CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Requirements for AI-Based Predictive Analytics Systems in an Industrial Manufacturing Context

Selection of the Most Suitable Data Fusion Method

Master's Thesis in Computer science and engineering

Isak Gustafsson, William Johnston

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2026

Master's Thesis 2026

Requirements for AI-Based Predictive Analytics Systems in an Industrial Manufacturing Context

Selection of the Most Suitable Data Fusion Method

Isak Gustafsson, William Johnston

UNIVERSITY OF
GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2026

Requirements for AI-Based Predictive
Analytics Systems in an Industrial
Manufacturing Context
Isak Gustafsson, William Johnston

Supervisor: Chih-Hong Cheng, Department of Computer Science and Engineering
Examiner: Robert Feldt, Department of Computer Science and Engineering

Master's Thesis 2026
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2026

Isak Gustafsson, William Johnston
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

The digitalisation of modern manufacturing has led to an increased volume of heterogeneous data, leading to operational problems such as disconnected systems, data overload, and a lack of data-driven insights. While AI-based predictive analytics systems attempt to solve these problems, they are often limited by the need to process homogeneous data, which clashes with the naturally heterogeneous data found within manufacturing; an issue that data fusion methods can rectify. This thesis, through its method combining requirement engineering and solution selection, outlines the process used for selecting the most appropriate fusion method for fusing heterogeneous data in manufacturing.
The fusion method was selected by first identifying a set of 10 requirements that the manufacturing industry has for such systems. These were identified through thematic analysis based on a combination of a literature review and interviews with industry professionals. 3 different fusion methods: early, intermediate, and late, were then evaluated against 6 of the derived requirements, which were deemed relevant to fusion method selection, and the most suitable method was selected.
Late fusion was selected with an aggregated score of 4.27 compared to the other alternatives' 3.71 and 2.60, largely due to its fulfilment of the industry's two key requirements: processing a large number of heterogeneous data sources, and allowing traceability to motivate the source of its predictions. The difference between late and intermediate fusion was small and is subject to change as the technology and industry evolve. Late fusion was implemented in a test-bed artefact able to process heterogeneous data sources. The test-bed can scale with growing sources, with latency being bounded by the slowest source and not increasing as other, faster sources are added. It also demonstrated the ability to provide both modality level and feature level traceability.

Keywords: Predictive Analytics System, Heterogeneous Data Fusion Methods, Industry Requirements, Manufacturing

Acknowledgements

We want to thank our partner Quay Technologies AB, specifically Andrew Johnston and Mårten Görnerup, for proposing this thesis and allowing us to work with them, for being understanding, providing us with support, data, and arranging the interviews.
We want to thank our supervisor, Chih-Hong Cheng, for supervising our work and keeping us on track throughout the thesis, for offering valuable guidance and support when needed.
We want to thank all the professionals who agreed to, and took valuable time from their days, to be a part of our interviews, sharing their experience and providing valuable insights with us.
Finally, we want to thank our friends and family who supported us during our six-month journey.

We acknowledge the utilisation of generative AI tools, including ChatGPT and Grammarly, for both language refinement and proofreading support. All text was written by us, and the aforementioned tools were simply used to assist with grammar and style to improve the overall readability and flow of the thesis.

Isak Gustafsson, Gothenburg, May 2026
William Johnston, Gothenburg, May 2026

Contents

List of Figures

List of Tables

1

Introduction

The digitalisation of complex systems such as modern manufacturing has led to high volumes of disparate data sources [1]. These systems produce so-called heterogeneous data, referring to data that comes in different formats, types, or sources. These can be difficult to process, analyse, and leverage effectively [2]. Some industrial artificial intelligence (AI) solutions offer systems that manage and utilise this data for improved decision-making [3]. However, these systems often struggle to handle the diversity and scale of the data inputs generated by the industry, leading to challenges such as data overload, disconnected systems, and in the end, a lack of data-driven insights.

This thesis aims to identify the manufacturing industry's requirements for AI-based predictive analytics systems, as well as methods for handling heterogeneous data within such systems. The thesis evaluates how well the methods satisfy the identified requirements and recommends the most suitable approach. Additionally, through the creation of an artefact, the thesis verifies how well the recommended method fulfils the identified industry requirements.

To achieve these goals, the project was structured in three parts. The first part consisted of gathering qualitative data regarding the challenges the manufacturing industry faces when implementing AI solutions, as well as specific requirements related to AI-based predictive analytics systems. This data was collected through a literature review and interviews with industry professionals from within the manufacturing domain. The collected data was then organised through a thematic analysis, which resulted in a report containing a set of industry requirements for predictive analytics systems. The requirements were then used to guide the following two parts. The second part consisted of conducting a literature review on current data fusion methods for handling heterogeneous data. These fusion methods were then evaluated based on how well they fulfil the previously identified industry requirements. In the third and final part, the most promising fusion method was implemented in an AI-based predictive analytics system. This system was then used to evaluate how well the selected fusion method supports the identified industry requirements.

1.1 Problem Description

The diversity in data types and formats found in industry presents significant integration challenges [4], as traditional systems often fail to handle such complexity. For instance, time-series data requires high-frequency sampling and real-time monitoring [5], while free-text data requires natural language processing techniques in order to extract actionable insights [6]. Additionally, some datasets must be cleaned and standardised [5, 7] for them to become useful and interoperable with data of other types and formats. This heterogeneity complicates data integration and reduces the effectiveness of many AI models, which generally perform their best on structured, homogeneous inputs [3].

AI solutions in the form of predictive analytics are especially valuable in industry [8]. However, for these systems to be considered effective, they need to be able to integrate heterogeneous data in a way that complies with the homogeneous needs of many AI models. To achieve this, data fusion methods can be used to merge heterogeneous data that model the same real-world entity into a unified, homogenous representation compatible with the requirements of AI models [9]. Without proper data fusion methods, data must be analysed in isolation, leading to incomplete insights, missed correlations, and reduced prediction accuracy. These shortcomings can increase the risk of critical issues, such as costly machine failures or safety incidents, from going undetected or being inaccurately predicted.

To put these challenges in context, this thesis is in cooperation with Quay Technologies AB, the company behind *Letshare* [10], a cloud-based platform made to support digital transformation in the manufacturing industry. *Letshare* serves as a unified workspace where heterogeneous data, both structured and unstructured, can be collected and organised in real-time through the use of their various components. The platform aims to make production data more accessible and actionable for manufacturing teams; however, while *Letshare* effectively makes production data more accessible, it cannot yet extract meaningful insights from the heterogeneous data sources connected to it, as it currently lacks the necessary AI functionalities.

What Quay Technologies AB, along with other companies with similar goals, need is to implement an AI-based predictive analytics system that can utilise heterogeneous production data to extract actionable insights. Therefore, to ensure that the development of such a system aligns with real industrial needs, it is important to have a clear set of requirements. With such a system in place, manufacturers would be better equipped to utilise the large volume of heterogeneous data that they generate to help address the following challenges:

- **Disconnected Systems.** Manufacturing factories often produce data from multiple, disparate systems. These systems often operate in isolation, making it difficult to get a real-time, accurate view of the entire operation, limiting the factory's ability to respond dynamically to challenges. Analysing each data stream in isolation rarely provides bigger picture operational insights, as such benefits are often achieved when disparate data sources are fused and analysed

together [11].

- **Data Overload**. Much of the generated data remains underutilised due to the lack of skilled personnel [12, 13], time, and tools to analyse it effectively [14]. For example, the Manufacturing Leadership Council surveyed a large number of manufacturing company executives and found that most of the companies that have a data-collection mechanism in place struggle to utilise and understand the data they collect [15].

- **Lack of White Box Decision-Making**. The companies that do have AI solutions capable of analysing operational data and generating insights, are often relying on black box solutions, such as deep learning networks incapable of tracing [16]. While these black box solutions can provide useful predictions, they usually do not provide any insights regarding how they reached their conclusions. This lack of explainability forces manufacturers to blindly trust the outputs or spend a lot of time trying to figure out their reasoning. Explainability has been documented as being one of the requirements industry has for future AI solutions [17, 18].

These challenges are rooted in the growing complexity of modern manufacturing environments and highlight the need for robust AI-based predictive analytics systems. Such systems must be capable of integrating and interpreting heterogeneous data types while remaining adaptable to the growing volume and variety of data sources. By addressing these problems, manufacturers can unlock the full potential of the large amount of data they generate through predictive analysis. This would free up valuable time and resources, reduce the risk of production stoppages and safety incidents, and allow teams to focus on continuous improvement.

1.2 Purpose of the Study

The purpose of this study is to recommend a suitable fusion method based on industry requirements. This was done by combining findings from relevant literature with observations from interviews with industry actors. Additionally, develop and present a software engineering framework, utilising requirements engineering (RE) and multiple-criteria decision-making (MCDM) analysis selection to identify relevant requirements and recommend the most suitable solution. This is intended to ensure that manufacturers get the most out of their operational data and software engineers receive the tools necessary to perform similar selection studies.

To support practical relevance, a subset of the identified requirements was evaluated through the design and analysis of a limited test-bed AI-based predictive analytics system artefact implementing the most suitable data fusion method. The artefact, including the implemented data fusion method, was evaluated based on its ability to meet industrial needs and its adaptability to evolving data types.

The resulting recommendations and process are intended to guide manufacturing companies interested in adopting or developing AI-based predictive analytics sys-

tems capable of combining heterogeneous data from diverse data sources into unified and informative outputs that can help improve operational efficiency and profitability. These findings can also serve as a basis for other parties who wish to conduct future research or incorporate them in works regarding other, similar needs.

1.3 Significance of the Study

This study addresses a gap in the manufacturing domain by providing recommendations for designing AI-based predictive analytics systems capable of integrating heterogeneous data from disparate sources. By identifying and recommending both the most suitable data fusion method and providing a process for identifying and utilising industry requirements, the findings can enable manufacturers to more confidently adopt or develop systems that transform complex data into actionable insights.

Practically, these insights can help improve decision-making in operational environments and mitigate issues such as production stoppages, equipment failures, and product waste. Effectively, improving operational efficiency and profitability, without requiring specialised technical expertise. By reducing the need for data specialists, small and medium-sized enterprises (SME) in manufacturing companies can more easily afford to utilise their operational data for analytic insights.

The created test-bed artefact was trained and evaluated on data closely imitating real-world industrial conditions. This ensures that the recommendations are not purely theoretical and are likely to have practical applicability.

In addition to the direct industrial relevance of the study's findings, they aim to provide the necessary foundation to support future research, whether inside or outside the manufacturing domain. By contributing to a better understanding of the selection process for data fusion methods in domain-specific predictive analytics systems, as well as the identification of requirements for such systems, this study supports engineers aiming to elevate operational efficiency, safety, and long-term competitiveness.

1.4 Research Questions

This study aims to answer the following three research questions to determine the most appropriate data fusion method and requirements for AI-based predictive analytics systems to be valuable in manufacturing contexts:

RQ1: *What are the key requirements for AI-based predictive analytics systems within the manufacturing industry?*

This question investigates the industrial needs of predictive analytical software. What are the key metrics which must be met for a deployable solution? To answer this, we conduct interviews with our partner Quay Technologies AB and other

industry actors, and review existing studies on the subject of industry requirements. These findings give us a set of requirements for industrial predictive analytics systems, which are used to determine the most suitable data fusion method from RQ2 to be implemented, as well as establish the evaluation criteria for the implementation in RQ3.

RQ2: *What data fusion method is the most suitable for handling heterogeneous manufacturing data according to the industry requirements from RQ1?*

What fusion methods are available for handling heterogeneous data? These are collected, and their expected performance for the industry requirements is evaluated. The most promising of these is carried over and implemented in RQ3. This research question gives an overview of possible avenues for handling heterogeneous data from disparate sources, as well as a recommendation for the most suitable fusion method to be implemented in an industrial predictive analytics system.

RQ3: *How can the most suitable fusion method for heterogeneous predictive analytic systems from RQ2 be implemented to support evolving data sources, while satisfying the industrial requirements from RQ1?*

This research question aims to evaluate the practical implementation of the selected data fusion method identified in RQ2 within an AI-based predictive analytics system designed to meet the real-world industrial requirements identified in RQ1. This creates data on the system's ability to maintain predictive accuracy as data sources evolve. The findings contribute to understanding how well the selected fusion method scales and can be adapted in realistic manufacturing settings. Additionally, the findings contribute to future work due to their role in guiding research into requirement areas where current system designs are deemed lacking when handling heterogeneous data and meeting industrial demands.

1.5 Limitations and Delimitations

This section outlines the key limitations and delimitations of this thesis, which together frame the scope and execution of the study.

1.5.1 Generalisability of Findings

The findings and recommendations presented in this thesis are primarily based on literature reviews and qualitative data in the form of interviews with a limited number of industry professionals. Because of this, the findings may not be fully generalisable across all subdomains within the manufacturing domain.

Since no large-scale survey or other quantitative data collection was conducted, there might be industry requirements that were left unidentified or identified requirements that do not apply at broader scales. Additionally, the lack of quantitative data regarding requirements might have influenced the relative importance of each requirement, which in turn could have affected the choice of the most suitable data fusion

method for industry AI-based predictive analytics systems.

As a consequence, the results should be seen as a useful industry template for implementing predictive analytics systems in manufacturing environments similar to those represented by the interviewees, rather than as universally representative findings. Future research with wider industrial coverage could be applied in this area to strengthen the generalisability and applicability of the findings of this study.

1.5.2 Artefact

The artefact developed in this thesis represents a simplified implementation of an AI-based predictive analytics system. Due to the inherent time constraints that come with writing a thesis, it was decided to limit the scope of the artefact. It only contains the core functionalities necessary to allow the evaluation of the selected data fusion method, as well as the industry requirements relevant to the artefact development and deployment.

As a result, the artefact does not include many of the features required for a system to be viable in a production environment, such as user-friendly interfaces, compatibility with existing industry infrastructure, and more advanced traceability. The limited scope led to the system having to be tested in a controlled, simulated environment instead of in a real-world setting, meaning further research may be needed to further validate this study's findings. This also means that there might be certain practical challenges with using the selected fusion method and following the identified industry requirements that went undetected.

1.5.3 Implementing Pre-existing AI Models

The goal of this thesis is neither designing new and groundbreaking AI models to be used in the artefact to achieve as high prediction accuracy as possible, nor implementing as many AI models as possible. Instead, the focus is to evaluate the data fusion method implemented in the artefact in how well it handles heterogeneous data. To do this, utilising pre-existing AI models suitable for each data modality was sufficient. These models are briefly mentioned in Results (Section 5.4).

1.5.4 Datasets

The datasets used for training and evaluation in this thesis are limited both in volume and variety. Given that the thesis focuses on industrial AI-based predictive analytics systems, as well as the handling of heterogeneous data, acquiring sufficient real-world data on undesirable outcomes such as production failures and safety incidents, and the associated telemetry which preceded them was challenging. As such, to address this limitation, synthetic data was generated to complement the limited acquired real-world data.

While this approach allows for more data points, it comes with an increased risk that the research becomes disconnected from the industry domain. To mitigate

this, anonymised production data was provided by Quay Technologies AB through selected consenting clients. Although most of the provided real-world data was incomplete, often lacking in undesirable outcomes, it was used to keep the generated data in line with real-world use cases.

1.5.5 Data Types

The thesis limited its scope to a selected subset of data types commonly found in manufacturing environments: **time-series data**, **text logs**, and **tabular data**. These data types were chosen because they are both representative of data found in real-world manufacturing systems and different enough to support meaningful evaluation of how the most suitable fusion method, identified in RQ2, handles heterogeneous data when addressing RQ3.

Other data types, such as images, audio, or video, were considered but were deemed to be beyond the scope of this study. This was partly due to the difficulty of acquiring and generating data for these types, as well as time constraints, which limited the scope of implementing additional data types for the selected fusion method to handle. While the excluded modalities were not used for evaluating the test-bed artefact, they may serve as grounds for future research.

1.6 Thesis Outline

The remainder of the thesis is organised as follows:

- **Background** provides the descriptions and context necessary to follow the study.
- **Related Work** reviews relevant literature and identifies existing research gaps related to AI-based predictive analytics in manufacturing.
- **Methods** explains the process followed in order to address the research questions.
- **Results** presents the findings from addressing the three research questions.
- **Discussion** discusses the process of addressing the research questions, relevant insights, implications, and future work.
- **Conclusion** contains final remarks and summary over what is covered and achieved in the thesis.

2

Background

This chapter defines and describes relevant terms and concepts that are used throughout the thesis. This includes heterogeneous data, predictive analytics systems, and fusion techniques.

2.1 Heterogeneous Data

Data sources provide the classic three problems of big data defined by Douglas Laney [19]. Data heterogeneity can be defined as the differences in **volume**, **velocity**, and **variety** between sources, making data sources different from one another. These differences will need to be handled in order to evaluate the data sources in a common analytical system [2].

- **Volume** is the amount of data a source provides. This may vary according to frequency and the amount of data each update contains. High frequency data sources such as automatic machine logs may overload analytical systems with too frequent input [20]. Low-frequency data sources, such as daily text logs manually written by humans, may be too infrequent to have temporal relevance to the situation being analysed. These differing volumes of data may require additional processing, such as subsampling and interpolation in order to be analysed together.

- **Velocity** is the timing of new data entering the system. Some data may be available in near real-time, while others may only arrive in spaced-out batches. The data may also have a limited time window in which they are relevant [20, 19]. These data sources require some kind of synchronisation mechanism in order to align and evaluate them together. The mismatched timings require the system to align the different timings while handling bigger batches without disturbances to the system in terms of latency.

