides optimal placements for microservices within the infrastructure.
- Resource estimation: Estimates the resource needs of microservices to optimize deployment strategies.

• **Monitoring Tasks:**

- Anomaly detection: Identifies abnormal behaviors or performance issues in microservices.

TABLE VIII: List of Selected PSs

Primary Study (PS)	Phase	Ref.
A DDD Approach Towards Automatic Migration to Microservices	I	[P63]
A Hierarchical DBSCAN Method for Extracting Microservices from Monolithic Applications	I	[P57]
A microservice decomposition method using a distributed representation of source code	I	[P56]
A Microservices Identification Method Based on Spectral Clustering for Industrial Legacy Systems	I	[P5]
A multi-model based microservices identification approach	I	[P6]
A New Decomposition Method for Designing Microservices	I	[P47]
An Automatic Extraction Approach: Transition to Microservices Architecture from Monolithic Application	I	[P81]
Anomaly Detection and Diagnosis for Container-Based Microservices with Performance Monitoring	M	[P38]
Autoencoder-Based Anomaly Detection in Microservices Using Distributed Tracing	M	[P59]
Automatic Microservices Identification Across Structural Dependency	I	[P7]
Automatic Microservices Identification from a Set of Business Processes	I	[P8]
Automatic Migration-Enabled Dynamic Resource Management for Containerized Workload	D	[P15]
Autonomous Selection of the Fault Classification Models for Diagnosing Microservice Applications	M	[P31]
BerthTLG: Graph-Based Microservice Anomaly Detection Through Sentence-BERT Enhancement	M	[P30]
BSDG: Anomaly Detection of Microservice Trace Based on Dual Graph Convolutional Neural Network	M	[P28]
CARGO: AI-Guided Dependency Analysis for Migrating Monolithic Applications to Microservices Architecture	I	[P61]
Cdascaler: A Cost-Effective Dynamic Autoscaling Approach for Containerized Microservices	D	[P19]
ChainsFormer: A Chain Latency-Aware Resource Provisioning Approach for Microservices Cluster	D	[P20]
Code Vectorization and Sequence of Accesses Strategies for Monolith Microservices Identification	I	[P9]
DeepRest: Deep Resource Estimation for Interactive Microservices	D	[P27]
Deepstitch: Deep Learning for Cross-Layer Stitching in Microservices	M	[P52]
DeepTraLog: Trace-Log Combined Microservice Anomaly Detection through Graph-based Deep Learning	M	[P29]
Dependencies-Based Microservices Decomposition Method	I	[P72]
Detecting Security and Privacy Risks in Microservices End-to-End Communication Using Neural Networks	M	[P35]
Dynamic and Static Feature-Aware Microservices Decomposition via Graph Neural Networks	I	[P53]
Enabling Practical Cloud Performance Debugging with Unsupervised Learning	M	[P21]
Enhancing Fault Localization in Microservices Systems Through Span-Level Using Graph Convolutional Networks	M	[P39]
Expert System for Automatic Microservices Identification Using API Similarity Graph	I	[P10]
Extracting Candidates of Microservices from Monolithic Application Code	I	[P4]
Facilitating the Migration to the Microservice Architecture via Model-Driven Reverse Engineering and RL	I	[P70]
From a Monolith to a Microservices Architecture: An Approach Based on Transactional Contexts	I	[P65]
From Legacy2Microservices: A Type-Based Approach for Microservices Identification Using ML and Semantic Analysis	I	[P11]
From Monolithic Architecture Style to Microservice One Based on a Semi-Automatic Approach	I	[P36]
Fuzzy Reinforcement Learning Based Microservice Allocation in Cloud Computing Environments	D	[P16]
GMA: Graph Multi-Agent Microservice Autoscaling Algorithm in Edge-Cloud Environment	M	[P24]
Graph Neural Network to Dilute Outliers for Refactoring Monolith Application	I	[P71]
Graph-Reinforcement-Learning-Based Dependency-Aware Microservice Deployment in Edge Computing	D	[P48]
GTMicro—Microservice Identification Approach Based on Deep NLP Transformer Model for Greenfield Developments	I	[P45]
Heterogeneous Data-Driven Failure Diagnosis for Microservice Industrial Clouds Towards Consumer Digital Ecosystems	M	[P32]
High-Performance Computing-Enabled Probabilistic Framework for Migration from Monolith2Microservices Using GAs	P	[P2]
