Licentiate Thesis

AI-Based Methods For Improved Testing of Radio Base Stations: A Case Study Towards Intelligent Manufacturing

Cristina Landin
Technology

Örebro Studies in Technology 102
Örebro 2023

AI-Based Methods For Improved Testing of Radio Base Stations: A Case Study Towards Intelligent Manufacturing

Örebro Studies in Technology 102

Cristina Landin

AI-Based Methods For Improved Testing of Radio Base Stations: A Case Study Towards Intelligent Manufacturing

Supervisors: Amy Loutfi
Martin Längkvist
Sahar Tahvili

Title: AI-Based Methods For Improved Testing of Radio Base Stations: A Case Study Towards Intelligent Manufacturing

ISSN 1650-8580

Abstract

Testing of complex systems may often require the use of tailored-made solutions, expensive testing equipment, large computing capacity, and manual implementation work due to domain uniqueness. The aforementioned test resources are expensive and time-consuming, which makes them good candidates to optimize. A radio base station (RBS) is a complex system. Upon the arrival of new RBS generations, new testing challenges have been introduced that traditional methods cannot cope with. In order to optimize the test process of RBSs, product quality and production efficiency can be studied.

Despite that AI techniques are valuable tools for monitoring behavioral changes in various applications, there have not been sufficient research efforts spent on the use of intelligent manufacturing in already existing factories and production lines. The concept of intelligent manufacturing involves the whole system development life-cycle, such as design, production, and maintenance. Available literature about optimization and integration of industrial applications using AI techniques has not resulted in common solutions due to the complexity of the real-world applications, which have their own unique characteristics, e.g., multivariate, non-linear, non-stationary, multi-modal, class imbalance; making it challenging to find generalizable solutions. This licentiate thesis aims to bridge the gap between theoretical approaches and the implementation of real industrial applications.

In this licentiate thesis, two questions are explored, namely how well AI techniques can perform and optimize fault detection and fault prediction on the production of RBSs, as well as how to modify learning algorithms in order to perform transfer learning between different products. These questions are addressed by using different AI techniques for test optimization purposes and are examined in three empirical studies focused on parallel test execution, fault detection and prediction, and automated fault localization. For the parallel test execution study, two different approaches were used to find and cluster semantically similar test cases and propose their execution in parallel. For this purpose, Levenshstein distance and two NLP techniques are compared. The results show that cluster-based test scenarios can be automatically generated from requirement specifications and the execution of semantically similar tests can reduce the number of tests by 95% in the study case if executed in parallel.

Study number two investigates the possibility of predicting testing performance outcomes by analyzing anomalies in the test process and classifying them by their compliance with dynamic test limits instead of fixed limits. The performance measures can be modeled using historical data through regression techniques and the classification of the anomalies is learned using support vector machines and convolutional neural networks. The results show good agreement between the actual and predicted learned model, where the root-mean-square error reaches 0.00073. Furthermore, this approach can automatically label the incoming tests according to the dynamic limits, making it possible to predict errors in an early stage of the process. This study contributes to product quality by monitoring the test measurements beyond fixed limits and contributes to making a more efficient testing process by detecting faults before they are measured. Moreover, study two considers the possibility of using transfer learning due to an insufficient number of anomalies in a single product.

The last study focuses on root cause analysis by analyzing test dependencies between test measurements using two known correlation-based methods and mutual information to find strength associations between measurements. The contributions of this study are twofold. First, test dependencies between measurements can be found using Pearson and Spearman correlation and MI; and their dependencies can be linear or higher order. Second, by clustering the associated tests, redundant tests are found, which could be used to update the test execution sequence and choose to execute only the relevant tests, hence, making a more efficient production process by saving test time.

Sammanfattning

Testning av komplexa system kräver ofta skräddarsydda lösningar, dyr utrustning, omfattande beräkningskapacitet och manuellt implementationsarbete på grund av dess domän unika egenskaper. De ovannämnda testresurserna är dyra och tidskrävande, vilka gör dem till bra kandidater att optimera. En radio basstation (RBS på engelska) är ett komplex system. Med nya RBS generationer introduceras nya utmaningar inom test vilka traditionella metoder inte kan hantera. För att optimera testprocesserna för RBS, så kan produktkvalitet och produktionseffektivitet studeras.

Trots att AI tekniker är värdefulla verktyg för övervakning av beteendemässiga förändringar i olika applikationer, har man inte studerat användandet av s.k. intelligent tillverkning i befintliga fabriker och produktionslinjer. Konceptet intelligent tillverkning berör hela systemets utvecklingsprocess, system development lifecycle (SDLC), såsom konstruktion, produktion och underhåll. Tillgänglig litteratur om optimering och integrering av industriella tillämpningar med AI tekniker har inte lett till gemensamma lösningar på grund av komplexiteten av praktiska tillämpningar vilka har sina egna unika egenskaper, t.ex. multivariata, ickelinjära, ickestationära och där data är multimodalt och obalanserat i fråga om klasser. Denna avhandling har som mål att överbygga gapet mellan teoretiska tillvägagångssätt och praktisk tillämpning i verkliga industriella applikationer.

Två frågor utforskas in denna avhandling, nämligen hur väl några AI-metoder kan utföra och optimera feldetektion och felprediktion tillämpat på produktionen av RBS:er, samt hur algoritmer för lärande kan modifieras för att överföra lärt beteende mellan olika produkter. Dessa frågor adresseras genom användande av olika AI-metoder för testoptimering och undersökts i tre empiriska studier som fokuserar på parallel testexekvering, feldetektering och -prediktion, samt automatiserad fellokalisering. För studien om parallell testexekvering användes två tillväqgagångssätt för att hitta och gruppera semantiskt liknande testfall och för att exekvera dessa parallellt. Levenshsteinavstånd och två metoder för naturlig språkbehandling jämförs för detta ändamål. Resultaten visar att gruppbaserade testscenarion kan genereras automatiskt från kravspecifikationer och exekveringen av semantiskt liknande testfall kan reducera antalet tester med 95% i det studerade fallet om testerna kan köras parallellt.

Den andra studien undersöker möjligheten att prediktera utfallet vid prestandatester genom att analysera avvikelser i testprocessen och klassificera dem genom de-

ras uppfyllande av dynamiska kravgränser istället för fasta kravgränser. Mätetalen för prestandatest modelleras genom att använda regressionsmetoder på historiska data och klassificeringen av avvikelserna grupperas genom s.k. support vector machines (SVM) och faltade neurala nätverk (CNN). Resultaten visar god överensstämmelse mellan faktisk och predikterad inlärd modell där kvadratiska medelvärdet når 0.00073. Vidare kan denna metodika automatiskt prediktera fel i tidiga steg av processen. Denna studie bidrar till att förbättra produktkvalitet genom användandet av dynamiska gränser vilka möjliggör prediktion av icke godkända tester utan att tidskrävande mätningar behöver genomföras. Denna studie utforskar även möjligheten att använda överfört lärande då det ofta saknas avvikande resultat när man begränsar sig till att studera en enskild produkt.

Den sista studien fokuserar på rotorsaksanalys genom att analysera beroenden mellan mätningar. Två välkända korrelationsbaserade metoder och mutual information användes för att undersöka sambanden mellan dessa olika mätningar. Studien bidrar i två aspekter. Den första är konstaterandet att beroenden i testprocessen kan hittas genom användandet av Pearsons och Spearmans korrelationsmetoder och även genom mutual information, samt att beroendena kan vara linjära eller av högre ordning. Den andra är att genom att gruppera olika tester så kan överflödiga tester hittas vilket kan användas för att uppdatera testsekvensen och endast exekvera de relevanta testerna. På så vis kan produktionsprocessen förbättras genom insparandet av testtid.

Acknowledgements

I would like to thank the University of Örebro, my main supervisor Prof. Amy Loutfi, and my co-supervisor Dr. Martin Längkvist, who gave me the opportunity to start this cooperation between the university and Ericsson AB, thanks for being there, thanks to your advises especially in the beginning of my studies; without your guidance, I would not have been able to pass the first year. My thanks go to the lab AASS, especially my group machine perception and interaction (MPI), thanks everyone, when I started my studies you were kind and made me feel welcome. Past and present colleagues: You are missed.

Thanks, Jonas Brattström for accepting me to participate in this collaboration, despite the heavy amount of work we have at the company. Thanks, Thomas Petterson (TP), we started this project together, and thanks for the constant support in keeping me there.

My thanks go to all my mentors, advisors, and supervisors during my years as a student in Peru and in Sweden. Thanks to them and their persistent teachings and high-standard demands, made of me the person I am. Thanks, M.Sc. Wildor Ferrel, Dr. Torbjörn Olsson, Eng. Robert Johansson and my industrial supervisor dear Dr. Sahar Tahvili, without You nothing of this would have been possible, I will always be thankful. Dear colleagues at Ericsson AB and co-authors, thanks a lot.

I thank a number of colleagues who graciously read early drafts of this thesis and who provided constructive advice that led to significant improvements. Among these, I mention in particular Dr. Shoaib Amin and Dr. Per Landin. Any shortcomings, errors, and misconceptions recorded here, however, are my own.

I wish to express my sincere gratitude for the support I received from the Swedish Knowledge Foundation (KKS) under grant number 20190128 and the Swedish Governmental Agency for Innovation Systems (VINNOVA) under grant number D-RODS (2023-00244), through the Industrial Graduate School Collaborative AI and Robotics (CoAIRob), for the research efforts I have dedicated to this subject in recent years.

