IntechOpen

IntechOpen Series
Industrial Engineering and Management,
Volume 12

Quality Control
Artificial Intelligence, Big Data, and New Trends

Edited by Sayyad Zahid Qamar, Nasr Al-Hinai, Sandeep Kumar, Shilpa Choudhary, Arpit Jain and Ankita Tiwari

Quality Control - Artificial Intelligence, Big Data, and New Trends

Edited by Sayyad Zahid Qamar, Nasr Al-Hinai, Sandeep Kumar, Shilpa Choudhary, Arpit Jain and Ankita Tiwari

Published in London, United Kingdom

Quality Control - Artificial Intelligence, Big Data, and New Trends
http://dx.doi.org/10.5772/intechopen.1006236
Edited by Sayyad Zahid Qamar, Nasr Al-Hinai, Sandeep Kumar, Shilpa Choudhary, Arpit Jain and Ankita Tiwari

Contributors
Abu Sarwar Zamani, Adisa Hasković Džubur, Alem Čolaković, Antonio Pérez-Torres, Ayoub Mhaouch, Bakir Karahodža, Jan Kamenický, Jose Costas, Luis E. Baquero-Rey, Marwa Fradi, Md. Mobin Akhtar, Miguel Hernández Bejarano, Mohsen Machhout, Nayyar Ahmed Khan, Rafael Pastor, Raj Bano Khan, René-Vinicio Sánchez, Sajjad Muhammad Khan, Sayyad Zahid Qamar, Sufyan Muhammad Khan, Susana Barceló-Cerdá, Věra Pelantová, Wafa Gtifa, Waqas Muhammad Khan

First published in London, United Kingdom, 2025 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Quality Control - Artificial Intelligence, Big Data, and New Trends
Edited by Sayyad Zahid Qamar, Nasr Al-Hinai, Sandeep Kumar, Shilpa Choudhary, Arpit Jain and Ankita Tiwari
p. cm.

This title is part of the Industrial Engineering and Management Book Series, Volume 12
Topic: Operational Excellence
Series Editor: Fausto Pedro Garcia Marquez
Topic Editor: Stuart So

Print ISBN 978-1-83634-536-7
Online ISBN 978-1-83634-535-0
eBook (PDF) ISBN 978-1-83634-537-4
ISSN 3029-0511

If disposing of this product, please recycle the paper responsibly.

IntechOpen Book Series
Industrial Engineering and Management
Volume 12

Aims and Scope of the Series

Industrial Engineering and Management (IEM) is a discipline that focuses on optimizing complex processes and systems within various industries. It involves the integration of engineering, business, economics, mathematics, and behavioral sciences to improve efficiency, productivity, quality, and overall performance in organizations. Key aspects of Industrial Engineering and Management include: Process Optimization; System Analysis and Design; Quality Control and Management; Supply Chain Management; Operations Management; Human Factors and Ergonomics; Project Management; Cost Analysis and Financial Management; Decision Analysis.

Overall, Industrial Engineering and Management aims to optimize resources, improve processes, enhance productivity, and ensure the effective and efficient utilization of all elements involved in the production or delivery of goods and services. It is crucial in today's competitive business environment for organizations to stay efficient and competitive.

Production Engineering and Operational Excellence are fields of study and practices that focus on optimizing and improving the manufacturing and production processes within an organization. It combines principles from engineering, management, and operational strategies to enhance productivity, efficiency, quality, safety, and sustainability in the production of goods and services.

Here are the key components of Production Engineering and Operational Excellence: Process Optimization; Operational Excellence; Manufacturing Systems Design; Quality Management; Supply Chain Optimization; Production Planning and Scheduling; Automation and Technology Integration; Health, Safety, and Environmental Management; Cost Management; Performance Measurement and Key Performance Indicators (KPIs); Continuous Improvement and Innovation.

Production Engineering and Operational Excellence are crucial for organizations aiming to stay competitive in the global market by achieving high levels of efficiency, quality, and customer satisfaction while optimizing resources and minimizing waste. It is a multidisciplinary approach that encompasses engineering principles, management strategies, and the effective use of technology to drive operational success.

Meet the Series Editor

Fausto Pedro Garcia Marquez is a Full Professor at UCLM, Spain, with accreditation since 2013. He also holds the position of Honorary Senior Research Fellow at Birmingham University, UK, and serves as a Lecturer at the Postgraduate European Institute. In addition to these roles, Fausto has experience as a Senior Manager at Accenture from 2013 to 2014. He earned his European Ph.D. with the highest distinction. Throughout his career, Fausto has received numerous awards and honors. These include the Nominate Prize (2022), Gran Maestre (2022), Grand Prize (2021), Runner Prize (2020), and Advancement Prize (2018), as well as Runner (2015), Advancement (2013), and Silver (2012) by the International Society of Management Science and Engineering Management (ICMSEM). He was also the recipient of the First International Business Ideas Competition 2017 Award. Fausto's contributions extend to academic publishing, with over 242 papers to his name. Notably, his work has been recognized in journals like "Applied Energy" (Q1, IF 9.746, Best Paper 2020) and "Renewable Energy" (Q1, IF 8.001, Best Paper 2014). His affiliations include the editorial and authorship roles in more than 50 books, with publications through respected publishers such as Elsevier, Springer, Pearson, Mc-GrawHill, IntechOpen, IGI, Marcombo, and AlfaOmega. He has authored over 100 international chapters and holds 6 patents. Fausto serves as the Editor of 5 International Journals and is a Committee Member for more than 70 International Conferences. His research portfolio encompasses being the Principal Investigator in 4 European Projects, 8 National Projects, and participating in over 150 projects involving universities and companies. His areas of expertise and research interests span Artificial Intelligence, Maintenance, Management, Renewable Energy, Transport, Advanced Analytics, and Data Science. Fausto is a recognized Expert in the European Union in AI4People (EISMD) and ESF. He also serves as the Director of www.ingeniumgroup.eu, holds the status of Senior Member at IEEE since 2021, and has been honored as an Honorary Member of the Research Council of the Indian Institute of Finance since 2021. Fausto is also the Committee Chair of The International Society for Management Science and Engineering Management (ISMSEM) since 2020.

Meet the Volume Editors

Prof. Dr. Sayyad Zahid Qamar is affiliated with the Mechanical Engineering Department at Sultan Qaboos University (SQU), Muscat, Oman. He has over 35 years of experience in academia, research, and industry. His research areas include applied materials and manufacturing, applied mechanics and design, reliability engineering, and engineering education. He has worked on funded projects exceeding 4 million USD in value. He has over 230 publications (books, book chapters, papers in international journals and conferences, and technical reports). Recently, he edited a book on Quality Assurance and Control for IntechOpen and contributed to the volume on Thermal Engineering of Steel Alloy Systems in the multi-volume series Comprehensive Materials Processing, 2nd edition, published by Elsevier. He serves on the editorial boards of several top-ranked research journals.

Dr. Nasr Al-Hinai is an Associate Professor in the Department of Mechanical and Industrial Engineering at Sultan Qaboos University. He is currently serving as the HoD of MIED. He earned his bachelor's degree in Mechanical Engineering from Sultan Qaboos University, Muscat; his MSc from the Department of Mechanical, Manufacturing, and Aerospace Engineering, UMIST, UK; and his Ph.D. in Industrial Engineering from the University of Manitoba, Canada. His academic and research interests include project management, production planning, operations management, supply chain, product design and analysis of bio-composites development processes. Dr. Al-Hinai has held several academic and administrative roles and is actively involved in organizing international conferences and workshops. He has authored numerous publications in peer-reviewed journals, international conferences and book chapters.

Dr. Sandeep Kumar is currently working as a Professor and Deputy Director at Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India. He is an expert in Machine Learning, Biometrics, Embedded Systems, and Computer Vision. Previously, he served as a Professor at SR University, Warangal. He has completed postdoctoral fellowships in India and Singapore, holds 34 granted patents, and has authored 21 books with top publishers. Listed in the top 2% of scientists worldwide by Stanford/Elsevier (2024), he has also received recognition from the USA Book of World Records. With over 200 publications, 17 awards, and multiple invited talks, Dr. Kumar is an active member of 22 professional societies and serves on the editorial boards of over 30 reputable journals and conferences.

Dr. Shilpa Choudhary is an Associate Professor in Computer Science and Engineering (AI and ML) and R&D Coordinator at the Neil Gogte Institute of Technology, Hyderabad, India. Her research interests include Image Processing, IoT, Machine Learning, and Deep Learning. She has published over 70 research papers in reputed journals and conferences, including IEEE and Springer. Dr. Choudhary holds 16 granted patents and 3 published patents. She completed her postdoctoral research at the Singapore Institute of Technology, Singapore. A prolific academic, she has co-authored five textbooks and edited five international books with leading publishers, including Taylor & Francis and Wiley. Her contributions reflect a strong commitment to research, innovation, and academic excellence in emerging technologies.

Dr. Arpit Jain is a Professor at Koneru Lakshmaiah University, Vijayawada, A. P., India. With over 17 years of experience in academics and research, he has been able to bring out the best in individuals and create a healthy work environment, adapting to changing work demands. He has written more than 40 research papers, including those published in SCI Journals, Scopus Journals, the UGC Care List, and Web of Science, on topics ranging from Digital Image Processing to Network-on-Chip Implementation, with a focus on improving chip performance and utility. Dr. Jain is a member of several professional societies, including the IEEE, the Computer Society of India, the ISTE, and the ACM, among others. He has also filed over 25 patents on national and International levels. His research areas include Digital Image Processing, Object Detection and Recognition, and Chip Implementation.

Dr. Ankita Tiwari is an Assistant Professor in the Department of Engineering Mathematics at Koneru Lakshmaiah Education Foundation (KLEF), Andhra Pradesh, India. She earned her Ph.D. in Mathematics from Vikram University, Ujjain, in 2015. Her research focuses on Fuzzy Metric Spaces, Fixed Point Theory, Mathematical Modelling, and Data Science. She has published 20 international journal papers, contributed to book chapters, and holds 8 patents (granted or applied). Dr. Tiwari serves as a reviewer and editorial board member for various reputed journals. She is also affiliated with professional organizations, including IEEE, IAENG, ISC, and IMS. Beyond research, she is actively involved in student mentorship, curriculum design, and institutional accreditation processes, reflecting her commitment to academic and professional excellence.

Contents

Preface

The rapid advancement of Artificial Intelligence (AI), Big Data, and the Internet of Things (IoT) has ushered in a new era for quality control. QC has now moved beyond traditional inspection and defect detection toward intelligent, adaptive, and predictive systems. This book, *Quality Control – Artificial Intelligence, Big Data, and New Trends*, was conceived to capture this transformative shift and provide a consolidated reference for academics, practitioners, and industry professionals navigating the evolving landscape of quality assurance. Each chapter in this volume explores a unique facet of this transition—from AI-driven software development and medical diagnostics using deep learning to smart factory solutions and enterprise-level quality architectures. Through real-world case studies, comparative analyses, and novel frameworks, the book highlights the opportunities and challenges of embedding intelligence into quality management processes. Our primary goal in this book is to stimulate critical thinking and innovation by bridging the gap between traditional quality principles and modern data-centric technologies. As quality remains a cornerstone of competitiveness and sustainability in digital industries, this book aims to inform, inspire, and guide those who are shaping the future of intelligent quality control.

The book is divided into three sections: Introduction, Advanced Topics in Software Engineering, and AI, Big Data, and New Trends. The chapters are briefly described below:

- "Introductory Chapter: Artificial Intelligence, Big Data, and New Trends in Quality Control". This chapter sets the stage by exploring how AI and Big Data are transforming quality control from traditional inspection to predictive, intelligent systems.

- "Investigation on Emotional Development and Story Structure in Fictional Dataset". This chapter examines the use of AI in analyzing emotional trajectories and narrative patterns within fictional datasets to enhance content quality and structure.

- "AI-Driven Software Development: A Paradigm Shift in Software Engineering". This chapter highlights how AI is reshaping software engineering processes, enabling smarter, faster, and more adaptive development cycles.

- "CNN Model-Based Denseblock Vs. BiLSTM Recurrent Network Model for Arrhythmia Class Detection". This chapter presents a comparative study of deep learning architectures for improving diagnostic accuracy in arrhythmia detection using quality-assured medical data.

- "An Innovative Process Engineering Framework: Pipes and Puddles". This chapter introduces a novel framework that visualizes process flows and inefficiencies, enhancing quality engineering through dynamic representation.

- "Severity of Failures in Spur Gearboxes by Vibration Signal Analysis". This chapter examines advanced signal processing techniques for assessing the severity of gear system failures, aiming to enhance maintenance and reliability.

- "Perspective Chapter: QoS-Aware IoT Framework for Performance Control and Resource Management". This chapter proposes an IoT-based framework that ensures quality of service through intelligent resource allocation and performance monitoring.

- "An Application of Machine Learning, Big Data, and IoT of Enterprise Architecture: Challenges, Solutions and Open Issues". This chapter examines the integration of machine learning (ML), Big Data, and the Internet of Things (IoT) into enterprise architecture, addressing key implementation challenges and outlining future directions.

- "Perspective Chapter: Quality Control in New Conditions". This chapter discusses the evolution of quality control under modern constraints, including remote work, digital production environments, and adaptive manufacturing.

We would like to thank all the authors for sharing their knowledge and research in this book. We are also grateful to the reviewers for giving helpful comments and to the editorial and production teams (especially Ms. Elvira Baumgartner) for their support in completing this project.

Sayyad Zahid Qamar and Nasr Al-Hinai
Department of Mechanical and Industrial Engineering,
Sultan Qaboos University,
Muscat, Oman

Sandeep Kumar
Symbiosis Institute of Technology,
Nagpur Campus, Symbiosis International (Deemed University),
Pune, India

Shilpa Choudhary
Department of CSE(AIML),
Neil Gogte Institute of Technology,
Hyderabad, India

Arpit Jain
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation (KLEF),
Andhra Pradesh, India

Ankita Tiwari
Department of Mathematics,
Koneru Lakshmaiah Education Foundation (KLEF),
Andhra Pradesh, India

Section 1

Introduction

Chapter 1

Introductory Chapter: Artificial Intelligence, Big Data, and New Trends in Quality Control

Sayyad Zahid Qamar

1. Introduction

At the very outset, it is important to differentiate between the two basic terms Quality Control (QC) and Quality Assurance (QA). QC refers to the operational techniques and activities used to fulfill quality requirements. It involves the detection and correction of defects in the finished product, and identifying which pieces to release for shipment and which to send back for repair or rejection. Though similar, QA is a system-level concept, a process-oriented approach that focuses on preventing defects by ensuring that the right processes and methodologies are followed during product development. Focus of QA is preventing defects, while of QC is identifying defects. In terms of approach, QA is process-oriented while QC is product-oriented. As to timing, QA is proactive (before production), while QC is reactive (after production). The goal of QA is to ensure that the entire process is correct, while QC makes sure that the product meets the specified standards.