Implementation of Domain-Oriented Microservices Decomposition Based on Node-Attributed Network	I	[P51]
Improving Industry 4.0 Readiness: Monolith Application Refactoring Using Graph Attention Networks	I	[P49]
Interpretable Failure Localization for Microservice Systems Based on Graph Autoencoder	M	[P33]
Learning Predictive Autoscaling Policies for Cloud-Hosted Microservices Using Trace-Driven Modeling	D	[P42]
Learning-Based Microservice Placement and Migration for Multi-Access Edge Computing	D	[P26]
MAAD: A Distributed Anomaly Detection Architecture for Microservices Systems	M	[P60]
Magnet: Method-Based Approach Using Graph Neural Network for Microservices Identification	I	[P12]
Method of Microservices Division for Complex Business Management System Based on Dual Clustering	I	[P54]
Microegrc: An Edge-Attention-Based Graph Neural Network Approach for Root Cause Localization in Microservice Systems	M	[P69]
MicroMatic: Fully Automated Microservices Identification Approach From Monolithic Systems	I	[P13]
Microservice Anomaly Detection Based on Tracing Data Using Semi-Supervised Learning	M	[P75]
Microservice Deployment in Edge Computing Based on Deep Q Learning	D	[P14]
Microservice Extraction Using Graph Deep Clustering Based on Dual View Fusion	I	[P62]
Microservices Backlog: Genetic Programming Technique for Identification and Evaluation of Microservices From User Stories	I	[P44]
Microservices Performance Forecast Using Dynamic Multiple Predictor Systems	M	[P40]
Migrating Monolith System to Microservices with Directed Graph Attention Neural Network	I	[P64]
Minimize Cost for Containerized Microservices Under SLO via ML-Enhanced Layered Queuing Network Optimization	M	[P17]
MIRAS: Model-Based Reinforcement Learning for Microservice Resource Allocation Over Scientific Workflows	D	[P22]
Mono2Micro: An AI-Based Toolchain for Evolving Monolithic Enterprise Applications to a Microservice Architecture	I	[P67]
Monolith to Microservices: Representing Application Software Through Heterogeneous Graph Neural Network	I	[P58]
Monolith to Microservices: VAE-Based GNN Approach with Duplication Consideration	I	[P55]
Multilayered Fault Detection and Localization With Transformer for Microservice Systems	M	[P76]
Partial Migration for Re-Architecting a Cloud Native Monolithic Application into Microservices and FaaS	I	[P66]
PUTraceAD: Trace Anomaly Detection with Partial Labels Based on GNN and PU Learning	M	[P74]
Sage: Practical and Scalable ML-Driven Performance Debugging in Microservices	M	[P34]
SCORE: A Resource-Efficient Microservice Orchestration Model Based on Spectral Clustering in Edge Computing	D	[P25]
Seer: Leveraging Big Data to Navigate the Complexity of Performance Debugging in Cloud Microservices	M	[P18]
Service Cutter: A Systematic Approach to Service Decomposition	I	[P3]
The Application of ChatGPT for Identification of Microservices	I	[P80]
Topology-Aware Self-Adaptive Resource Provisioning for Microservices	D	[P23]
Towards an Automatic Identification of Microservices from Business Processes	I	[P43]
Towards an Ontology-Driven Approach to Model and Analyze Microservices Architectures	I	[P37]
Towards Migrating Legacy2Microservice-Based Architectures: A Data-Centric Process for Microservice Identification	I	[P50]
Trace Anomaly Detection for Microservice Systems via Graph-Based Semi-Supervised Learning	M	[P78]
Tracegra: A Trace-Based Anomaly Detection for Microservice Using Graph Deep Learning	M	[P41]
TraceModel: An Automatic Anomaly Detection and Root Cause Localization Framework for Microservice Systems	M	[P68]
Twin Graph-Based Anomaly Detection via Attentive Multi-Modal Learning for Microservice Systems	M	[P77]
Unsupervised Detection of Microservice Trace Anomalies through Service-Level Deep Bayesian Networks	M	[P79]
Unsupervised Learning Approach for Web Application Auto-Decomposition into Microservices	I	[P46]
Unsupervised Microservice Log Anomaly Detection Method Based on Graph Neural Network	M	[P73]
Visualization Tool for Designing Microservices with the Monolith-First Approach	P	[P1]

* P: Pre-migration; I: Identification; D: Deployment; M: Monitoring.