Thanks to my beloved family here in Sweden, Leo, Christian, Per, Lily, Lilja, Olle, and my beloved family in Peru, dear mom Avelina, Alfredo, Hector, Carmencita, Diego, my aunts, and her families, for all their support and understanding during this journey. There have been tough times especially during the pandemic when I could not be with you and mostly when we lost our father. Dear Dad: I will keep all your teachings as treasures in my heart. Thanks to my parents who gave me the courage to pursue my dreams, without limitations and also gave me the freedom to choose on my own. Thanks to my friends here in Sweden, and my Peruvian and Latino community, all these years I feel well accompanied. Thanks for making this trip a happy trip.

And finally, I would like to thank the reader, this is my first book, so be kind. It has been the efforts of many years and I left part of my soul (also some tears) here. Thanks and remember: Today can be the first day of the rest of your life!

List Of Publications

This licentiate thesis is a compilation of papers. The research findings and contributions have been published in several conference and workshop papers and are referred to in the text using the following labels:

Paper I Cristina Landin, Sahar Tahvili, Hugo Haggren, Martin Längkvist, Auwn Muhammad and Amy Loutfi, *"Cluster-Based Parallel Testing Using Semantic Analysis"*, IEEE AITest 2020, The Second IEEE International Conference On Artificial Intelligence Testing, August 2020.

Paper II Cristina Landin, Leo Hatvani, Sahar Tahvili, Hugo Haggren, Martin Längkvist, Amy Loutfi and Anne Hákansson, *"Performance Comparison of Two Deep Learning Algorithms in Detecting Similarities Between Manual Integration Test Cases"*, The Fifteenth International Conference on Software Engineering Advances (ICSEA), October 2020.

Paper III Cristina Landin, Jie Liu, Sahar Tahvili, *"A Dynamic Threshold-Based Approach for Detecting the Test Limits"*, The Sixteenth International Conference on Software Engineering Advances (ICSEA), October 2021.

Paper IV Cristina Landin, Xinrong Zhao, Martin Längkvist and Amy Loutfi, *"An Intelligent Monitoring Algorithm to Detect Dependencies between Test Cases in the Manual Integration Process"*, The 16th IEEE International Conference on Software Testing, Verification and Validation Workshop (ICSTW), April 2023.

Paper V Cristina Landin, Jie Liu, Katerina Katsarou and Sahar Tahvili, *"Time series anomaly detection using Convolutional Neural Networks in the Manufacturing Process of RAN"*, The 5th IEEE International Conference on Artificial Intelligence Testing (AITest), July 2023.

Additional publication by the author during her studies but not included in the Licentiate thesis:

Patent I Sahar Tahvili, Alzahraa Salman, Cristina Landin and Vincent Huang, *"Test script generation from test specifications using natural language processing"*, Patent No. WO2022028721A1. Application filed by Telefonaktiebolaget Lm Ericsson (Publ), 2020.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

Throughout human history, there have been four major industrial revolutions transforming industrial processes into more effective ones. These technological developments have greatly reduced manual labor by enabling automation and improving production efficiency while reducing production cost.

The Fourth Industrial Revolution (4IR) major goal is to shift from automated manufacturing toward intelligent manufacturing. Intelligent manufacturing is a term used for describing the efforts to make the manufacturing process optimal, in the sense of efficiency, and effectiveness by using the advantages of currently available technologies [1], such as artificial intelligence (AI), sensing, networking, and fast communication. These technologies and the easy access to data have made 4IR very attractive, which has driven an explosion of diverse applications to improve all kinds of manufacturing processes using data-driven solutions [2]. However, despite these attempts, there are challenges remaining in the application of the aforementioned strategies in real-world industrial applications [3], such as the production of complex systems. That is understandable due to the industrial applications' unique characteristics. On one hand, the processes can be multidimensional, non-linear, and non-stationary. On the other hand, the produced data can be multi-modal, heterogeneous, and severely imbalanced.

This thesis has specifically studied the testing process in the production of radio base stations(RBSs). These are complex systems that are used in mobile radio communication systems such as long-term evolution (LTE)/4G and 5G. An RBS contains a plethora of analog electronic components intended for radio communication, optical components for connection to the core network, power converters for powering all components, microprocessors for signal conditioning, transmission, and reception, as well as in-house and commercially available software applications. RBSs are required to fulfill stringent requirements, especially related to the radio aspects, before they are put into operation, such as those specified by 3GPP standard regulator [4] as well as company internal requirements.

The system development life-cycle (SDLC) contains three important stages: design, execution, and deployment. Enhancing the testing process can be approached by optimizing the SDLC's stages using AI techniques. However, RBS's test optimization requires knowledge of the test flow to approach the most urgent needs. Moreover, for practical test optimization, several factors must be considered, such as physical systems' constraints, coverage, scalability, and generalizability. The available literature about RBSs and AI are mostly limited to the effect of exposure to electromagnetic fields and how to improve the power and capacity efficiency of RBSs once they are deployed [5]. There is not much literature on how to optimize the testing process of such complex systems using smart solutions likely due to the very specific domain [6] and commercial confidentiality. The potential of the findings could benefit other stakeholders sharing the same data properties, open the floor for further discussions, and perhaps initiate collaboration.

In this licentiate thesis, several methods for solving the test optimization problem in the production of RBSs are proposed in the following forms: parallel test execution, test scheduling, test suite minimization; and fault detection and prediction of performance outcomes. All of the proposed approaches in this thesis are applied and evaluated in a set of empirical studies at the company using data from the production test of 4G and 5G RBSs and focuses on product quality and optimization of the production process at the integration level. By achieving these objectives, testing approaches can keep pace with industry advancements and contribute to the overall success of new technologies.

1.1 Problem Statement

The SDLC process of RBSs has been extensively optimized in accordance with the arrival of new radio generations. The optimization processes include statistical approaches to reduce the execution of stable tests, manual test generation and reuse of legacy test models, and manual data analysis to reduce time in the test execution and in the fault diagnosis. Upon the arrival of 5G RBSs, traditional test optimization methods are not tractable anymore. Because the number of ports has increased significantly due to the need for beamforming and massive MIMO used to increase capacity and cell coverage, as well as increased requirements on bandwidth and number of signal scenarios that are to be tested. As 5G RBSs possess these new properties, the production process and test must be extended proportionally. A linear scaling of the testing time and cost with the increase of RBS complexity is not economically sustainable Therefore, the search for better approaches using data-driven solutions that can optimize the current test process is an appealing and necessary option for the production test development unit at Ericsson.

However, the direct use of AI approaches is not straightforward. Similar approaches with the same characteristics of complex systems, such as RBSs, are difficult to find in the literature. The processes are multi-modal, have high dimension sequential data, heterogeneous, non-linear, and have been created using manual approaches. The two possible areas to optimize are to reduce the manual work (design and analysis) and the usage of fixed criteria for favourable outcomes. However, to gain a fuller understanding of what are the AI approaches applicable to the design and development stages in the production of complex systems, in-depth quantitative research focused on optimization and performance of fault detection and prediction is also needed. It can potentially also identify other stages to optimize or conflict with in the production process of RBSs that may explain positive (or negative) outcomes. Furthermore, due to the inherent properties of industrial applications, transfer learning may be utilized to manage the main problems traditional AI techniques cannot handle, such as insufficient amount of labeled data, computational complexity, and distribution mismatch of different radio products.

The primary objective of this licentiate thesis is to optimize the testing process, aiming to minimize manual efforts and improve the efficient utilization of testing resources. To achieve this, the research begins by conducting an extensive literature review to gain insights into the latest developments in the field of AI regarding test optimization. The knowledge acquired from the literature review is applied to real-world use cases within the company. It is crucial to acknowledge that industry use cases are dynamic and often present numerous constraints. These constraints may arise from factors such as time limitations or infrastructure restrictions. Consequently, the application of state-of-the-art algorithms can become challenging in such practical scenarios. The research will address these challenges by seeking practical and feasible solutions that balance the advancements in AI with the practical constraints faced in the industry. By adapting and customizing AI techniques to suit the specific requirements of the company's use cases [7], the thesis aims to find practical, effective, and efficient approaches to optimize the testing process, despite the existing constraints.

1.2 Research Questions

The problem statement mentioned in Section 1.1 is addressed by specifically looking at the following research questions:

- *RQ1: To what extent can AI techniques optimize and perform fault detection and prediction on the production test of RBSs?*

  Using the waterfall model for system verification and validation, where each step in the process depends on the previous steps, will frequently cause development losses. For instance, when an error is found in a later stage of the process, the whole system must be redesigned. Therefore, we aim to use dynamic methods to find dynamic solutions using the available data from the production of 5G (and older generations) radio products at Ericsson AB. Currently, the available data is analyzed using statistical approaches and if changes are needed to improve the efficiency, they are done manually, or sample-based. Firstly, it is important to understand to what extent can the available literature within AI be applied to optimize the production process of an existing industrial application, where computational simplicity, heterogeneous data handling, and real-time data processing are needed. To facilitate the search, this research question is divided into two sub-research questions as follows:

- *RQ1.1:*

  The first part aims to find the most important features (test cases), that can predict a successful outcome (fault-free) product using different AI techniques, test results, and historical data. **Paper III** and **Paper V** aims to monitor different performance indicators (e.g., yield and final test result) and anomaly detection using data-driven solutions instead of fixed thresholds with the purpose of predicting future events.