The International Organization for Standardization [1] says that "Quality Assurance is focused on providing confidence that quality requirements will be fulfilled, whereas Quality Control is focused on fulfilling quality requirements." Juran's Quality Handbook [2] defines that "QA is a managerial tool that ensures planned and systematic activities, while QC involves operational techniques and activities used to verify quality." According to the American Society for Quality [3], "While QA focuses on preventing defects, QC involves identifying and fixing them after they have occurred."

Traditional QC largely relies on statistical process control and end-of-line inspection. This is evident from almost all traditional design and manufacturing studies, such as the work on design and manufacture of swell packers by Qamar et al. [4]. But in today's digitalized manufacturing environments, these methods fail to keep pace with real-time demands and complexity. Artificial intelligence (AI) and Big Data now enable QC systems capable of early defect detection, adaptive testing, and self-optimizing processes. This shift represents not just a technological upgrade, but a paradigm shift toward continuous, data-centric Quality 4.0 [5]. For further details of some recent developments in Quality Assurance and Quality Control, please see the book [6].

In this chapter, we explore how AI and Big Data redefine QC, examine successful implementations, and highlight challenges and future innovations.

2. Definitions and the Quality 4.0 paradigm

Quality Control (QC) is the systematic inspection, measurement, and evaluation of products, processes, or services to ensure compliance with predefined standards and specifications. As a core component of the broader Quality Management (QM) framework, QC plays a vital role in identifying, isolating, and correcting defects to maintain product integrity and customer satisfaction. While QC is primarily concerned with the detection of non-conformities at various production stages, it works in tandem with Quality Assurance (QA), which focuses on proactive strategies to prevent such issues from occurring in the first place [7].

In the era of digital transformation, Quality 4.0 marks a significant evolution in the practice of QC by embedding cyber-physical systems (CPS), Artificial Intelligence (AI), the Internet of Things (IoT), and Big Data analytics into quality workflows. These technologies empower real-time process monitoring, self-optimizing production systems, predictive defect detection, and data-driven decision-making. Quality 4.0 not only enhances operational efficiency and traceability but also aligns with the strategic objectives of Industry 4.0, fostering a more intelligent, adaptive, and interconnected manufacturing environment [8]. This paradigm shift paves the way for transformative advances in quality strategies, positioning QC as a driver of continuous innovation and competitiveness [9].

3. Role of big data in QC

In today's data-driven manufacturing landscape, Big Data plays a pivotal role in transforming traditional Quality Control (QC) practices into intelligent, predictive, and adaptive systems. Big Data's three Vs (volume, velocity, variety) present significant potential, yet also impose quality challenges. Big data fuels intelligent QC approaches—anomaly detection, predictive maintenance, and root-cause analysis. But quality assurance itself now must extend to data pipelines to ensure accuracy, timeliness, and trustworthiness [10–12].

Volume and velocity: High-frequency sensor streams (e.g., vibration, temperature, visual data) from CPS require scalable storage and real-time analysis. Modern manufacturing environments, particularly those enabled by Cyber-Physical Systems (CPS) and Industrial Internet of Things (IIoT), generate massive volumes of high-frequency data. These include continuous streams from vibration sensors, thermal cameras, pressure monitors, acoustic detectors, and real-time visual inspection tools. Managing such large-scale data influx demands robust, scalable storage systems and low-latency computing frameworks that support real-time analytics. Rapid processing enables immediate feedback loops for in-line quality control, early fault detection, and operational optimization [13].

Variety: Sensor, video, log, and legacy system data must be integrated, processed, and normalized for coherent analytics. Big Data in QC encompasses a diverse range of sources—structured data from programmable logic controllers (PLCs), semi-structured logs from production systems, unstructured images and videos from inspection cameras, and even historical records from legacy equipment. This heterogeneous data must be effectively integrated, pre-processed, and normalized before meaningful analysis can occur. Data fusion techniques and advanced ETL (Extract, Transform, Load) pipelines are essential to ensure a coherent and unified view of the manufacturing process.

Big Data technologies empower intelligent quality control through advanced analytics techniques such as machine learning-based anomaly detection, predictive maintenance algorithms, defect classification using deep learning, and automated root-cause analysis. These methods not only improve product consistency but also minimize downtime and reduce material waste. However, the expansion of Big Data into quality functions also introduces a new dimension of responsibility: data quality assurance. Just as physical product quality is scrutinized, so too must the quality of the data streams be ensured. This includes validating sensor calibration, detecting corrupted or missing data, ensuring the timeliness of data ingestion, and preserving data lineage for auditability. Thus, the scope of QC now extends to the digital domain, where reliable data pipelines are critical for trustworthy insights and effective decision-making [14].

4. AI techniques in QC

Advances in Artificial Intelligence (AI) have significantly transformed Quality Control (QC) by enabling more accurate, adaptive, and automated inspection systems across diverse industries. The following subsections highlight key AI-driven techniques—ranging from computer vision and predictive maintenance to process monitoring and explainable AI—that are redefining how quality is ensured in modern manufacturing environments [15, 16].

Computer vision and deep learning: AI-powered computer vision (CV) systems are extensively used for real-time visual QC in factories—on assembly lines, printed-circuit boards, textiles, food packaging, and more. Convolutional Neural Networks (CNNs) can detect complex anomalies and reduce false positives compared to rule-based AOI systems [17]. Ettalibi et al. [8] review many industrial deployments, noting both visual and non-visual inspection use cases. While CNNs deliver high accuracy, challenges remain in explainability and adaptation to new product variants.

Predictive and prescriptive maintenance: Intelligent Maintenance Systems (IMS), empowered by AI, analyze sensor streams (e.g., vibration, temperature) to predict failures and prescribe corrective actions. Predictive maintenance reduces downtime and cost. Aquant, Gecko Robotics, and Waites exemplify this trend by coupling AI with robotics and LLM-based guidance for technicians [18, 19].

Process monitoring and rare-event detection: Deep AI models can detect ultra-rare defect events in manufacturing processes. For instance, Escobar [20] introduced a Process Monitoring-for-Quality (PMQ) framework, using hyper-dimensional classification to detect rare defects in real-time SCADA environments.

Explainable AI (XAI): Transparent AI decisions are essential in QC, especially in regulated industries. Research by Sofianidis et al. [21] and others emphasizes integrating interpretability modules such as Grad-CAM and SHAP in QC tools. XAI can bolster trust, aid root-cause analysis, and support regulatory compliance.

5. Smart Metrology and Digital Twins

In modern manufacturing, advances in digital technologies have transformed traditional quality measurement and monitoring into dynamic, data-driven practices through Smart Metrology and Digital Twins, enabling unprecedented precision, real-time insights, and predictive capabilities.

Smart Metrology extends traditional measurement systems through big data use, integrating uncertainty assessment and traceability based on AI-enhanced decision-making. This allows metrology to evolve from equipment calibration to decision assurance [22]. Smart Metrology significantly advances traditional measurement systems by leveraging the power of big data analytics combined with artificial intelligence (AI). Instead of focusing solely on equipment calibration and manual measurements, Smart Metrology incorporates comprehensive uncertainty assessment and traceability frameworks that are continuously updated through AI-enhanced decision-making processes. This evolution transforms metrology into an active, adaptive discipline that not only verifies measurements but also provides real-time decision assurance to quality control teams and process engineers. By integrating sensor networks, cloud computing, and advanced data analytics, Smart Metrology enables dynamic calibration strategies and predictive maintenance, reducing downtime and improving overall measurement reliability and accuracy. The result is a metrology framework that is more resilient, context-aware, and aligned with Industry 4.0 paradigms, supporting complex manufacturing environments with agility and precision [23].

Digital Twins, virtual replicas synchronized with sensor data, enable predictive QC and process optimization in real time. They simulate process variables, identify anomalies, and support "what if" analyses—advancing QC to a prescriptive stage. Digital Twins represent a transformative innovation in quality control and process management by creating highly detailed, virtual replicas of physical systems that are continuously synchronized with live sensor data. These digital counterparts serve as interactive platforms for real-time monitoring, enabling manufacturers to simulate process variables and environmental conditions under different scenarios. By employing advanced algorithms and machine learning models, digital twins can detect subtle anomalies early, forecast potential failures, and optimize production parameters dynamically. This capability not only enhances predictive quality control but also supports "what-if" analyses that allow engineers to explore alternative process adjustments without physical trials, saving time and resources. Consequently, digital twins shift the quality control function from reactive troubleshooting to proactive, prescriptive management, driving higher process efficiency, reduced waste, and improved product consistency [24].

6. Some case studies

A few success stories (case studies) of the use of big data and AI in different industrial and other applications are mentioned below.

AOI in electronics manufacturing: Automated Optical Inspection (AOI) employs AI-CNN models to inspect PCBs at multiple stages—bare board, pre-solder, post-solder. This reduces defect escape and accelerates new product introductions [17].

Fabric quality scanning: Yıldız et al. [25] used thermal imagery and KNN to detect fabric porosity defects. More recent work integrates deep learning and particle swarm optimization to enhance detection accuracy.

Food industry applications: AI systems detect microbial and chemical contaminants using vision and spectral data. However, food QC faces challenges such as data scarcity and inconsistent labels, prompting calls for synthetic dataset creation [26].

Industrial infrastructure maintenance: Gecko Robotics deploys inspection robots and ultrasonic sensors in powerplants and pipelines; AI platforms analyze data to detect corrosion and structural degradation before failures unfold [18]. This exemplifies AI-based QC outside traditional manufacturing.

7. Implementation challenges

As Big Data and AI technologies evolve rapidly, their implementation across industries is met with a range of technical, ethical, and organizational challenges. From data integrity issues to workforce readiness and system integration, these hurdles must be addressed to fully realize the transformative potential of intelligent systems [27–29].

Data quality and class imbalance: Poor data quality, including missing values, noisy inputs, and mislabeled data, severely hinders the performance of AI models. Class imbalance further exacerbates the issue, leading to biased predictions where minority classes are underrepresented and often misclassified. Addressing this requires robust pre-processing, resampling techniques, and adaptive algorithms. Rare defect events lead to skewed data distributions. Techniques like SMOTE, anomaly detection, and oversampling can help—but training requires caution to avoid false positives.

Integration with legacy systems: Many organizations struggle to incorporate modern AI solutions into existing legacy infrastructure, which may be incompatible or lack the capacity for real-time data handling. This challenge increases costs and complicates deployment timelines, often requiring middleware or complete system overhauls. AI models must connect seamlessly with existing CPS, MES, and ERP systems. Issues include legacy constraints, data silos, and compliance requirements.

Interpretability and trust: As AI models grow more complex, especially with deep learning, their "black box" nature raises serious concerns regarding decision transparency. Stakeholders—including regulators, users, and executives—demand explainable AI to ensure accountability, fairness, and ethical compliance. Opaque AI models are difficult to trust. In critical industries (semiconductors, pharmaceuticals), Explainable AI and expert augmentation are increasingly essential.

Workforce skills and organizational adoption: A significant gap exists between AI capabilities and the workforce's ability to understand, manage, or apply them effectively. Organizations often face cultural resistance, a lack of training, and insufficient leadership buy-in, slowing down AI adoption across sectors. Introducing AI tools faces resistance from traditional QC personnel. Success requires cross-functional training, human-centric design, and change management [20, 29, 30].

Data governance and privacy concerns: Implementing AI at scale raises complex issues related to data ownership, consent, and cross-border compliance. Without strong governance frameworks, organizations risk violating regulations such as GDPR or exposing sensitive user data.

Scalability and real-time processing: AI applications increasingly demand real-time analytics on vast and fast-moving datasets. Ensuring computational scalability without sacrificing performance or accuracy remains a core technical and infrastructure challenge, especially for edge and IoT-enabled deployments.

8. Future directions and new trends

A brief overview of new developments and future trends in big data and AI is presented below.

Smart Metrology: Smart Metrology integrates AI and big data analytics into measurement processes, enabling real-time, adaptive quality control in manufacturing and engineering. It enhances precision by learning from process data, reducing human error, and optimizing inspection strategies dynamically [31].

Digital Twins: Digital twins are virtual replicas of physical systems that use real-time data and AI to simulate, monitor, and predict the performance of assets. This technology is revolutionizing industries by enabling predictive maintenance, scenario testing, and enhanced decision-making across product lifecycles [32].

Edge AI deployment: Edge AI brings data processing and intelligent decision-making closer to the source (e.g., sensors, machines), reducing latency and bandwidth demands. It is crucial for time-sensitive applications like autonomous vehicles, smart manufacturing, and remote monitoring in IoT environments [33].

Industry 5.0: Industry 5.0 emphasizes human-centric collaboration between humans and intelligent machines, combining the efficiency of automation with human creativity. It focuses on sustainability, resilience, and personalized production, moving beyond the purely efficiency-driven goals of Industry 4.0 [34].

Federated learning: Federated learning enables AI models to be trained across multiple decentralized devices without exchanging raw data, preserving privacy and reducing central processing loads. It is especially impactful in sectors like healthcare and finance where data sensitivity is critical [35].

Explainable AI (XAI): Explainable AI focuses on making AI decisions transparent and interpretable, helping stakeholders understand, trust, and effectively audit automated systems. It is essential for high-stakes domains like healthcare, finance, and legal systems where accountability is vital [36].

9. Strategic recommendations

To successfully adopt AI-driven Quality Control (AI-QC), organizations should consider a multi-faceted approach that addresses technical, organizational, and human factors. Key strategies include the following [37–42].

Self-assess data readiness: Begin by evaluating the quality, completeness, and accessibility of existing data. Invest in robust data pipelines and ensure sensor calibration is precise to capture reliable real-time data streams essential for effective AI model training and operation.