- Performance analysis: Evaluates the performance metrics of deployed microservices.
- Detection of failure types: Classifies failure types to facilitate troubleshooting.
- Fault diagnosis: Diagnoses root causes of failures to enable faster resolution.
- Root cause analysis: Pinpoints the underlying issues causing anomalies or failures.
- Privacy risks detection: Identifies potential privacy concerns in microservice interactions.
- Sanity check: [\[R4C9\]Automated validations to confirm system states and outputs remain within expected bounds.](#)

DEFINITIONS OF KEY TERMS FOR RQ2

Input Types

- **Domain Artifacts:** High-level inputs that capture business and functional requirements of the system. These include:
 - **API Documentation:** Details about available APIs and their interactions, often used to modularize microservices.
 - **QoS Constraints:** Quality of Service parameters (e.g., response time, availability) that must be maintained during migration.
 - **Architecture Recommendations:** Expert suggestions for migration strategies and component separation.
 - **Functional Description:** High-level descriptions of system functionalities, often captured in ontologies.
- **Runtime Artifacts:** Inputs derived from the system's runtime behavior, including:
 - **Resource Metrics:** Data on resource consumption (e.g., CPU, memory usage).
 - **Monitoring Metrics:** Data collected from system monitoring tools.
 - **Performance Metrics:** Measurements of system performance (e.g., response time, latency).
 - **Trace Logs:** Detailed records of system execution flow.
 - **Workloads:** Descriptions of system usage scenarios and patterns.
- **Model Artifacts:** Abstract representations of the system, including:
 - **Business Processes:** Workflows and processes supported by the system.
 - **UML Diagrams:** Visual representations of system design.
 - **User Stories:** Descriptions of functionalities from the user's perspective.
- [\[R4C4\]Source Artifacts:](#) Inputs related to the actual software and its configuration, including:
 - **Source Code:** Implementation of the system, including modules and components.
 - **Configuration Files:** Settings required for system operation.
 - **Data Files:** Files used or generated by the system for operations.
- **Technical Artifacts:** Supporting information for system operation and migration planning, including:

- **Server Information:** Details about the infrastructure supporting the system.

Input Granularity

- **Application Level:** Examines inputs at a higher level of abstraction, focusing on files, APIs, and use cases.
 - **Use Case:** Scenarios describing user interactions with the application.
 - **Files:** Application files (e.g., configuration or resource files).
 - **URI:** Resources or endpoints within the application.
 - **API:** Application Programming Interfaces for interacting with functionalities.
- **System Level:** Investigates components and interactions within the larger system architecture.
 - **Services:** Functional units delivering specific capabilities.
 - **Microservices:** Independent units designed for specific tasks.
 - **Node:** Individual units in distributed systems (e.g., servers, containers).
 - **Component:** [\[R4C6\]A logical grouping of code units \(e.g., classes, modules\) within a larger architectural unit.](#)
 - **Performance Metrics:** System-wide measurements (e.g., throughput, latency).
- **Process Level:** Examines activities and operations representing workflows within the system.
 - **Operation:** Specific tasks or actions performed within the system.
 - **Activity:** Groups of operations into higher-level workflows.
- **Code Level:** Focuses on the smallest units of the software system, such as classes and methods.
 - **Class:** Foundational building blocks of object-oriented programming.
 - **Method:** Specific functionalities or operations within a class.
- **Data Level:** Focuses on stored information supporting data modeling and analysis.
 - **Table:** Structured data collections in databases.
 - **Entity:** Logical representations of real-world objects or concepts.

Data Sources

- **Open Source:** Publicly available datasets derived from open projects, repositories, or collaborative platforms.
- **Real World:** Data collected directly from operational systems or real-life environments.
- **Synthetic:** Data generated to simulate real-world scenarios when real-world data is scarce or sensitive.

Preprocessing Tasks

- **Data Extraction:** Extracts relevant information from raw data sources (e.g., logs, traces).
 - **Extraction of Operation and Dependency Information:** Captures workflows and dependencies.