- *RQ1.2:*

  The second part tries to optimize the test case schedule and find the most efficient test sequence to guarantee good test performance and test coverage. This part intends to use the defined product's test requirement specifications (TRSs) and the results from the first part to propose the most efficient scheduling. **Paper I** and **Paper II** aim to optimize the test case generation and test execution. Typically, TRSs are given in natural language and test case generation is performed manually. By using different natural language processing (NLP) techniques, automatic test case generation and parallel test execution can be a reality, however, there may exist challenges for deployment, such as test case dependencies, as well as lack of sufficiently detailed information from the TRS documents.

  Paper IV studies the possibility of using the most important features for dynamic test case scheduling and test suite minimization. It also reviews automated fault localization by analyzing test results and test dependencies between test cases. This not only mitigates the arduous manual work for test debugging but also finds the most relevant and redundant test cases which may vary with the evolution of the process. Therefore, this paper covers both parts of RQ1.

- *RQ2: How do we modify learning algorithms in order to do transfer learning between different RBS versions and generations?*

  Once a model has been designed and tuned for a specific product, there is an intent to extend it to cover other products within the same domain, i.e., other RBSs. Though this RQ may seem challenging, we have seen the possibility of using transfer learning (TL) between 4G and 5G RBSs as shown in **Paper III** and **Paper V**. However, it is important to observe that this work is in an early stage and that future radio generations might be relatively different. However, there is potential to use this technology for continuous learning in the industry, where training data is limited.

1.3 Mapping of Contributions to the Papers

This work follows an iterative process, which starts with a literature review and proceed with three studies to answer the research questions listed in Section 1.2. Table 1.1 provides a summary of the studies, maps the corresponding papers developed throughout this thesis to the research questions, and gives a starting point for the purpose of each study. Figure 1.1 summarizes the thesis.

Study 1 was developed to answer which AI methods can be applicable to the studied industrial use case, can handle text data inputs, and reduce the resources required for test case generation. Test optimization in general has been in focus at the beginning of the study but changed the focus to efficient test debugging in *Study 3*. Both studies cover production optimization. On the other hand, product quality enhancements can be obtained by anomaly detection and prediction of performance indicators. *Study 2* was proposed to study the applicability of using those definitions in real-world applications, such as heterogeneous and imbalanced data sets.

Study	Paper	RQs	Purpose
1	I, II	RQ 1.2	Parallel test execution
2	III, V	RQ 1.1, 2	Anomaly detection and prediction
3	IV	RQ 1	Test case dependencies and automated fault localization

Table 1.1: Summary of contributions and mapping of the studies developed in this thesis.

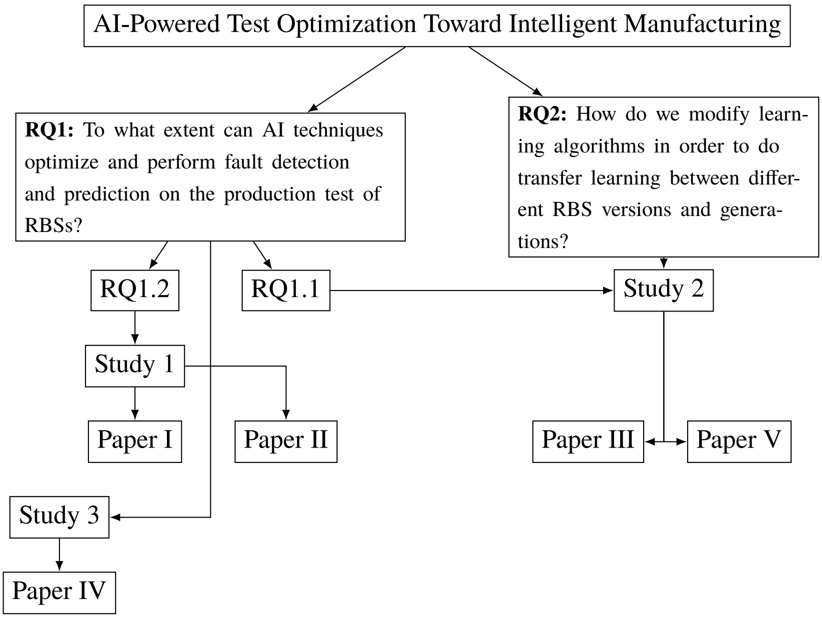

Figure 1.1: A holistic overview of how the studies included in this licentiate thesis support the research goals.

1.4 Research Process and Methodology

The methodology employed in this thesis can be summarized as follows: case studies and various data collection methods, including unstructured interviews, document

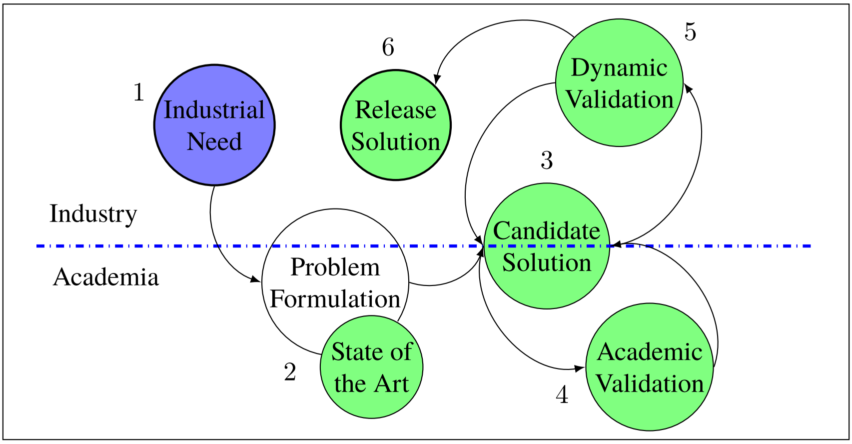

Figure 1.2: The research model and technology transfer overview used in this thesis.

analysis, and observations. This approach was chosen to ensure a comprehensive understanding of the research topic and to obtain valid answers to the research inquiries. The collaboration between industry and academia was a crucial aspect of this industrial research, as it allowed for the assessment of academic research outcomes in a real industrial context, thereby enhancing the industrial development process. A strong and dynamic partnership between researchers and practitioners was emphasized throughout the research process, as it is considered key to achieving success in industrial research. By adopting an appropriate research strategy and employing a range of data collection methods, this licentiate thesis aimed to contribute valuable insights to the field.

The research process adopted in this thesis is depicted in Figure 1.2, which is inspired by Tahvili [8] and is an adaptation of the technology transfer model proposed by Gorschek et al.in [9]. This model serves as a visual representation of the steps involved in conducting the research in this thesis. It outlines the stages of problem identification, review of the state of the art, ongoing issue examination, engagement with testing experts, selection of specific challenges, formulation of research objectives and questions, proposal of solutions, and evaluation through empirical evaluations.

The forthcoming steps delve into the specifics of the research methodology employed in this thesis, offering a more comprehensive understanding of its implementation.

- **Step 1: Identify an industrial need**

  It is critical to consider the demands of the industry before designing research questions. Thus, we started our research by observing the industrial setting and current solutions. During this process, several potential areas of improvement at various testing levels at Ericsson AB were identified, of which the integration testing level has been selected as a viable candidate for improvement. Moreover, the test optimization problem in the form of test case scheduling and minimization, has been recognized as a real industrial challenge at Ericsson AB production during our process assessment and observation activities.

- **Step 2: Problem formulation**

  Based on the identified needs in the previous step and by collaborating closely with the testing experts at Ericsson AB, the problem statement was formulated. The testing department at Ericsson AB consists of several testing teams including software developers, integration engineers, team leaders, and middle managers. The researchers have regular meetings with the Ericsson testing group. Furthermore, to identify challenges, a comprehensive review of the current state of the art is conducted. This involves examining ongoing issues and engaging in semi-structured interviews with testing experts at Ericsson AB in Sweden, which established a common vocabulary of the research area and the system under test (SUT) between researchers and testing experts.

- **Step 3: Formulation of a candidate solution** A set of solutions is proposed to address the identified research goals effectively. These solutions are designed to provide practical and innovative approaches that can contribute to resolving the research challenges identified previously. In a continuous collaboration with the testing teams at Ericsson, a set of candidate solutions for the improvement of the integration testing process was designed. In this step, Ericsson covered the role of keeping the proposed solutions compatible with their testing environment. On the other hand, we as the research partners took on the main responsibility for keeping track of state of the art in the test optimization domain and applying the proposed solutions with a combination of new ideas. We designed a support system for anomaly detection. The main purpose of the proposed solution was to monitor the anomalies and predict future failures. With an agreement with Ericsson, the proposed solution was selected as a promising solution for test optimization at the integration testing level at Ericsson AB.

- **Step 4: Academic Validation**

  In the principal technology transfer model proposed by Gorschek et al. in [9], several steps are considered for evaluating the candidate solutions proposed in the previous step. In this licentiate thesis, we employed academic and dynamic validation methods to evaluate the proposed solution for solving the test optimization problem at Ericsson AB. Our scientific work was evaluated by international review committees from the venues where we published our research results, of which five papers have been selected and presented in this licentiate thesis. In this step, the limitations of the various approaches are identified and certain solutions for addressing these limitations are provided as future work.