Pilot AI solutions in controlled environments: Implement AI systems initially in low-risk production areas. This controlled deployment enables thorough measurement of technical performance, such as defect detection accuracy, operational cost savings, and equipment uptime improvements. It also facilitates ROI evaluation across multiple KPIs before full-scale rollout.

Embed explainable AI (XAI) Tools: Incorporate explainability and transparency features in AI models from the development phase. This fosters stakeholder trust, supports regulatory compliance, and facilitates troubleshooting when unexpected outputs occur.

Reskill and empower staff: Invest in comprehensive training programs to upskill frontline quality control inspectors, operators, and maintenance personnel. Empowered employees with AI literacy can effectively interpret AI outputs, make informed decisions, and maintain the AI systems, enhancing adoption success.

Foster academia-industry collaboration: Establish partnerships with academic institutions to stay abreast of cutting-edge AI research and ensure that practical challenges faced on the factory floor inform ongoing algorithm development. This symbiotic relationship accelerates innovation and contextualizes AI solutions.

Implement continuous model monitoring and adaptation: Actively monitor AI models for data drift or concept drift, which can degrade performance over time. Regularly retrain models with updated data, validate results, and fine-tune parameters to maintain high accuracy and reliability.

Ensure data privacy and security: Develop strict data governance policies that comply with relevant standards and regulations. Protect sensitive production and customer data from breaches or misuse to build confidence among stakeholders and avoid legal repercussions.

Promote cross-functional collaboration: Encourage collaboration between quality engineers, IT professionals, data scientists, and operations teams. Such multidisciplinary coordination ensures AI-QC initiatives align with broader business objectives and operational workflows.

Leverage scalable cloud and edge computing: Adopt scalable infrastructure solutions that enable flexible processing of big data. Utilize cloud platforms for centralized data storage and analytics, while deploying edge computing to deliver real-time inference and rapid decision-making on the shop floor.

Plan for change management: Develop clear communication strategies and change management plans that address employee concerns, outline new workflows, and highlight benefits of AI-QC. Successful cultural integration minimizes resistance and accelerates adoption.

By following these strategic recommendations, organizations can maximize the benefits of Big Data and AI in quality control, leading to improved product consistency, operational efficiency, and competitive advantage.

10. Conclusions

This introductory chapter starts with the observation that AI and Big Data are revolutionizing quality control (QC) by shifting it from traditional, reactive inspection methods toward proactive, intelligent, and adaptive systems. Leveraging AI-driven tools such as machine vision, real-time anomaly detection, and predictive maintenance, manufacturers are now able to achieve greater precision, minimize defects, reduce material and energy waste, and respond more swiftly to process deviations. These technologies also enable continuous learning from production data, fostering ongoing improvement in both product quality and process efficiency. Yet, the full potential of these advancements depends on successfully addressing several critical implementation barriers. These include ensuring the accuracy and completeness of data, managing class imbalance, integrating AI with legacy infrastructure, building systems that are interpretable and trustworthy (especially through Explainable AI), and equipping the workforce with the digital and analytical skills required to operate and trust these intelligent systems.

Emerging paradigms such as Smart Metrology and Industry 5.0 further suggest a future where QC systems are not only automated and intelligent but also deeply collaborative, sustainable, and tailored to human needs. These new frameworks emphasize transparency, human-machine synergy, and resilience, aiming to embed quality as a dynamic and predictive capability within entire production ecosystems. Ultimately, effective adoption will rely on strategic approaches such as small-scale pilot projects, modular deployment, strong data governance, and targeted upskilling or reskilling programs. With thoughtful integration, AI and Big Data will not just enhance QC—they will redefine it as a cornerstone of future-ready, human-centric manufacturing systems.

Author details

Sayyad Zahid Qamar
Mechanical and Industrial Engineering Department, Sultan Qaboos University, Muscat, Oman

*Address all correspondence to: sayyad@squ.edu.om

References

[1] ISO 9000:2015. Quality management systems – Fundamentals and vocabulary. Available from: https://www.iso.org/standard/45481.html [Accessed: May 2025]

[2] Juran JM, DeFeo JA. Juran's Quality Handbook. 7th ed. London: McGraw-Hill Education; 2016

[3] American Society for Quality (ASQ). Glossary. 2025. Available from: https://asq.org/quality-resources/quality-assurance-vs-control [Accessed: May 2025]

[4] Qamar SZ, Pervez T, Akhtar M, Al-Kharusi MSM. Design and manufacture of swell packers: Influence of material behavior. Materials and Manufacturing Processes. 2012;**27**(7):721-726

[5] Christiaanse R. Quality 4.0: Data Quality and Integrity – A Computational Approach. London, UK: IntechOpen; 2024. DOI: 10.5772/intechopen.108213

[6] Qamar SZ, Al-Hinai N, editors. Quality Control and Quality Assurance - Techniques and Applications. London, UK: IntechOpen; 2024. DOI: 10.5772/intechopen.111297

[7] Hoffmann R, Reich C. A systematic literature review on artificial intelligence and explainable artificial intelligence for visual quality assurance in manufacturing. Electronics. 2023;**12**(22):4572. DOI: 10.1109/TII.2022.3201234

[8] Ettalibi A, Elouadi A, Mansour A. AI and computer vision-based real-time quality control: A review of industrial applications. Procedia Computer Science. 2024;**231**:212-220. DOI: 10.1016/j.procs.2023.12.195. ISSN 1877-0509

[9] Quadir AM, Jha K, Haneef S, Sivaraman AK, Tee KF. A review on data-driven quality prediction in the production process with machine learning for industry 4.0. PRO. 2022;**10**(10):1966

[10] Zhang P, Cao W, Muccini H. Quality assurance technologies of big data applications: A systematic literature review. 2020. arXiv preprint arXiv:2002.01759

[11] Jagatheesaperumal SK, Rahouti M, Ahmad K, Al-Fuqaha A, Guizani M. The duo of artificial intelligence and big data for industry 4.0: Applications, techniques, challenges, and future research directions. IEEE Internet of Things Journal. 2021;**9**(15):12861-12885

[12] Wang C, Yang Z, Li ZS, Damian D, Lo D. Quality assurance for artificial intelligence: A study of industrial concerns, challenges and best practices. 2024. arXiv preprint arXiv:2402.16391

[13] Kockum F, Dacre N. Project management volume, velocity, variety: A big data dynamics approach. Advanced Project Management. 2021;**21**(1)

[14] Ji S, Li Q, Cao W, Zhang P, Muccini H. Quality assurance technologies of big data applications: A systematic literature review. Applied Sciences. 2020;**10**(22):8052

[15] Jain D. Artificial intelligence in quality control systems: A cross-industry analysis. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2024;**10**(6):1321-1333. Available from: ijsrcseit.com+1researchgate.net+1

[16] Times of India. AI-Driven Quality Control Systems and Mfg Solutions Focus of CICU Workshop. Ludhiana; 2025. Available from: timesofindia.indiatimes.com

[17] Wikipedia AOI. Automated optical inspection. Wikipedia. 2025 [Accessed: May 2025]

[18] Business Insider. How AI and robotics can help prevent breakdowns in factories and save manufacturers big bucks. 2025. Available from: https://www.businessinsider.com/artificial-intelligence-robotics-predictive-maintenance-manufacturing-factory-solutions-2025-5?utm_source=chatgpt.com [Accessed: May 2025]

[19] Wikipedia IMS. Intelligent maintenance system. Wikipedia, the Free Encyclopedia; 2025. Available from: https://en.wikipedia.org/wiki/Intelligent_maintenance_system

[20] Escobar C. Process monitoring for quality – A big data and artificial intelligence-focused manufacturing quality control [doctoral thesis]. Mexico: Tecnológico de Monterrey; 2019. p. 2019

[21] Sofianidis G, Rožanec JM, Mladenic D, Kyriazis D. A review of explainable artificial intelligence in manufacturing. In: Trusted Artificial Intelligence in Manufacturing. 2021;**24**:93-113

[22] Wikipedia SM. Smart Metrology. Wikipedia; 2025. Available from: https://en.wikipedia.org/wiki/Smart_Metrology [Accessed: May 2025]

[23] Wright L, Davidson S. Digital twins for metrology; metrology for digital twins. Measurement Science and Technology. 2024;**35**(5):051001

[24] Rathore MM, Shah SA, Shukla D, Bentafat E, Bakiras S. The role of AI, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. IEEE Access. 2021;**9**:32030-32052

[25] Yildiz K, Buldu A, Demetgul M, Yildiz Z. A novel thermal-based fabric defect detection technique. The Journal of The Textile Institute. 2015;**106**(3):275-283

[26] Naseem S, Rizwan M. The role of artificial intelligence in advancing food safety: A strategic path to zero contamination. Food Control. 2025;**175**:111292

[27] Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. Journal of Medical Internet Research. 2019;**21**(7):e13659

[28] Wylde V, Prakash E, Hewage C, Platts J. Ethical challenges in the use of digital technologies: AI and big data. In: Digital Transformation in Policing: The Promise, Perils and Solutions. Cham: Springer International Publishing; 2023. pp. 33-58

[29] Zhuang YT, Wu F, Chen C, Pan YH. Challenges and opportunities: From big data to knowledge in AI 2.0. Frontiers of Information Technology and Electronic Engineering. 2017;**18**:3-14

[30] Ojeda A, Valera J, Medina E, Samadian H, Padilla R. AI implementation in big data: Shaping data analysis for business decisions. Issues in Information Systems. 2024;**25**(4)

[31] Lazzari A, Pou JM, Dubois C, Leblond L. Smart metrology: The importance of metrology of decisions in the big data era. IEEE Instrumentation and Measurement Magazine. 2017;**20**(6):22-29

[32] Juarez MG, Botti VJ, Giret AS. Digital twins: Review and challenges. Journal of Computing and Information Science in Engineering. 2021;**21**(3):030802

[33] Singh R, Gill SS. Edge AI: A survey. Internet of Things and Cyber-Physical Systems. 2023;**3**:71-92

[34] Leng J, Sha W, Wang B, Zheng P, Zhuang C, Liu Q, et al. Industry 5.0: Prospect and retrospect. Journal of Manufacturing Systems. 2022;**65**:279-295

[35] Wen J, Zhang Z, Lan Y, Cui Z, Cai J, Zhang W. A survey on federated learning: Challenges and applications. International Journal of Machine Learning and Cybernetics. 2023;**14**(2):513-535

[36] Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P, et al. Explainable AI (XAI): Core ideas, techniques, and solutions. ACM Computing Surveys. 2023;**55**(9):1-33

[37] Davenport TH, Ronanki R. Artificial intelligence for the real world. Harvard Business Review. 2018;**96**(1):108-116. Available from: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

[38] Jiang F, Li J. Continuous monitoring and adaptation of AI models for industrial quality control. Computers in Industry. 2024;**146**:103843. DOI: 10.1016/j.compind.2023.103843

[39] Lee J, Bagheri B, Kao H-A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters. 2015;**3**:18-23. DOI: 10.1016/j.mfglet.2014.12.001

[40] Scarton C, Ribeiro R, Silva M. Strategies for AI adoption in industrial quality control: A case study approach. Journal of Manufacturing Systems. 2025;**67**:130-145. DOI: 10.1016/j.jmsy.2024.11.005

[41] Scarton G, Formentini M, Romano P. Automating quality control through an expert system. Electron Markets. 2025;**35**:14. DOI: 10.1007/s12525-025-00766-y

[42] Wang L, Xu X, Zhang L. Explainable AI for quality control in smart manufacturing: Techniques and applications. IEEE Transactions on Industrial Informatics. 2023;**19**(2):1207-1215

Section 2

Advanced Topics in Software Engineering

Chapter 2

Investigation on Emotional Development and Story Structure in Fictional Dataset

Raj Bano Khan, Sajjad Muhammad Khan,
Sufyan Muhammad Khan and Waqas Muhammad Khan

Abstract

This study explores the evolution of sentiment expressed in a fictional dataset by characters and inspect the correlation between specific events and important change in sentiment. We employ completely sentiment analysis framework to examine carefully selected set of text in order to capture the emotional journeys that the characters take throughout the story. Using VADER and TextBlob, among other techniques, we measure changes in sentiment at different story points and pinpoint the critical areas that elicit strong emotion reaction. Our results shows that significant character interactions, including disputes and reconciliations, and large plot twists, like unexpected disclosures, are highly correlated with the significant changes in sentiment. Tjis study sheds the light on the complex interplay between character emotions and story structure, offering insightful observations about the dynamics.

Keywords: data mining, event identification, narration, temporal segmentation, sentiment analysis, statistical analysis, visualizations

1. Introduction

Sentiment analysis is a highly significant field that analyzes and identifies the sentiment expressed in text [1]. The subtle of characters' emotions greatly enhances the emotional fabric of a story in the fields of literature and storytelling [2]. As a literary beings, characters change and react to plot the developments in a dynamic way creating a rich emotional web that draws readers in and influence the narrative experience as a whole [3]. In order to answer the crucial question, "how does the sentiment articulated by characters evolve over the course of the narrative, and are the specific events or character interactions that correlate with significant change in the sentiment?" this research sets out on an engrossing journey to analyze and comprehend the complex evolution of sentiment expressed by character in fictional datasets.

The comprehension of character emotional arcs is crucial for improving our knowledge of storytelling technique as well as for providing new perspectives on human psychology, empathy, and the significant influence of narrative structures on readers' perceptions [4, 5]. This study aims to quantify the ebb and flow of sentiments across various characters and their interactions by utilizing sophisticated natural language processing techniques and sentiment analysis tools to uncover the emotional undercurrent contained within the text.

Character sentiment analysis is more than just a literary analysis exercise; it is a mean of gaining insight in to the fundamental processes that shape the emotional dynamics of the story [6]. We seek to uncover the crucial moments that mold the emotional terrain of a narrative by pinpointing particular occurrences or character interactions that align with notable shifts in mood. These discoveries have ramifications for the study of human emotion more broadly, computational linguistics, human-computer interaction, and literature in addition to narrative theory and literature [7].

Our goal is to shed light on the complex interactions between characters, events, and emotions by offering a thorough framework of character sentiment analysis in fictional stories through this research. We hope to advance our knowledge of storytelling mechanisms, expand the toolkit for literary analysis, and open up new avenues for creative applications in natural language processing and other fields by shedding light on the emotional pathways within fictional worlds.