- **Variety** is the differences in structure between sources. One of the main differences in structure is the format. Images, text, and numerical data are all structured in differing ways, requiring preprocessing to be analysed together. The other major difference is the rigidity of the structure itself. This is commonly divided into three categories. Structured, semi-structured, and unstruc-

tured data, which are all generated in manufacturing environments and need to be analysed. Structured data is organised in a predefined format with fixed length and contents [21]. This is the most commonly used format in industry due to the ease of managing it [22]. Semi-structured data is organised in a predefined format; however, the format is hard to organise into a predefined scheme of fixed length, such as trees, graphs, or XML documents [21, 22]. Unstructured data lacks a rigid structure or fixed length. Examples of this are bitmap images, text, and emails [22]. These are also commonly generated in manufacturing [21, 23], but are not often analysed as doing so is very labour intensive, due to the lack of predefined structure needed for automatic analysis.

2.2 Predictive Analytics Systems in Manufacturing

Predictive analytics is a term that refers to a set of business intelligence technologies that fall under the category of advanced analytics [24, 25]. Predictive analytics systems generally utilise advanced mathematics, statistics, and techniques from both AI and machine learning, as well as data mining, to make predictions about future behaviours and events.

By utilising big data and predictive analytics systems, companies can become proactive instead of reactive, allowing them to anticipate problems and trends before they occur, as well as helping the optimisation of existing processes [24, 25]. Investing in predictive analytics can thereby give great returns [24].

In the context of manufacturing, although many sectors have achieved high levels of efficiency, manufacturers are still under constant pressure to identify ways to further improve productivity [8]. Predictive analytics is a natural next step in meeting this demand. In manufacturing, there are many ways to utilise operational and sensor data for predictive analytics, such as failure prediction, fault detection, and product demand forecasting [8].

Despite the value that predictive analytics can provide, there are still many SME manufacturing companies that are not yet utilising it [8]. This is because they often lack the necessary infrastructure to support such a system, the knowledge to collect, store, and process the data they generate [8], or simply are not aware that such systems exist.

2.3 Fusion Techniques

Fusion techniques are methods of combining heterogeneous data into a single output. They are usually divided into three main categories: **Early fusion**, **Intermediate fusion**, and **Late fusion**. Each with a collection of different ways to implement them, which may blur the lines between the methods as they borrow characteristics

from multiple categories.

2.3.1 Early Fusion

Early fusion, also known as feature level fusion, involves combining the heterogeneous sources into a homogenous representation before feeding it into a combined model. In a theoretical data analytics system, sensor readings, for example (temperature, vibration and machine text logs), could be concatenated into a single input vector and passed into a singular model. By fusing the data early, the model can instantly start to cross-analyse the low-level relations between the combined modalities [26].

Early fusion allows the inputs to be analysed as a whole, allowing the model to see patterns between inputs early on and act upon them immediately. This can, for example, be used in image analysis where the surrounding pixels usually are highly connected and produce useful patterns when analysed together [27]. A major downside of early fusion is the need to be able to combine all of the data into a singular format compatible with a single model. This may work fine for heterogeneous data coming from different sensors of the same type, but it becomes more difficult as the types start to differ in more ways, requiring a complicated transformation of each source into a common format.

2.3.2 Intermediate Fusion

Intermediate fusion, also known as joint fusion, combines late- and early fusion. The first step is to analyse the sources individually and extract features from them. The second step is to combine the new feature sets from all the separate sources into a homogeneous model. This means that there is an initial analysis of the source, transforming it into features before cross-analysis can be done with the extracted features [27].

This allows the feature analysis portion of individual features to be independent and modular while allowing for cross-analysis of the sources later on when the features are extracted. Additionally, this allows sources which are heterogeneous in structure to be analysed individually in a way that is more relevant to them. For example, the analysis of a spreadsheet and an image vary in their feature analysis, but the relations between these still need to be cross-analysed.

2.3.3 Late Fusion

Late fusion, also known as decision-level fusion, combines the output of separate models into a singular output by the use of a meta-model. This allows different data sources to be entirely analysed by independent models, making the system more modular as these can be replaced without impacting the rest of the system [26].

One advantage of late fusion is its flexibility: if a prediction from an individual model is unavailable, its contribution to the final model can be effectively omitted

by assigning it zero weight [26]. Additionally, another benefit of late fusion is the simplicity possible with the final combining model. It may be as simple as a single mathematical function to combine the different predictions into one [27].

However, a major downside of late fusion is that the separate nature of the models means that analysis between sources is highly limited [26]. Each individual model does not reveal to the meta-model how it came up with its prediction, and the only inter-source analysis available is the comparison of the individual predictions.

3

Related Work

AI has been widely recognised as a key enabler of Industry 4.0, with numerous studies highlighting its transformative potential in manufacturing [28, 29, 30]. This new use of AI requires the collection of vast datasets from industry operations. These datasets are naturally highly heterogeneous [31], due to multiple reasons, such as timing differences and sensor variety. This heterogeneous data requires data fusion before its use in AI, which requires homogeneous input. At the same time, data fusion techniques, ranging from early signal-level integration to late decision-level fusion, have been extensively studied, particularly in the context of heterogeneous data usage for robust prediction [32, 27, 33]. These works provide tools and models to use but do not properly evaluate their applicability in practical, industry settings. Although data heterogeneity has been identified as a major obstacle to AI adoption in industry, a major potential user for predictive analytics systems, works specifically solving practical problems utilising heterogeneous data, overlook industry-specific needs and instead focus on accuracy and latency [34, 35].

In parallel, there is a growing body of work on the engineering of AI systems, especially in safety-critical and data-intensive domains. Surveys such as [36, 26, 37] catalogue a wide variety of fusion algorithms and architectural choices, while domain-specific proposals [38] present innovative techniques for handling heterogeneous data. These solutions can be generally categorised into three main groups: early, intermediate, and late fusion [36, 37, 26, 32]. Works discussing these three groups often focus on the technical benefits of combining multiple data sources, such as improved accuracy or robustness, but tend to overlook the practical implications and reasons for deploying such methods in real-world manufacturing environments. Meanwhile, industrial requirements for AI systems have been explored mainly through surveys and quantitative data. This data has been analysed to produce lists of current challenges facing the manufacturing industry [29, 17, 18]. The empirical software engineering community has also developed methodologies for capturing industrial needs and evaluating the practical constraints of implementing advanced technologies [39, 40]. Such approaches help identify not only what can be done technically, but also what is desirable and feasible from an industrial perspective.

What is missing, however, is a synthesis of these threads: applying empirical software engineering methods to extract actionable industry requirements from real industrial needs in the manufacturing sector, and then using them to assess which data fusion

strategies align best with industry needs. This study addresses this gap by conducting a thematic analysis based on a systematic literature review and interviews with industry professionals to produce a grounded set of industry requirements, which are used to identify and select the most suitable fusion method for AI-based predictive analytics systems in a manufacturing environment. The selected fusion method is then evaluated in a synthetic test-bed artefact with respect to its ability to satisfy identified requirements. Unlike prior studies that emphasise theoretical or technical merits, this work offers a structured and empirically informed approach to fusion strategy selection in manufacturing settings.

4

Method

The study is structured into three distinct phases, each focusing on addressing a separate research question. Starting at RQ1, each subsequent research question builds on the previous, totalling in 7 parts as seen in the methodology overview in Figure 4.1:

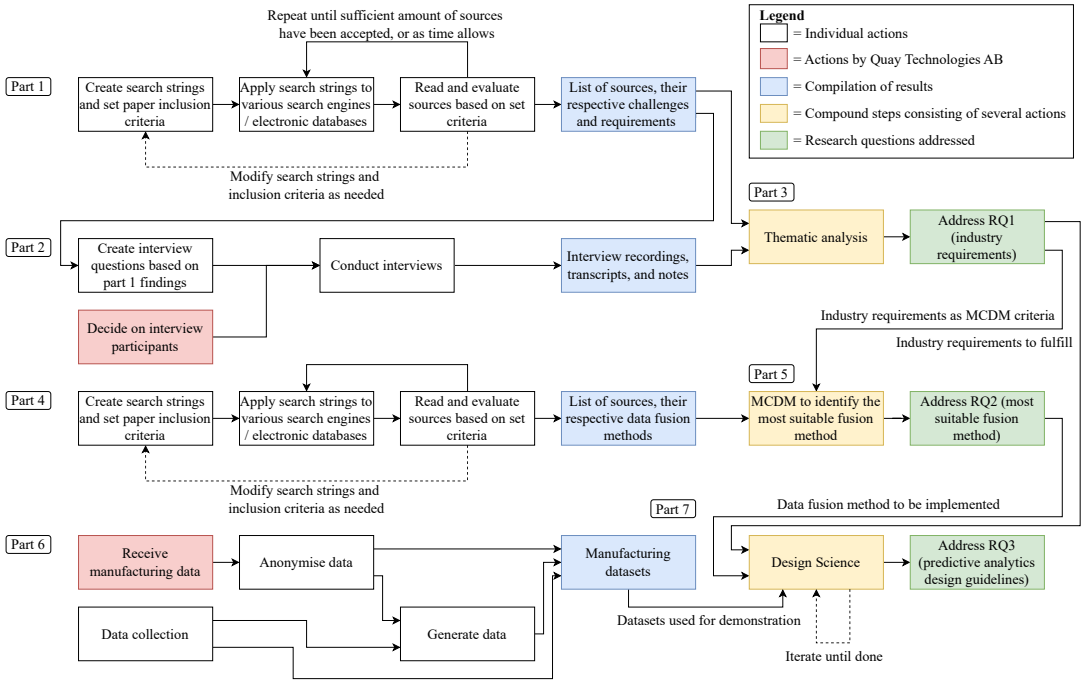

Figure 4.1: Overview of the methodology addressing the research questions. Parts 1-7 show the order in which actions are taken, and will be outlined in further detail in Chapter 4.

Part 1 consists of performing a systematic literature review with the aim of identifying the industrial challenges and requirements of AI-based systems.

Part 2 consists of creating and conducting interviews based on the findings from Part 1, getting a realistic view of industry requirements for predictive analytics systems.

Part 3 consists of combining the findings from Parts 1 and 2, creating an analytical narrative to address RQ1.

Part 4 consists of conducting a systematic literature review, identifying data fusion methods that can handle heterogeneous manufacturing data in predictive analytics systems.

Part 5 consists of using a MCDM approach with the industry requirements from RQ1 as criteria for determining the most suitable data fusion method.

Part 6 consists of collecting and generating manufacturing data, which will be used to assess industry applicability of the artefact developed in Part 7. The process of collecting manufacturing data has been run in parallel with the previous steps.

Part 7 consists of using a Design Science (DS) research approach to create an artefact to address RQ3.

4.1 Thematic Analysis - RQ1

Addressing RQ1 consisted of collecting data through two stages: a systematic literature review and interviews. The collected data was then analysed and converted into a narrative that addresses the research question using the thematic analysis [41] approach.

4.1.1 Research Strategy

In order to collect the data necessary to answer RQ1, this study uses a RE perspective, with the goal of eliciting stakeholder requirements for AI-based predictive analytics systems in manufacturing. The RE process used in this thesis was inspired by the four-step process outlined by Pohl [42]. In particular, the first three steps of elicitation, negotiation, and specification and documentation are followed as part of the research strategy outlined here. The final step, verification and validation, is executed across several parts of the study. Requirement validation is performed throughout the process, primarily through interviews with industry stakeholders to ensure that the requirements reflect the industry's needs, as well as through discussions with stakeholders at Quay Technologies after the thematic analysis was finalised. Requirement verification, on the other hand, is primarily performed in RQ3 through the artefact evaluation, where a test-bed system is experimentally assessed against derived requirements from addressing RQ1.

It is important to note that while the concepts of verification and validation are widely used within software engineering, multiple definitions for each term exist in literature. Hence, this study follows the distinction proposed by Boehm [42] to ensure clarity and consistency. This distinction compares verification to the question "are we building the product right?" and validation to "am I building the right product?".

The research strategy used is a mixed-methods, knowledge-seeking approach inspired by the ABC framework proposed by Stol and Fitzgerald [43]. While the research strategy is not strictly structured around actors, behaviour, and context, it uses the framework's idea of looking at how these elements interact. Specifically, the study explores how industry stakeholders (actors) express expectations and requirements (behaviour) for AI-based predictive analytics systems within the manufacturing domain (context). The data elicited through this strategy is then analysed and synthesised into a set of requirements through a thematic analysis (see Section 4.1.2).

The research consists of two stages, aligned with established methods from both empirical software engineering and RE. These stages align with the ABC framework's emphasis on combining multiple methods and sources of knowledge to mitigate individual methods' weaknesses and increase realism:

1. **Systematic Literature Review:** The first stage involves conducting a systematic literature review, which is a structured and repeatable process used to find and assess existing research related to a well-defined topic or research question [44]. The ABC framework categorises systematic literature reviews as secondary studies, aiming to synthesise or summarise research findings from primary studies [43], such as the one conducted in the following stage. For this study, the systematic literature review will be used to help inform the design of the field study, serving as a precursor to the RE methods used in the second stage, and will contribute to answering RQ1 through a thematic analysis of the results from the two stages.

2. **Field Study (Interview):** The second stage involves conducting a field study, which, according to the ABC framework, is a strategy used to investigate software engineering phenomena in real-world environments, without manipulating the setting or influencing participants [43]. For this study, the field study will be in the form of semi-structured interviews, as it is one of the most effective elicitation techniques [45], allowing for in-depth, context-rich insights into what requirements manufacturers have for AI-based predictive analytics systems.

Step 1 - Systematic Literature Review. The literature review conducted in this study was inspired by the structured process outlined in Keele's guidelines for performing systematic literature reviews in software engineering research [44]. Search strings were created based on RQ1 and searched in Google, Bing, Google Scholar, IEEE Xplore, ScienceDirect, and ACM Digital Library. The following search string was used:

- ('AI' OR 'artificial intelligence') AND ('manufacturing' OR 'production' OR 'industry') AND ('challenges' OR 'difficulties' OR 'problems')

The search string has three criteria. The first is to keep the search inside our research area of AI. The second limits it to the manufacturing domain. The third is used to

Participant	Role	Company
Andrew Johnston	CEO	Quay Technologies AB
Mårten Görnerup	Product Owner	Quay Technologies AB
Mattias Ottermark	CEO	Dahls Bageri
Robert Wiktorén	CTO	Absolent Air Care Group
Jimmy Sterner	Consultant (CEO)	Learn4Future
Erik Persson	Production Planner	Linxens
Emmanuel Mbanugo	Managing Director	Global Opex Limited

Table 4.1: Overview of Interview Participants

find the challenges and, therefore, the requirements that the industry has for these solutions.

The following criteria were used in selecting sources to be included:

1. Be related to the manufacturing industry.
2. Be a compilation of multiple viewpoints (>10).
3. Be written in English.

The search gave 74 sources which passed the criteria. These were then manually inspected to see if they mentioned multiple needs of the industry for AI solutions. After inspection, 7 sources were deemed to be comprehensive and detailed enough to draw conclusions from.

Step 2 - Field Study. The field study was conducted in the form of semi-structured interviews with key stakeholders at Quay Technologies AB, along with other knowledgeable manufacturing professionals. To design the interviews, a semi-structured format was utilised. Semi-structured interviews combine specific questions with open-ended questions to elicit both foreseen and unforeseen information [46], allowing for both consistency across participants and the flexibility to explore unanticipated insights as they come up. The interview questions were based on the findings from the systematic literature review (Step 1), which helped identify recurring themes and gaps related to RQ1. The findings were then translated into open-ended questions (see Appendix A for the full interview guide).

Quay Technologies AB helped identify and recruit potential participants. Seven individuals with relevant expertise in manufacturing were selected for the interviews using purposive sampling [47], based on their professional roles and domain knowledge. The interviewees and their respective roles can be seen in Table 4.1.

Two of the interviewees, the CEO and product owner at Quay Technologies AB, were selected as interviewees due to their central roles in the development of *Letshare*, its need for an AI-based predictive analytics system, and their deep complementary expertise with both software development and manufacturing operations. The product owner has over 25 years of experience working in the metals and manufac-

turing sectors, including leading the development of software solutions specialised for iron and steel production. His background gives him a strong understanding of the integration challenges, system reliability expectations, and operational constraints that predictive analytic systems must address. The CEO has decades of experience working with developing and scaling software as a service solutions aimed at improving companies' efficiency, including manufacturing companies, with a focus on usability, performance, and client adoption. His insights help clarify the reliability expectations from a systems and customer perspective.

Together with the participants from Quay Technologies AB, the broader group of interviewees, including senior leaders and consultants from other companies, provided a well-rounded view of both the technical and organisational requirements that predictive analytics systems must meet to be valuable in manufacturing settings.

The interviews were conducted remotely via Microsoft Teams, one at a time, and lasted approximately 24 minutes each. Before the interviews, the interviewees were given an overview of the topics that were to be discussed, to give them time to prepare meaningful and thought-through answers. With the interviewees' consent, the interviews were recorded and transcribed for further analysis. The interviewees were also asked whether they wanted to remain anonymous, which no one did. Additionally, as a safety precaution for data loss, one researcher led the interviews while the other took notes throughout the sessions.

4.1.2 Thematic Analysis

Once the two stages outlined in Subsection 4.1.1 had been completed, the data collected from the literature review and interviews were analysed through a thematic analysis. From a RE perspective, the thematic analysis process corresponds to both requirement negotiation as well as specification and documentation by transforming elicited stakeholder input together with data from the systematic literature review into a structured set of requirements. The thematic analysis used follows the six-phase process outlined by Braun and Clarke [41]:

1. **Familiarising Yourself with Your Data:** The first step involves becoming familiar with the collected dataset by reading through it several times, reflecting on possible patterns [41].

2. **Generating Initial Codes:** The second step consists of systematically labelling (coding) features across the dataset to organise the data in a meaningful way [41].

3. **Searching for Themes:** Once all the data have been coded, the third step groups related codes into broader themes and sub-themes that capture important aspects of the data [41].

4. **Reviewing Themes:** The fourth step consists of refining the themes by assessing how well they represent the coded data and the dataset as a whole [41].

5. **Defining and Naming Themes:** The fifth step involves specifying the essence of each theme by writing a detailed analysis and labelling them appropriately to capture their core ideas [41].