- **Step 5: Dynamic Validation**

This step has been performed through three research studies. According to the project's plan, a physical weekly meeting needs to be held at Ericsson between industrial and academic partners who are involved in the mentioned research projects. The results of the conducted case studies, prototypes, and experiments are presented by researchers during the meetings. Moreover, some small workshops are organized by us for the team members of different internal testing projects within Ericsson. The industrial partners gave valuable feedback, some of which are applied in this step. Tool support was the main feedback for the proposed solutions that we received from Ericsson.

- **Step 6: Release solution**

  Exploration and development of new technologies, methodologies, or frameworks are undertaken to contribute to solving the research challenges in the industry. These solutions must be flexible to adapt and tailor-made the solutions for specific applications. The value of the research results is focused on usability in the industry. After receiving and analyzing the feedback from the academic and dynamic validation steps, the proposed solutions are then implemented as actual supportive and monitoring tools. The initial versions of the proposed solutions are then implemented by our master thesis students and Ericsson engineers.

  Full implementation of the released solutions in this licentiate thesis is pending results from the prototypes. However, our ultimate goal is to integrate the monitoring systems to predict anomalies and to incrementally release it as a supportive tool to the company's big monitoring system.

By following the above-mentioned steps and approaches, the research in this licentiate thesis progresses systematically, contributing to the advancement of knowledge and addressing the research inquiries on the topic. Furthermore, in this thesis, both collaboration and continuous learning between academia and industrial partners are considered to play an important role in guiding and validating the research process. Collaboration and partnerships with industry experts, practitioners, and other researchers are pursued to exchange knowledge, share resources, and benefit from complementary expertise, while continuous learning keeps abreast of the latest advancements and trends in the field through continuous learning and active participation in conferences, workshops, and professional networks.

1.5 Thesis outline

This licentiate thesis comprises two main parts: the thesis summary and the included papers. This chapter presents an overview of the research, including the problem statement, research questions, technical and industrial contributions, and the research methodology employed during the studies.

- **Chapter 2** provides background information on the research conducted in the field of automation testing, the test optimization use cases approached in this thesis and the latest AI techniques applied in the industry. To facilitate the understanding of the testing process, a section on RBSs is also included.

- **Chapter 3** provides a summary of the findings and answers to the research questions. It also opens some discussions on the applicability of the findings in real use cases.

- **Chapter 4** presents the concluding remarks, limitations, possible societal ethical impacts, and outlines potential future work.

The second part of the thesis consists of five investigations (Papers I-to-V) that detail the research findings.

Chapter 2
Background and Related work

This chapter introduces the reader to the concepts behind test optimization for 4IR, specifically RBSs and gives an overview of the related work to this thesis. Section 2.1 describes in detail an RBS. Section 2.2 explains the system development life-cycle in product testing. Section 2.3 develops the concept of test optimization from an industrial perspective. Section 2.4 continues with the AI-based solutions for test optimization studied in the literature and Section 2.5 focuses on applied use cases in the industry and the measures to evaluate the performance for manufacturing optimization.

2.1 Radio Base Station

RBSs are radio transceivers of electromagnetic waves to enable mobile radio communication. In 4G and older generations mobile communication, the antenna was not explicitly included in the design of the RBSs. Unlike 5G, where the antennas are an integrated component of the product. RBSs work at high frequencies, e.g., 800 MHz for global system for mobile communications (GSM) and 40 GHz for 5G NR technologies. Figure 2.1 depicts the block diagram of a RBS. In the case of communication to the mobile phone, transmission (Tx), data from the network arrives at the RBS and then is transferred as a digital signal to the ASIC. The digital information is converted into an analog signal using digital-to-analog converters (DACs). The analog signal can then up-converted to higher frequencies, amplified in the power amplifier (PA) and filtered using the filters to remove unwanted signals before being transmitted to the field through the antennas. Conversely, the signal from the mobile phone toward the RBS, reception (Rx), follow the same principle but in reverse and with lower power levels.

The large quantity of analog components, radio frequency technology and the use of multiple software applications make the product testing of RBSs challenging. Figure 2.2 shows a general view of the test flow at Ericsson, where the TRS, a set of instructions on what and how to measure, arrives as non-formal text. The test execution follows the next sequence: general test, calibration test, and performance test before the product complies with the standards and sends to the customer. The test

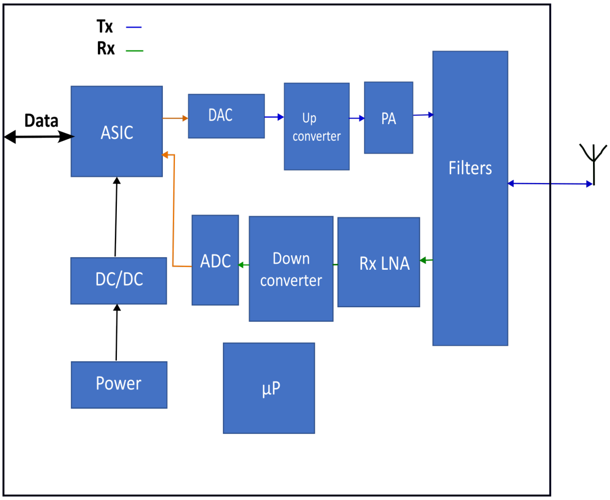

Figure 2.1: RBS Block Diagram.

sequence order for execution is up to the test developers and the empirical knowledge of the integration engineers, based on the optimal use of resources, common settings and constraints, e.g., performance test after calibration test.

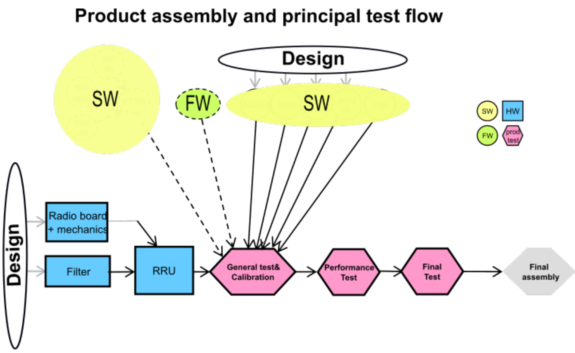

Figure 2.2: RBS test flow at Ericsson.

2.2 System Development Life Cycle

Product testing stands as an indispensable component within the system development life-cycle (SDLC). Various models or methodologies are available for adoption during the software development process, each bearing its distinct set of pros and cons. The array of software testing models encompasses the V-model, waterfall model, agile model, spiral model, and iterative model. The selection of a specific testing model should be contingent upon the products' intended outcomes and the intricacy inherent within it [7]. An overall overview of the V-model (Validation and Verification model) is followed for the SDLC, depicted in Figure 2.3. The model comprises the left side, representing the Development Life Cycle, and the right side, representing the Test Life Cycle. The V-model derives its name from the visual resemblance of the figure to the letter "V". It is an evolved version of the waterfall model, which involves sequential development and testing processes [10].

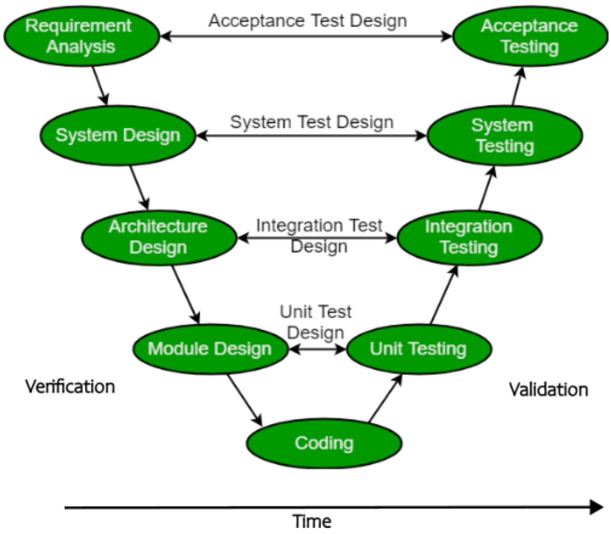

Figure 2.3: A graphical representation of the V-model.

Once the product has gone from design to implementation, the next step is integration testing. In integration testing, the already tested modules are integrated and the total system is tested to uncover the possible errors produced by interfacing those modules. Nevertheless, this view is a simplified version of the SDLC because the new technologies are more complex, demand faster responses, and are non-stationary. Therefore, the industry struggles to comply with all steps of the V-model, however, there are some principles that remain applicable today: define stakeholder requirements, design according to those requirements and validate the product according to the requirements.

2.3 Optimization of the Testing Process

Test verification and test validation are the simplified parts of an SDLC as shown in Figure 2.3. As the new applications demand optimal solutions to make efficient use of the test resources without compromising the test coverage, starting from design to the test validation processes, there is an increasing need to optimize the process, even more, to adapt to the always competitive market but still assure a good performance, which could lead to have better quality products. As of today, SDLC can take up to 50% of the total product development cost [11, 12], which makes it a good candidate to optimize. Though test automation has already been implemented during the validation part at our company through a huge effort to reduce the stable tests using statistical approaches and smart use of the resources by test planning and analysis, we believe that the AI techniques could open new possibilities to manage the large set of available data and find insightful information. Furthermore, new products will make the current solutions intractable, which verifies the need for new test optimization solutions [13].