In an effort to better understand the complex emotional dynamics that underlie storytelling, character sentiment analysis in fictional narratives has drawn more attention from the fields of computational linguistics and literary studies. To grasp the mechanics of narrative production reader engagement, one must be aware of how character attitude vary across the course of a story and recognize the occasions or encounters that correspond with noteworthy shifts in sentiments.

The field of sentiment analysis in literature has seen a revolution with the introduction of computational methodologies and natural language processing techniques [8]. Classifying text as positive, negative, or neutral was the primary emphasis of early investigations on sentiment classification [8, 9]. More advance techniques, on the other hand, use machine learning algorithms and sentiment analysis tools such as TextBlob [10] and VADER to extract subtle emotional under tones from literary text [10, 11].

Character sentiment refers to the emotional journeys that characters take under on during the course of the story and is closely associated with narrative arcs [12]. Prominent academics including Christopher Booker, Joseph Campbell, and Vladimir Propp have put forth a number of archetypal narrative frameworks, emphasizing the significance of emotional growth for character and plot development [13, 14]. Character feelings, which reflects the internal tensions wants, and transformations of the characters, is an essential element of these narrative arcs [15, 16].

Numerous research works have investigated the relationship between narrative occurrences and emotional shift in characters [17]. Automatic techniques for the assessing the sentiment dynamics between characters in plays are presented in a number of research [15]. Monitoring the emotional trajectories of interpersonal interactions is the aim of the Nalisnick and Baird study [18]. By summing the valence values of words that occur in the continuous direct speech and are present in the lexicon of affective standards, they may determine who is speaking to whom thanks to the structured format of the dialog this allow of mining of character to character feelings [19]. Similarly, Elkins and Chun [20] analyzed the novel through sentiment analysis, and identified character interactions and plot structure that drive emotional fluctuations.

The findings of Langure and Zareei [21] demonstrate the efficacy of specially designed models in capturing minute emotional changes, especially in intricate narrative contexts. These findings are consistent with the observed emotional alteration dataset. The feelings and thoughts convey in five distinct Chinese translations of Charles Dickens' David Copperfield were the subject of a study done in 2024 by Li, Afzaal, and Yin. The study sought to comprehend the ways in which these translations vary over time in the way they express emotions and portray emotions [22]. The use of sentiment arcs in literary works as a metric of evaluating of literary quality is examined by Bizzoni et al. [23], with a focus on Nobel laureates in literature. The study investigates the fractal nature of sentiment arcs and how they relate to literary texts perceived quality.

2. Methodology

The process for analyzing how character sentiment changes over a fictional dataset progress over the narrative and identify character interactions or specific events that influence in sentiment changes:

2.1 Data selection

A dataset of a script of "Harry Potter and the Philosopher's Stone Script" in which a fictional narratives cover various narrative styles. The script is on "https://tomfeltonandmore.tripod.com/home/id9.html". The dataset is designed with the information of character names, dialogs\ scenes, and events.

- *Character:* the character name that conveys emotion.
- *Text:* the passage that corresponds with the character at the specific moment in the narrative (full script).
- *Distinguish text:* distinguish in to conversation (dialog) and incident (scene) in text.
- *Event:* denotes an important conversation or incident in the narrative.
- *Chapter:* gives a chronological sequence or grouping that illustrates the story's development.

The sample of dataset shown in form of **Table 1** is as follows:

2.2 Data preprocessing

It is essential stage in sentiment analysis process for cleaning and organizing the text data to gain significant meaning that represents the expressed sentiment. The accuracy and reliability of sentiment analysis algorithms are enhanced by this procedure. Normalize the text, tokenize sentences and words, and eliminate stop words and other superfluous information [24]. Determine and extract dialog and exchanges between specified characters from the story.

Sentiment analysis requires refining text input in order to get the greatest possible outcomes. NLP libraries like NLTK can help to simplify the preprocessing process [25].

Character	Text	Distinguish text	Chapter	Event
	A neighborhood on a street called Privet Drive. An owl, sitting on the street sign, flies off to reveal a mysterious appearing old man walking through a forest near the street. He stops at the start of the street and takes out a mechanical device and zaps all the light out of the lampposts. He puts away the device and a cat meows. The man, Albus Dumbledore, looks down at the cat, which is a tabby and is sitting on a brick ledge.	Scene	Starting	Reveal
Dumbledore	I should have known that you would be here…Professor McGonagall.	Dialog	Starting	Encounter
	The cat meows, sniffs out, and the camera pans back to a wall. The cats shadow is seen progressing into a human. There are footsteps and Minerva Mcgonagall is revealed.	Scene	Starting	Transformation
McGonagall	Good evening, Professor Dumbledore. Are the rumors true, Albus?	Dialog	Starting	Inquiry
Dumbledore	I'm afraid so, Professor. The good, and the bad.	Dialog	Starting	Confirmation

Table 1.
The sample of dataset.

2.3 Sentiment analysis

Sentiment analysis for narrative texts is the process of using methods to infer a text's emotional content form the words it uses [26]. This can be beneficial for understanding character relationships, and the emotional development of a story, characters interactions, and the event's impact on the character. The analysis can be performed by the following:

- Apply sentiment analysis methods to measure the sentiment expressed by characters in the story convey.

- Assign sentiment scores to character interactions and dialog using sentiment lexicons [26].

- Machine learning approach used for sentiment analysis, to forecast sentiment labels for characters and story segments, and train machine algorithms (e.g., support vector machine [SVM] or Naive Bayes classifier) on labeled sentiment data.

- To capture contextual nuances and temporal relationships in character attitudes, use deep learning architectures such as recurrent Neural Network or Long Short Term Memory (LSTM) networks.

2.3.1 Temporal segmentation

Now the format and structure of fictional dataset can depend on the temporal segmentation technique selection. The process of breaking a story into temporal segments, such as chapters, scenes/dialogs, or other significant units of time, in order to examine the sentiment evolution of the characters within a fictional dataset requires temporal segmentation [27]. The process of temporal segmentation used in this research has including steps:

- To examine how sentient changes over time, divide the story into temporal segments (such as chapters or scenes/dialogs).
- Aggregate sentiment scores for every temporal segment to represent how the characters feelings have changed throughout the course of the story.

2.3.2 Event identification

Finding particular instances or exchanges in the story that might have an impact on a character's sentiment shift is known as event identification when examining the sentiments of characters in a fictional dataset. The main events in the text are as follows:

- Point out certain incidents or character exchanges that correspond with notable shifts in sentiment.
- Identify emotional turning points, plot twists, or significant character interactions that may affect sentiment shift by automated technique.

2.3.3 Correlation analysis

Statistical approach can be utilized examine the evolution of sentiment expressed by characters in a fictional dataset and to pinpoint particular events or character interactions that are associated with noteworthy shifts in sentiment. Key statistical analysis is used in this methodology:

- To ascertain whether narrative events and shifts in character sentiment are correlated and perform statistical analysis.
- Determine descriptive statistics across time, such as sentiment polarity distribution, mean sentiment scores, and sentiment trends.

2.3.4 Qualitative analysis

The fictitious dataset's characters such as emotional shifts, character interactions, and narrative events are all quantitatively revealed to academics using statistical studies. Researchers can gain a better grasp of character emotions and narrative dynamics by using statistical tools to find important patterns, correlations, and trends in sentiment evolution. Quantitative analysis is used in this methodology:

- To contextualize sentiment changes, and conduct a qualitative analysis of story events and character interactions.

- Analyze how significant events affect the feelings of the characters and the plot's development.

2.3.5 Visualizations

To demonstrate sentiment research findings with narrative events and sentiment patterns using visualizations such as line chart and bar chart.

3. Result and discussions

3.1 Transformation of sentiment throughout the story

- A dynamic progression of the sentiment exhibited by the characters throughout the story is revealed by analysis. As the characters are introduced and the location is created, the initial feelings are usually neutral or somewhat favorable. The feelings shown in form of **Table 2** are given below:

- The story's developing character arcs, emotional dynamics, and plot advancements are all reflected in the increasingly noticeable mood swings that occur throughout.

- Sentiments peaks and troughs align with significant narrative events, revealing the influence of thematic components, character interactions, and plot twists on character feelings as shown in **Figure 1**.

3.2 Detecting important occurrences and interactions

- It is discovered that a number of particular occurrences and character interactions are related to a notable shifts in emotions.

- Characters react intensely emotionally to major plot twists such as surprise revelations or betrayals, which cause noticeable changes in attitude.

- Character interactions—especially those involving conflict or reconciliation—are linked to observable emotional swings, underscoring the critical role that interpersonal relationships play in influencing character emotions as shown in **Figure 2**.

1	Character	Text	Distinguish	Chapter	Event	Sentiment
2		A neighbor	Scene	Starting	Reveal	Negative
3	Dumbledore	I should ha	Dialog	Starting	Encounter	Neutral
4		The cat me	Scene	Starting	Transforma	Neutral
5	McGonagal	Good even	Dialog	Starting	Inquiry	Positive

Table 2.
Transformation of sentiment throughout the story.

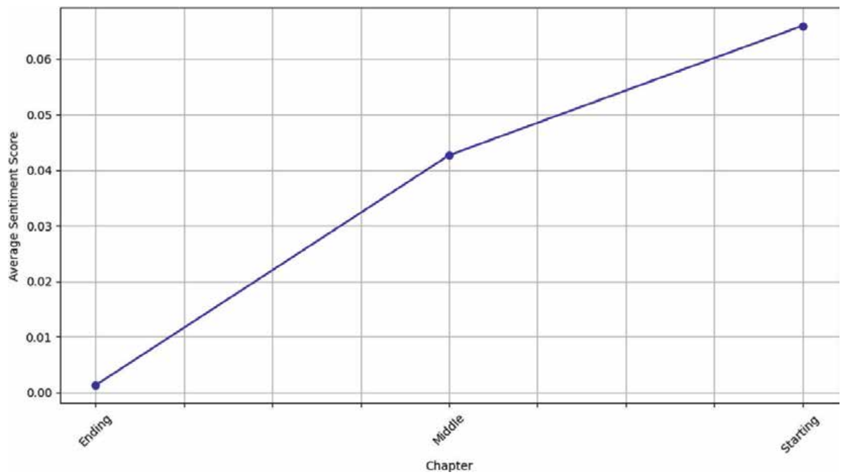


Figure 1.
Sentiment evolution across chapters.

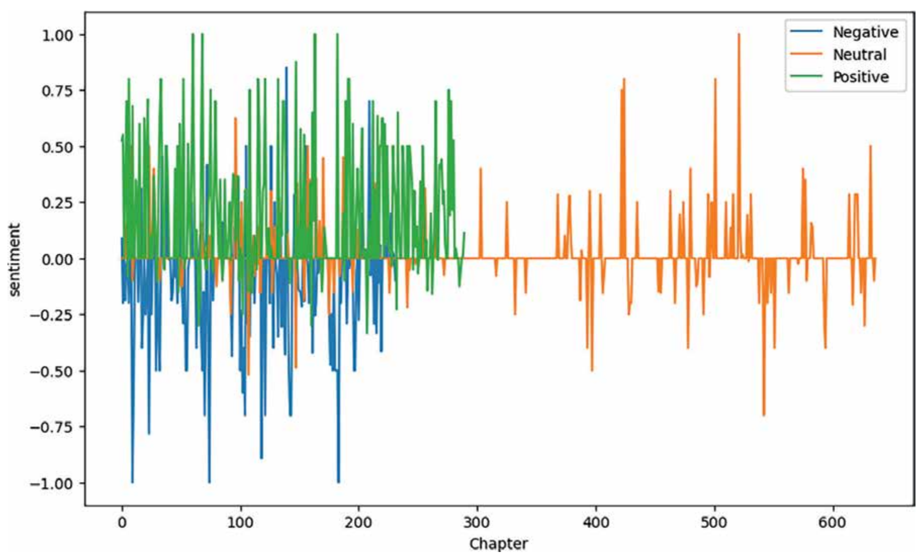


Figure 2.
Detecting important occurrences and interactions.

3.3 Temporal analysis of sentiment trends

- Finding trends in sentiment across time is made easier by the narrative's temporal division.

- Sentiment usually increases at the story's climax points, reflecting the character' growing stakes and growing tension.

- Resolve conflicts and character findings closure, on the other hand, are marked by a slow return to equilibrium throughout resolution phases as shown in **Figure 3**.

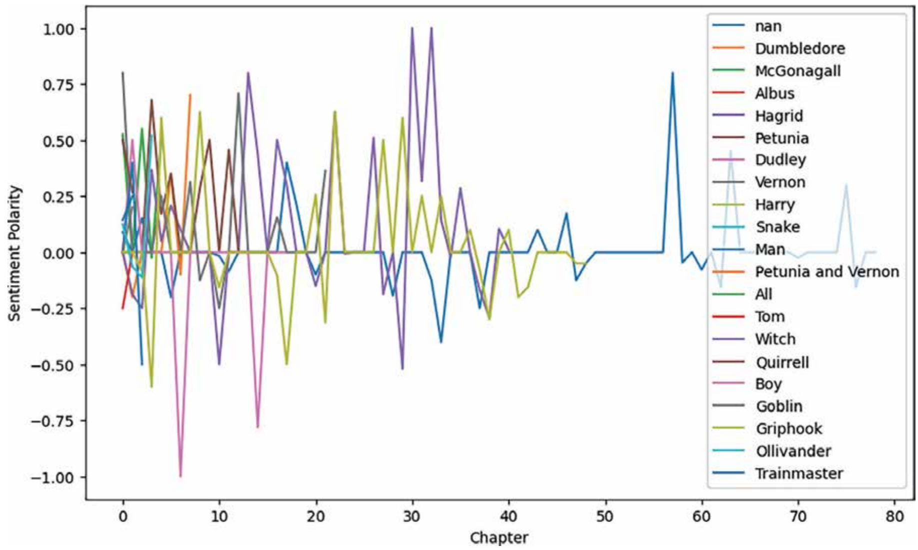


Figure 3.
Sentiment analysis over time.

3.4 Particular sentiment dynamics for character

- Distinctive sentiment trajectories for every character can be found through analyzing their distinct storyline.

- Protagonists frequently experience extreme emotional shifts, exhibiting shifts between success and misfortune, and hope and despair.