6. **Producing the Report:** Finally, the sixth step addresses the research question by synthesising the themes and their supporting data into a logical narrative [41].

Step 1 - Familiarising Yourself with Your Data. To become familiar with the collected data, different approaches were taken depending on the data collection tool. For the data collected through the systematic literature review, it was read and re-read several times, while the data collected from the interviews were in the format of recorded videos and transcripts. The interview videos were watched, and their transcripts read. Throughout this entire phase, initial observations and potential patterns found in the data were noted to support later stages of the thematic analysis.

Step 2 - Generating Initial Codes. After becoming familiar with the data, the data, supported by the notes taken from the previous step, was systematically examined to identify and label relevant features. These initial codes were applied across the entire dataset, capturing statements related to stakeholder needs, expectations, constraints and insights relevant to RQ1. The coding was done manually, with the focus on being inclusive and thorough, allowing for as many potential codes as possible.

Step 3 - Searching for Themes. Once the dataset was coded, the search for themes and sub-themes began. Codes were grouped based on their similarity or recurring patterns. To make the process of organising codes and creating themes easier to follow, it was visualised using a mind map. In the mind map, ovals represented themes, while rectangles represented codes. The lines between the entities showed their relationship and connection to each other. At this stage, when coming up with themes, the focus was not solely on the frequency of related codes, but also on their relevance to answering RQ1. The outcome of this step was a preliminary set of themes that would then be reviewed and refined in the following step.

Step 4 - Reviewing Themes. With a preliminary set of themes, these were carefully examined to assess how well the encapsulated codes fit together, as well as to make sure that no themes significantly overlapped with each other. This process consisted of checking whether themes consisted of relatively few codes or if the codes within a theme were too diverse to be meaningfully grouped. Themes that lacked enough supporting codes or were too broad were either refined, split into smaller themes or sub-themes, or removed altogether. Similarly, themes that overlapped or represented the same thing were merged together.

Step 5 - Defining and Naming Themes. Once the set of themes had been reviewed and the mind map looked seemingly complete, each theme was further analysed. The analysis for this step consisted of determining exactly what each

theme represented, writing a detailed description for each theme and how each theme contributed to addressing RQ1 by defining requirements based on their sub-themes and underlying codes. To ensure consistency between how requirements were written, the syntactic structure proposed in *Easy Approach to Requirements Syntax (EARS)* [48] was followed. Additionally, each theme was given a clear, descriptive label that captured its core idea and underlying codes.

This step represents the culmination of the RE's requirement negation step, as while defining and naming the themes, negotiations took place with stakeholders from Quay Technologies to ensure that the formulated requirements accurately reflected their expectations.

Step 6 - Producing the Report. Finally, the last step consisted of connecting the themes into an analytical narrative that directly addressed RQ1. The report integrated the detailed theme descriptions from the previous step, along with relevant code extracts, providing a structured response to the research question. From a RE perspective, this step finalises the specification and documentation step.

As mention in Subsection 4.1.1, once the report containing the identified requirements was completed, it was presented to stakeholders at Quay Technologies, as a requirement validation step, where they could offer feedback to ensure the requirements accurately represent industry needs.

4.2 Selecting the Optimal Data Fusion Method via Multi-Criteria Decision-Making - RQ2

In order to answer RQ2, a second systematic literature review had to be conducted. Once again, the review was inspired by the structured process outlined in Keele's guidelines [44]. After the literature review had been completed, the fusion methods were narrowed down to the most suitable by comparing how well each method held up to or could support the industry requirements found in RQ1.

4.2.1 Systematic Literature Review

To collect relevant literature, structured search strings were designed based on RQ2 and searched in Google Scholar, IEEE Xplore, ScienceDirect, and ACM Digital Library. The following search strings were used:

- ('data fusion' OR 'information fusion) AND ('machine learning')
- ('heterogeneous') AND ('data fusion' OR 'information fusion) AND ('machine learning')
- ('multimodal') AND ('machine learning')

The search string, compared to the ones from RQ1, does not contain any mention of industry or manufacturing. This was done as fusion methods do not belong

to any one field but belong to the field of machine learning. These methods are general enough to be applied to any field. The first string was used to gather papers generally in the area of fusion methods. The second string was used to focus the search on heterogeneous data, giving papers more specifically in the area of this thesis. The third string was created later after "multimodal", a highly relevant word to heterogeneous data, was repeatedly identified in other papers, focusing the search even more on the heterogeneous aspect.

The following criteria were used in selecting sources to be included:

1. Be related to the manufacturing industry.
2. Be a compilation of a high number of sources (>10).
3. Mentions benefits or drawbacks of at least 2 of the fusion methods.
4. Be written in English.

The search resulted in 30 sources which passed the criteria. These were then manually inspected to see if they mentioned benefits or drawbacks of multiple fusion methods. After inspection, 4 sources were deemed to be comprehensive and detailed enough to draw conclusions from.

4.2.2 Ranking the Fusion Methods using MCDM

Once the systematic literature review had concluded, a set of fusion methods was identified. In order to limit the project's scope and align the selected techniques with the industry's needs, the requirements identified from addressing RQ1 were used as inclusion criteria to filter and assess the fusion methods, determining the most suitable method regarding the industry's needs.

To conduct the review, a structured MCDM approach was implemented, following the guidelines outlined by Taherdoost and Madanchian [49]. A MCDM approach was selected for its ability to accurately and systematically compare and prioritise alternatives based on multiple, potentially conflicting requirements.

The MCDM approach used consisted of the following process:

1. **Define a Set of Alternatives:** The alternatives in this case were the fusion methods identified during the most recent literature review.
2. **Define a Set of Criteria:** The initial set of criteria consisted of the requirements collected from addressing RQ1. Through the literature review conducted to help address RQ2, the set of criteria was narrowed down to only contain those that can be used to evaluate fusion methods. The new set of criteria was then used to evaluate the alternatives from the previous step.
3. **Assign Weights to Each Criterion:** This was done by utilising the Saaty

scale [50], performing pairwise comparison between each criterion, considering how important each criterion is to the industry, the frequency of their mentions in the literature review and interviews, and in comparison to each other. Criteria were given a number between 1-9 depending on their level of importance, with reciprocal values for the inverse comparisons. After all weights had been set, they were then normalised to sum to 1 to allow for consistent aggregation in step 5.

4. **Evaluate Each Alternative Based on Each Criterion:** This was done by ranking to what extent the alternatives support the criteria on a scale of 1-5.

5. **Select the Most Suitable Alternative:** This was done by, for each alternative, summing the scores multiplied by their respective weights, selecting the alternative with the highest sum.

As a sanity check, both the weight assignment and evaluation of alternatives were validated through expert review conducted by Quay Technologies AB.

4.3 Design Science Research - RQ3

To address RQ3, the findings from addressing RQ1 and RQ2 were crucial. The most suitable fusion method derived from RQ2 was implemented within a predictive analytics test-bed artefact and evaluated based on its ability to handle heterogeneous manufacturing data while satisfying the industrial requirements derived from RQ1. The test-bed artefact was also used to perform the last step of the RE process mentioned in Subsection 4.1.1, which primarily involves verifying the set of derived requirements.

4.3.1 Design Strategy

To guide this process, a common software engineering research approach, DS was utilised. DS was chosen due to its focus on the structured iterative development and evaluation of artefacts intended to solve domain-specific problems in real-world contexts and to deliver the results to relevant stakeholders [51]. This approach enables the evaluation stage of this study, as no ready-to-use AI-based predictive analytics system utilising the selected fusion method was available. Therefore, a system had to be developed from the ground up to serve as a limited test-bed artefact, enabling the evaluation of the selected fusion method with regard to the derived requirements.

The DS research approach utilised in this study was inspired by the six-step process outlined by Peffers et al. [51]:

1. **Problem Identification and Motivation:** This step was addressed through RQ1, where industrial requirements for predictive analytics systems were identified via the thematic analysis based on the literature review and interviews. Additionally, the need and motivation for an AI-based predictive analytics

system was made clear by Quay Technologies AB when they proposed this thesis.

2. **Define the Objectives for a Solution:** Already addressed through findings identified in RQ1 and RQ3's description. The objectives included: (1) creating a basic AI-based analytic system using the most suitable heterogeneous data fusion method, (2) ensuring system accuracy should be preserved as data sources evolve, and that (3) the system aligns with the manufacturing industry's requirements.

3. **Design and Development:** The design and development of an AI-based predictive analytics system implementing the most appropriate fusion method, capable of handling heterogeneous manufacturing data, identified in RQ2.

4. **Demonstration:** The created predictive analytics system was given synthetic, real-world inspired, manufacturing data to assess industry applicability.

5. **Evaluation:** Artefact performance was evaluated based on the requirements derived from RQ1, as well as to what extent the system could handle evolving data sources.

6. **Communication:** Results and insights were compiled and communicated through the Results (Chapter 5) and Discussion (Chapter 6) chapters in this thesis to both academic and industrial audiences.

While the process is defined as a series of sequential steps, Peffers et al. [51] emphasise that researchers may start at different steps of the process depending on the nature of their research, rather than having to strictly follow a linear progression. As steps 1 and 2 were completed through previously conducted research and definitions, the DS process in this thesis begins at step 3. The resulting process inspired by Peffers et al. [51] can be seen in Figure 4.2.

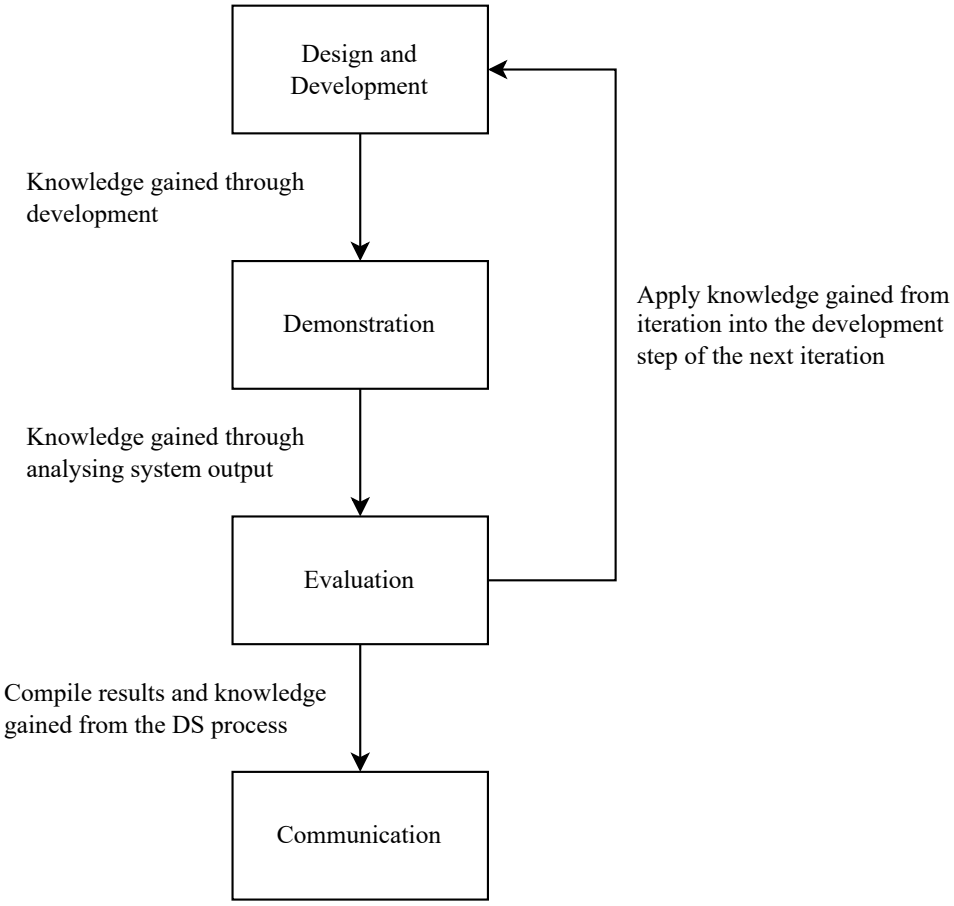

Figure 4.2: The DS research approach used to address RQ3.

4.3.2 Artefact Development

In order to evaluate the theoretical practicality of the fusion method identified in RQ2, a basic artefact consisting of an AI-based predictive analytics system had to be developed. As mentioned in 4.3.1 and visualised in Figure 4.2 the system was developed in an iterative manner.

The following development tools and frameworks were used during the design and development phase:

- **Visual Studio Code:** The IDE used for development.
- **Python 3.10:** The primary programming language used. Version 3.10 was chosen over newer releases to ensure compatibility with all required libraries.
- **Libraries:** Pandas, numpy, sklearn, torch, xgboost, and transformers.
- **Microsoft Visual Studio Live Share:** An extension that allowed for real-time pair programming.

- **Git and Github:** For version control and collaboration.
- **Trello:** Used for project management and tracking progress of each iteration.
- **MoSCoW:** Framework for prioritising the features that should be implemented during the artefact development.

It is worth noting that the importance of developing the artefact does not lie in creating a system that performs well in classic machine learning metrics, but rather it comes from the ability to serve as a tool for evaluating to what extent the selected fusion method can fulfil objectives 2 and 3 outlined in the DS process as well as verifying derived requirements from RQ1.

The system was designed to handle heterogeneous input data from three common manufacturing data modalities: time-series data, text-log data, and tabular data. Although not covering all modalities found within the manufacturing industry, this multi-modal approach reflects data found within a typical industry environment.

No new machine learning models were created during the development of the artefact. Instead, widely used and independently validated models were integrated into the system to keep the focus of the study on the data fusion method implemented and its ability to handle heterogeneous data. The performance of the system was therefore only done in relation to itself as to eliminate the effect of model selection.

4.3.3 Data Collection

Throughout the study, heterogeneous manufacturing data were collected to support the development and evaluation of the artefact. The data collection process combined data from three different sources: real-world manufacturing companies, public manufacturing datasets, and synthetically generated data.

The data from real-world manufacturing companies was provided through Quay Technologies AB, who contacted their existing clients and contacts to request access to relevant datasets. To encourage participation, companies were assured that all their data would be anonymised to protect sensitive or identifiable data. Despite assuring that their data would be anonymised, many companies were still hesitant or unwilling to share their data. Out of the companies that were willing to share their data, most lacked either necessary information, such as records of production stoppages and safety incidents, or did not have enough data for meaningful analysis, limiting the usefulness of their input.

To complement the company-sourced data, additional datasets were collected from publicly available repositories. The process of finding suitable datasets online was similar to that of a systematic literature review. The inclusion criteria consisted of the datasets having to be free-to-use, containing real-world manufacturing data, and having relevant information for predictive analytics.

In cases where the real-world datasets were insufficient, either in quantity or in

coverage of specific scenarios, heterogeneous industrial data was generated using tools such as *ChatGPT* [52] or Python scripts. In order to improve the realism of the synthetic data, the datasets were created using prompts or scripts based on the anonymised company data and publicly available datasets. This helped fill data gaps by complementing datasets that were lacking in necessary information and expanding coverage where real-world records were lacking.

4.3.4 Artefact Evaluation

At the end of the thesis, the created test-bed system was evaluated in order to determine whether the selected fusion method from RQ2 fulfils the requirements from RQ1. The evaluation was done in 3 steps.

1. **Dataset Generation:** The data required was generated by a Python script running a simulation of an industrial process. This simulation produced multiple modalities of varying structure and content. Structure and contents were inspired by datasets collected from Quay Technologies AB and public resources.

2. **System Performance Evaluation:** The dataset was split into two subsets, training and testing, with 80% and 20% of the data, respectively. The system was then trained and evaluated using the two subsets. This was repeated with different combinations of the modalities to capture the differences in the system having access to more or less data sources. Metrics were collected based on the requirements to be evaluated.

3. **Metrics Analysis:** The final step was to analyse the collected metrics and decide whether they hinted at the system being sufficient for manufacturing industry use. Each metric was presented, and its connection to the requirements was explained.

This provides the experimental portion of the thesis, where the theoretical applicability is demonstrated in a limited experiment. This experiment does not conclusively answer anything but rather gives indications of how well the theoretical answer provided by RQ1 and RQ2 can be applied to a real-world inspired test environment.

5

Results

The Results chapter is divided into four sections, split across the three research questions:

- The first two sections, Industry Requirements (Section 5.1) and Thematic Analysis (Section 5.2), create the industry requirements, thus answering RQ1.
- The third section, Fusion Methods (Section 5.3), describes the available fusion methods as well as how well they fulfil the industry requirements from RQ1, thus answering RQ2 by selecting the most suitable method.
- The fourth section, Fusion Method Validation (Section 5.4), demonstrates how well the most promising fusion method from RQ2 fulfils the requirements from RQ1, thus answering RQ3.

5.1 Industry Requirements - RQ1

This section is structured into two parts: findings from conducting the literature review and findings from conducting the interviews. The combined findings of these two parts were then analysed in a thematic analysis to answer RQ1 in Section 5.2.

5.1.1 Literature Review Findings

The first step to acquiring the industry requirements was a systematic literature review to gather the current understanding of industry requirements in academia and industry. During the review process, a total of 7 sources were selected to represent the industrial requirements. From these 7 sources, 5 categories of challenges were identified. These categories and the papers which discussed them were collected in Table 5.1.

Challenge	Description	Sources
Simplicity	The complexity of a solution regarding the needed specialists to implement.	[39, 17, 29, 53, 54, 18]
Price	The financial cost of developing and deploying a solution.	[39, 29, 53, 54, 18]
Computation	The computational need of a solution in computation is needed, as well as latency.	[17, 53, 54, 18]
Traceability	The ability to trace why, as well as from where a solution got its result.	[17, 55, 29, 18]
Integration	The difficulty of connecting the solution to both new and old data.	[55, 53, 18]

Table 5.1: Categories of challenges in implementing AI solutions in manufacturing.