- **Parsing Logs and Traces:** Extracts structured information from logs and traces.
- **Feature Extraction:** Identifies meaningful features for analysis.
- **Data Collection:** Aggregates data from diverse sources.
- **Data Cleaning:** Improves data quality by removing noise and inconsistencies.
 - **Noise Reduction:** Eliminates irrelevant or redundant information.
 - **Normalization:** Scales and transforms data for uniformity.
- **Code Analysis:** Analyzes the structural and semantic aspects of the codebase.
 - **Semantic Analysis:** Examines the meaning and functionality of code components.
 - **Structural Analysis:** Examines relationships and hierarchies within the codebase.
- **Dependency Modeling:** Structures and transforms data to model dependencies between components.
 - **Vectorization and Embedding:** Converts textual or categorical data into numerical formats.
 - **Dependency Graph Generation:** Creates graphical representations of component dependencies.
 - **Modeling and Task Structuring:** Structures data for specific migration tasks.
 - **Matrix Structuring:** Converts dependencies into matrix formats for computational analysis.
 - **Graph Transformation:** Transforms dependency graphs to highlight relevant relationships.

DEFINITIONS OF KEY TERMS FOR RQ3

Machine Learning Models

1) Classical Machine Learning:

- **Support Vector Machines (SVM):** Used to classify components based on their attributes, facilitating precise service decomposition.
- **Random Forest:** An ensemble method effective in identifying service boundaries and dependency classifications by aggregating interaction patterns.
- **Naive Bayes:** Useful for analysing textual data, such as code comments and documentation, to extract insights about service functionalities and dependencies.
- **k-Nearest Neighbors (kNN):** Classifies components based on proximity in the feature space, aiding in optimization by grouping related entities.
- **Logistic Regression:** Estimates the probability of a component being part of a microservice based on extracted features.
- **Decision Trees:** Provide interpretable paths for classifying components based on specified rules, used for functionality and interaction-based classification.
- **Multilayer Perceptron (MLP):** A simple neural network model applied to learn complex patterns for classification tasks.

2) Deep Learning:

- **Autoencoders:** Neural networks that learn to compress input data into a lower-dimensional space (encoder) and then reconstruct the original data (decoder).
- **Variational Autoencoders (VAE):** Extend autoencoders by modelling the latent space probabilistically, allowing for the generation of new data similar to the training set.
- **Convolutional Autoencoders (CAE):** Use convolutional layers to process images, effectively capturing spatial features.
- **Recurrent Networks:** Process sequential data by maintaining memory of previous inputs.
- **Long Short-Term Memory (LSTM):** Addresses the vanishing gradient problem in RNNs, allowing them to capture long-range dependencies in sequences.
- **Gated Recurrent Units (GRU):** Simplify LSTMs by combining forget and input gates, making them efficient for sequential tasks.
- **Transformers:** Use self-attention mechanisms to process sequences, enabling parallelisation and improved training speed.
- **BERT:** Learns word context by considering both left and right contexts, excelling in tasks like sentiment analysis.
- **Sentence-BERT (SBERT):** A variation of BERT optimized for sentence-level tasks.
- **CodeBERT:** A transformer model designed for leveraging similarities in programming languages.
- **ChatGPT:** Based on the GPT architecture, supports developers in understanding and refactoring legacy systems.

3) Graph-Based Models:

- **Graph Convolutional Networks (GCN):** Extend CNNs to graph data by aggregating features from neighbouring nodes.
- **Graph Attention Networks (GAT):** Incorporate attention mechanisms into graph learning, improving representation learning.
- **Graph Isomorphism Networks (GIN):** Capture graph structural properties for effective representation learning.
- **Relational GCN (RGCN):** Extends GCNs to multi-relational graphs, handling various types of relationships among entities.
- **Variational Graph Autoencoders (VGAE):** Combine GCNs with variational inference to learn latent representations of graphs.

4) Reinforcement Learning:

- **Fuzzy Q-Learning (FQL):** Combines traditional Q-learning with fuzzy logic principles to handle uncertainty.
- **Deep Q-Learning (DQL):** Combines traditional Q-learning with deep neural networks for handling complex environments.
- **Deep Deterministic Policy Gradient (DDPG):** A model-free, off-policy reinforcement learning algorithm for continuous action spaces.
- **Multi-Agent Deep Deterministic Policy Gradient (MADDPG):** Extends DDPG for multi-agent environments.