- More consistent sentiment patterns can be seen in antagonists, which may be reflections of their wavering commitment to evil goals or evil conflicts as shown in **Figure 4**.

3.5 Interpreting the results in story telling

- Authorial purpose, character motives, and thematic themes are taken in to account when interpreting the observed sentiment dynamics within the larger narrative framework.

- Readers are guided by the narrative signposts of emotional highs and lows as they traverse the story's emotional landscape.

- Plot structure, character development, and reader involvement are intricately intertwined, as seen by the correlation found between sentiment and narrative events as shown in **Figure 5** and the values of correlation are shown in **Table 3**.

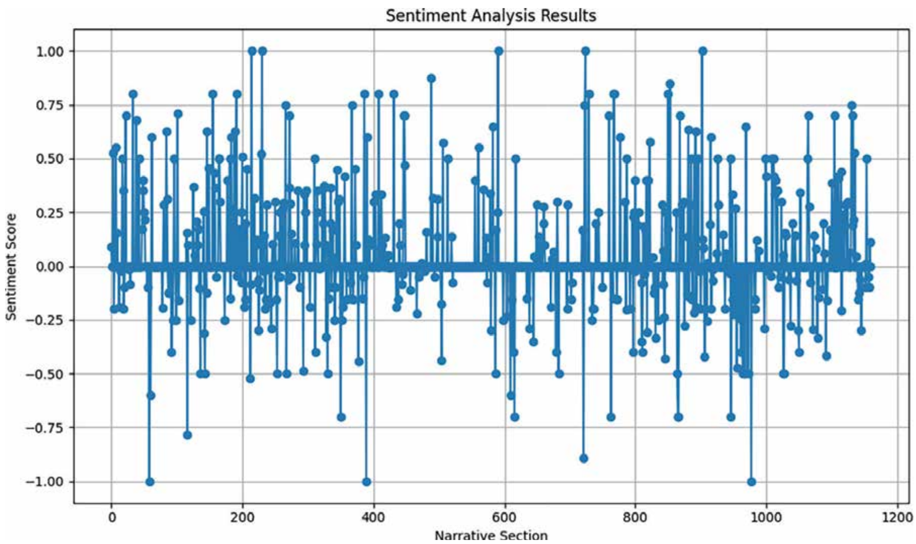


Figure 4.
Particular sentiment dynamics for character.

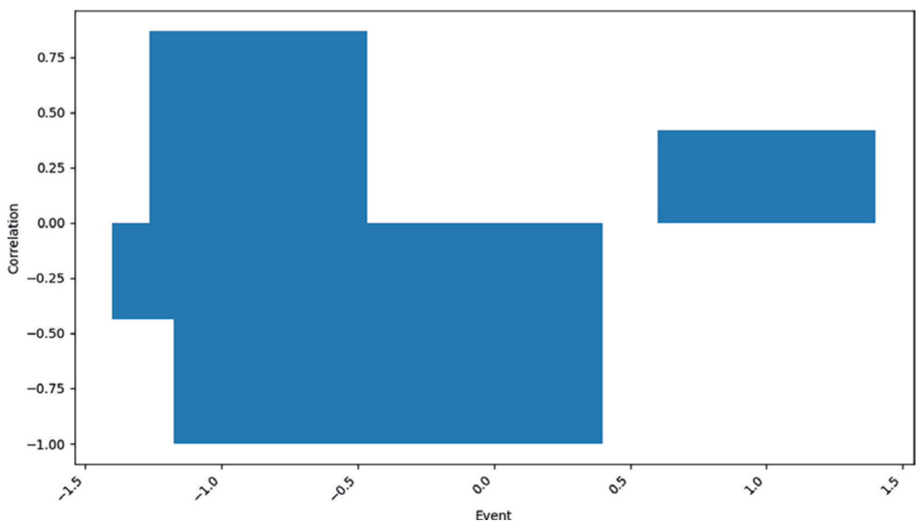


Figure 5.
Correlation between events and sentiment scores.

3.6 Relevance to narrative theory and literary analysis

- By illuminating the mechanisms promoting emotional connection and reader immersion, the findings advance our knowledge of how character emotions evolves in fictional stories.

	Event	Correlation
0	Inquiry	−0.774597
1	Confirmation	1.000000
2	Arrival	0.591503
3	Questioning	−0.272033
4	Acknowledgment	NaN
..	...	...
77	The Intense Wizard's Chess Match	−1.000000
78	The Sacrifice	NaN
79	The Revelation	0.402651
80	The Truth Revealed	NaN
81	The Temptation	−1.000000

Table 3.
Correlation values.

- Sentiment analysis sheds light on literary analysis and narrative theory, contributing to conversations about thematic resonance, character dynamics, and narrative structure.

- In order to better understand the nuances of character emotions and narrative expression, future study may examine the use of sentiment analysis tools across a variety of literary genres, cultural contexts, and narrative formats.

In order to gain a greater grasp of storytelling dynamics and character emotions, methodically examine the evolution of character sentiment in fictional narratives and find patterns, correlations, and discoveries. Kim and Klinger [28] look in to the numerous ways the dialog, narrative descriptions, and character actions in fan fiction can convey emotions and analysis of feelings fan fiction communication channels offer insight information about how emotionally charged stories are written. Using these insights authors and storytellers may craft more immersive, captivating, and emotionally impactful narratives that attract readers and have a lasting effect [28].

In a fictional dataset, the sentiment is expressed by the characters changes during the story. Temporal segmentation is the process of breaking the story up in to the meaningful temporal units, such as scenes, chapters, or other relevant time intervals, in order to track the evolution of sentiment. The study by Tu and Brown [13] focuses on how character mediates plot structure, highlighting how crucial characters are to be development of a story. It also indicates the characters are important for influencing the overall narrative structure as well as for advancing the plot and this view point highlights the importance of characters in storytelling, which has important consequences for narrative theorists [13].

With a particular emphasis on the use of machine learning algorithms for sentiment analysis in literary texts, Anvar Shathik and Krishna Prasad's [29] literature review offers a broad overview of this field work and in-depth investigation of sentiment analysis methods used on literary texts, covering the theoretical underpinnings as well as real world applications (literature, marketing, and social media) of sentiment analysis . In order to analyze the sentiment represented in literary works and derive valuable insights from textual data, it emphasizes the significance of utilizing machine learning techniques and sentiment lexicons [29].

In this research, machine learning techniques are applicable for examining how the sentiment expressed by characters in fictional dataset changes across the story and pinpointing character interactions that correspond with notable shifts in sentiment. In the context of examining the sentiment espoused by characters in a fictional dataset, event identification refers to the process of identifying particular interactions that may have an impact on narrative shifts in character sentiment and the statistical analysis can be employed.

In their study, Rebora and Pianzola [30] look at Wattpad reader comments, emphasizing how user interacts with and remarks on digital writing and has implications for how writers might use reader feedback on digital platforms such as Wattpad to increase emotional connection with their stories, even though it does not explicitly address emotional storytelling. Writing more emotionally resonant narratives can be achieved by authors by honing their storytelling techniques through the analysis of reader feedback and they feelings they arouse [30].

It can provide a thorough study that places the work in the larger framework of literary analysis in Vishnubhotla et al.'s publication [31]. This method can enhance conversation and offer insightful information about the development of sentiment and the dynamics of storytelling in made up datasets [31].

The findings of the investigation highlight sentiment's important role in forming the narrative experience, shedding light on the interaction between character emotions, plot dynamics, and reader interpretations. This study enhances our understanding of the narrative skill and its significant impact on human emotions and imaginations by unraveling the vast fabric of passion portrayed by the characters.

4. Conclusion

The study's conclusions are described below, along with how they relate to literary scholarship. The significance of sentiment analysis in comprehending story dynamics and character psychology is emphasized. The conclusion underlines the continued use of sentiment analysis in literary studies and offers potential directions for further research.

With an emphasis on the interaction between narrative advances and emotional alterations, the study offers a thorough examination of the sentiment evolution conveyed by characters within a fictitious dataset. By utilizing tools such as VADER and TextBlob, we have effectively employed sentiment analysis techniques to map the emotional trajectories of characters throughout the story, emphasizing crucial events and interactions that lead to notable shifts in sentiment.

Major story twists cause strong emotional reactions and observable changes in character mood, according to our findings. In similar vein, noticeable emotional fluctuations are closely linked to character interactions, particularly those that involve conflict or reconciliation. The significance of narrative structure and interpersonal dynamics in determining the emotional terrain of a story is highlighted by these study.

The study's findings add to our understanding of how narrative devices affect the feelings of characters and, in turn, reader involvement. For writers, literary critics, and academics interested in the mechanics of storytelling, and this research offers insightful information by measuring sentiment shifts and connecting them to particular narrative occurrences.

Subsequent investigations could build on this work by examining the evolution of sentiment in various genres or cultural context, or by enhancing sentiment analysis with more advanced machine learning models. In summary, this research highlights the efficacy of sentiment analysis as a means of revealing the intricate effective foundations of fictional stories and providing a strong structure for future studies pertaining to the affective dynamics of storytelling.

Author details

Raj Bano Khan[1*], Sajjad Muhammad Khan[2], Sufyan Muhammad Khan[3] and Waqas Muhammad Khan[4]

1 Alkhair University Bhimber Azad Jammu and Kashmir, India

2 National Center of Excellence in Geology, University of Peshawar, Pakistan

3 Sindh Madrassatul Islam University, Pakistan

4 International Islamic University, Islamabad, Pakistan

*Address all correspondence to: rajbanokhandm@gmail.com

References

[1] Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal. 2014;**5**(4):1093-1113

[2] Sylvia EC. Exploring the symbiosis of romanticism and social critique in French literature through les miserables. Nnadiebube Journal of Languages and Literatures. 2023;**1**(1)

[3] Banerjee QS. Embracing the Uncharted: Unveiling the Resurgence of Historical Forms in Contemporary Indian Theatre through Girish Karnad's Tughlaq and Nagamandala

[4] Tanenbaum J, Tomizu A. Narrative meaning creation in interactive storytelling. International Journal of Computational Science. 2008;**2**(1):3-20

[5] Patange DC. Crafting literary realms: Exploring narrative techniques in the works of Namitha Gokhle and Anitha Desai

[6] Burke M. Literary Reading, Cognition and Emotion: An Exploration of the Oceanic Mind. Routledge; 2010

[7] Rapp A, Curti L, Boldi A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. International Journal of Human-Computer Studies. 2021;**1**(151):102630

[8] Mäntylä MV, Graziotin D, Kuutila M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review. 2018;**1**(27):16-32

[9] Melville P, Gryc W, Lawrence RD. Sentiment analysis of blogs by combining lexical knowledge with text classification. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 28 June 2009. pp. 1275-1284

[10] Mohammad SM. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In: Emotion Measurement. Woodhead Publishing; 2016. pp. 201-237

[11] Kreiter S. The Role of Social Media in Political polarisation-Mapping Twitter Topic Clusters and Focusing on the Sentiment of a Political Statement

[12] Udelson J. Character Arcs: Mapping Creative Writers' Trajectories into the Composition Classroom

[13] Tu C, Brown S. Character mediation of plot structure: Toward an embodied model of narrative. Frontiers of Narrative Studies. 2020;**6**(1):77-112

[14] Ianniello M. The leftovers (rehated): Examining the layering of story and character archetypes in the television drama series: The leftovers. Comunicazioni Sociali. 2019;**3**(2019):460-474

[15] Kim E, Klinger R. A survey on sentiment and emotion analysis for computational literary studies. arXiv. 2018

[16] Riedl MO, Young RM. Narrative planning: Balancing plot and character. Journal of Artificial Intelligence Research. 2010;**29**(39):217-268

[17] Busselle R, Bilandzic H. Measuring narrative engagement. Media Psychology. 2009;**12**(4):321-347

[18] Nalisnick ET, Baird HS. Extracting sentiment networks from Shakespeare's

plays. In: 2013 12th International Conference on Document Analysis and Recognition. IEEE; 2013. pp. 758-762

[19] Nalisnick ET, Baird HS. Character-to-character sentiment analysis in Shakespeare's plays. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Vol. 2: Short Papers. 2013. pp. 479-483

[20] Elkins K, Chun J. Can sentiment analysis reveal structure in a plotless novel? arXiv. 2019

[21] Languré AD, Zareei M. Evaluating the effect of emotion models on the generalizability of text emotion detection systems. IEEE Access. 2024

[22] Li Y, Afzaal M, Yin Y. A lexicon-based diachronic comparison of emotions and sentiments in literary translation: A case study of five Chinese versions of David Copperfield. PLoS One. 2024;**19**(2):e0297101

[23] Bizzoni Y, Moreira P, Thomsen MR, Nielbo KL. The fractality of sentiment arcs for literary quality assessment: The case of Nobel laureates. Journal of Data Mining & Digital Humanities. 2023

[24] Kadhim AI. An evaluation of preprocessing techniques for text classification. International Journal of Computer Science and Information Security (IJCSIS). 2018;**16**(6):22-32

[25] Pandey M, Williams R, Jindal N, Batra A. Sentiment analysis using lexicon based approach. IITM Journal of Management and IT. 2019;**10**(1):68-76

[26] Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review. 2022;**55**(7):5731-5780

[27] Lienhart R, Wernicke A. Localizing and segmenting text in images and videos. IEEE Transactions on Circuits and Systems for Video Technology. 2002;**12**(4):256-268

[28] Kim E, Klinger R. An analysis of emotion communication channels in fan fiction: Towards emotional storytelling. arXiv. 2019

[29] Anvar Shathik J, Krishna Prasad K. A literature review on application of sentiment analysis using machine learning techniques. International Journal of Applied Engineering and Management Letters (IJAEML). 2020;**4**(2):41-67

[30] Rebora S, Pianzola F. A new research programme for reading research: Analysing comments in the margins on Wattpad. DigitCult-Scientific Journal on Digital Cultures. 2018;**3**(2):19-36

[31] Vishnubhotla K, Hammond A, Hirst G, Mohammad SM. The emotion dynamics of literary novels. arXiv. 2024

Chapter 3

AI-Driven Software Development: A Paradigm Shift in Software Engineering

Miguel Hernández Bejarano and Luis E. Baquero-Rey

Abstract

Next-generation software development (NGSD), driven by the integration of artificial intelligence (AI), has radically transformed efficiency and accuracy in software creation. Innovative tools, such as ChatGPT, LLM, GitHub Copilot, Google AI AutoML, Amazon CodeGuru, Microsoft IntelliCode, OpenAI Codex, and more, have streamlined the software development lifecycle. Not only do these tools automate repetitive tasks, but they also provide intelligent suggestions and code completion, allowing developers to focus on more creative and complex aspects of development. The implementation of these technologies has given rise to new roles within the field of programming and software development. Developers should now be familiar with using and integrating AI tools into their daily workflows. In addition, continuous training in these technologies has become essential to keep up with industry trends and advancements. In short, AI has revolutionized software development, improving developer productivity and efficiency, and creating new opportunities and roles in the industry. The adoption of AI tools is crucial for any development team looking to stay competitive in today's market.