Simplicity. A major challenge, mentioned by 6 out of 7 sources, is the complexity of potential solutions [39, 17, 29, 53, 54, 18]. This call for simplicity does not come from a lack of ability to implement AI-powered solutions, especially in developed economies [39]. Instead, multiple studies point to a lack of specialised staff needed to develop and operate current solutions [17, 29, 54]. A desired solution to this complexity is an increase in standardised solutions which require fewer developers at individual companies [17, 53, 18]. Such a standardised solution would need to be easily applicable to new, but similar, problems.

Price. A major challenge, mentioned by 5 out of 7 sources, is the financial cost of implementing AI solutions [39, 29, 53, 54, 18]. Financial cost is a particular concern for SMEs where the implementation of a tailor-made solution represents a large financial risk [39, 29]. The financial cost is not only in developing the solution but also in deploying it, as the cost of needed equipment upgrades in data collection may be seen as a limiting factor [54]. Another is the ability to financially motivate an AI solution investment, as the benefits often are difficult to estimate [53].

Computation. A challenge for a lot of newer, more complex modern AI solutions, mentioned in 4 out of 7 sources, is computational needs [17, 53, 54, 18]. This comes in multiple forms. One is the need for general computation power, not only to be able to complete the training and prediction, but to do so in an energy-efficient manner, while being able to scale up and decentralise the computation [17, 53, 54]. A focus on computation is also needed to keep the latencies for the end user within acceptable limits [18].

Traceability. A challenge, mentioned by 4 out of 7 sources, is the traceability of available solutions [17, 55, 29, 18]. The traceability requirement comes from the need to validate decisions and evaluate the reasons for a specific outcome. An industry focus on "black boxes", capable but untraceable systems, has caused a distrust of AI-powered solutions due to them not revealing the reason for their outcome to the end user [17, 18]. This lack of traceability has led to challenges in motivating AI implementation due to a lack of trust in the system output [53, 54].

Integration. A challenge with any new IT solution, but especially with data-oriented AI ones, mentioned by 3 of 7 sources, is the integration of older as well as new infrastructure [55, 53, 18]. A possible AI solution should be able to handle the heterogeneous, often unclean, data found in manufacturing environments [17, 53]. This integration should also be able to handle many different data sources, as there are many older sources which need to be connected [55, 53, 18].

5.1.2 Interview Findings

To further identify industry requirements, a series of semi-structured interviews were conducted. A total of 7 interviews were conducted with experts from the manufacturing domain. These interviews contextualised the challenge categories identified in the literature review, as well as explored areas not mentioned in the review, such as reliability and other domain-specific requirements not previously discussed. The interviewees represented a diverse range of perspectives: some spoke on behalf of multiple companies or industries, while others provided deep company-specific insights. Together, the insights gained from the interviews provided a well-rounded and realistic understanding of what industry professionals expect from AI-based predictive analytics systems.

Reliability: When asked how they would judge if a predictive analytics system is good enough to trust, all seven interviewees provided insights. Five interviewees emphasised that trust is built over time. Out of these five, one interviewee stated that the system must prove itself capable of understanding its given goal, such as predicting production stoppages or identifying unacceptable quality levels and be able to interpret what an acceptable outcome is over an extended period of time. One interviewee expressed that the system needs to be able to uncover difficult-to-find relationships between workplace factors to be deemed trustworthy. One interviewee noted that their trust in a system depends less on the system itself and more on the quality of data they will feed it. Two interviewees stated that the system has to demonstrably reduce breakdowns and incidents to earn trust.

> "If the system does not actually understand the goal first, then I will be quite concerned that it is going to give me the right results." - Andrew Johnston

> "You would judge it [the predictive analytics system] by its accuracy over time, and see if it really helps us or really predicts the future." - Mattias Ottermark

All seven interviewees provided insights regarding what mistakes would be considered unacceptable for a predictive analytics system. Four out of the interviewees agreed that the most unacceptable mistakes for predictive analytics systems are those involving wrongly classifying or failing to detect serious incidents, such as those that could harm people or significantly disrupt costly operations. Four interviewees noted that not being able to maintain a consistent level of accuracy without an excessive amount of false alarms or consistently missing obvious issues is un-

acceptable. One interviewee highlighted that occasional wrong predictions would not be a major issue, as manual judgement would always be used to validate the system's predictions. Another interviewee emphasised that simple questions should give simple answers.

> "If it [the predictive analytics system] is overseeing or overlooking extremely serious future accidents or breakdowns that are very costly or even risk people's lives, it would be extremely unacceptable." - Mårten Görnerup

Six out of the interviewees provided insights regarding the performance or accuracy metrics required before a predictive analytics system would be considered deployable. Three interviewees emphasised that the system must meet operational standards comparable to those expected of a human, including aligning with company-specific definitions of uptime, quality, and acceptable risk. Two interviewees highlighted that a system must prove its value by delivering measurable business outcomes, particularly by reducing waste, downtime, or delivering a clear return on investment (ROI), ideally within one year of deployment. Two interviewees noted that system transparency is essential, and that probabilistic outputs should be provided (for instance, "I am 89% certain X will happen"), which helps users understand the system's limitations and use it as a decision-support tool rather than an unquestioned authority. Another interviewee emphasised that deployment readiness depends heavily on whether the foundational data and processes are in place, particularly for SMEs. Finally, one interviewee pointed out that a useful system must predict not only successes but also failures, and be able to manage data at a depth and scale beyond what humans can handle.

> "If we decide to invest in something, we want to see some kind of ROI, and I think that for most companies the standard is 12 months, and we like to see results pretty early." - Jimmy Sterner

Integration: Out of the seven interviewees, four stated that their company or clients are currently not fully ready to integrate AI and analytics software. One interviewee mentioned that their company is technologically ready to integrate AI, but that most of their clients are not. Another interviewee mentioned that one of their clients had implemented a basic AI system into their operations; however, it was not smart enough to predict outcomes. One of the interviewees did not provide any insights regarding this. The main challenges preventing the integration of AI-based predictive analytical software include insufficient or poorly structured digital data, outdated or disconnected business systems, or a lack of understanding of AI's practical applications.

> "I think there is a big challenge because we do not really know what it [the predictive analytics system] is yet, we do not know if it can give us the information, and we do not know how to use it. That is why we are not even trying to use it." - Erik Persson

Simplicity: All interviewees provided insights regarding the availability of technical expertise to implement predictive analytics systems. Four of the interviewees explicitly stated that their companies or clients do not have the necessary expertise to implement such systems. Additionally, four interviewees mentioned that SMEs often lack the technical personnel needed to implement analytical systems, while larger companies often have in-house developers or staff with the technical capability to work together with external consultants for implementation.

> "Because AI is developing and moving so fast, there is a shortage in terms of capabilities and people who could do the work. And the people that they [SMEs] have at the moment might not have the standard and knowledge to do certain tasks, which causes a lot of delay and frustration, and is one of the reasons why some organisations do not move on with AI." - Emmanuel Mbanugo

Five interviewees provided insights into who in their companies would use a predictive analytics system if implemented. The following roles were mentioned, along with the number of interviewees who mentioned each: production manager (1), quality manager (1), specialised technicians/engineers (4), process development department (1), maintenance teams (1), planners (2), and sales teams (1). Together, these responses suggest that predictive analytics systems should support decision-making across a wide range of company departments and roles within manufacturing.

> "The main users are probably found in the process development department, those are the guys who work with, for example, new investments for machines and planning production. I think that department would be central in this." - Robert Wiktorén

All interviewees provided perspectives on whether companies should go for simpler, low-cost predictive analytics systems or more advanced, company-specific solutions. Three interviewees preferred a more advanced customised solution, given that it provides a greater ROI. One of these acknowledged that SMEs might struggle to afford high-cost systems upfront and may prefer starting with a simpler, low-cost solution. Two interviewees mentioned that the decision often depends on the CEO's mindset: a leader with an entrepreneurial background is more inclined to invest in the best solution possible in the hope that it will generate a greater ROI, whereas a CEO with a financial background might opt for the cheaper, more conservative option. One interviewee stated that they would prefer to start with a simpler solution and scale up as they master it properly. Finally, one interviewee stated that due to the smaller size of their company, they would like to begin with a simpler, more affordable system, as they believe any improvement is valuable, with the option to scale up later if the need arises.

> "I do not think I would expect a low-cost solution to give benefits, I would expect that if we really changed something here, it would be at least a level above a low-cost solution. Some customisation would be needed." - Robert Wiktorén

Price: Six of the seven interviewees provided insights on whether the price of implementing a predictive analytics system would be justified by its benefits. Overall, the interviewees agreed that the price could be justified, provided that the solution provides clear long-term financial benefits. One interviewee expressed uncertainty regarding the full range of potential benefits such a system could bring. Another emphasised that customisation and the system's ability to integrate well into their current setup are more important than the cost of the system itself.

> "The biggest problem is that investment in predictive analysis tools is expensive. The good news is that it pays off, not in the short term, but in the long term." - Emmanuel Mbanugo

When asked what specific benefits would justify such an investment, all interviewees provided insights. The following range of outcomes was mentioned, along with the number of interviewees who mentioned each: improved quality (6), increased productivity (5), fewer production errors (5), reduced stoppages (4), reduced waste (2), increased profits (2), increased competitive advantage (1), increased safety (1), and improved planning (1). One interviewee emphasised that even small efficiency gains within industries with tight margins, such as the food industry, can potentially multiply profits by two to ten times.

> "At the bottom of that buckle, it is all about money. If we earn more money, everything becomes possible. We can then break it down to less stops, less quality issues, less bad planning, and almost everything improves. But in the end, it all comes down to money." - Jimmy Sterner

Computation: All seven interviewees provided insights on the importance of latency for predictive analytics systems. Overall, five interviewees viewed latency as context-dependent: while important, it is not the top priority, especially not initially. Two of the interviewees mentioned that simply getting predictions is more important than the speed it takes to generate them, though expectations may shift towards lower latency over time. One interviewee noted that while immediate insights are not essential, it should not take days to generate predictions. Another two interviewees mentioned that the latency requirements are highly dependent on the type of business and industry. Out of the two interviewees who did not view latency as context-dependent, one strongly emphasised that fast insights are crucial for decision-making and removing time-consuming non-value-adding tasks, while the other expressed that latency should not be a major concern for predictive analytics systems.

> "I think that to begin with, people will be satisfied with just having that solution or technology [the predictive analytics system]. Latency will not be the first thing, as initially they will be quite tolerant, as they do not have anything that can do this [generate predictive insights]. Although latency will become important later on, as when people get used to something, then they start to get impatient." - Andrew Johnston

Six of the seven interviewees shared perspectives on how often a predictive analytics system should deliver insights. The responses varied based on use case and practical application, with most interviewees specifying different frequencies. One interviewee mentioned that the system should generate insights daily for production and weekly for quality and waste, while another said that they probably do not need daily insights and that weekly would be enough. One interviewee wanted the system to generate insights as frequently as possible without generating false predictions, while two others would be satisfied as long as it is faster than a human. The final interviewee highlighted that it is highly contextual and that it completely depends on their clients and what they want to achieve.

> "With production, I would say daily, and with quality and waste, perhaps weekly. - Andrew Johnston

Traceability: All participants provided insights regarding whether they trust AI solutions. Everyone agreed that they trust AI in a general sense, but the extent to which they trust, depends on the context and application of the solution. Three out of seven interviewees highlighted that deeper trust is dependent on the user's expertise within the field in which the AI solution is to be deployed, as they are more likely able to judge and verify outputs, and spot inaccuracies. Three out of seven interviewees expressed that they would never be able to fully trust AI, but see it as a great additional perspective. Two interviewees agreed that an AI solution with 99% accuracy would be good enough to support the majority of operational decisions. Three interviewees mentioned there being a learning curve in understanding AI, and that their trust develops as the respective AI solutions prove accurate over an extended period, emphasising that trust is built over time. One interviewee expressed a strong trust in AI and has been impressed with what they have seen so far.

> "People know their production equipment, normally, and what they can get from a predictive analytics system is quicker access to things that they probably can figure out, but would take them more time. So I do not think you can fool a factory team, they will know if this tool is saying strange and stupid things. Yeah, I think it would be trusted." - Robert Wiktorén

All participants provided insights regarding whether understanding how a predictive analytics system came to its prediction is necessary. Six out of seven interviewees expressed that understanding how a system arrived at its prediction is crucial for building trust and making important decisions. The interviewee who did not find it crucial noted that simply seeing the prediction is enough for him; however, he acknowledged that for the people working on fixing problems, having the system's reasoning would be beneficial. One interviewee highlighted that seeing how the system came to its prediction is especially important in the early stages after its implementation, and that over time, when the AI provides evidence and reasoning that aligns with the user's domain knowledge, being able to understand the predictions becomes less important. Two interviewees agreed that demonstration is key in

building system trust; however, they noted that the required degree of step-by-step details is highly dependent on context.

> "Yes of course, I think a few hints, and especially if it could say 'parameter A, B, C, and D are looking like this and this is a very abnormal pattern, and last time I saw this pattern...', you know trying to at least elaborate on the result would help a lot rather than just having a red light flashing." - Mårten Görnerup

Other: After all questions had been covered, the interviewees were asked if they had anything else to add in terms of non-negotiable requirements for predictive analytics systems to be considered deployable in a manufacturing setting, and whether these requirements would evolve over time. Four out of seven interviewees commented on non-negotiable requirements. Out of these interviewees, one stated that the system should be seen as an AI support, meaning that it should only be able to give support and not perform any actions. The same interviewee specified that the system has to be able to show correlations, as that would help build trust. Two interviewees specified that the system has to be able to adapt to or be customised for individual factories or companies. Out of the two, one specified that the system should not be too strict, and the other noted that the system has to be able to learn and improve over time. Another interviewee highlighted affordability, especially for SMEs, transparency, and being user-friendly as non-negotiable requirements. When asked whether they could see the requirements change over time, three of the interviewees provided their insights. They all agreed that the requirements will evolve, as AI becomes better, or as their company grows and requires better data support for improved data-driven decisions.

> "It [the predictive analytics system] should not be too strict, so that you can adapt and develop it over time. It should also be adaptable to your business or facility." - Mattias Ottermark

5.2 Thematic Analysis - RQ1

This section presents the findings from the thematic analysis conducted on the results from the systematic literature review and interviews. Based on the generated codes, four themes were identified, each with its respective subthemes. Each theme encapsulates a different area of requirements that manufacturers have for AI-based predictive analytics systems, which together answer RQ1: 'What are the key requirements for AI-based predictive analytics systems within the manufacturing industry?'.

5.2.1 Identified Themes

This subsection covers and describes the four themes identified during the thematic analysis, along with their subthemes and requirements derived from the identified codes. See Figure 5.1 for the themes and their respective subthemes.

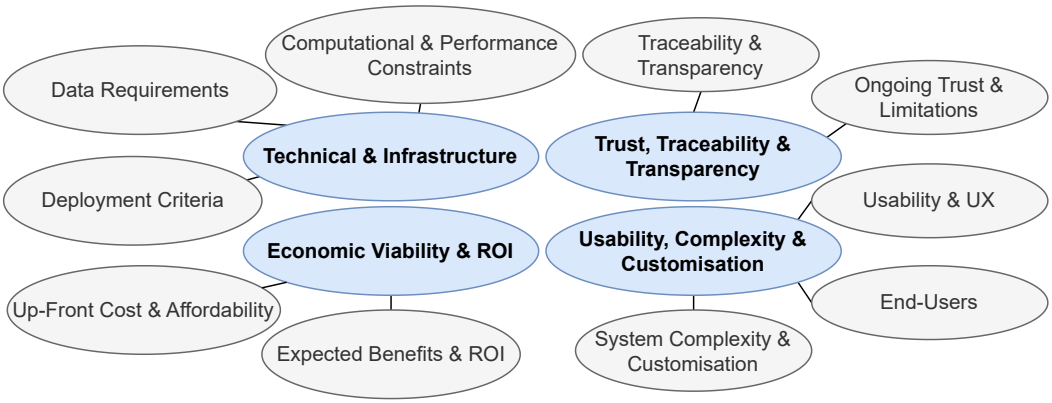

Figure 5.1: The four identified themes along with their related subthemes.

Theme 1: Technical and Infrastructure Requirements. This theme captures the general system-level expectations that manufacturers have for AI-based predictive analytics systems. To be usable in complex industrial environments, these systems must seamlessly integrate with existing infrastructure, be capable of handling heterogeneous data from multiple modalities, and operate efficiently within performance and resource constraints.

The following subthemes were identified:

- **Deployment Criteria** focuses on the specific criteria that must be fulfilled in order to call the deployment of a predictive analytics system a success. *Requirement R1*: When the company is internally ready to implement (for example, foundational data and processes are in place), the system shall act as a decision-support tool, not a replacement.

- **Data Requirements** focuses on systems being capable of handling data from multiple heterogeneous sources. *Requirement R2*: The system shall be able to process multi-source heterogeneous data, including data from real-time and legacy systems.

- **Computational and Performance Constraints** focuses on how systems must perform in terms of energy efficiency, latency, and prediction output frequency to be viable in manufacturing settings. *Requirement R3*: The system shall manage latency expectations, scale with operations, and remain energy-efficient.

Theme 2: Trust, Traceability, and Transparency Requirements. This theme covers how traceability, transparency, and consistent system performance contribute to both building and maintaining user trust in predictive analytics systems. Manufacturers expect the system predictions to be both explainable and verifiable. Additionally, user trust is something that is earned over time through the system demonstrating an ability to understand what is considered an acceptable outcome and being accurate, particularly in safety-critical and high-stakes environments.

The following subthemes were identified:

- **Traceability and Transparency** focuses on the system's need to be transparent. Transparency, along with traceability, is crucial for building user trust in the system, as well as informing important decisions. *Requirement R4*: The system shall clearly show how its predictions are generated, including traceable decision pathways and displaying the likelihood of predictions happening.