Learning Paradigms

- **Supervised Learning:** Trains models on labeled datasets, primarily for tasks such as anomaly detection and resource optimization.
- **Unsupervised Learning:** Identifies patterns and clusters in datasets without explicit labels, widely applied in microservices identification.
- **Semi-Supervised Learning:** Combines a small amount of labeled data with a larger pool of unlabeled data to enhance learning accuracy.
- **Self-Supervised Learning:** Generates pseudo-labels from raw data, allowing models to learn meaningful representations without requiring explicit labels.
- **Reinforcement Learning:** Trains agents to make sequential decisions by interacting with an environment and optimizing cumulative rewards.

Selected Features

5) Structural Features:

- **Class Dependencies:** Include inheritance, composition, and association relationships used to analyze modularity.
- **Method Calls:** Reveal functional overlaps and interaction patterns between components.
- **Data Dependencies:** Show how data flows within a system and between its components.
- **Transactional Dependencies:** Identify sequences of operations that must occur together.
- **Call Graph Dependencies:** Capture invocation relationships between methods across components.
- **Graph-Based Analysis:** Uses graph structures to analyze dependencies in microservice architectures.
- **Control Dependencies:** Highlight the logical flow of control within a system.

6) Behavioral Features:

- **Invocation Paths:** Represent the sequence of method calls during program execution.
- **Transactional Similarities:** Refer to repeated patterns of operations or interactions between components.
- **Response Times:** Measure the duration taken by a system to respond to a request.
- **Log Events:** Capture system activities for monitoring, debugging, and anomaly detection.
- **Contextual Log Entries:** Enhance log data with additional information.
- **Event Frequency:** Tracks the number of times specific events occur.
- **Timestamps:** Record the exact time of events.
- **Temporal Patterns:** Identify time-based trends in system behavior.
- **Fault Patterns:** Describe recurring errors or failures within the system.

7) Performance Features:

- **CPU Usage:** Tracks the percentage of processor time consumed by the system.
- **Memory Usage:** Measures the amount of RAM utilized.
- **Network Traffic:** Captures the volume of data exchanged between services.

- **Response Time:** Measures how quickly a system or service responds to requests.
- **Availability:** Measures the uptime or accessibility of a system or service.
- **Reliability:** Reflects a system's ability to perform without failure.
- **Compliance:** Ensures that the system adheres to predefined performance standards.

8) Semantic Features:

- **Semantic Embeddings:** Represent components as numerical vectors based on their contextual meaning.
- **Function Names:** Provide a high-level description of a method's purpose.
- **Method Embeddings:** Encode methods as vectors, capturing their functionality and relationships.
- **API Descriptions:** Provide semantic information about system components.
- **Business Logic:** Captures the operational rules and processes that define system functionality.

DEFINITIONS OF KEY TERMS FOR RQ4

Classification and Prediction Metrics

- **Precision:** Measures the accuracy of positive predictions.
- **Recall:** Assesses the model's ability to identify all relevant instances.
- **F1-Score:** Combines precision and recall for a balanced measure.
- **Accuracy:** Represents the overall correctness of predictions.
- **AUC:** Evaluates the quality of binary classification models.
- **MCC:** Measures the quality of binary classifications, considering true and false positives/negatives.

Clustering Metrics

- **Dunn Index:** Measures the separation between clusters.
- **Silhouette Score:** Evaluates the similarity of objects within their cluster.
- **Newman-Girvan Modularity:** Assesses the strength of network division into clusters.
- **Non-Extreme Distribution:** Ensures balanced cluster sizes.
- **Maximum Cluster Size:** Limits the largest cluster size to avoid granularity issues.
- **Number of Singleton Clusters:** Counts clusters with a single element.

System Behavior Metrics

- **Response Time:** The time taken for the system to respond to a request.
- **Resource Utilization:** Measures the effectiveness of resource use.
- **Energy Consumption:** Assesses the total energy required for operations.
- **SLA Violation Rate:** The percentage of time a system fails to meet SLA obligations.

- **Scalability:** The system's capacity to maintain or improve performance under increased workload.

Software Design Metrics

- **Cohesion:** Evaluates the relatedness of functionalities within a service.
- **Coupling:** Measures dependencies between services.
- **Granularity Metric:** Evaluates service size and scope.
- **Structural Modularity:** Measures the degree of system decomposition into independent components.
- **Cognitive Complexity:** Assesses code understandability for developers.

Developer-Centric Metrics

- **Developer Validation:** Assesses developer feedback on migrated services.
- **Closeness to Manual Expert Analysis:** Compares automated classifications to expert evaluations.