Keywords: artificial intelligence (AI), automation, innovation, next-generation software development (NGSD), innovation, optimization

1. Introduction

Software is essential to the operation of almost every organization and can bring value to all areas of the business. Its development involves a series of IT activities, ranging from creation and design to implementation and support of the software. In order to offer effective solutions to the needs of customers, these processes are integrated to ensure a result that adequately responds to organizational requirements [1].

In turn, the software development life cycle (SDLC) is a comprehensive methodology for creating applications in organizations. It details the tasks required at each stage of development, ensuring the generation of high-quality software that meets customer expectations and is delivered on time and on budget [2].

In this sense, SDLC is the most widely used method for building software, encompassing proper analysis, planning, execution, and support. The SDLC provides a detailed plan that describes how to develop, program, and maintain specific software. It acts as an organizational framework that details the tasks to be performed at each stage of the software development process [3].

In addition, the SDLC consists of several key phases. Planning involves gathering requirements and drawing up a detailed plan with cost and resource estimates. In the design phase, technological solutions are defined and the integration of the software is planned. Implementation focuses on product coding, followed by rigorous testing to detect errors and ensure compliance with requirements. Finally, deployment puts the software in a production environment and maintenance is concerned with fixing bugs, resolving problems, and continuously improving the software [4, 5].

Likewise, in software development, developers perform a series of technical and collaborative tasks on a daily basis. These include activities such as implementing new features, fixing bugs, and configuring environments, in addition to tasks such as gathering information needs and requirements analysis, architecture design, code development, module integration, and documentation. Activities also include the creation of manuals, software testing, and user training. These responsibilities combine technical and communication skills to ensure the quality and functionality of the software [6, 7].

On the other hand, AI optimizes software development, improves user experience, and increases industrial productivity, thus driving economic and technical advancement. As an innovator in technology, AI is transforming businesses and everyday aspects. Its application in software production, using algorithms and machine learning, has enabled unprecedented creativity, reflecting the rapid technological progress in this field [8].

In turn, the construction of software within the SDLC and in view of the large volume of data that could be processed at each stage, implies a growing complexity of the software that slows down its processing, which with AI techniques such as neural networks, machine learning and data mining, are significantly improving the activities of the SDLC, including requirements engineering, architecture design, coding, software testing, and automation of the software development process [9].

Also, white papers aimed primarily at specialists in these fields highlight potential opportunities to integrate AI into various areas of software development and in aspects, such as maintenance, quality management, supply chain, engineering design, production, marketing, climate forecasting, banking, digital telecommunications, and healthcare [10].

According to a study by Siemens, carried out in collaboration with Longitude Research Ltd., on the growth and transformation of artificial intelligence in industrial corporations, 70% of companies declare that they achieve this by establishing links with AI experts and consultants; 38% implement their own development; 31% integrate their AI into the distributor's software; and 18% use ready-made solutions to use open-source models or AI services from cloud service providers [10].

To this extent, AI is becoming a fundamental pillar in technology, transforming both our daily lives and business practices. In software testing, AI introduces significant change by automating repetitive tasks, improving accuracy through advanced data analytics, and predicting future needs based on historical data. These innovations streamline the testing process, enabling faster and more efficient deployment of high-quality software [11].

Recent advances in code generation using large language models (LLMs) have significantly increased the ability to automate tasks in software engineering. LLMs can improve human skills by modularly automating parts or even all of the software development tasks, thereby streamlining the process and increasing efficiency [12].

2. Software development with generative AI

The software development industry is going through a disruptive transformation with the adoption of generative artificial intelligence (GenAI) assistants for programming. Although AI is already used in various areas of software, technologies, such as GitHub Copilot and ChatGPT, are capturing attention and generating both enthusiasm and concern among professionals in the sector [13].

NGSD, driven by the integration of AI, is transforming the industry by enabling developers to create more efficient and customized solutions. The introduction of large GPT-based language models and services is accelerating software development and maintenance, optimizing tools and workflows, and pushing traditional methods, such as Agile and DevOps. Generative AI stands out as a key disruption, improving productivity, costs, and quality through automation and process optimization [14].

3. Generative artificial intelligence technologies

GenAI tools, powered by AI and machine learning techniques, have great potential for software development by automating tasks and optimizing procedures. These tools can increase the efficiency and effectiveness of developers at all levels, facilitating their training and performance. In addition, they can greatly increase developer productivity by offering code completion and suggestions, automating testing, assisting in code refactoring, and optimizing software. Some of these tools are discussed below.

AI-powered code wizards are advanced tools that can simplify and speed up programming. They use machine learning and natural language processing algorithms to collaborate with developers, suggesting code snippets (such as method implementations) that can be integrated into projects. These assistants are gaining popularity for their ability to improve productivity and code quality. AI-based code assistants, such as GitHub Copilot, Tabnine, ChatGPT, and Gemini, have been evaluated on their ability to generate methods, focusing on their ability to produce accurate, correct, and efficient code. The results of this evaluation indicate that while these wizards are useful and complementary tools, they rarely generate fully out-of-the-box code [15, 16].

ChatGPT has proven to be advantageous in various fields, such as education, design, architecture, advertising, marketing, media, cybercrime, climate change, among others. It is capable of generating human-like texts, debugging software, writing essays, music and fairy tales, etc. In the field of health, it has been used to evaluate and simplify the quality of medical reports [17, 18].

ChatGPT is a sophisticated language model created by OpenAI, which employs the innovative generative pretrained transformer (GPT) architecture. He stands out for his remarkable skills in natural language processing, obtained through exhaustive pretraining with the GPT framework. ChatGPT along with OpenAI have achieved wide acceptance across industries. In the field of computer vision, they are used for tasks, such as object identification, image creation and processing, encouraging innovative breakthroughs. In the video game industry, they facilitate the development of virtual characters and the elaboration of stories and dialogs, offering more immersive user experiences [19].

GitHub Copilot is an AI-based code assistant that helps developers write code more efficiently. Using advanced language models, such as OpenAI's, Copilot suggests lines of code and complete functions in real time, adapting to the context of the project. While it does not always generate production-ready code, it is a valuable tool that speeds up the development process and improves productivity. Its ability to learn and adapt to different coding styles makes it an indispensable resource for modern programmers [20].

OpenAI Codex is the model behind GitHub Copilot, mastering more than a dozen programming languages and can now interpret commands in natural language, enabling natural language interfaces for existing applications. OpenAI Codex, derived from GPT-3, is trained on natural language and public source code, excelling in Python and other languages, such as JavaScript and PHP. Its ability to generate functional code from English commands makes programming easier, especially in repetitive tasks [21].

IBM watsonx code assistant uses generative AI to streamline development, keeping trust, security, and compliance top of mind. Developers and IT operators can accelerate application modernization and automate processes to scale IT environments quickly. It offers AI-based code recommendations, either from natural language or existing code, making it easier to write new code or convert high-quality code by reducing cognitive effort. Developers can review and compare code hints to the original source before adopting them [22].

DeepCode is a software security tool designed to assist developers in detecting and resolving vulnerabilities in their code. The platform integrates seamlessly with existing development workflows, allowing teams to ensure quality without compromising speed and efficiency. DeepCode identifies security flaws in open-source dependencies, connects with version control systems, such as GitHub and Bitbucket, and continues to monitor dependencies after deployment to detect new vulnerabilities. In addition, it facilitates the management of open-source licenses to avoid legal inconveniences.

Amazon CodeGuru Application Security Static Analysis (SAST) tool that uses machine learning (ML) and automated reasoning to detect vulnerabilities in code. It provides suggestions on how to fix the vulnerabilities found and monitors their status until they are resolved [23, 24].

Tabnine is a powerful tool that can revolutionize the way code is developed. If you want to improve productivity and code quality, it is definitely worth the attention. This AI-based code wizard streamlines and simplifies the programming process, while ensuring that code remains private, protected, and compliant [25, 26].

Extensive language models (LLMs) represent an advanced evolution of pre-trained models (PLMs), developed by increasing their size, training with large data sets, and complex calculations. Due to their large scale and training in large volumes of text, LLMs display exceptional abilities, achieving outstanding performance in various natural language processing tasks without requiring specific training. These models, which are a novelty in AI, have attracted considerable attention in academic and industrial environments for their outstanding efficiency in natural language processing (NLP) [27].

Google AI AutoML ML tool set that makes it easy to train high-quality, custom models with minimal effort and without the need for prior experience in the area, emerged to increase productivity and efficiency when automating [28, 29].

For error correction by automatically generating one-line patches, that is, replacing a faulty line with a correct line. While these patches are specific, the purpose is to make fixes across the board, without relying on language, based on ML. This approach aims to establish a connection between program repair and ML, allowing the community to benefit from broader datasets and continuous improvements in machine learning algorithms and libraries [30, 31].

SonarQube is a static analysis tool that detects and evaluates the technical debt of code, referring to poorly written code that violates coding standards or best practices (e.g., duplicate code). SonarQube can be used to identify and measure this technology or code debt. It is an open-source software tool that identifies and quantifies these bad practices, widely used in the industry and in software developer communities [32, 33].

Continuous integration and continuous delivery (CI/CD) is a software development method that allows teams to automate the build, test, and deployment of applications, resulting in faster and more efficient software delivery. However, the CI/CD workflow can become a potential target for vulnerabilities and attacks, leading to serious consequences for the software development process and overall system security. In this sense, Snyk is a tool that reports a variety of failures, including insecure dependencies, outdated libraries, and widespread code injection failures [34].

4. Social context and the code

The development and adoption of AI have raised issues, such as algorithmic bias, privacy, and legal and ethical issues. To address these challenges, governments, international organizations, and tech giants, such as Microsoft and IBM, promote trustworthy AI. In 2018, Microsoft introduced six principles for the responsible use of AI, including fairness and security, while IBM has pushed the concept of trustworthy AI to ensure trust in systems [35].

However, the use of these technologies also presents significant challenges for the industry and software developers. Keeping up with emerging technologies can be a daunting and expensive task. In addition, the incorrect implementation of AI can lead to unintended or biased results, which raises ethical and safety concerns.

Generative AI can create and spread false information in text, images, audio, and video, which can be used maliciously to deceive people and distort reality. This potential for "deep counterfeiting" poses significant risks, including the generation of false public statements and the spread of misinformation, which can negatively affect business leaders and the decision-making of the public and investors. In addition, there is an urgent need to adapt and train software developers to handle these ethical challenges and reduce the impact on employment [36].

In turn, a software developer does not operate in isolation, and his success is conditioned by various factors, such as interactions with the team and the use of tools and resources to obtain information. You can work on a piece of code alone or in collaboration with others, with QA and code reviews being crucial elements in the development process [37].

Today, the use of AI has become essential in numerous disciplines, including software engineering. By examining various data sources generated in this field, AI can offer key insights into user behavior, product performance, failures, and errors, among other aspects [38].

Recent technological advances have fostered a growing use of AI in various sectors, such as medicine. In the field of clinical surgery, the incorporation of AI could significantly transform the preoperative, intraoperative, and postoperative phases of orthopedic surgery [39].

The use of ChatGPT and other AI-based code generation tools, such as GitHub Copilot, OpenAI's Codex, and Tabnine, to get code recommendations is becoming very common among developers. However, this use comes with risks as generative models can produce vulnerable code, and there are controversial debates about potential copyright and licensing violations. Despite this, these tools can assist in various tasks, from writing code according to a natural language specification to designing software architecture, generating tests or correcting errors [40].

In turn, the LLM era began with OpenAI's GPT-3 model, and its popularity has grown tremendously with the advent of models, such as ChatGPT and GPT-4. These extensive language models (LLMs) have attracted great interest in academia and industry due to their outstanding performance in NLP tasks. Fundamentally, they are transformer-based deep learning models, trained on large textual datasets and adjusted to human preferences through meta-training [27].

As the use of AI in cybersecurity grows, so does the demand for secure and decentralized AI systems to protect against cyber threats. Blockchain technology has emerged as the ideal approach to improve the security and privacy of these systems, providing decentralized and immutable data storage. Secure software development helps developers from security design to testing, allowing threats to be identified, countermeasures to be selected, and their effectiveness to be verified through static evaluations and specific tests [41, 42].

In addition, OWASP Zed Attack Proxy (ZAP) is an efficient, effective, and integrated penetration testing tool that detects vulnerabilities in web applications. It is designed for ease of use for new developers and penetration testers, as well as security auditors [43].

Likewise, some developers do not see AI as an immediate threat to their job security, although they expressed concerns about future stability. Despite this, those who observe an increase in productivity thanks to AI tools also showed greater confidence in these technologies, although at the same time they felt a greater sense of job insecurity [44].

5. Role of the software developer

Software development is a collaborative endeavor that encompasses various roles and responsibilities, all essential to the creation and delivery of products. These roles include technical leaders, IT department heads, support staff, and engineering students. In addition, there are other positions, such as developers, principal architects, and vendor project managers [45].

In turn, companies that want to maintain their current codes and applications, as well as update software in highly regulated sectors, such as finance, healthcare, law, government, and telecommunications, relying solely on an AI software engineer for these tasks may be less attractive [46].

AI tools, such as OpenAI Codex and ChatGPT, have emerged, ushering in a new era of AI-based assistance for programmers. The performance of these tools is encouraging. For example, Codex, with 12 billion parameters and trained on 54 million GitHub repositories, demonstrates great potential. Recent research indicates that tools, such as ChatGPT, can offer code suggestions and solve programming problems. Using ChatGPT helps instructors improve the critical thinking skills of novice programmers [27, 47].