- **Ongoing Trust and Limitations** highlights that while a lot of manufacturers trust AI, there are still some who are sceptical towards trusting AI solutions. It also highlights that the degree of user trust is context-specific, and what predictive analytics systems need is to be capable of building user trust over time. *Requirement R5*: The system shall produce predictions whose accuracy aligns with human expert assessments over time.

Theme 3: Usability, Complexity, and Customisation Requirements. This theme explores what manufacturers require from a predictive analytics system in terms of usability, the appropriate level of system complexity, and its ability to be customised and scaled.

The following subthemes were identified:

- **Usability and UX** focuses on what is required for a system to be usable, both in terms of understanding tasks and user experience. *Requirement R6*: The system shall be user-friendly, capable of understanding its given goals, and provide clear, actionable outputs with minimal complexity.

- **System Complexity and Customisation** highlights the various needs of companies. Depending on size and budget, some companies prefer cheaper and more generic predictive analytics systems that can be scaled up over time as requirements change, while others want more advanced company-specific solutions. *Requirement R7*: The system shall offer scalable complexity, allowing customisation and advanced features for larger or more financially stable companies.

- **End-Users** highlights the variety of users that might benefit from using a predictive analytics system. *Requirement R8*: The system shall be suitable for end-users across various departments, including engineering, maintenance, production, and planning.

Theme 4: Economic Viability and ROI Requirements. This theme reflects the financial considerations manufacturers have when deciding whether to invest in an AI-based predictive analytics system. It highlights concerns around up-front affordability and the clear expectation of measurable ROI.

The following subthemes were identified:

- **Up-Front Cost and Affordability** focuses on the cost concerns that SMEs

have, especially for those that do not already have the necessary equipment to collect data. *Requirement R9*: The system shall be affordable for SMEs, either through offering feature add-ons or simplified entry-level versions.

- **Expected Benefits and ROI** highlights the productivity and economic benefits users might expect after implementing a predictive analytics system. *Requirement R10*: The system shall deliver clear operational and financial benefits, such as improved productivity, reduced waste, fewer production errors and stoppages, and better product quality, ideally within a year.

5.2.2 Thematic Report

This subsection combines insights from the four identified themes into a set of interrelated requirements that directly answer RQ1. The findings show that successful adoption of an AI-based predictive analytics system depends on foundational technical readiness, trust-building transparency features, a user-centred design, and clear economic viability.

First, manufacturers require a technical infrastructure capable of collecting and storing relevant data. The predictive analytics system needs to be able to process heterogeneous inputs from various sources while meeting context-specific performance demands. Without these foundational elements, no system can deliver reliable or actionable insights.

Once the foundational elements are in place, the system needs to integrate features that promote transparency and traceability in order to build user trust. Through predictions with consistent accuracy over time and explainable outputs, manufacturers can be assured, especially in safety-critical and high-stakes settings.

While trust enables manufacturers to have confidence in the system, it needs to be usable and adaptable to promote user engagement. With intuitive interfaces, scalable complexity, and support for end-users across different departments, the system is ensured to be both useful and valuable.

Finally, all of the aforementioned requirements need to align with economic viability. This can be done with affordable entry points and by demonstrating ROI within, ideally, a year. The system being affordable is essential for organisations, especially SMEs, to invest.

In summary, the thematic analysis demonstrates that the manufacturing industry's requirements for AI-based predictive analytics systems go beyond technical capability alone. Instead, they have several areas of interdependent requirements: infrastructure readiness enables the system to function properly and make predictions based on heterogeneous company data, system transparency builds trust for the end-users, usability and customisation allow for meaningful adoption across multiple companies and departments, and economic viability ensures long-term ROI. The full list of requirements that the manufacturing industry has for predictive analytics

systems, along with which theme and subtheme they originate from, can be seen in Table 5.2.

Theme	Subtheme	Requirements
Technical and Infrastructure	Deployment Criteria	[**R1**] When the company is internally ready to implement, the system shall act as a decision-support tool.
	Data Requirements	[**R2**] The system shall be able to process multi-source heterogeneous data.
	Computational and Performance Constraints	[**R3**] The system shall manage latency expectations, scale with operations, and remain energy-efficient.
Trust, Traceability, and Transparency	Traceability and Transparency	[**R4**] The system shall clearly show how its predictions are generated.
	Ongoing Trust and Limitations	[**R5**] The system shall produce predictions whose accuracy aligns with human expert assessments over time.
Usability, Complexity, and Customisation	Usability and UX	[**R6**] The system shall be user-friendly, capable of understanding its given goals, and provide clear, actionable outputs with minimal complexity.
	System Complexity and Customisation	[**R7**] The system shall offer scalable complexity, allowing customisation and advanced features for larger or more financially stable companies.
	End-Users	[**R8**] The system shall be suitable for end-users across various departments.
Economic Viability and ROI	Up-Front Cost and Affordability	[**R9**] The system shall be affordable for SMEs.
	Expected Benefits and ROI	[**R10**] The system shall deliver clear operational and financial benefits, ideally within a year.

Table 5.2: The manufacturing industry's requirements for AI-based predictive analytics systems, along with their respective theme and subtheme. Requirement IDs are written in bold square brackets.

5.3 Fusion Methods - RQ2

This section presents the findings from conducting a literature review identifying potential data fusion methods as well as utilising MCDM to select the most suitable fusion method for an industrial AI-based predictive analytics system. Together, these findings answer RQ2: 'What data fusion method is the most suitable for handling heterogeneous manufacturing data according to the industry requirements from RQ1?'.

5.3.1 Potential Fusion Methods

The first step of gathering fusion methods is defining what types there are. The absolute majority of sources describe three categories of data fusion: early, intermediate, and late fusion [36, 37, 26, 32]. The categories and which sources discuss them are shown in Table 5.3. These categories contain multiple possible implementations which all share commonalities in how the fusion happens. In some practically oriented papers [56, 57], the line between fusion types becomes hard to define as fusion may happen multiple times at different stages.

Level of fusion	**Description**	**Sources**
Early (signal)	Signal fusion is combining raw or minimally processed data. This is critically done before any analysis, which might obfuscate the raw input, removing information which models may use to learn patterns from.	[36, 37, 26, 32]
Intermediate (feature)	Feature fusion allows for some modelling to be done with raw data before combination. This modelling extracts features from the raw data, which are then combined and used for further analysis.	[36, 37, 26]
Late (decision)	Decision fusion keeps the analysis of separate sources completely separate until the end, when multiple outputs are concatenated and evaluated together.	[36, 37, 26, 32]

Table 5.3: Categories of data fusion, their descriptions, and which sources describe them. Both commonly used names are used for clarity.

Although this classification is widely used, it has also become increasingly outdated as newer, more advanced algorithms are developed. These algorithms do not divide themselves easily into the above-stated categories but may operate fusion steps in multiple ways, in multiple stages of execution [37]. This has led to other categorisation systems being proposed [36] and used [58], although these new systems are not widespread enough to be useful as categories between works. Others have chosen to instead focus on using multiple categories to categorise parts of methods instead [26].

The identified fusion methods were analysed with regard to the relevant requirements from RQ1. The relevant requirements were those which were deemed to be applicable when selecting the most suitable fusion method. Requirements **R2**, **R3**, **R4**, **R5**, **R7**, and **R9** were kept and **R1**, **R6**, **R8**, and **R10** were excluded.

[R2] The system shall be able to process multi-source heterogeneous data. Dataset availability is a severe problem in designing multimodal systems [37]. This can be alleviated by the use of late fusion, where missing modalities can be ignored

by not considering the missing modality's output into the final prediction [37]. This does not work for earlier (early or intermediate) fusion methods, where every input is needed, as there is no default value which provides no input. Instead, these missing values may be generated by interpolation between values. This interpolation needs to both generate a value, and possibly determine how accurate this value is [26].

If there are many inputs from similar devices recording the same process, the use of early fusion may help in smoothing out individual sensors' errors. This has been utilised in 3D motion analysis to reduce tracked objects' jitter by combining multiple views [36]. Higher-level (intermediate or late) fusion assumes that previous levels individually have enough correct data to extract features, while early fusion allows for cross-reference between multiple sources.

[R3] The system shall manage latency expectations, scale with operations, and remain energy-efficient. A lack of computational efficiency may be a problem with late fusion, as it may use a superfluous number of networks. Some modalities may be entirely discarded due to noisiness or irrelevance of data [36]. If some modalities are unable to contribute, then their calculations have been entirely wasted, as they could instead have been used to further analyse another, more useful modality. This waste becomes greater if modalities are very scattered and unable to individually come to any meaningful conclusion. Intermediate, in comparison to late fusion, does not allow early exits of its calculation as its generated features are required for the following parts to analyse [37]. If these are not useful, then they may also be discarded, although earlier than with late fusion. These have a greater chance of being useful to the following parts by not being a final answer, but rather features which may, together with other features, describe the final answer.

In situations where the ability to transmit data is limited by physical distance, bandwidth, or other means, it may be required to do feature extraction to lower the amount of data needed to transfer to the next stage of analysis. In [56], they describe a scenario where an unmanned vehicle may be unable to transmit the entirety of its raw data and, as such, runs a model locally, subsamples the output, and transmits the smaller subsample to a cloud server for processing. By transferring only features instead of the raw data, the total number of transfers can be limited. This processing of signals into features may also need to be balanced with battery limitations if the edge sensor is powered independently [36].

[R4] The system shall clearly show how its predictions are generated. Early fusion may be, if the models used are not completely traceable, harder to trace than intermediate and late fusion. As imperfect tracing methods have to cover more steps, the tracing may become worse. With early fusion, all steps need to be traced, from signal to decision. With intermediate, only the steps from extracted features to the decision need to be traced. In contrast, tracing with late fusion is much simpler, not only due to there being fewer steps, but also due to the simplicity of commonly used voting methods. Although intermediate and late fusion are simpler, they only allow tracing back to the modality level and not to the signal level. All methods can be traceable if traceable models are used internally. This distinction is more

applicable when using good but not perfect tracing methods, as you do not have to trace as far to get back to a specific modality.

[R5] The system shall produce predictions whose accuracy aligns with human expert assessments over time. The accuracy of the system is highly important to its usefulness. In order to achieve this accuracy, cross-analysis between modalities may be required. Late fusion lacks this ability to cross-analyse, which is present in both intermediate and early fusion. Early and intermediate fusion are therefore generally more accurate, but differ in how flexible and easy to train they are. Early fusion may take longer to train and re-train as modalities change, as the feature extraction phase needs to be re-trained as well. Intermediate fusion, on the other hand, does not need to re-train feature extraction, as that is a separate step. This allows more of the training time to be used to improve the analysis. Most of the modern AI solutions utilising fusion strategies that focus on accuracy performance use a form of intermediate fusion with separate feature extraction steps.

[R7] The system shall offer scalable complexity, allowing customisation and advanced features for larger or more financially stable companies. Early and late fusion are simpler designs, requiring fewer distinct steps to analyse data. They differ in model training as early fusion requires re-training as new data sources are added in order to incorporate them. With late and intermediate fusion, only the final combination needs to be re-trained as the number of data sources is expanded. The limitations on how complex an early or late fusion design can be make intermediate fusion the leader in high-end designs. This can be seen with modern implementations often being intermediate solutions, as the added crossover points of data in parts of the design allow for advanced cross-analysis.

[R9] The system shall be affordable for SMEs. The complexity issues of designing fusion systems are different between early and intermediate in comparison to late fusion. Late fusion can be used in incremental systems [26], as the analysis of individual modalities is completely separate from each other. In contrast, early and intermediate systems are monolithic, needing integration of new modalities at all or some levels of the architecture, respectively [26]. This ease of integrating new modalities makes late fusion attractive when considering dynamic, evolving systems.

5.3.2 Fusion Method Selection

To determine the most suitable fusion method MCDM was used. In order to create the raw values for the MCDM selection, aspects of the fusion methods found in Section 5.3.1 were used to motivate how well they fulfil the requirements derived from RQ1 (see Table 5.2). The created values are shown in Table 5.4.

Requirement	Early	Intermediate	Late
[**R2**] The system shall be able to process multi-source heterogeneous data.	3	4	5
[**R3**] The system shall manage latency expectations, scale with operations, and remain energy-efficient.	2	4	3
[**R4**] The system shall clearly show how its predictions are generated.	1	3	5
[**R5**] The system shall produce predictions whose accuracy aligns with human expert assessments over time.	4	5	2
[**R7**] The system shall offer scalable complexity, allowing customisation and advanced features for larger or more financially stable companies.	2	4	5
[**R9**] The system shall be affordable for SMEs.	5	1	3

Table 5.4: Raw scores for the fusion strategies of early, intermediate, and late on a scale from 1 (poorly supports requirement) to 5 (strongly supports requirement).

Weights for the requirements were created by analysing the frequency of their mentions in interviews as well as the literature review. Table 5.5 shows the comparison between requirements, describing how important they are compared to other requirements. Rows with large numbers and fractions describe requirements which were deemed important according to the interviews conducted, as well as the literature review in RQ1. Requirements R2 and R4 are overall more important than the other requirements, implying that any fusion method which fulfils them have an advantage in the selection process.

	R2	**R3**	**R4**	**R5**	**R7**	**R9**
R2	1	5	3	1	5	5
R3	1/5	1	1/5	1/5	1/3	1
R4	1/3	5	1	2	2	5
R5	1	5	1/2	1	1/2	2
R7	1/5	3	1/2	2	1	1
R9	1/5	1	1/5	1/2	1	1

Table 5.5: Comparison matrix containing raw weights describing the importance of the row requirement compared to the column requirement, on a scale from 1/9 to 9, with higher being better.

The weights for all the requirements were summed row-wise and then normalised so that all weights add up to 1, to create the final weights per requirement, as shown

in Table 5.6. Larger numbers represent more important requirements. Requirement [**R2**] **The system shall be able to process multi-source heterogeneous data** was found to be the most important to the industry in relation to the other requirements. This is in line with the study's purpose which is to find methods for handling heterogeneous data. Requirement [**R4**] **The system shall clearly show how its predictions are generated** had the second largest weight, implying that the industry holds traceability in high regard. Together, R2 and R4 constitute 59% of the total weights.

R2	**R3**	**R4**	**R5**	**R7**	**R9**
0.3340	0.04899	0.2561	0.1670	0.1286	0.0651

Table 5.6: Normalised weights for the relative importance of requirements.

Multiplying the normalised weights for the requirements from Table 5.6 with the raw scores from Table 5.4 gives the final scores of each fusion strategy as shown in Table 5.7. Late fusion has the largest score and is therefore considered the most applicable to manufacturing industry use. This is the fusion type which will be implemented and validated in RQ3. The difference between late and intermediate fusion is narrow with only a 0.56 point difference. Early fusion has a larger 1.66 point difference to late fusion.

Early Fusion	**Intermediate Fusion**	**Late Fusion**
2.6074	3.7154	4.2706

Table 5.7: Final aggregated scores for the fusion strategies of early, intermediate, and late.

5.4 Fusion Method Validation - RQ3

This section presents the findings from the evaluation of the most suitable fusion method identified in RQ2: late fusion. Late fusion was evaluated using a test-bed artefact and a generated dataset in terms of how well the fusion method fulfils accuracy, traceability, and computational needs, which together answer RQ3: 'How can the most suitable fusion method for heterogeneous predictive analytic systems from RQ2 be implemented to support evolving data sources, while satisfying the industrial requirements from RQ1?'.

5.4.1 Artefact Design

The data flow of the artefact is split into 3 parts. **Preprocessing**, **Model**, and **Meta model**, as shown in Figure 5.2. These can be expanded with new data sources with their own separate preprocessing and model.

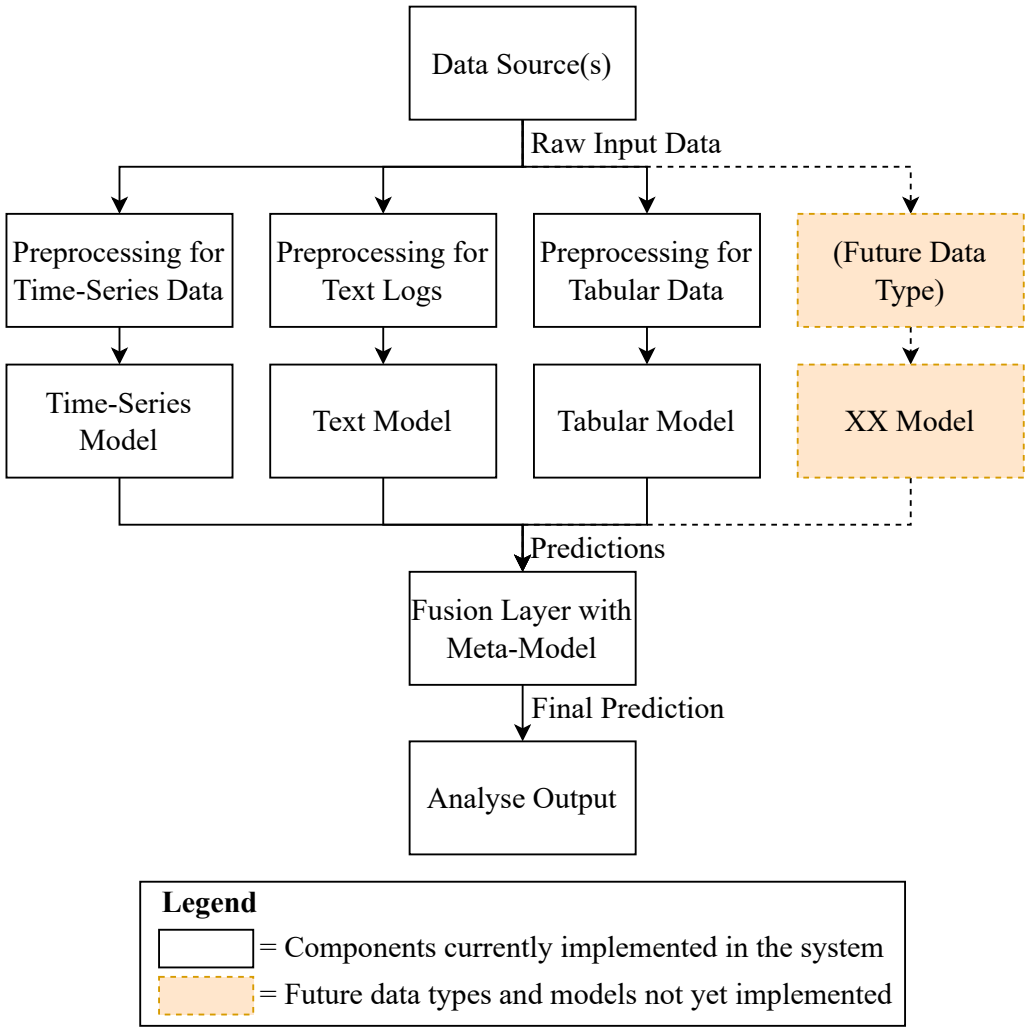

Figure 5.2: Overview of the created AI-based predictive analytics system artefact utilising late fusion.