Using AI techniques to empower the production process can improve efficiency by monitoring and predicting the quality of products. Simultaneously, it can enhance the abilities of the practitioner by showing only the important measures to focus on or sending a high-level message to the user, thus, one can decide what actions to take before resuming the testing. As has been shown in **Papers III, V**, the output performance can be improved by monitoring the normal behavior of the process and warning when an anomaly appears. On the other hand, a test case is a series of operations or events developed according to the TRSs of the stakeholder, e.g., regulator standards. Generally, the content of the test cases is a description of what and how to test the test criteria, and the corresponding output (pass or fail) together with the output results of the test. Normally, the TRS is given in natural language statements, i.e., human-readable.

Table 2.1 gives an example of a test case specification for validation testing at a radio base station (RBS) in Ericsson. The test execution result column may also contain the measurement results, e.g., 48 dBc for step 1.4. Moreover, a test suite is a group of relevant test cases arranged in a specific mode.

2.3.1 Test Case Generation

Test case generation refers to the systematic process of producing test suites that are used to assess the functionality and quality of a specific system [7], based on given criteria, i.e., model-based or specification-based criteria[14]. Model-based testing (MBT) is an advanced approach employed to generate these test suites, wherein a model representing the system under test is utilized as the basis for generating test cases [15]. MBT enables efficient and effective test case generation by leveraging the descriptive power of the model, which abstracts the system's behavior, interactions, and specifications. However, despite MBT creating useful and flexible test automation, MBT requires formal models of the system to perform the testing. On the other hand, Specification-based test case generation has the advantage of being able to generate test cases early in the development life-cycle and the errors can be caught early as well. However, specification-based test case generation is limited by the quality of the specifications [16].

Test case name	Measure adjacent power	
Test case ID	TC 1	
Test case description	Measure the Adjacent channel leakage power ratio (ACLR) of product A Configuration Procedure Pass criteria $> 45\,\mathrm{dBc}$	
Steps	**Description**	**Test Execution Result**
1.1	Send the right settings to the product	☐ Pass ☐ Fail
1.2	Set up the carrier	☐ Pass ☐ Fail
1.3	Send the right settings to the instrument to start measuring the ACLR	☐ Pass ☐ Fail
1.4	Measure the ACLR	☐ Pass ☐ Fail
1.5	Compare the results to the pass criteria	☐ Pass ☐ Fail

Table 2.1: A test case specification example.

2.3.2 Test Case Selection

Test case selection is a process to select a subset of test cases from a test suite, whose test will reveal most of the errors [8]. Here, the difference between hardware and software testing is more discernible. Our aim is to select the minimum amount of test cases for the successful production of a unit, not to reveal the faulty ones. While the principle is the same the aim is quite the opposite. Test case selection usually works together with test case prioritization, however, the literature in this particular area is concerned with independent test cases. Unfortunately, that is not the case for applications where the test cases must follow a certain sequence [13].

2.3.3 Test Suite Minimization

As aforementioned, the testing process is the most expensive part of a system development process. As systems become more complex, it becomes unfeasible to execute all test cases. Test suite minimization or test case reduction is the process of reducing the number of test cases without compromising the fault detection coverage, i.e., minimizing the testing cost in terms of execution time, resources, etc [17]. There are many techniques to approach this process, including some: heuristics, genetic algorithms, and clustering. Unfortunately, there is a trade-off between test minimization techniques and fault detection capability, thus, it is difficult to say which technique performs best [18].

The TRSs given by our domain standard regulators, dictate certain compliance levels we must follow, which will become test cases. Up till now, all test cases considered important to comply with the TRS have been generated. However, there is scope for improvement by finding and selecting the most relevant test cases for a fault-free product during the validation process.

2.3.4 Test Case Scheduling

This process is the execution of the minimum number of selected test cases in the best possible sequence to optimize the testing process [19]. Generally, test case scheduling is done by monitoring the test results [20]. At the beginning of a project, test developers write the first test plan (test suite) covering all points given at the TRSs. The test plan is based on both, the TRSs and the infrastructure constraints. Once the project is verified and the product is implemented and unit tested, the integration test can be done by monitoring the test results and applying different AI approaches, then, the test case scheduling can be updated accordingly [8].

2.3.5 Automatic Data Analysis

Data mining techniques can be used to extract knowledge, either automatically or semi-automatically, from large amounts of data and make it visually easy to decipher by our human understanding. Generally, data mining is entirely machine-driven to discover automatically hidden patterns or find models out of data. In traditional data mining, the results are often presented as simple tables with no interface to the practitioners. Previous works in the literature have studied the possibility of improving knowledge discovery by using feedback loops [21]. Namely, the models can be enhanced by the feedback of the users, who may be able to tune the parameters of the models or change the algorithms according to their domain knowledge and make better sense of the findings. Currently, these kinds of interactions are practically impossible.

One of the points studied in this thesis is the automatic data analysis of test results to find dependencies between test cases and redundant test cases, thus, finding when exactly some test cases start showing abnormal behavior before the next test case fails due to non-compliance with the fixed limits.

2.3.6 Verification and Validation Testing

Verification and validation are the methods for confirming a system complies with its design specifications and meets its intended purpose respectively. In other words, verification is a quality control process and validation is a quality assurance process. We will use these terms differently throughout this thesis since we are talking about different processes. As systems become more complex and are designed by large engineering teams across the world, they require robust methodologies in terms of design verification and validation such as the V-model for verification of complex engineering products [22]. However, for physically complex products, the validation part is also important because hardware and software require interaction between different interfaces, i.e. integration testing. Therefore, we follow the V-model for product testing in this thesis.

2.4 AI-Based Solutions for Test Optimization

AI applications have shown to be advantageous due to the fast learning progress and the ability to learn from data [23]. Automated analysis and monitoring tools based on AI are desired to enhance human capabilities. There are different test optimization methods regarding intelligent manufacturing and the ones studied in this thesis are described below.

2.4.1 Fault Prediction

Fault detection and diagnosis has been studied for many years in machine and industrial processes and its functions are monitoring the process and detecting and analyzing the real source of errors [24]. On the other hand, fault prediction aims to predict possible failures based on different approaches, knowledge-driven, data-driven, and value-driven [25]. Where new fault prediction methods need to handle big data produced in complex industrial applications because the gathered data has evolved and increased.

Fault diagnosis and prognosis are terms used mostly in deployed systems, e.g., real-world scenarios, once the systems are already operational, and aim to find the source of error after it has already occurred.

Despite the research community has studied extensively methodologies that aim for fault-free products, fully applied fault-prediction is not achieved at the industrial scale. There may be some reasons behind this statement, such as compatibility, scalability, lack of generalizability (tailored made solutions), unclear performance visibility, non-continuous delivering environment, and the limitations due to non-disclosure data set, among others [26].

In this licentiate thesis, fault prediction and fault prognosis concepts will be used together because the prediction of future outcomes in the testing process using historical data is the target.

2.4.2 Test Case Dependencies

There are many factors that can contribute to test case dependencies, e.g., execution order, shared resources, and functional dependencies (data and prerequisites). A good understanding of this matter is very important for test planning and effective resource allocation and could enhance the test optimization process. There are different approaches to find dependencies between test cases and this thesis will focus on the AI applications given in the literature. Among other approaches, natural language processing (NLP) can extract meaningful information that implies dependencies from test descriptions written in natural text [27], machine learning (ML) techniques can learn from historical data and find features to predict dependencies between test cases, and trace analysis can identify and uncover implicit correlation between test cases. During **Papers I, II, IV**, AI techniques are analyzed to find dependencies between test cases in the test case generation and the test debugging process. The papers explain some challenges these approaches may have due to imbalance learning and the need for manual labeling based on domain knowledge since there is no ground truth which can confirm test case dependencies.

2.4.3 Automated Fault Localization

Test time optimization during troubleshooting also includes the time a practitioner takes to find the real source of an error. Fault diagnosis is the process of identifying, isolating, and estimating a fault once it has occurred using systematic analysis. This analysis usually contains data gathering, deep analysis, fault localization, and fault repair. All of these activities are time-consuming and difficult to implement, therefore automating these processes, such as fault localization, is desired. According to the literature, automated fault localization implies to *automatically and accurately identifying the cause of the failure in order to provide the developer with a report that pinpoints the parts of the code that need to be repaired.*. Unfortunately, long dependence chains between different variables of the complex systems, make it impossible to comply with this definition. Therefore, the interpretation of the previous ideal definition of automated fault localization has been modified to *identify some program event(s) or state(s) that cause or correlate with the failure in order to provide the developer with a report that will aid fault repair.* [28].

However, in many applications, it is difficult to distinguish the real cause of an error since it can be the successive or congruence of tests that cause the error. One technique to implement automated-based localization is based on test-driven techniques such as dependence analysis, causal inference, and information theory. In this work, we have focused on all three techniques mentioned above which will be developed in **Paper IV**.

2.5 AI-Based Solutions Applied to Use Cases in the Industry

AI is extensively employed across various industrial domains to handle multiple tasks simultaneously, tackle intricate challenges, automate workflows, derive meaningful insights from data, and augment decision-making processes. This section outlines the applications of AI techniques in the industry and how it is revolutionizing factories. Specifically, it delves into AI solutions adopted in manufacturing and supply chain operations, including predictive maintenance, quality control, and production optimization.