Low-code tools are gaining popularity in enterprise software development for their speed and simplicity. With these platforms, developers write less code and focus on the goal of the system being developed. There is a perceived requirement for certified professionals with experience in these platforms. This implies that companies with greater resources can attract and retain these specialists. These platforms facilitate software design and development through a graphical interface and predefined components, allowing users to create complete solutions by combining elements, such as user interfaces, business objects, and data into one intuitive interface [48].

In addition, state-of-the-art language models, such as the GPT series, and their potential applications in various fields have advanced significantly. Crucial innovations, such as extensive pre-training that absorbs global knowledge, instruction-based fine-tuning, and human feedback reinforcement learning (RLHF) have markedly improved the flexibility and performance of these models. A notable advancement in this area is InstructGPT, a framework that facilitates the fine-tuning of pre-trained language models using RLHF [49].

In this sense, NGSD is undergoing a significant transformation with AI integration. Programmers and developers are taking on new roles that combine knowledge in programming, AI, data, and ethics. In this context, their role is becoming more multidisciplinary, focusing on designing intelligent algorithms, integrating AI systems, optimizing performance, and ensuring ethics and safety in the use of AI. These roles are essential to creating future technologies that are efficient and aligned with business goals.

NGSD is undergoing a radical transformation thanks to the integration of AI. Programmers and software developers are taking on new roles and responsibilities in this new paradigm. Some of the essential roles:

The design and development of AI models in the creation of ML algorithms for specific tasks, such as classification, prediction, and content generation. This requires data preparation and cleansing, which includes collecting, organizing, and transforming data to feed models. In addition, model training and evaluation focuses on optimizing these algorithms for the best possible performance [50].

As a software developer with a background in artificial intelligence, he specializes in integrating AI capabilities into new and existing applications, creating intuitive user interfaces that facilitate natural interaction with AI-based systems. In addition, it optimizes performance to ensure that applications work efficiently and effectively [51, 52].

Specialist in the design of scalable AI systems, creating software architectures capable of handling large volumes of data and workloads. With evaluation and selection of the most suitable AI technologies and frameworks for each project, in addition to managing and maintaining the infrastructure necessary to run AI applications efficiently [38, 53, 54].

Specialist in data management and analysis, ranging from collection and storage to analysis of large data sets. I create visualizations that make it easier to understand the data and results of AI models and transform the data into formats suitable for AI model training [55, 56].

The development of ethical AI systems, ensuring transparency and accountability. It identifies and eliminates bias in data and AI models, as well as implements security measures to protect users' privacy [57–59].

In addition to technical knowledge in programming and AI, developers must develop skills such as critical thinking to solve complex problems, creativity to design unique solutions, collaboration to work in multidisciplinary teams, and continuous learning to keep up with the latest trends in AI.

Knowledge about the active use of AI benefits both users and businesses. It is used to improve cybersecurity and software development, addressing the challenges and risks involved in implementing algorithms and learning methods applied to security technologies and processes. There are several AI models in cybersecurity, such as ML, deep learning (DL), generative AI (GenAI), and NLP, which are also used in the development of secure software [60].

6. Discussion

Among the findings in the literature review are: the integration of AI techniques, such as artificial neural networks and fuzzy logic, to automate software design, allowing useful knowledge to be extracted throughout the SDLC. As well as the development of software supported by AI includes the significant improvement in the efficiency and accuracy of coding processes. AI allows errors to be detected and corrected automatically, which reduces development time and improves software quality. In addition, the optimization of resources through ML allows a more effective allocation of them, boosting the productivity of development teams [9], and that AI-powered software development highlights the importance of addressing ethical issues, such as bias mitigation and algorithmic transparency. In governance, it is essential to establish regulatory frameworks that ensure responsibility and accountability. In cybersecurity, developers must implement robust measures to protect data and detect vulnerabilities in AI-generated code.

Software developers working with generative AI must acquire skills in ethics, governance, and cybersecurity to ensure the responsible and safe development of advanced technologies. These competencies will be essential to address the emerging challenges and risks in the use of AI.

In ethics, it is crucial to identify and mitigate biases, promote fairness and transparency in algorithms, and understand the social repercussions of their developments. Ethical issues related to generative AI include bias, plagiarism, and false textual content. Ethical considerations linked to advanced technologies encompass issues of fairness, accountability, and transparency, and developers should focus on the ethical implementation of generative AI to improve organizational performance [61, 62].

In governance, the rapid advancement of generative AI has captured global attention and posed challenges to today's data governance frameworks. Growing technical complexity and massive use of data complicate the regulation of artificial intelligence and pose significant ethical challenges. It is essential to establish a regulatory framework that guarantees security, privacy, fairness and transparency, continuously adapting to the rapid evolution of technology. It has been argued that generative AI presents multifaceted risks, including political, military, cultural, ethical, and data-related concerns, which has created widespread apprehension and increased attention to AI risk governance. Therefore, it is necessary to participate and promote the creation of policies that ensure transparency and accountability, implement regulatory frameworks and promote responsible practices [63, 64].

In cybersecurity, generative AI as a support tool has become common in software development, highlighting the importance of its responsible use. LLMs are now widely adopted in software engineering, but professionals have expressed concerns about the security of these tools. Code generated by LLMs can contain vulnerabilities and pose risks related to data privacy and security in AI [65–67].

7. Conclusions

NGSD, driven by the integration of AI, has marked a before and after in the software industry. The adoption of innovative tools, such as Chat GPT, LLM, GitHub Copilot, Google AI AutoML, Amazon CodeGuru, Microsoft IntelliCode, and OpenAI Codex, has radically transformed the software development lifecycle, improving both efficiency and accuracy at every stage of the process.

These technologies have not only streamlined traditional programming tasks but have also redefined the roles of software developers. Programmers can now focus on more creative and strategic aspects of development, delegating repetitive and less value-added tasks to AI tools. This not only increases productivity but also encourages innovation and the quality of the software produced.

The integration of AI into software development is not just a passing trend, but a necessary evolution that will continue to shape the future of the industry. Developers and organizations that adopt these technologies will be better positioned to meet the challenges and take advantage of the opportunities offered by this new paradigm.

Secure software development with AI for the next generation of software (NGSD) focuses on the integration of advanced ML algorithms and predictive analytics to identify and mitigate vulnerabilities in real time. This methodology not only improves security but also optimizes development performance and efficiency, enabling the creation of robust and reliable applications from their initial phases.

For this work, some tools used include SonarQube for static code analysis, OWASP ZAP for penetration testing, GitHub Advanced Security for detecting vulnerabilities in repositories, and Snyk for dependency management and security in open source.

Notes/acknowledgements/other statements

The authors thank the Central Technical Institute Technological School (ETITC) for the support granted for the realization of this paper within the framework of the ongoing research project.

Finally, we would like to thank the professors of the PhD program in Systems Engineering and Computer Science at the University of Zaragoza (Spain) for the guidance and training received during our research. Their support has been instrumental in substantially improving our writing in scientific publications.

Conflict of interest

The authors state that they have no competing financial interest or known personal relationships that would have appeared to influence the work presented in this article.

Author details

Miguel Hernández Bejarano* and Luis E. Baquero-Rey
K-demy Research Group, Escuela Tecnológica Instituto Técnico Central, Bogotá D.C., Colombia

*Address all correspondence to: mhernandezb@itc.edu.co

References

[1] Madhavaram S, Hunt SD, Arnett DB. Building capabilities for software development firm competitiveness: The role of intellectual capital and intra-firm relational capital. Information and Management. 2023;**60**(2):103744. DOI: 10.1016/j.im.2022.103744

[2] Olusanya OO et al. A neuro-fuzzy security risk assessment system for software development life cycle. Heliyon. 2024;**10**(13):e33495. DOI: 10.1016/j.heliyon.2024.e33495

[3] Sunil G, Gayathri N. Study of the software development life cycle and the function of testing. In: 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC). 2022. pp. 1270-1275. DOI: 10.1109/IIHC55949.2022.10060231

[4] AWS. What is the software development life cycle (SDLC)? Amazon: AWS; 2023. Available from: https://aws.amazon.com/es/what-is/sdlc/ [Accessed: August 14, 2024]

[5] Vamsi K, Gopinath G. Software development life cycle for web application by using traditional methodology vs agile methodology. In: 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). 2024. pp. 1-6. DOI: 10.1109/ICONSTEM60960.2024.10568596

[6] Masood A, Khan S, Ali R. Like, dislike, or just do it? How developers approach software development tasks. Information and Software Technology. 2022;**150**:106963. DOI: 10.1016/j.infsof.2022.106963

[7] Davis DK, Zhu F. Analysis of software developers' coding behavior: A survey of visualization analysis techniques using eye trackers. Computers in Human Behavior Reports. 2022;**7**:100213. DOI: 10.1016/j.chbr.2022.100213

[8] Kumar R, Singh A, Gupta P. Game-changing intelligence: Unveiling the societal impact of artificial intelligence in game software. Entertainment Computing. 2025;**52**:100862. DOI: 10.1016/j.entcom.2024.100862

[9] Divanshi PW. Wangoo DP. Artificial intelligence techniques in software engineering for automated software reuse and design. In: 2018 4th International Conference on Computing Communication and Automation (ICCCA). 2018. pp. 1-4. DOI: 10.1109/CCAA.2018.8777584

[10] Dzhusupova R, Bosch J, Olsson H. Choosing the right path for AI integration in engineering companies: A strategic guide. Journal of Systems and Software. 2024;**210**:111945. DOI: 10.1016/j.jss.2023.111945

[11] Yogendra K. An enhanced role of AI in SDLC in implementing maintenance system. In: 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). 2024. pp. 1064-1068. DOI: 10.1109/ICACITE60783.2024.10616905

[12] Mira L, Gail M. On automated assistants for software development: The role of LLMs. In: 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). 2023. pp. 1737-1741. DOI: 10.1109/ASE56229.2023.00035

[13] Christopher B, Ahmed K. Generative artificial intelligence assistants in

software development education: A vision for integrating generative artificial intelligence into educational practice, not instinctively defending against it. IEEE Software. 2024;**41**:52-59. DOI: 10.1109/MS.2023.3300574

[14] Sauvola J, Mikkonen T, Leppänen V. Future of software development with generative AI. Automated Software Engineering. 2024;**31**:26-31. DOI: 10.1007/s10515-024-00426-z

[15] Vincenzo G, Rossi M, Bianchi L, Corso et al. Generating Java methods: An empirical assessment of four AI-based code assistants. In: 2024 IEEE/ACM 32nd International Conference on Program Comprehension (ICPC). 2024. pp. 13-23

[16] Vincenzo G, Rossi M, Bianchi L, Corso et al. Assessing AI-based code assistants in method generation tasks. In: 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion). 2024. pp. 308-381. DOI: 10.1145/3639478.3643122

[17] Mohammad A, Ali B, Hassan S. ChatGpt: Open possibilities. Iraqi Journal For Computer Science and Mathematics. 2023;**4**:62-64

[18] Handa A, Sharma R, Gupta S. Exploring the role of ChatGPT in medical image analysis. Biomedical Signal Processing and Control. 2023;**86**:105292. DOI: 10.1016/j.bspc.2023.105292

[19] Zhang H, Shao H. Exploring the latest applications of OpenAI and ChatGPT: An In-depth survey. CMES Computer Modeling in Engineering and Sciences. 2023;**138**(3):2061-2102. DOI: 10.32604/cmes.2023.030649

[20] GitHub. What is GitHub Copilot? GitHub Copilot: GitHub, Inc.; 2024. Available from: https://docs.github.com/en/copilot [Accessed: August 15, 2024]

[21] OpenAI. Códice OpenAI. 2021. Available from: https://openai.com/index/openai-codex/ [Accessed: August 16, 2024]

[22] IBM. Meet Watsonx. 2024. Available from: https://www.ibm.com/products/watsonx-code-assistant [Accessed: August 20, 2024]

[23] AWS. Amazon CodeGuru Security. Amazon: AWS; 2023. Available from: https://aws.amazon.com/es/codeguru/ [Accessed: August 15, 2024]

[24] Cifuentes C, Roychoudhury A, Padioleau Y. The role of program analysis in security vulnerability detection: Then and now. Computers & Security. 2023;**135**:103463. DOI: 10.1016/j.cose.2023.103463

[25] Tabnine. What is Tabnine? Available from: https://docs.tabnine.com/main [Accessed: August 16, 2024]

[26] Savary-Leblanc M, Le Pallec X, Gérard S. The modeling assistant recommender: A UML class diagram recommender system. Science of Computer Programming. 2024;**234**:103090. DOI: 10.1016/j.scico.2024.103090

[27] Subramanyam KK. A survey of GPT-3 family large language models including ChatGPT and GPT-4. Natural Language Processing Journal. 2024;**6**:100048. DOI: 10.1016/j.nlp.2023.100048

[28] Google. AutoML. Google Cloud. Available from: https://cloud.google.com/automl?hl=es [Accessed: August 15, 2024]

[29] Salehin I et al. AutoML: A systematic review on automated machine learning

with neural architecture search. Journal of Information and Intelligence. 2024;**2**(1):52-81. DOI: 10.1016/j.jiixd.2023.10.002

[30] Zimin A, Kim J, Lee S. SequenceR: Sequence-to-sequence learning for end-to-end program repair. IEEE Transactions on Software Engineering. 2024;**47**(9):1943-1959. DOI: 10.1109/TSE.2019.2940179

[31] Michihiro Y, Percy L. Break-it-fix-it: Unsupervised learning for program repair. In: Proceedings of the 38th International Conference on Machine Learning. Vol. 139. 2021. pp. 11941-11952

[32] Alfayez F, Alshammari M, Alshammari R. How SonarQube-identified technical debt is prioritized: An exploratory case study. Information and Software Technology. 2023;**153**:107147. DOI: 10.1016/j.infsof.2023.107147

[33] Baldassarre M, Di Nucci D, Palomba F. On the diffuseness of technical debt items and accuracy of remediation time when using SonarQube. Information and Software Technology. 2020;**128**:106377. DOI: 10.1016/j.infsof.2020.106377

[34] Sushma R, Kumar A, Singh P. 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)2023, 2023. pp. 1-10. DOI: 10.1109/CSITSS60515.2023.10334136