Data preprocessing takes the raw data from each modality and prepares it for the individual model. This involves aspects such as interpolating the data of less frequent modalities. This was done in order to simplify the evaluation. The alternative would be to omit modalities with missing data. This was not done as the number of modalities in our experiment is limited. This stage also transforms the data into the preferred format of the assigned model, which evaluates the specific modality.

Individual model evaluates the preprocessed data by passing it to a model capable of processing it. The main models used were Random Forest for tabular, LSTM for time-series, and a combination of BERT followed by a Linear Regression for handling text. The model makes its own prediction and passes it on to the next part.

Meta model concatenates all the predictions from all the individual modality stages. The string of predictions is fed into a final meta model, which decides how the different individual predictions contribute to the final combined prediction with its own Random Forest model.

All of these steps can be easily expanded to include more data sources. By simply adding a preprocessing step, preparing it for a model, running the preprocessed data through the model, and concatenating it to the other predictions. The meta model in the fusion layer is then trained on the additional input from the new modality. This process enables the design to handle additional heterogeneous data sources, therefore fulfilling the first part of RQ3.

5.4.2 Dataset Generation

The datasets used for evaluation are based on a drill machining dataset released by Bosch Research [59] which is presented in Table 5.8. This dataset was used early on in development to test the validity of our program on a single modality. It was also used later on as the inspiration for the three modality dataset used in the validation of our system's ability to merge multiple modalities.

Machine	Machine ID of the machine running the process.
Process	Process ID of process being ran.
X	X coordinate of drill.
Y	Y coordinate of drill.
Z	Z coordinate of drill.
Label	"good" or "bad" depending on if vibration is anomalous.

Table 5.8: The dataset from Bosch Research upon which our new generated dataset is based on.

From the original dataset described in Table 5.8, a simulation was created which generated three separate modalities, all of which describe different perspectives of the same simulation.

The **time dataset** described in Table 5.9 describes the current x, y, and z values. This dataset was made to resemble the original dataset with real-time coordinate values. These coordinates were reported every minute and represent the highest frequency of all of the generated datasets. The rationalisation of inclusion is to provide frequent data comparable to the original dataset, which can then be corroborated with additional modalities.

The **text dataset** described in Table 5.10 describes an infrequent free text report describing experienced vibrations. This is reported every 30 minutes, representing infrequent data. The rationalisation of inclusion is to provide an additional, less frequent input which can be used to complement the other datasets. The format

being text provides more utilisation of the preprocessing step described in Figure 5.2 by having to convert the free text strings into a numerical representation.

The **tabular dataset** described in Table 5.11 describes the drift factor. This factor influences how fast the x, y, and z values change. This was reported every 5 minutes, representing data which is somewhat frequent but not always available. The rationalisation of inclusion is that a high drift factor may mean that values are able to be kept outside of acceptable ranges without saying anything about the actual x, y, and z values.

These three datasets were used in validating our solution with regard to the requirements most relevant to the technical implementation: accuracy, traceability, and computation.

Timestamp	Timestamp at which this data was recorded.
Machine ID	Machine ID of the machine running the process.
Operation	The operation being ran.
X	X coordinate of drill.
Y	Y coordinate of drill.
Z	Z coordinate of drill.
Vibration state	The state defining if the state is anomalous.

Table 5.9: The time modality generated based on 5.8.

Timestamp	Timestamp at which this data was recorded.
Machine ID	Machine ID of the machine running the process.
Operation	The operation being ran.
Text log	A text log describing the approximate vibrations of the drill.
Vibration state	The state defining if the state is anomalous.

Table 5.10: The text modality generated based on 5.8.

Timestamp	Timestamp at which this data was recorded.
Machine ID	Machine ID of the machine running the process.
Operation	The operation being ran.
Drift	The measured drift velocity of the drill.
Vibration state	The state defining if the state is anomalous.

Table 5.11: The tabular modality generated based on 5.8.

5.4.3 Accuracy

The design was evaluated by running the models for each modality and then combining the results with the meta model. The late fusion method's ability to utilise any combination of these three heterogeneous modalities shows that requirement [**R2**] **The system shall be able to process multi-source heterogeneous data** is fulfilled. This set of modalities can theoretically be expanded with new sources, preprocessing, and models. This ease of integration is especially important considering the heterogeneity of industry data and is a large contributing factor to late fusion being the selected method.

Table 5.12 presents which modalities were utilised as well as the coefficient of determination ($\mathbf{R^2}$) and Mean Squared Error (MSE). This shows the accuracy performance of differing combinations of modalities. This shows that accuracy does not degrade as modalities are added. No combination of modalities performs worse by introducing an additional modality, they either perform better or remain relatively stable. If late fusion goes from only processing Text, a comparatively bad performing modality with an **MSE** of 0.1364 and an $\mathbf{R^2}$ of 0.7287, to having Text and Tabular data, **MSE** is lowered to 0.0477 and $\mathbf{R^2}$ increased to 0.9050. This shows that the system's ability to capture variance in the data has improved while the prediction error has reduced, or at least maintained, by adding additional heterogeneous sources. Accuracy as a requirement is harder to motivate as being completely fulfilled, as accuracy is highly variable between models. However, the late fusion system was able to accurately model the binary classification of "bad" and "good" vibrations with a very low (0.0399) MSE and high (0.9205) $\mathbf{R^2}$ when processing all three modalities. These arguments point to requirement [**R5**] **The system shall produce predictions whose accuracy aligns with human expert assessments over time** being fulfilled.

MSE	$\mathbf{R^2}$	**Timeseries**	**Text**	**Tabular**
0.0618	0.8770	X		
0.1364	0.7287		X	
0.0477	0.9051			X
0.0615	0.8775	X	X	
0.0399	0.9204	X		X
0.0477	0.9050		X	X
0.0399	0.9205	X	X	X

Table 5.12: $\mathbf{R^2}$ and MSE scores for different combinations of the three modalities. X shows that a modality was included.

5.4.4 Traceability

The importance of individual features in each of the modalities is shown in the Tables 5.13, 5.14, and 5.15. The text modality utilised a model consisting of a pretrained BERT [60] model combined with a linear regressor. The BERT model is not

easily traceable, removing the possibility of tracing causality within the modality in this relatively small artefact. The traceability within other modalities and between the individual modalities and the meta model is importantly unharmed. Within Table 5.15, the importance of machine ID and operation are 0. This is by design, as these were not implemented to have any importance in the simulation.

Feature	Importance
y	0.8866
timestamp	0.0415
x	0.0398
z	0.0320
machine ID	0.0000
operation	0.0000

Table 5.13: The individual feature weights used in the time modality.

Feature	Importance
timestamp	N/A
machine ID	N/A
operation	N/A
text log	N/A
vibration state	N/A

Table 5.14: The individual feature weights used in the text modality.

Feature	Importance
timestamp	0.8857
drift	0.1142
machine ID	0.0000
operation	0.0000

Table 5.15: The individual feature weights used in the tabular modality.

The importance of combining multiple modalities in the meta model voting algorithm is shown in Table 5.16. The higher importance score of tabular data (0.6963) shows that it is the most information-rich data source and therefore more utilised. The text modality (0) is barely used in the final prediction, with a score near zero. This ability to trace the cause of a final prediction clearly satisfies [**R4**] **The system shall clearly show how its predictions are generated**. This is achieved by always having modality traceability as well as potentially providing traceability inside of modalities if the model allows it.

Feature	Importance
time	0.3036
text	0.0000
tabular	0.6963

Table 5.16: The modality weights used in the voting stage when including all 3 modalities.

5.4.5 Computation

The computation needed for evaluation is difficult to quantify and generalise as the needs can vary significantly depending on the context. The test-bed artefact developed to validate late fusion in this study is likely much simpler than a real-world system implementing late fusion. In practice, both the available hardware and the specific industry problems to be solved may change the total processing needs. Instead, the overall characteristics may be described to give guidelines as to how to design analytics systems utilising late fusion.

The time needed for the computation of the artefact implementation can easily be estimated. The individual models are evaluated entirely separately, meaning that the latency of the individual stage of evaluation is equal to the slowest modality. The meta model is then evaluated with all the individual modalities' output concatenated as the input. This means that the evaluation time is $max(modalities) + metamodel$. This simple expression ignores the time to transfer data between modalities and the meta model. This was done as the individual modality evaluation and the meta model evaluation were done on the same machine with the calculations, rather than data transfer, taking up nearly all of the elapsed time. This can be seen in the latency measurements shown in Table 5.17, where the total time elapsed is almost equal to the time taken by each modality plus the meta model. The reason the elapsed time is equal to $sum(modalities) + metamodel$ (13415ms) instead of $max(modalities) + metamodel$ (13396ms) is due to the artefact being single-threaded. The individual modalities have entirely separate data and could easily be processed in parallel. This parallelism would not be of much use in this example, as the difference between the average and highest latency is very large.

Evaluation	Latency
time	8ms
text	13391ms
tabular	8ms
meta model	5ms
total	13415ms

Table 5.17: The latencies of different individual stages as well as the meta model stage.

The scaling properties of late fusion is shown better with modalities with similar evaluation time, such as time and tabular as shown in Table 5.18. With both taking around 8ms to finish and the meta model 5ms. Sequential execution is done in 22ms. A parallel execution of the modalities would theoretically take around $max(8, 8) + 5 = 13ms$. This means that if the modalities take a similar amount of time, the speed-up of the individual stage of n modalities can reach n, meaning the algorithm scales constantly with unlimited processing cores. This beneficial scalability means that requirement **[R3] The system shall manage latency expectations, scale with operations, and remain energy-efficient** is fulfilled.

Evaluation	**Latency**
time	8ms
tabular	8ms
meta model	5ms
total	22ms

Table 5.18: The latencies of a reduced number of individual stages as well as the meta model stage.

6

Discussion

This chapter discusses the findings presented in Results (Chapter 5) regarding each of the three addressed research questions, including the insights gained during the process of addressing the questions, their implications, as well as future work.

6.1 RQ1 - Industry Requirements

Research Question 1 asked: 'What are the key requirements for AI-based predictive analytics systems within the manufacturing industry?'. The question aimed to uncover what manufacturers want from AI-based predictive analytics systems, as well as what is required for a successful adoption of it. This section discusses the process of contextualising and expanding upon the literature, some inconsistencies between findings, and the implications of the identified themes and requirements derived from addressing RQ1.

6.1.1 Contextualising and Expanding on Literature

To address RQ1, information was gathered through two phases: a systematic literature review and semi-structured interviews. The literature review identified five categories of challenges manufacturers commonly face when implementing AI solutions: Simplicity, Price, Computation, Traceability, and Integration. These challenge categories were then contextualised specifically for AI-based predictive analytics systems by conducting interviews with knowledgeable individuals from the manufacturing domain. The interviews also helped uncover additional concerns not highlighted in the literature, providing a more comprehensive understanding of industry requirements.

In addition to contextualising the five challenge categories identified in the literature, the interviews uncovered several domain-specific concerns that were either under-explored or missing from prior research. For example, while the literature within the category *Price* covers the manufacturing industry's financial concerns for implementing and deploying AI solutions, it does not account for the industry's expectation that predictive analytics systems should demonstrate clear ROI, often within one year of deployment. This highlights that financial justification is not simply about aligning costs with benefits but also about timing, as short-term financial

performance is seen as something crucial for adoption, especially when working with tight budgets. Similarly, other interview findings expanded the identified categories, revealing concerns such as building trust over time, a lack of knowledge regarding AI solutions, and wanting a scalable system that can be expanded over time.

6.1.2 Inconsistencies Between Findings

While the interviews helped contextualise and expand on the findings from the literature review, they also revealed some inconsistencies, both between interviewees and when compared to the literature. For instance, looking back at how two interviewees expressed that they want AI-based predictive analytics systems to provide ROI within a year, most interviewees expressed that the price of implementing such a system can be justified based on whether it provides clear long-term benefits. While interviewees might have different time frames in mind when talking about long-term benefits, this inconsistency highlights some tension between manufacturers wanting quick ROI and those that recognise that some benefits, such as increased productivity and fewer production errors, happen over time. This inconsistency also suggests that requirements regarding expected benefits and ROI could vary depending on organisational priorities, risk tolerance, and budget.

Similarly, three interviewees explicitly mentioned that they prefer more advanced, company-specific, AI-based predictive analytics systems, while two interviewees specified that they would like to start with simpler, low-cost solutions. While the interviewees differ in opinion, their preferences appear to depend on the company's size and mindset. SMEs often struggle to afford paying for high-cost systems upfront and therefore prefer more affordable systems that can be upgraded over time. In contrast, some manufacturers are sceptical that simpler, low-cost solutions would have the necessary features and capabilities to be viable in complex production environments. These inconsistencies suggest that requirements regarding system complexity and customisation could vary depending on company size, budget constraints, and leadership style, an observation which was echoed by two interviewees who emphasised the role of the CEO's mindset.

Prior research discussed in the literature showed that meeting computational demands, such as latency, is a common challenge for implementing complex AI solutions. However, the interviews emphasised a contrasting perspective. Most interviewees expressed that such performance metrics are not a primary concern in the initial stages of implementing AI-based predictive analytics systems. Interviewees expressed that, initially, people will be satisfied with the ability to access new insights, regardless of minor delays. However, interviewees noted that as companies get used to such systems, their expectations tend to change. Most interviewees agreed that latency will be of greater priority over time, as users get impatient and perhaps predictions are used for more time-dependent decisions. Nevertheless, as long as the predictions are delivered within a reasonable time frame, such as hours rather than days, interviewees generally viewed the latency as acceptable. The inconsistencies between literature and interviews suggest that, while meeting computational needs is important, priorities may change depending on the stage of adoption and use of

AI-based predictive analytics systems, as well as user expectations evolving over time.

6.1.3 Thematic Report and Requirements

The findings gathered from the literature review and interviews were analysed through a thematic analysis and compiled in a report. The report highlighted four different themes along with ten subthemes, each representing a different requirement for AI-based predictive analytics systems to be viable in manufacturing settings.

The theme *Technical and Infrastructure* shows that manufacturers require a technical infrastructure capable of collecting and storing relevant data, while the system itself needs to be able to process heterogeneous data from various modalities and provide support through actionable insights, all while adhering to various computational demands. A key implication from requirement R1 is that manufacturers are against handing over decision-making authority to predictive analytics systems, insisting that human operators remain in charge. They prefer that systems should not be able to perform actions by themselves, for instance, try to mitigate predicted future incidents from happening unless explicitly instructed to do so. This could be interpreted as manufacturers lacking trust in the current state of AI technology, after all, even highly accurate systems can make wrong predictions. Additionally, they could be concerned about operational risk and legal liability, for instance, if the system made a decision that led to a production stoppage or safety incident, assigning responsibility could be very difficult.

A key implication from requirement R2 is that AI-based predictive analytics systems should be designed in a modular and extensible fashion, allowing the handling of multimodal data in various formats. As the manufacturing industry continuously generates large volumes of heterogeneous data [1], manufacturers expect systems to be able to ingest and create predictions based on the data they already produce. Creating a system that can handle all types of heterogeneous data is both technically challenging and computationally intensive. Therefore, designing systems to be modular and extensible allows manufacturers to simply utilise data from the sources they currently have, with the option to add support for more sources and data formats in the future. This reduces the unnecessary complexity and performance latency that could come with supporting irrelevant or unused data formats.

A key implication from requirement R3 is that predictive analytics systems must meet context-specific computational and performance constraints. These constraints vary depending on the system's intended use cases. For instance, systems that handle real-time data to prevent short-term disruptions will likely have strict low-latency requirements, while systems used on weekly intervals might tolerate higher latency. Similarly, companies that process a lot of data might have a greater energy consumption in comparison to those that process less, making energy efficiency more important.

The theme *Trust, Traceability, and Transparency* shows that for AI-based

predictive analytics systems to be valuable in manufacturing contexts, the end-users need to be able to both trust the systems they use and verify their outputs. Requirement R4 implies that such verifiability comes from avoiding black box system designs, meaning that systems must be designed so that users can see and follow along with how outputs are generated, as well as know how accurate the outputs are. Being able to trace and verify outputs helps the users build trust in the system, encouraging its adoption in industry.

Similarly, a key implication from requirement R5 is that trust in AI-based predictive analytics systems must be earned and maintained over time. While some manufacturers might initially approach AI solutions with scepticism, seeing that systems consistently meet accuracy requirements and align with human expert assessments over time can help build user trust in the system. Over time, as systems prove themselves, users can come to rely on the predictions more and perhaps not require verifying the outputs as frequently or as thoroughly, freeing up human resources. As such, traceability and transparency features are necessary for ensuring and keeping long-term user trust and promoting system adoption.