Within the manufacturing and supply chain context, AI is harnessed to optimize existing production lines through two main approaches: enhancing product quality and improving the production process itself. In this thesis, the focus will be on investigating AI-driven methods to achieve both product quality enhancement and production optimization.

The innovation of using AI is that everything starts with observations instead of a theory or assuming how things work or should work. In this way, we could discover new knowledge and new means to solve problems on a much larger automated scale than ever before, because these technologies also allow us to handle large amounts of data. Despite, the aforementioned advantages, there are instead some challenges that have limited the implementation of these solutions in the industry which need to be addressed, e.g., poor data quality, unstructured data, lack of level of expertise to scale up AI solutions, cyber-security issues, etc. Furthermore, a study by Roland Berger shows that improvements in AI in the utilities industry have been slow [29]. Only 23% of survey respondents said they have a clearly defined AI strategy, and just 17% describe their data availability as good.

Product Quality and Production Optimization

Though the application of AI techniques is broadly adapted in other domains, it is not well established in the production quality context. Krauß et al. [30] studied a tangible example at a German manufacturing company. To predict the product quality, the authors follow the CRISP-DM (cross-industry standard process for data mining) and explain that the chain process shows how advantageous would be to discover and predict coming steps which will run out of specifications. They also highlight the necessity to have a support system or decision-making tool (DMT) to help in the selection of the optimal solutions to the specific applications. However, they emphasized that domain expertise will not be replaced [30].

On the other hand, Maschler et al. [31] discuss why despite the huge potential of AI techniques in automation, there are only a few available examples of the application, among others, in real-world predictive quality for manufacturing. Using AI methods in predictive quality in manufacturing is challenging due to the continuous changes in the process. The authors studied four use cases to find the reason for this problem and propose industrial transfer learning, a combination of continuous learning, transfer learning (TL), and the capability of "learning without forgetting". However, this solution would imply technical and organizational changes.

Though the potential of using AI in optimization would be enormous, the literature probes the contrary. Weichert et al. [32] studied the correlation between the machine learning approaches, the data properties, the optimizer, and the applications regarding the optimization of production processes. The authors came to the conclusion that such correlation does not exist and there is not a clear connection between the complexity of the data and the complexity of the problem, which in turn can cause certain skepticism to use machine learning techniques in the manufacturing industry. Therefore, it is of vital importance to understand to nature of the problem and know the data type and the application to venture into knowledge discovery by using AI techniques (machine learning and data mining) [33]. The authors discussed how relevant is to have ML knowledge and domain expertise to find the most optimal solution to the selected application, hence, it is advisable a collaborate between both domains. In this thesis, we have combined the previous manufacturing and product knowledge and AI exploration and we can coincide with authors in the literature.

Chapter 3
Summary of Contributions

This chapter encapsulates the investigation detailed within the enclosed papers and furnishes responses to the research inquiries that directed the efforts of this thesis. Summaries of the studies were given in Section 1.3. Elaboration on the experimental data and outcomes is expounded upon within the papers themselves. The pivotal points and valuable inputs of the five papers are delineated as follows:

- **Paper I** investigates the automatic test case generation from test case descriptions given in natural language. It clusters test cases by the semantic similarities using Levenshtein distance and empirical thresholds. This approach confirms the possibility of using ML and editing distance metrics for test scheduling and parallel test execution purposes.

- **Paper II** extends Paper I, where more advanced NLP methods are used and compared. For this study, Word2vec obtained better performance compared to SBERT. Furthermore, a data-driven threshold is found using the Levenshtein distance of the clustered groups which indicates that similar test cases have lower thresholds than the empirical value defined in the previous paper.

- **Paper III** studies the final test yield as the most important production performance measure, which is directly influenced by the given fixed test limits generated from the test specifications. This paper models the final test yield (FY) using historical data and regression methods and monitors the arriving inputs (test cases), which might cause yield problems. Moreover, it advises the practitioner when divergent points surpass dynamic limits, hence, better decision-making actions can be done before an error occurs.

- **Paper IV** studies the fault diagnosis as a process to find an error once has occurred and repair the damage. Fault localization is the process of finding the source of error. This paper studies the test case dependencies, something that *Study 1* and *Study 2* did not discuss. This paper works with test measurement results to cluster associated test cases and reveals the relevant and redundant features given their dependency degrees. It studies the test case scheduling, features important, and test suite minimization for test optimization.

- **Paper V** extends Paper III, where only FY was modeled and monitored, and complements *Study 2*. In this paper, the test cases themselves are studied as time-series data, and the data-driven thresholds are designed to label new incoming tests to predict and avoid future failures. Transfer learning (TL) is used to confirm good harmony between theoretical and empirical approaches.

In the following sections, the research questions which initialize this work are answered and mapped to the papers, detailed insights are explained, and some challenges found on the search for test time optimization for the next radio generation production are mentioned. Some generalizations throughout the algorithms and the TL approaches are addressed, however, some limitations are mentioned here, which will be developed in detail in Section 4.2.

3.1 Research Question 1

The first research question focused on understanding the applicability of AI techniques in the manufacturing process of complex systems: *To what extent can AI techniques optimize and perform fault detection and prediction on the production test of RBSs?*

The first step was the literature review of the current AI techniques used in industrial applications. Despite, the multiple review papers[32], [34], which bring insightful information on AI techniques used in optimization for industrial applications, the conclusions are nearly the same, there are enough algorithms but the implementation in the industry is challenging [33]. As aforementioned, traditional solutions, such as Six Sigma (6σ) and sample testing cannot cope with nowadays complex, multi-modal, and heterogeneous data. The new solutions for production process optimization, in addition, need to be able to guide the practitioners, suggest the best applicable algorithm (depending on the data and application), and help with the interpretation of the results. Since our application optimizes the existing production process, there are a number of constraints found during the journey. We have seen some of them in our studies, however, we believe further development (time) will help to alleviate the introduction of new solutions for real-time processing and modeling. First, industrial application data is often unstructured. To apply AI techniques, there are some requirements the data needs to have before pre-processing to avoid manual selection. Because our data set contains several thousands of test cases, manual labeling is not a doable option. For instance, each test case was manually named. The difficulties arise when this task is executed by teams around the world, where not everyone uses the same terminology to name the same test cases, e.g., BrA, Br1, PortA, and Port1 may all refer to radio port 1. Thus, to have a complete automated data gathering as part of pre-processing is not feasible to this point and may need manual labeling each time a new product revision arrives.

In order to find the most important features or test cases to predict and guarantee a fault-free product, the whole process must be studied as such. The first step is to recognize the type of data inputs before suggesting pre-processing and learning options. This step is very important since there are tests of different data types in a radio product. According to the data type, some algorithms can be suggested to the user.

Studies 1-3 aim to find AI techniques that can enhance the total test optimization, this means, an efficient usage of the resources needed to test radio products from different generations. Among these resources, in this thesis, we focused on four: 1-test time, 2-test infrastructure, 3-test case generation, and 4-test debugging analysis.

Study 3 investigates the usage of automated fault localization by analyzing the test case dependencies for test suite minimization using data from a 5G radio product. The results show that it is possible to find unique test cases which can describe the future behavior of dependent test cases, and therefore, predict future outcomes (**Paper IV**). These conclusions may help in test case scheduling and suggest a test plan containing only the relevant test cases. Furthermore, this prototype can be used in the test debugging process, where the method shows a list of the relevant and redundant test cases to the data analysts, hence, reducing the amount of manual work to find the source of error. The advantage of this method is that the learning process is executed once and can find higher-order dependencies than only linear dependencies.

- RQ1.1:

  Study 2 researched the possibility of modeling important features using historical data and AI algorithms. Furthermore, the data inputs (test cases) can be automatically labeled as normal or abnormal and predict the final performance, hence, finding possible problems beforehand. In this study, all test cases are equally important and there is no search to minimize the test suite. **Paper III** used support vector machines (SVMs) to classify the inputs after they have been labeled with the purpose of predicting and avoiding yield drops. The final yield (FY) was modeled using several regression methods, of which XGBoost obtained the best performance for the studied data set. This paper proposed an interesting solution to predict an important performance measure, such as FY. It normalizes, scales, removes noise, and balances the inputs before classifying them as normal or abnormal, which makes the approach of this paper a flexible tool and can be used in other applications. Moreover, the use of data-driven thresholds was attempted by following the AI-modeled data. This solution is a reasonable improvement to statistical approaches being still used by the industry. **Paper V** extends Paper III and focuses on the test cases themselves, the SVM is replaced by CNN which can learn more advanced features of the anomalies and can be used as a pre-trained model and improve learning performance as new data arrives.

- RQ1.2:

  On the other hand, test optimization of the SDLC can be reached by automatic test case generation, consequently, it can save time and testing resources. Currently, test case generation in the industry is often done manually from the TRSs.

However, TRSs come in the natural text to be interpreted by experts in the subject. As is known, this process is time-consuming, often incomplete, and follows the waterfall model, which is not tractable in all applications [35]. To facilitate this process, *Study 1* used NLP techniques to analyze and cluster similar test cases based on their similarities for more effective test execution, hence, reducing and automating the manual test case generation and test case scheduling. We show our first results in **Paper I**, where the test descriptions for five different 4G RBSs were studied. The goal was to develop a more efficient test execution with the purpose of minimizing the total testing costs. This paper addressed a very important matter studied within hardware and software fields but it is in the hardware domain which may find limited deployment due to two reasons: First, the TRSs give hardly enough detailed information to cluster the test cases in groups based on their semantic similarities and common settings for parallel test execution. Second the dependencies between test cases will make this approach more difficult to implement because, for real-world applications, the order of execution is often of vital importance[36].