The theme *Usability, Complexity, and Customisation* highlights that manufacturers require predictive analytics systems that are adaptable to company-specific requirements and easy to use. A key implication from requirement R6 is that usability should be a design priority when developing such a system. This includes creating a clear and interactive user interface so that it is easy for manufacturers to select what they want the system to predict, without needing excessive technical knowledge. Additionally, the outputs generated by the system should be presented in a readable and easily understandable manner so that less technical individuals can take action.

A key implication from requirement R7 is that systems should be designed with a target company or companies in mind. This means that, depending on the size and budget, complexity and customisation should be adjusted. As advanced systems can be very expensive to develop, SMEs with limited resources might prefer simpler, more generic systems that cost less to develop and manage. In contrast, companies with greater resources might require more advanced systems that are tailored to handle their specific data and integrate better with their existing workflows. To accommodate this diversity, systems should be designed to be extensible, with a set of core features that can be expanded on or adapted over time. This allows companies that start with a generic solution to scale up and tailor their system to fit their needs as their budget increases or priorities change.

A key implication from requirement R8 is that predictive analytics systems should be designed with a target audience in mind. The target audience consists of end-users, who, depending on the company and applications of its systems, could come from various manufacturing departments, such as engineering, maintenance, and production. These users could all have different goals and technical backgrounds, and therefore need the system to be crafted in a way similar to what requirement R6 implies, ensuring that they find the system user-friendly and can navigate it

well enough to achieve their goals. This could be done by creating role-specific views or dashboards that only display data and insights relevant to the current user's responsibilities. For instance, an engineer might require access to detailed performance metrics, while a manufacturer working within maintenance might only need high-level status reports and forecasts.

The theme *Economic Viability and ROI* highlights the expectations manufacturers have when deciding whether to invest in AI-based predictive analytics systems, as well as how they expect to benefit from using such systems. A key implication from requirement R9 is that systems must be financially accessible, especially for SMEs that might not have the budget to invest in complex solutions. This can be done by offering affordable entry points, such as generic base versions or having subscription-based models with a more affordable up-front payment, assuming that manufacturing companies are purchasing solutions instead of developing them in-house. Even in the cases where companies are developing solutions themselves, requirement R9 remains important. It highlights the importance of designing systems in a way that limits initial development costs by, for example, prioritising core functionalities first or utilising open-source tools as much as possible.

Although requirement R9 is similar to R7, in the sense that they both acknowledge the different needs of companies and their respective circumstances, R9 differs by focusing less on the technical aspects. While R7 focuses on scalability and customisation, R9 focuses more on being able to practically adopt systems. It ensures that manufacturing companies, regardless of size or resources, can realistically start utilising such systems.

A key implication from requirement R10 is that predictive analytics systems must demonstrate noticeable benefits within a set time frame. While the exact time frame can vary depending on context, some interviewees suggested that benefits should be clear within a year to justify the investment in the system. Although this requirement can sound somewhat vague, it highlights the importance of systems being able to achieve measurable outcomes, such as reducing waste and production stoppages. It also suggests that the system should be able to learn and adapt over time to become accurate enough to justify its investment. If the system fails in delivering noticeable outcomes within the expected time frame, it risks appearing as ineffective or not worth the cost.

6.1.4 Implications of Requirements

While the previous subsection highlighted the ten requirements along with their implications for designing AI-based predictive analytics systems that are useful in manufacturing contexts, it is important to acknowledge that these requirements are not prescriptions set in stone. Given the many use cases for predictive analytics systems, especially within the manufacturing industry, the outlined set of requirements should be viewed more as a flexible template that can be adjusted to fit individual needs and application contexts. For instance, requirement R3 addresses latency expectations without specifying exact thresholds. R8 mentions end-users, who can

vary depending on the organisation. R10 suggests demonstrating ROI within a year, while the actual time frame can be much longer or shorter depending on company size, budget, and priorities.

6.2 RQ2 - Fusion Methods

Research Question 2 asked: 'What data fusion method is the most suitable for handling heterogeneous manufacturing data according to the industry requirements from RQ1?'. This section discusses the process of gathering fusion methods and selecting a suitable candidate for later implementation in the artefact of RQ3 which fulfils the requirements from RQ1.

6.2.1 Fusion Method Categories

There are many ways of categorising fusion methods. Many papers found during the literature review had their own systems of categorisation by focusing on different aspects. Most focus on dividing methods into what underlying techniques they use. For example, a paper frequently used during the literature review [37] divided methods into decoder-encoder, kernel, graphical, and attention-based models. These types of classification systems may be of more use as they are related to specific models, but they are too undefined and not widely used enough for cross-referencing between papers. This lack of standardised categories leads to the use of the more classic, widely used, early, intermediate, and late classifications.

Early, intermediate, and late fusion, while widely used and therefore widely applicable, group disparate implementations with different characteristics. There were, therefore, moments where a fusion method had general weaknesses which specific implementations had found solutions for. For example, late fusion suffers from excessive computation by analysing modalities which may not be used in the final evaluation. This can, and has been, mitigated by not analysing weakly connected modalities or using easily evaluated modalities as an indicator of whether further analysis is required. This overgeneralisation was deemed necessary to get enough separate data sources to motivate a selection of method, but requires further analysis being needed to select a specific solution.

6.2.2 Scoring

Scoring fusion methods fulfilment of requirements could be done quite easily for certain technical requirements, such as accuracy, where fusion methods' performance has already been explicitly evaluated. Some requirements were comparatively hard to score. Requirements such as **[R9] The system shall be affordable for SMEs** required finding tangential but still related motivations and manually relating them to a specific requirement. This may lead to less objective scores due to personal evaluations of how those motivations relate to the requirement, influencing the final scores. This could be alleviated by having more previous research into the specific requirements and how they relate to fusion methods.

6.2.3 Weights

This thesis uses the conducted interviews as well as the RQ1 literature review to ground the weights in real-life requirements. This grounding is usually done in MCDM evaluations with quantitative data, in comparison to this study which had a focus on quality rather than quantity research. A greater focus on quantitative data may be better in defining the slight differences in how many industry professionals see certain requirements as necessary. This could be done with surveys or other quantitative data. The weight scoring would be made with more quantitative data to complement the qualitative data from the interviews. The qualitative data can, as it is done in this thesis, better define current problems and requirements. The quantitative data can, after requirements have been defined, be used to define how severely this problem needs to be fixed. Using qualitative data to identify problems and quantitative data to define the severity can enable them to jointly better define problems. This severity can then be compared to other requirements' severities and used to create the weights.

In order to counteract the lack of qualitative measurements for our specific requirements, a secondary screening of scores was done with industry professionals at Quay Technologies. This was done to keep the scores general despite being derived from the qualitative data from the interviews. It was hoped that the one-to-many sources at Quay Technologies would compensate as a substitute for proper qualitative data.

6.2.4 Implications of Final Scores

Late fusion was selected as the "most suitable" fusion method according to the MCDM selection in RQ2. This result is not an absolute fact but is determined by the weights. According to the results from Table 5.6, some requirements are deemed more important than others. Due to the study's reliance on qualitative data from interviews, this evaluation is highly relevant to current issues. Issues such as the current disparity of systems and the need for simple solutions led to late fusion being favoured.

As a counter to this conclusion, the point of changing requirements as the field matures could be raised. As trust is earned and inter-modal analysis becomes more needed, solutions where data is cross-analysed, such as in both early and intermediate fusion may become more attractive. With intermediate fusion only being 0.56 points away from late fusion, there is a possibility that it may become favoured. This would align with much of the current research being into this type of fusion [36, 37, 26]. This means that the requirements will need to be reevaluated later on, at which point a study following a similar methodology to the thematic analysis in this thesis could be recommended.

6.3 RQ3 - Validation

Research Question 3 asked: 'How can the most suitable fusion method for heterogeneous predictive analytic systems from RQ2 be implemented to support evolving

data sources, while satisfying the industrial requirements from RQ1?'. This section discusses how the implementation of the selected "most suitable" fusion method from RQ2 impacts the fulfilment of the industry requirements from RQ1. The section also covers details relating to the evaluation artefact, as well as its implications for wider industrial adoption.

6.3.1 Dataset Availability

A major limiter during the work has been a lack of widely available heterogeneous datasets representing the manufacturing industry. We have had access to sets of data showing the general structure and trends of data in manufacturing, but none which could be used for the validation phase during this thesis. The Bosch Research dataset [59] which inspired the simulation used to generate data is homogenous and therefore unusable in our prototype. However, combining it with the information regarding regularly collected data from industry actors leads to a heterogeneous dataset which reflects the industry needs. This lack of available data also corroborates the identified lack of digitised datasets for analytical systems to utilise.

6.3.2 Relation to Requirements

The artefact's evaluation provided the opportunity to gain additional insights into four of the identified industrial requirements.

[R2] The system shall be able to process multi-source heterogeneous data. All of the fusion methods discussed in this thesis are able to handle heterogeneous data in some way. Late fusion was preferred due to the separated preprocessing and the simple fusion step of multiple data sources. Another likely candidate in the future is intermediate fusion. Both have a separate preprocessing step, which makes them better suited to multi-source heterogeneous data in comparison to early fusion. This can be seen in Table 5.16 where the origin of the result is traced back to each modality. With an importance score split around 0.7 to 0.3 for the tabular source and the time source, the use of multiple sources is proven. This suggests that although late fusion does not cross-analyse modalities, it is still able to use multiple when calculating a prediction. This is and will continue to be a highly desired requirement, as it seems to be in the nature of data, in a field as diverse as manufacturing, to be highly heterogeneous.

[R3] The system shall manage latency expectations, scale with operations, and remain energy-efficient. Although requirement R3 was not assigned a high weight in the MCDM evaluation, it is highly connected to the future of fusion methods. As the field continues to mature and systems are deployed, the ways in which to scale them up need to advance as well. All of this while keeping within latency limits. Some benefits of working with late fusion systems became clear during the work, which may be of use in this area. The benefits and drawbacks of fusion methods regarding computation speed and latency are shown in Subsection 6.3.4.

[R4] The system shall clearly show how its predictions are generated. One

of the largest benefits of late fusion is the simplicity of tracing back the cause of a prediction to the different modalities. Even though the analysis inside a modality may be complicated, and therefore hard to trace, the voting systems of a meta model do not need to be as complicated. This results in easily traceable meta models where the blame for a prediction can be easily placed between modalities. This information can then be used by the system's operator to act upon. Providing interpretability to the system.

[R5] The system shall produce predictions whose accuracy aligns with human expert assessments over time. Accuracy was not a major focus, although it was included as a requirement. This was done as there were other requirements which were deemed more relevant, as the field is quite new and small-scale adoption is of a bigger concern. This lack of focus on specifically accuracy can also be noticed in the areas where late fusion performs well compared to early or intermediate fusion. Late fusion performs well with redundant or complementary modalities. The exact numbers of the accuracy are not important, as the exact accuracy is highly dependent on the models used, which is not the focus. The focus is instead on the trends in accuracy when different modalities are included or excluded. In the accuracy table for the artefact (see Table 5.12), the three modalities all described the same simulation with highly independent data. These do not strictly require as much cross-analysis as, for example, cooperative or synergistic sensors would. These would describe the same simulation, but from completely different perspectives or dimensions. This form of fusion would require the cross-analysis afforded by other forms of earlier (early and intermediate) data fusion methods. Even though the artefact used late fusion with limited cross-analysis, both the $\mathbf{R^2}$ and MSE scores improved. This highlights the need for more data sources even though the underlying method can not do full cross-analysis by itself.

6.3.3 Design Complexity

The designed test-bed artefact is quite simple. This was done due to the limited development time assigned to the thesis. A lack of large IT departments and limited early investments were both identified as barriers to the adoption of analytical systems utilising AI. This shows that analytical systems, and specifically late fusion-oriented ones, do not need to be very complicated. The simpler systems, such as this one, may later evolve into more advanced forms should internal expertise increase. These simpler systems are critical for the early stages of adoption, which the industry is currently in.

6.3.4 Increasing Computational Demand

Computational demand was listed in the requirements as **[R2] The system shall be able to process multi-source heterogeneous data**. Although it was included, it received a small weight due to the currently perceived near irrelevance of performant algorithms. This is due to any predictions being seen as novel, not requiring quick predictions to satisfy the demand. There were considerations that, as the technology matures, the performance requirements will become more relevant.

Two avenues of computation optimisation were considered, depending on the chosen fusion method.

Edge computing is the use of computers at the source of data to do a portion of the calculations. This is a form of distributed computing which helps parallelise the total computation needed. Late fusion is very well adapted to the deployment of edge computing as the individual modalities are entirely separate. These modalities can be calculated separately, with only the result needing to be transferred to and combined in a central computer, saving on data transportation costs. The reduced traffic could make late fusion attractive in environments where data traffic is reduced, such as in war zones, as indicated by the literature review. The edge computing may also require less powerful hardware, as each computer only needs to evaluate a single modality. This may be less efficient computationally compared to highly efficient large-scale GPUs or TPUs, as they are not able to be deployed to every data source, as they may lack the scale necessary. The scaling factor afforded by implementing edge computing in a late fusion system can become highly sought after as the scale and number of modalities increase. With hundreds or thousands of sources, the improved scaling can become a deciding factor, as logical and therefore possible spatial separation limits the reliance on data centres.

Large-scale accelerators are a major source of performance improvement in modern analysis, due to the high levels of accelerator adoption. Some fusion methods are more suitable for accelerator use. Late fusion with its independent modalities limits the scale of accelerators to the individual modalities. Early, as well as intermediate fusion after feature extraction, analyses the entire collection of modalities. This scale allows for the use of large-scale accelerators in energy-efficient data centres by analysing large amounts of data simultaneously. This type of centralisation scaling is more monetarily efficient as the process has had a long time to mature, with large data centres having been deployed for decades.

6.3.5 Implications of Evaluation

The accurate evaluation of fusion methods is difficult due to the large scales at which real production systems operate. In this thesis, the evaluation was done with a small system with a limited number of modalities. The system did need to be able to show the path to scaling up operations, which it does by providing an easy point of expansion, as is shown in Figure 5.2. The evaluation also shows the promising aspects of late fusion, with its satisfaction of four requirements (R2, R3, R4, R5). The other two requirements (R7, R9), while relevant to the selection of fusion method, were not deemed to be motivated thoroughly enough by the implemented artefact alone. Another relevant aspect is the predicted increased performance demand as these systems are more widely adopted and therefore need to be streamlined. Solutions like edge computing and accelerators were discussed; however, future research will also be needed to explore the fusion methods' performance characteristics as the technology moves forward.

6.3.6 Recommendations For Industrial Prediction Systems

According to the importance of requirements shown in Table 5.6 and the experience gained creating the test-bed for them in RQ3, there are some key takeaways for engineers designing prediction systems utilising late fusion.

Keep single point of integration. Heterogeneity is prevalent in industrial data and requires simple integration of new sources. One approach is the test-bed's approach with entirely separate pre-processing. This enabled modalities to be added or removed without significantly affecting the rest of the system. This modular approach reduces system coupling and simplifies maintenance as modality requirements change. This is not the only possibility as any solution which enables the system to expand without having to update multiple parts is to be preferred.

Consider system-specific performance demands. While defining what the requirements were in RQ1, performance metrics were often mentioned. However, due to the thesis' wide scope of evaluating general requirements over the entirety of industry, exact requirements were impossible to define. Engineers creating systems should therefore confirm that any specific performance requirements in their environment are fulfilled by the designed system. Further discussion regarding future computation scaling related to late fusion is provided in Subsection 6.3.4.

Ensure prediction traceability when creating prediction systems. Predictions without justification lack persuasiveness, particularly in environments where operators must understand or justify automated decisions. To fulfil this, the test-bed implemented feature level and modality level traceability to argue why a prediction was made. This level of traceability may not be needed for every system, but providing reasoning behind the prediction should be considered.

Align system predictions with expert judgement and manage expectations. Prediction systems are typically used to support domain experts rather than replace them. As a result, the system's predictions should not only be evaluated through traditional machine learning metrics but also by comparing them with expert assessments over time. If the system produces results that contradict experienced operators without being able to convince them with traceability, trust in the system can degrade. Engineers should therefore seek expert feedback as ongoing validation to ensure the system is in step with experts. In practice, this means positioning prediction systems as decision-support tools rather than autonomous decision making systems. Over time, monitoring the agreement between experts and the model's predictions can help identify disconnects, guiding further model refinement.

If these recommendations are followed, the development and maintenance of late fusion based systems are more likely to fulfil our more general industry requirements, while supporting system specific variation.

6.4 Synthesis

The goal of the thesis was to generate "Requirements for AI-Based Predictive Analytics Systems in an Industrial Manufacturing Context" and "Selection of the Most Suitable Data Fusion Method". This is at its core a RE problem which solves a selection of software architecture. To accomplish this three main questions needed to be answered.

The first question, "What are the industry requirements?" is addressed in RQ1. This is Requirements Elicitation, Negotiation, Specification, and Validation where the industry requirements are defined and weighted according to their relative importance. This forms the largest part of the method, the result of the entire thesis, and leads into the other two parts.

The second question, "What are the possible solutions?" is addressed in RQ2 and builds on the answer for the first question. This is an analysis of potential fusion methods and how well they satisfy the requirements. This is not so much RE in and of itself but an application of it. However, it forms a vital part for the third question, where RE returns.

The third question, "Does the selected best solution satisfy the requirements?" is addressed in RQ3. This is primarily Requirements Verification of the requirements identified by addressing the first question, where a small test-bed of the selected fusion method identified by selecting the most promising solution from addressing the second question, is used to verify the original requirements. This verification step gives legitimacy to the previous work on requirements in RQ1 and fusion method selection in RQ2 by demonstrating their applicability in a test environment.

By addressing the three questions together, requirements are created, a fusion method is selected based on those requirements, and both processes' results are verified.

6.5 Future Work

This thesis sets the foundation for future research regarding requirements for AI-based predictive analytics systems and data fusion methods, both in the manufacturing domain as well as other related fields.