Paper II extends the study of the latter paper and compares two more advanced NLP techniques on the same data set. Doc2vec and SBERT are techniques that not only compare edit distance metrics but also semantic similarities between phrases or even documents. Besides, this paper finds a data-driven threshold from the clustered groups. This is an improvement to the findings of Paper I, which needed an empirical threshold to distinguish between similar or non-similar test cases. This is the next level to further automate the test case generation, from TRS given in non-formal natural text, into groups that share similar tasks, settings, or artifact usage. The application of the findings of *Study 1* are many, such as text mining, sentiment analysis, and test log analysis.

3.2 Research Question 2

The second research question addressed in this thesis concerns the usage of TL of the studied algorithms on different products: *How do we modify learning algorithms in order to do transfer learning between different RBS versions and generations?*

TL technique reuses a model trained in one task into another similar task, thus, improving generalization. It is useful in cases when there is not enough data and still good performance is required. For the case of anomaly detection, where data scarcity is a problem, TL the same definition can be used.

Study 2 contemplates the possibility of transferring the knowledge of the learned models between different tasks, e.g., different radio generations and different test cases. **Paper III** used the results obtained after training the classifier from different data sets from several 4G radio products and testing on a 5G product. At the time of the study, 5G was an emerging technology, and not enough production data was available. The first assumption was that both technologies carry sufficient similarity between them, hence, knowledge transfer would be feasible. The results show high performance in the unseen data, which opens the possibility of using the same model, without any modification, in a similar kind of data. Furthermore, this solution is also flexible for different dynamic thresholds and it scales and normalizes all inputs, making it ideal for different real-world applications. Nevertheless, this paper does not analyze the influences of dependencies and non-normal distribution of the input variables.

On the other hand, this study also aims to understand to which point TL is possible by using other methods than classical ML, such as deep learning. The different features tested in each product make this study challenging since the inputs are multimodal(from different sources) and heterogeneous (numerical, categorical, Boolean). Besides, there is a certain dependency between consecutive data points tested in the same infrastructure (time-series data), which makes SMOTE used in previous work [see: **Paper III**], not suitable for time-series data. **Paper V** used one-dimensional CNN to learn the anomalies in one test case and then tuned and tested in another correlated test case. The results show that TL has good performance after little tuning, but for cases where the anomalies overlap or the dynamic limits are too close, the results are not as good as expected. A remark is that the CNN was implemented from scratch since the data set comes from an internal database within the company and the anomaly features might be different from other standard anomalies.

Study 2 concludes confirming that TL is possible between different radio generations and between correlated test cases, with none or barely small changes. This was exactly what we were seeking for the online-mode application since there is no need to train the model every time new data arrive, merely change some outer layers (task-specific) and execute the classification, which in turn will require much less training data and time, and less computing resources.

To improve generalizability, the challenges are many, whether the target task keeps the same data distribution as the source task, or whether they have sufficient correlation. Otherwise, the performance of the TL will show some limitations in cases of over-fitting and negative transfer (domain mismatch). Among other interesting themes about TL, Niu et al. [37] discuss the open challenges of TL at the algorithm and application levels, which may limit the usage of this technology in both industry and academia, whose conclusions only serve to confirm our findings of *Study 2*.

Chapter 4
Conclusions and Future Work

This section discusses possible future directions of this study and explains some concepts we believe can limit the application of this thesis. Besides, it is considered important to emphasize the ethical impacts this work may have in society due to the use of AI techniques and this chapter concludes with some final words.

4.1 Future Research Directions

In the sight of imminent AI advances in every known field, this technology will become pervasive in people's lives. Therefore, the literature advises enterprises to use this technology to remain competitive in the always-changing market.

In the future, we aim to continue studying fault prediction to enhance production efficiency. Up to now in the learned models, we have used batched data to model and update the parameters based on newly accumulated data. In the future and when our database allows it, we plan to use stream data (real-time) processing to model and update the learned model to benefit real-time decision-making.

For product quality, the study will be focused on the different kinds of anomalies and not only the ones outside certain thresholds, i.e., additive outliers [38]. It would be interesting to know to what extent CNNs can learn to represent all important features of other kinds of anomalies in our process and how this can improve production efficiency.

For the automated fault localization, we have used association rule-mining (ARM). For future studies, we would like to study parallel sequential pattern mining (PSPM). Thus, manufacturers can respond to time-related situations and predict future activities without being bound by big data properties and the usage of resources [39]. Furthermore, we would like to investigate how the production performance will be influenced by executing only the relevant features.

4.2 Limitations

There were some assumptions taken during the studies and development of this thesis, which we consider worth mentioning here. These were discussed early during this thesis and are written in specific sections in the papers.

First, all the studies are based on data from our internal database at Ericsson AB from 4G and 5G radio products. We did not make any further comparison with other data sets due to the applications were designed specifically for industrial use cases within Ericsson AB. To this point, I consider our findings can be applicable to the other kinds of data for the following reasons. The solution approaches studied in this thesis were intended for working in an online environment, which requires flexible, scalable, and low computational resources. However, we did not make any comparison and it is difficult to generalize the results due to the limited data set used during the studies.

Second, the assumption of data normally distributed might change the performance numbers if this assumption is not fulfilled. For instance, in *Study 2* the inputs were normalized because of heterogeneous data. In *Study 3* correlation-based methods and Sturges' rule (histogram bin size for mutual information) assume data inputs are normally distributed.

Third, smoothing methods were used in *Study 2* for noise removal. There have been discussions that smoothing methods are good for reducing false positives in anomaly detection cases [40]. However, we are not aware this solution will last in cases where the smoothing processes remove relevant information.

Fourth, *Study 3* has some limitations regarding the priority of the results given in the form of lists, how to decide which test cases have more priority or are influencing the most the final outcome? This study assumes the first row in the list contains the most important test case to execute since we usually follow sequential test execution and neglect the next test cases which are assumed redundant. After further analysis, some results were influenced by a combination of test cases (not only the first one), but the weight and the order were not analyzed.

Finally, there have been used empirical thresholds along the studies, given by the subject matter expert (SME) based on the experiences due to the lack of ground truth. Despite the long experience working in the field, those values are entirely based on previous knowledge and can be subject to discussion.

4.3 Societal and Ethical Impacts

There are three potential societal impacts on the current research area that we are focusing on in this thesis. First, the automation process aims to make better use of the resources by minimizing the number of test executions done by AI methods, which base their insights on probabilities. One of the possible consequences could be that we deliver faulty products, which do not comply entirely with the system and then they may cause a real problem in sensitive areas where good performance is not sufficient, e.g., where either the life or the safety of a person is at risk. The second potential impact is the loss of many qualified jobs, where manual analysis and implementation of these test cases will require fewer job positions, which in turn, will influence the job market and then society in general. The aim of my research is to automate as much as possible the test case generation, test case execution, and test case analysis of the outputs by using AI-assisted methods, which as the literature shows, are faster and have more reliable performance than manual processes in fields where repetitive tasks are done. As McKinsey report published in 2016, the most automatable activities are the ones that do not require any creativity and perform their activities in predictable environments, the technical feasibility for automation in those jobs is up to 78% [41]. It is believed that repetitive tasks can be exhausting, prone to errors, and sometimes endanger our health. Moreover, repetitive work can cause the technician not to pay enough attention to some specific tasks due to the monotonous repetition which can negatively influence the output. This brings up the third possible impact on society, total surveillance. Until what point are the co-workers willing to give their information to increase production efficiency in the factory? It might be necessary to have cameras, sensors, and other kinds of trackers to help in the automation process. Thus, the production can be monitored for changes and then the algorithms update accordingly. In that case, special care must be taken in the data gathering, where non-relevant information (voice, video) is recorded and then processed as part of the learning process. We are not there yet but I can assume that data can be safely saved internally, although, careful proceedings must be taken to mitigate the risk of data leakage.

4.4 Final Words

The focus of this thesis revolves around optimizing test time in the production of 5G radio networks. These networks comprise advanced components, such as digital, analog, high and low-frequency components, as well as software. The effectiveness of production performance can be gauged by the yield performance, which is directly proportional to the number of successfully produced units and inversely proportional to the number of retested and failed units. To achieve favorable yield levels, meticulous attention is paid to the design, verification, integration, and deployment of all stages in the product cycle. This process commences with test requirement specifications and concludes with the installed product, carrying forward into the maintenance phase. Consequently, cost and time optimization are crucial objectives for the industry to maintain competitiveness. The advent of AI techniques has opened up numerous possibilities to leverage advanced learning methods for identifying patterns, modeling processes based on historical data, and classifying features based on shared characteristics. This thesis harnesses AI techniques and applies them to the internal production data of 4G and 5G products, thereby optimizing production methods at Ericsson AB. Additionally, it involves developing monitoring tools to augment existing solutions. I encourage further research towards more automated production processes with constant monitoring and advanced prediction methods to break the bridge between theory and practice, as well as close cooperation between industry and academia to design insightful tools in the ever-changing telecommunication market.