6.5.1 Industry Requirements

A large contribution of this thesis is the definition of industry requirements derived from interviews with industry professionals and a literature review. This is an area in which there is a distinct lack of previous research. Although this thesis makes the claim that the derived requirements are in line with actual industry needs, it is also limited in the quantity of data upon which they are founded. Future research may conduct larger-scale research into industry needs, in order to generate more

generalisable requirements. These may then be filtered depending on the specific use case, as this thesis has done with the focus on fusion methods, limiting the applicable requirements to the requirements R2, R3, R4, R5, R7, and R9. The approach of relying on qualitative data helps to gather requirements without preconceived notions about what the problem is infecting the research. The original 10 requirements mentioned in this thesis are specifically related to the manufacturing industry's adoption of AI. These can then be reduced. Having a larger pool of issues can help identify subproblems such as "AI adoption in the manufacturing industry" covered in this thesis.

6.5.2 Possible Additional Verification of Weights

Verification of RQ1 could be split into two distinct areas: requirement scoring and requirement scoring verification. Despite the thesis' aim to be as objective as possible, the research process may have introduced some subjectivity through the authors' interpretation of interviews and literature review. Hence, future iterations of the research process outlined in this thesis could benefit from the introduction of processes eliminating individual bias.

To improve requirement scoring, a process such as Delphi [61] could remove the authors' subjectivity when deriving quantitative scores from qualitative sources such as interviews or literature. The Delphi method is particularly suitable in broad areas where experts might not posses all the necessary expertise needed to make an informed decision, such as industry. Another alternative is using analytic hierarchy process (AHP) [62] to derive a hierarchical description of how experts opinions diverge. This method has the added benefit of providing a consistency ratio which can be used to evaluate if a result is contentious. With Delphi being more consensus oriented and AHP being more quantitative the choice can also be influenced by the ratios of quantitative vs qualitative data which is available. With this thesis' exact methodology focusing on qualitative data, Delphi might be more applicable. That said, implementing either process would further reduce scoring bias.

To improve requirement scoring verification, a process such as inter-rater agreement [63] or scenario-based evaluation [64] could be used. These approaches are more focused on verifying completed weights rather than generating them. Inter-rater agreement's main goal is to provide feedback as to how reliable the created weights are, rather than correcting them. The alternative, scenario-based evaluation, verifies the created weights by testing them against real use cases. This is more in line with software architecture evaluation and requires information regarding what the end use case is, something which is too specific for this thesis. This could be a good method if there is a specific use case in mind as it would better incorporate practical limitations and requirements. As this thesis is aimed at the broader industry rather than any specific use case, inter-rater agreement would be an interesting addition to the work. While the thesis had done some expert verification by having industry experts at Quay Technologies give feedback on the scores, using a standard inter-rater reliability process would further increase reliability and reproducibility of the final result.

6.5.3 Standardised Categories of Fusion Methods

During the literature review for RQ2, the lack of standardised categorisation of fusion methods led to the use of early, intermediate, and late fusion as categories. These may not be specific enough to help the developers implementing systems choose which fusion method is most appropriately suited. This was done as separate papers discussing fusion methods used various methods to define categories. A standardised set of categories would help with discussing fusion methods across papers.

6.5.4 Test-Bed Artefact

A limiting factor in the evaluation of the artefact is its scope. Due to time constraints, as well as the primary focus on addressing RQ1 and RQ2, several of the identified industrial requirements were not implemented, and hence could not be included in the evaluation. Additionally, the artefact could be tested in handling a greater volume of, as well as more varied, data types and modalities. It could also be tested by being integrated into a real-world environment to better evaluate its practical applicability. Future research could be applied to mitigate these shortcomings in the current artefact evaluation.

7

Conclusion

In this study, the selection of fusion methods for heterogeneous data in AI-based predictive analysis systems was explored through literature review, interviews, and thematic analysis. Ten industrial requirements, representing the current-day needs of the manufacturing industry when developing or adopting predictive analytics systems, were identified. Six of which were directly relevant to the fusion method selection. Late fusion, which was found to best theoretically support the fulfilment of requirements, with an aggregated score of 4.27 compared to intermediate's 3.71 and early's 2.60, was implemented in a limited test-bed artefact and showed promising alignment with these requirements. Of particular note was late fusion's fulfilment of the key requirements of scalability, and traceability.

In addition to the methodological results, this study highlights the importance of understanding and addressing the target domain's needs when designing AI systems. The process outlined in this thesis demonstrates that combining findings from relevant literature with insights from experts within the targeted domain improves the alignment between technical choices and practical requirements. From a software engineering perspective, the study demonstrates how requirement engineering combined with a decision framework such as MCDM can guide complex software architectural design choices and support the execution of a DS approach. While not the primary focus of this study, the developed process may be transferable to other domains beyond manufacturing and to other software engineering problems where requirement-driven design and MCDM are needed.

Looking forward, the process and insights derived from this study can be used to further guide and inform future research and decisions in requirement-driven software architecture design areas. Although the result of the selection process for fusion methods should be further validated due to being validated only in a limited test-bed. As data sources, industrial processes, and the manufacturing industry as a whole evolve, the identified industrial requirements may come to change, potentially requiring additional iterations of the process to update the fusion method selection and architecture decisions. By providing a systematic process to identify and align domain-specific needs with technical decisions, this work proposes a solution for ongoing improvements in complex software system design and engineering practice.

Bibliography

[1] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial artificial intelligence in industry 4.0 - systematic review, challenges and outlook," *IEEE Access*, vol. 8, pp. 220121–220139, 2020.

[2] L. Wang, "Heterogeneous data and big data analytics," in *ACIS*, vol. 3, pp. 8–15, 2017.

[3] S. Kamm, S. S. Veekati, T. Müller, N. Jazdi, and M. Weyrich, "A survey on machine learning based analysis of heterogeneous data in industrial automation," *Computers in Industry*, vol. 149, p. 103930, 2023.

[4] I. M. Putrama and P. Martinek, "Heterogeneous data integration: Challenges and opportunities," *Data in Brief*, p. 110853, 2024.

[5] C. T. Brownlees and G. M. Gallo, "Financial econometric analysis at ultra-high frequency: Data handling concerns," *Computational statistics & data analysis*, vol. 51, no. 4, pp. 2232–2245, 2006.

[6] A. Ittoo, A. van den Bosch, *et al.*, "Text analytics in industry: Challenges, desiderata and trends," *Computers in Industry*, vol. 78, pp. 96–107, 2016.

[7] M. Y. Saeed, M. Awais, R. Talib, and M. Younas, "Unstructured text documents summarization with multi-stage clustering," *IEEE Access*, vol. 8, pp. 212838–212854, 2020.

[8] D. Lechevalier, A. Narayanan, and S. Rachuri, "Towards a domain-specific framework for predictive analytics in manufacturing," in *2014 IEEE International Conference on Big Data (Big Data)*, pp. 987–995, IEEE, 2014.

[9] J. Bleiholder and F. Naumann, "Data fusion," *ACM computing surveys (CSUR)*, vol. 41, no. 1, pp. 1–41, 2009.

[10] Quay Technologies AB, "Letshare." https://letshare.se/, 2023. Accessed: 2025-05-21.

[11] F. Herrmann, "The smart factory and its risks," *Systems*, vol. 6, no. 4, p. 38, 2018.

[12] Gartner, Inc., "Gartner survey shows 57% of manufacturing leaders feel their organization lacks skilled workers to support smart manufacturing digitization plans," 2021. Accessed: 2025-07-05.

[13] B. Kudzmanaitė, R. Gabaliņa, and O. Yevdokymova, "Accelerating european industry digitalisation: challenges ahead," *PPMI News and Insights*, 2022.

[14] A. Corallo, A. M. Crespino, V. Del Vecchio, M. Lazoi, and M. Marra, "Understanding and defining dark data for the manufacturing industry," *IEEE Transactions on Engineering Management*, vol. 70, no. 2, pp. 700–712, 2021.

[15] R. Homman, "Harnessing the data overload," *ML Journal August 2021*, 2021.

[16] S. Yoo and N. Kang, "Explainable artificial intelligence for manufacturing cost estimation and machining feature visualization," *Expert Systems with Applications*, vol. 183, p. 115430, Nov. 2021.

[17] S. J. Plathottam, A. Rzonca, R. Lakhnori, and C. O. Iloeje, "A review of artificial intelligence applications in manufacturing operations," *Journal of Advanced Manufacturing and Processing*, vol. 5, no. 3, p. e10159, 2023.

[18] J. Kutz, J. Neuhüttler, J. Spilski, and T. Lachmann, "Implementation of ai technologies in manufacturing - success factors and challenges," 01 2022.

[19] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Computers and Electronics in Agriculture*, vol. 143, pp. 23–37, 2017.

[20] J. Wieczorkowski and P. Polak, "Big data: Three-aspect approach," *Online Journal of Applied Knowledge Management*, vol. 2, p. 2014, 01 2014.

[21] A. C. Eberendu *et al.*, "Unstructured data: an overview of the data of big data," *International Journal of Computer Trends and Technology*, vol. 38, no. 1, pp. 46–50, 2016.

[22] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Journal of Manufacturing Systems*, vol. 48, pp. 157–169, 2018. Special Issue on Smart Manufacturing.

[23] S. Kamm, N. Jazdi, and M. Weyrich, "Knowledge discovery in heterogeneous and unstructured data of industry 4.0 systems: Challenges and approaches," *Procedia CIRP*, vol. 104, pp. 975–980, 2021. 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0.

[24] W. W. Eckerson, "Predictive analytics," *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, vol. 1, pp. 1–36, 2007.

[25] V. Kumar and M. Garg, "Predictive analytics: a review of trends and techniques," *International Journal of Computer Applications*, vol. 182, no. 1, pp. 31–

37, 2018.

[26] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.

[27] S. Y. Boulahia, A. Amamra, M. R. Madi, and S. Daikh, "Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition," *Machine Vision and Applications*, vol. 32, p. 121, Sep 2021.

[28] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Enabling flexible manufacturing system (fms) through the applications of industry 4.0 technologies," *Internet of Things and Cyber-Physical Systems*, vol. 2, pp. 49–62, 2022.

[29] Z. Jan, F. Ahamed, W. Mayer, N. Patel, G. Grossmann, M. Stumptner, and A. Kuusk, "Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities," *Expert Systems with Applications*, vol. 216, p. 119456, 2023.

[30] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance," *IEEE Access*, vol. 5, pp. 23484–23491, 2017.

[31] J. Wang, W. Zhang, Y. Shi, S. Duan, and J. Liu, "Industrial big data analytics: Challenges, methodologies, and applications," 2018.

[32] K. Gadzicki, R. Khamsehashari, and C. Zetzsche, "Early vs late fusion in multimodal convolutional neural networks," in *2020 IEEE 23rd international conference on information fusion (FUSION)*, pp. 1–6, IEEE, 2020.

[33] J.-H. Choi and J.-S. Lee, "Embracenet: A robust deep learning architecture for multimodal classification," *Information Fusion*, vol. 51, pp. 259–270, 2019.

[34] Z. A. Sahili and M. Awad, "The power of transfer learning in agricultural applications: Agrinet," 2022.

[35] C. A. Aguilar-Lazcano, I. E. Espinosa-Curiel, J. A. Ríos-Martínez, F. A. Madera-Ramírez, and H. Pérez-Espinosa, "Machine learning-based sensor data fusion for animal monitoring: Scoping review," *Sensors*, vol. 23, no. 12, 2023.

[36] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Information Fusion*, vol. 57, pp. 115–129, 2020.

[37] S. Li and H. Tang, "Multimodal alignment and fusion: A survey," 2024.

[38] M. Elhefnawy, M.-S. Ouali, A. Ragab, and M. Amazouz, "Fusion of heterogeneous industrial data using polygon generation & deep learning," *Results in Engineering*, vol. 19, p. 101234, 2023.

[39] L. Oldemeyer, A. Jede, and F. Teuteberg, "Understanding the challenges of ai implementation: Insights from an empirical study of manufacturing companies," in *19th International Conference on Wirtschaftsinformatik*, 09 2024.

[40] I. Sommerville and J. Ransom, "An empirical study of industrial requirements engineering process assessment and improvement," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 14, no. 1, pp. 85–117, 2005.

[41] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[42] K. Pohl, *Requirements engineering: An overview*. RWTH, Fachgruppe Informatik Aachen, 1996.

[43] K.-J. Stol and B. Fitzgerald, "The abc of software engineering research," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 27, no. 3, pp. 1–51, 2018.

[44] S. Keele *et al.*, "Guidelines for performing systematic literature reviews in software engineering," tech. rep., Technical report, ver. 2.3 ebse technical report. ebse, 2007.

[45] A. Davis, O. Dieste, A. Hickey, N. Juristo, and A. M. Moreno, "Effectiveness of requirements elicitation techniques: Empirical results derived from a systematic review," in *14th IEEE International Requirements Engineering Conference (RE'06)*, pp. 179–188, 2006.

[46] S. E. Hove and B. Anda, "Experiences from conducting semi-structured interviews in empirical software engineering research," in *11th IEEE International Software Metrics Symposium (METRICS'05)*, pp. 10–pp, IEEE, 2005.

[47] M. D. C. Tongco, "Purposive sampling as a tool for informant selection," 2007.

[48] A. Mavin, P. Wilkinson, A. Harwood, and M. Novak, "Easy approach to requirements syntax (ears)," in *2009 17th IEEE international requirements engineering conference*, pp. 317–322, IEEE, 2009.

[49] H. Taherdoost and M. Madanchian, "Multi-criteria decision making (mcdm) methods and concepts," *Encyclopedia*, vol. 3, no. 1, pp. 77–87, 2023.

[50] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *Journal of mathematical psychology*, vol. 15, no. 3, pp. 234–281, 1977.

[51] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[52] OpenAI, "Chatgpt: A large language model." https://openai.com/chatgpt, 2023. Accessed: 2025-04-01.

[53] M. Sharma, S. Luthra, S. Joshi, and A. Kumar, "Implementing challenges of artificial intelligence: Evidence from public manufacturing sector of an emerging economy," *Government Information Quarterly*, vol. 39, no. 4, p. 101624, 2022.

[54] M. L. Council, "The future of industrial ai in manufacturing," Jun 2023. Accessed: 2025-05-09.

[55] S. Sinha and Y. M. Lee, "Challenges with developing and deploying ai models and applications in industrial systems," *Discover Artificial Intelligence*, vol. 4, p. 55, Aug 2024.

[56] E. Blasch, T. Pham, C.-Y. Chong, W. Koch, H. Leung, D. Braines, and T. Abdelzaher, "Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges," *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 7, pp. 80–93, 2021.

[57] T. Segreto and R. Teti, "Data quality evaluation for smart multi-sensor process monitoring using data fusion and machine learning algorithms," *Production Engineering*, vol. 17, pp. 197–210, Apr 2023.

[58] S. Kumar, S. Rani, S. Sharma, and H. Min, "Multimodality fusion aspects of medical diagnosis: A comprehensive review," *Bioengineering*, vol. 11, no. 12, 2024.

[59] M.-A. Tnani and M. Feil, "GitHub - CNC machining data: Data set for process monitoring on CNC machines." `https://github.com/boschresearch/CNC_Machining`, 2022. [Accessed 6-03-2025].

[60] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.

[61] N. Dalkey and O. Helmer, "An experimental application of the delphi method to the use of experts," *Management science*, vol. 9, no. 3, pp. 458–467, 1963.

[62] E. H. Forman and S. I. Gass, "The analytic hierarchy process—an exposition," *Operations research*, vol. 49, no. 4, pp. 469–486, 2001.

[63] K. Gwet, "Handbook of inter-rater reliability," *Gaithersburg, MD: STATAXIS Publishing Company*, pp. 223–246, 2001.

[64] N. Looker, D. Webster, D. Russell, and J. Xu, "Scenario based evaluation," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, pp. 148–154, IEEE, 2008.

A

Interview Guide

A.1 Introduction

- A brief introduction and overview of the study will be given.
- The purpose of the interview will be explained.
- The interview is not a technical test, and there are no wrong answers.

- Do we have your permission to record this session?

A.2 Background and Context

- Can you tell us a little about yourself and your current work role?
    - What are your current work responsibilities?
- What types of data does your company currently collect or use (e.g. time-series data, logs, or maintenance records)?
- What's the current status of AI at your company?
    - (If they are using AI) What were the biggest challenges during the implementation?
    - (If they have considered using AI) What is stopping the implementation of AI?
    - (If they have not considered using AI) Why was AI not implemented?
- Have you ever used or heard about systems that, e.g. monitor or forecast machine health, production trends, or quality issues?

A.3 Core Questions

Reliability

- If you had a system that could make predictions about, e.g. production stoppages or safety incidents, how would you judge if it's good enough to trust?
- What kind of mistakes would be unacceptable?
- What specific performance or accuracy metrics are necessary before a predictive system is considered deployable?

Integration

- Is your current IT infrastructure (e.g. MES, ERP, sensors) ready to integrate with analysis software like AI?
- What other systems or data sources would a prediction system need to connect to?
    - Who would need to be involved to get such a system working on your site?

Simplicity

- Do you believe you have access to the expertise needed to implement a system like this?
- Who in your company would use a system like this?
- If you had to choose between a simple, low-cost solution and a more advanced, custom one, which would you prefer?
    - Why?

Price

- Do you believe that the price involved in implementing a predictive analytics system would be justified by its benefits? Are the price and benefits clear to you?
- What kind of benefits would justify an investment in your view? (e.g. less downtime, better quality, fewer errors)

Computation

- Are aspects like latency relevant to you with a predictive system?
- How often do you need the predictive system to generate insights?

Traceability

- Do you trust an AI solution?

  - Why?

- Would you need to understand how the system came to its prediction?
  - Why?

A.4 Wrap-Up

- In your view, what are the non-negotiable requirements for predictive analytics systems to be accepted in manufacturing environments?
- How do you see these requirements evolving over the next few years?
- Is there anything else you think is critical that we haven't covered?