References

[1] B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, and T. Freiheit, “Smart manufacturing and intelligent manufacturing: A comparative review,” *Engineering*, vol. 7, no. 6, pp. 738–757, 2021. (Cited on page 1.)

[2] K.-D. Thoben, S. Wiesner, and T. Wuest, “"industrie 4.0" and smart manufacturing - a review of research issues and application examples,” *International Journal of Automation Technology*, vol. 11, pp. 4–19, 01 2017. (Cited on page 1.)

[3] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, “Machine learning for industrial applications: A comprehensive literature review,” *Expert Systems with Applications*, vol. 175, p. 114820, 2021. (Cited on page 1.)

[4] 3GPP, “Group Radio Access Network; NR; Base Station (BS) conformance testing Part 1: Conducted conformance testing,” Technical Specification (TS) 38.141-1, 3rd Generation Partnership Project (3GPP), 06 2023. Version 18.2.0. (Cited on page 1.)

[5] C. Zhang, P. Patras, and H. Haddadi, “Deep learning in mobile and wireless networking: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019. (Cited on page 2.)

[6] M. Fayazi, Z. Colter, E. Afshari, and R. Dreslinski, “Applications of artificial intelligence on the modeling and optimization for analog and mixed-signal circuits: A review,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 6, pp. 2418–2431, 2021. (Cited on page 2.)

[7] S. Tahvili and L. Hatvani, *Artificial Intelligence Methods for Optimization of the Software Testing Process With Practical Examples and Exercises*. Elsevier, July 2022. (Cited on pages 3, 13, and 14.)

[8] S. Tahvili, *Multi-Criteria Optimization of System Integration Testing*. PhD thesis, Mälardalen University, December 2018. (Cited on pages 6 and 16.)

[9] T. Gorschek, P. Garre, S. Larsson, and C. Wohlin, “A model for technology transfer in practice,” *IEEE Software*, vol. 23, no. 6, pp. 88–95, 2006. (Cited on pages 6 and 7.)

[10] S. Balaji and M. S. Murugaiyan, "Waterfall vs. v-model vs. agile: A comparative study on sdlc," *International Journal of Information Technology and Business Management*, vol. 2, no. 1, pp. 26–30, 2012. (Cited on page 13.)

[11] M. Pezzè and M. Young, *Software Testing and Analysis: Process, Principles, and Techniques*. Wiley India Pvt. Limited, 2008. (Cited on page 14.)

[12] S. Tahvili, M. Bohlin, M. Saadatmand, S. Larsson, W. Afzal, and D. Sundmark, "Cost-benefit analysis of using dependency knowledge at integration testing," in *The 17th International Conference On Product-Focused Software Process Improvement*, 2016. (Cited on page 14.)

[13] M. Felderer, E. P. Enoiu, and S. Tahvili, *Artificial Intelligence Techniques in System Testing*, pp. 221–240. Singapore: Springer Nature Singapore, 2023. (Cited on pages 14 and 16.)

[14] M. Prasanna, S. Sivanandam, R. Venkatesan, and R. Sundarrajan, "A survey on automatic test case generation," *Academic Open Internet Journal*, vol. 15, no. 6, 2005. (Cited on page 14.)

[15] M. Camilli, R. Mirandola, P. Scandurra, and C. Trubiani, "Towards online testing under uncertainty using model-based reinforcement learning," in *European Conference on Software Architecture*, pp. 233–245, Springer, 2022. (Cited on page 14.)

[16] D. Richardson, O. O'Malley, and C. Tittle, "Approaches to specification-based testing," *SIGSOFT Softw. Eng. Notes*, vol. 14, pp. 86–96, nov 1989. (Cited on page 14.)

[17] S. Wang, S. Ali, and A. Gotlieb, "Cost-effective test suite minimization in product lines using search techniques," *Journal of Systems and Software*, vol. 103, pp. 370–391, 2015. (Cited on page 16.)

[18] R. Singh and M. Santosh, "Test case minimization techniques: a review," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 12, 2013. (Cited on page 16.)

[19] S. Arlt, T. Morciniec, A. Podelski, and S. Wagner, "If a fails, can b still succeed? inferring dependencies between test results in automotive system testing," in *2015 IEEE 8th International Conference on Software Testing, Verification and Validation (ICST)*, pp. 1–10, IEEE, 2015. (Cited on page 16.)

[20] C. Lyu, S. Ji, C. Zhang, Y. Li, W.-H. Lee, Y. Song, and R. Beyah, "{MOPT}: Optimized mutation scheduling for fuzzers," in *28th USENIX Security Symposium (USENIX Security 19)*, pp. 1949–1966, 2019. (Cited on page 16.)

[21] E. Bertini and D. Lalanne, “Surveying the complementary role of automatic data analysis and visualization in knowledge discovery,” in *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration*, VAKD ’09, (New York, NY, USA), pp. 12–20, Association for Computing Machinery, 2009. (Cited on page 17.)

[22] P. G. Maropoulos and D. Ceglarek, “Design verification and validation in product lifecycle,” *CIRP Annals*, vol. 59, no. 2, pp. 740–759, 2010. (Cited on page 17.)

[23] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: advantages, challenges, and applications,” *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23–45, 2016. (Cited on page 17.)

[24] Y.-J. Park, S.-K. S. Fan, and C.-Y. Hsu, “A review on fault detection and process diagnostics in industrial processes,” *Processes*, vol. 8, no. 9, p. 1123, 2020. (Cited on page 17.)

[25] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, “Industrial big data for fault diagnosis: Taxonomy, review, and applications,” *IEEE Access*, vol. 5, pp. 17368–17380, 2017. (Cited on page 17.)

[26] E. J. Weyuker, T. J. Ostrand, and R. M. Bell, “Comparing the effectiveness of several modeling methods for fault prediction,” *Empirical Software Engineering*, vol. 15, pp. 277–295, 2010. (Cited on page 18.)

[27] S. Tahvili, L. Hatvani, E. Ramentol, R. Pimentel, W. Afzal, and F. Herrera, “A novel methodology to classify test cases using natural language processing and imbalanced learning,” *Engineering Applications of Artificial Intelligence*, vol. 95, pp. 1–13, August 2020. (Cited on page 18.)

[28] W. Masri, “Chapter three - automated fault localization: Advances and challenges,” in *Advances in Software Engineering*, vol. 99, pp. 103–156, Elsevier, 2015. (Cited on page 19.)

[29] R. B. GMBH, “Artificial intelligence: A smart move for utilities,” 2018. [Online]. Available from: https://www.rolandberger.com/en/Media/Energy-Sector-Just-23-Percent-of-Utilities-Have-a-Clearly-Defined-Strategy-for.html 2023.09.01. (Cited on page 19.)

[30] J. Kraus, M. Frye, G. T. D. Beck, and R. H. Schmitt, “Selection and application of machine learning-algorithms in production quality,” in *Machine Learning for Cyber Physical Systems: Selected papers from the International Conference ML4CPS 2018*, pp. 46–57, Springer, 2019. (Cited on pages 19 and 20.)

[31] B. Maschler, H. Vietz, H. Tercan, C. Bitter, T. Meisen, and M. Weyrich, "Insights and example use cases on industrial transfer learning," *Procedia CIRP*, vol. 107, pp. 511–516, 2022. (Cited on page 20.)

[32] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, and S. Wrobel, "A review of machine learning for the optimization of production processes," *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 5-8, pp. 1889–1902, 2019. (Cited on pages 20 and 22.)

[33] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, p. 114060, 2021. (Cited on pages 20 and 22.)

[34] G. Koksal, I. Batmaz, and M. C. Testik, "A review of data mining applications for quality improvement in manufacturing industry," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13448–13467, 2011. (Cited on page 22.)

[35] S. Balaji and M. S. Murugaiyan, "Waterfall vs. v-model vs. agile: A comparative study on sdlc," *International Journal of Information Technology and Business Management*, vol. 2, no. 1, pp. 26–30, 2012. (Cited on page 24.)

[36] S. Tahvili, L. Hatvani, M. Felderer, W. Afzal, and M. Bohlin, "Automated functional dependency detection between test cases using doc2vec and clustering," in *IEEE International Conference on Artificial Intelligence Testing*, 2019. (Cited on page 24.)

[37] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010 - 2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020. (Cited on page 25.)

[38] T. Wen and R. Keyes, "Time series anomaly detection using convolutional neural networks and transfer learning," *ArXiv*, vol. abs-1905.13628, 2019. (Cited on page 27.)

[39] W. Gan, J. C.-W. Lin, P. Fournier-Viger, H.-C. Chao, and P. S. Yu, "A survey of parallel sequential pattern mining," *ACM Trans. Knowl. Discov. Data*, vol. 13, jun 2019. (Cited on page 27.)

[40] M. Grill, T. Pevny, and M. Rehak, "Reducing false positives of network anomaly detection by local adaptive multivariate smoothing," *Journal of Computer and System Sciences*, vol. 83, no. 1, pp. 43–57, 2017. (Cited on page 28.)

[41] M. Xu, J. M. David, S. H. Kim, *et al.*, "The fourth industrial revolution: Opportunities and challenges," *International journal of financial research*, vol. 9, no. 2, pp. 90–95, 2018. (Cited on page 29.)