[35] Tao Z, Yi Q, Qiang L. Trusted artificial intelligence: Technique requirements and best practices. In: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2021. pp. 1458-1462. DOI: 10.1109/TrustCom53373.2021.00207

[36] Wang J, Li H, Zhang Y. Legal risks and governance paths for generative AI—A case study of ChatGPT. In: 2023 13th International Conference on Information Technology in Medicine and Education (ITME). 2023. pp. 58-62. DOI: 10.1109/ITME60234.2023.00023

[37] France SL. Navigating software development in the ChatGPT and GitHub Copilot era. Business Horizons. 2024;**67**(5):649-661. DOI: 10.1016/j.bushor.2024.05.009

[38] Figalist M, Müller R, Schmidt T. Breaking the vicious circle: A case study on why AI for software analytics and business intelligence does not take off in practice. Journal of Systems and Software. 2022;**184**:111135. DOI: 10.1016/j.jss.2021.111135

[39] Geda MW, Ming Y, Lee CK. Applications of artificial intelligence in orthopaedic surgery: A systematic review and meta-analysis. Engineering Applications of Artificial Intelligence. 2024;**133**:108326. DOI: 10.1016/j.engappai.2024.108326

[40] Nguyen T, Tran D, Le H. GPTSniffer: A CodeBERT-based classifier to detect source code written by ChatGPT. Journal of Systems and Software. 2024;**214**:112059. DOI: 10.1016/j.jss.2024.112059

[41] Shamsan Saleh AM. Blockchain for secure and decentralized artificial intelligence in cybersecurity: A comprehensive review. Blockchain: Research and Applications. 2024;**5**(3):100193. DOI: 10.1016/j.bcra.2024.100193

[42] Casola V, Rak M, Villano U. Secure software development and testing: A

model-based methodology. Computers & Security. 2024;**137**:103639. DOI: 10.1016/j.cose.2023.103639

[43] Chahal K, Singh J, Kaur H. A proactive approach to assess web application security through the integration of security tools in a security orchestration platform. Computers & Security. 2022;**122**:102886. DOI: 10.1016/j.cose.2022.102886

[44] Kuhail M, Alshammari R, Alshammari M. “Will I be replaced?” Assessing ChatGPT’s effect on software development and programmer perceptions of AI tools. Science of Computer Programming. 2024;**235**:103111. DOI: 10.1016/j.scico.2024.103111

[45] Gunatilake et al. The impact of human aspects on the interactions between software developers and end-users in software engineering: A systematic literature review. Information and Software Technology. 2024;**173**:107489. Available from: https://www.sciencedirect.com/science/article/pii/S0950584924000946

[46] VB, Venture Beat. [Online]. 2024. Available from: https://acortar.link/6t2beo [Accessed: August 19, 2024]

[47] Bucaioni et al. Using ChatGPT to help instructors improve the critical thinking skills of novice programmers. Machine Learning with Applications. 2024;**15**:100526. DOI: 10.1016/j.mlwa.2024.100526

[48] Martins J, Branco F, Mamede H. Combining low-code development with ChatGPT to novel no-code approaches: A focus-group study. Intelligent Systems with Applications. 2023;**20**:200289. DOI: 10.1016/j.iswa.2023.200289

[49] Liu et al. Summary of ChatGPT-related research and perspective towards the future of large language models. Meta-Radiology. 2023;**1**:100017. DOI: 10.1016/j.metrad.2023.100017

[50] Moroz E, Grizkevich V, Novozhilov IM. The potential of artificial intelligence as a method of software developer’s productivity improvement. In: 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). 2022. DOI: 10.1109/ElConRus54750.2022.9755659

[51] Takefuji Y. Generative AI for analysis and identification of Medicare improper payments by provider type and HCPC code. Exploratory Research in Clinical and Social Pharmacy. 2023;**12**:100387. DOI: 10.1016/j.rcsop.2023.100387

[52] Rasmussen E et al. RadDeploy: A framework for integrating in-house developed software and artificial intelligence models seamlessly into radiotherapy workflows. Physics and Imaging in Radiation Oncology. 2024;**31**:100607. DOI: 10.1016/j.phro.2024.100607

[53] Osama HN. Artificial intelligence’s effects on design process creativity: “A study on used A.I. Text-to-Image in architecture”. Journal of Building Engineering. 2023;**80**:107999. DOI: 10.1016/j.jobe.2023.107999

[54] Malavolta I, Muccini H, Ozkaya I. Software architecture and artificial intelligence. Journal of Systems and Software. 2022;**193**:111436. DOI: 10.1016/j.jss.2022.111436

[55] An et al. Modeling agent decision and behavior in the light of data science and artificial intelligence. Environmental Modelling and Software. 2023;**166**:105713. DOI: 10.1016/j.envsoft.2023.105713

[56] Safitra F et al. Advancements in artificial intelligence and data science: Models, applications, and challenges. Procedia Computer Science. 2024;**234**:381-388. DOI: 10.1016/j.procs.2024.03.018

[57] Vikrant KA. Artificial intelligence technology readiness for social sustainability and business ethics: Evidence from MSMEs in developing nations. International Journal of Information Management Data Insights. 2024;**4**:100250. DOI: 10.1016/j.jjimei.2024.100250

[58] Liang et al. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. Automation in Construction. 2024;**166**:105369. DOI: 10.1016/j.autcon.2024.105369

[59] Reka S et al. Big data analytics and artificial intelligence aspects for privacy and security concerns for demand response modelling in smart grid: A futuristic approach. Heliyon. 2024;**10**:e35683. DOI: 10.1016/j.heliyon.2024.e35683

[60] Kulkarni AV, et al. Chapter 101—Use of artificial intelligence (AI) in cybersecurity. In: Computer and Information Security Handbook. 2025. pp. 1617-1624. DOI: 10.1016/B978-0-443-13223-0.00101-6

[61] Rana et al. Assessing the nexus of generative AI adoption, ethical considerations and organizational performance. Technovation. 2024;**135**:103064. DOI: 10.1016/j.technovation.2024.103064

[62] Riemer K, Peter S. Conceptualizing generative AI as style engines: Application archetypes and implications. International Journal of Information Management. 2024;**79**:102824. DOI: 10.1016/j.ijinfomgt.2024.102824

[63] Ye et al. Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective. Telecommunications Policy. 2024:102851. DOI: 10.1016/j.telpol.2024.102851

[64] Li K, Wu H, Dong Y. Copyright protection during the training stage of generative AI: Industry-oriented U.S. law, rights-oriented EU law, and fair remuneration rights for generative AI training under the UN's international governance regime for AI. Computer Law and Security Review. 2024;**135**:106056. DOI: 10.1016/j.clsr.2024.106056

[65] Sivana H, Marcelo D, Laurie M. Just another copy and paste? Comparing the security vulnerabilities of ChatGPT generated code and StackOverflow answers. In: 2024 IEEE Security and Privacy Workshops (SPW). 2024. pp. 87-94. DOI: 10.1109/SPW63631.2024.00014

[66] Maanak G. From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. IEEE Access. 2023;**11**:80218-80245. DOI: 10.1109/ACCESS.2023.3300381

[67] Jianan Y. Learning nonlinear loop invariants with gated continuous logic networks. Association for Computing Machinery. 2020;**55**:106-120. DOI: 10.1145/3385412.3385986

Chapter 4

CNN Model-Based Denseblock Vs. BiLSTM Recurrent Network Model for Arrhythmia Class Detection

Marwa Fradi, Wafa Gtifa, Ayoub Mhaouch and Mohsen Machhout

Abstract

Artificial intelligence is spreading worldwide, achieving huge success in medical image and signal analysis, coming with various deep learning algorithms. In this light, this work proposed two deep algorithm models for electroencephalogram ECG arrhythmia class detection. Therefore, two deep learning DL models have been put forward such as a CNN-based Denseblock module and BiLSTM recurrent neural network. Two different algorithm architectures are implemented on the Mit-bih-ECG database. The first one is going more deeper than the BiLSTM, which is lightweight and presents less complexity. The achieved results for both DL models are superior performance compared to previous works. The proposed Denseblock architecture has achieved a training accuracy of 98.82% and an F1-score equal to 98.78. With the aim of getting more improved results, a lightweight recurrent neural network model has been proposed, reaching accurate performances in terms of F1-score with 100 and 99.99%, respectively, for training and validation. Predicted test results show the class of such tested signal, with an accuracy of 100%. In addition, learning curves show a small fit between the training and validation learning models, where the loss reached zero error. Furthermore, normal ECG samples, as well as arrhythmic ECG signals are well detected. Thus, the performance and effectiveness of BiLSTM against CNN-based Denseblock module.

Keywords: ECG, BiLSTM, Denseblock, accuracy, F1-score, class detection

1. Introduction

Nowadays, artificial intelligence (AI) has been widespread in the medical field due to its success in assisting clinicians in getting true diagnostics at an early stage. AI research is developing more and more, coming with various deep learning algorithms such as Convolutional neural networks and recurrent neural networks. These DL models show a huge effect in ECG arrhythmia class detection with the increasing number of deaths caused by cardiac diseases. Recent networks have shown promising results in signal analysis, more especially in medical signal analysis. Many works have used recurrent neural network algorithms for ECG classification [1]. These RNNs involved many network models such as LSTM and BiLSTM. These models have been known to have huge success in various applications involving cardiac disease prediction [2, 3]. Hou et al. [4] and Gao et al. in [5] have applied LSTM in different ways to improve the efficiency and accuracy of ECG signal classification. DenseNet has achieved excellent development, being the best among most DL models thanks to its layers of connectivity with each other [6]. It presents a very huge role in medical image analysis. Subsequently, TCN has been proposed, implementing a dilated causal 1D convolution and outperforming all existing ML methods with an accuracy of 96.12% and an F1 score of 84.13% [7]. However, in [8], authors have implemented an end-to-end neural network using a combination of CNN and LSTM Gated Recurrent Unit (GRU), reaching excellent performances coming to 94.36, 89.4, 98.36, and 91.97% as recall, precision, accuracy and F1-score [8]. Authors in [9] have used the MIT-BIH database, proposing a method combining pre-processing step with a deep CNN architecture. Achieved results show promise for the training, validation, and testing accuracies, reaching 99.5, 99.06, and 99.34%, respectively. In [10], a lightweight model architecture based on the Densely Connected Convolutional Networks have been proposed to classify cardiac arrhythmia, getting promising results. Moreover, transfer-learning algorithms have been implemented from 2D-DCNN features to ECG signals for four classes of detection. Achieved results reached more than 97% [11]. These results show that the DBLSTM-WS model gives a high recognition performance of 99.39%. It has been observed that the wavelet-based layer proposed in the study significantly improves the recognition performance of conventional networks [12]. The experiment adopted ten-fold cross-validation and obtained an accuracy rate of 89.3% and an F1-score of 0.891 [13]. First, the arrhythmia sparse data is augmented by generative adversarial networks. Then, aiming at the identification of different types of arrhythmias in long-term ECG, a spatial information fusion model based on ResNet and a temporal information fusion model based on BiLSTM are proposed [14].

2. Proposed CNN method

The proposed solution for ECG classification to detect arythmia signals has put forward two deep learning models such as CNN-based Denseblock and LSTM recurrent networks models, with the aim to classify ECG signals into two classes: Normal and Arrythmia, then to predict the class of each tested image as described in **Figure 1**.

2.1 MIT-Bih dataset preparation

The used dataset is Mit-Bih available from physionet Bank [15]. Our dataset consists of 11,500 1D ECG signals, which are original. Each signal represents a vector of 1*187. Each vector represents a temporal segment of 1.49 s. We have worked only on two classes, where the first represents normal with the label 0, and the second represents supraventricular arrhythmia ECG signals with the label 1. The total number of datasets has been divided into 50, 25, and 25% for train, validation, and test, respectively.

2.2 DL classifiers algorithms

2.2.1 CNN-based Denseblock module

As depicted in **Table 1**, the Denseblock structure connects all layers directly to each other's. Each layer obtained from the precedent layer its inputs, where feature

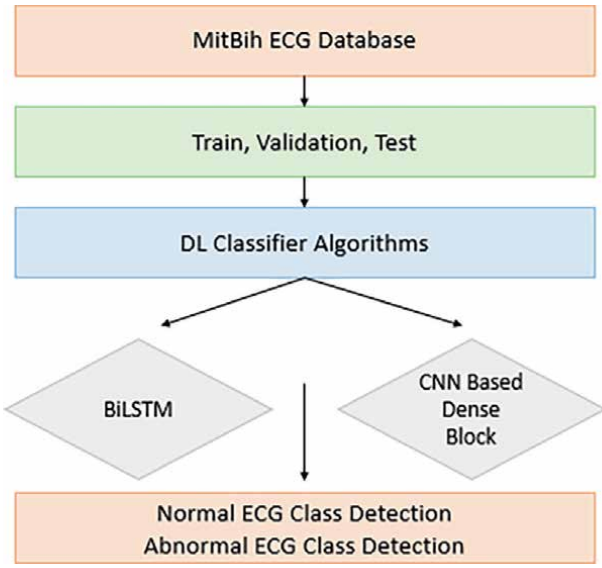


Figure 1.
Proposed methodology.

Layers	Output shape	Parameters
Input _LSTM	(None, 178,1)	0
dense	(None, 178,32)	64
Bidirectional LSTM	(None, 256)	164,864
dropout	(None, 256)	0
Batch_normalization	(none, 256)	1024
Dense_1	(None, 64)	0
dropout_2	(None, 64)	0
Batch_normalization_1	(None, 64)	256
Dense_6	(None, 2)	130

Table 1.
LSTM layers architecture.

were concatenated. Thus each layer contains the features maps of the preceding layers. The proposed implemented denseblock structure includes many layers as described in **Table 2**. thus many parameters and more complexity and finally more computational time.

2.2.2 BilSTM model

With a aim to reduce DL model complexity, a *BiLSTM* has been proposed, defined as a Bidirectional Long Short term memory recurrent neural network, has involved two LSTMs as depicted in **Figure 2**: The first take the input in a forward direction, whereas the second is going on backward direction. As described in **Figure 3**, each LSTM consists of a cell and an LSTM unit involving three principal gates such as

Layer	Output shape	Parameters	Connected to
Input:	[(None, 186)]	0	
batch_normalization	[(None, 186)]	744	input[0][0]
batch normalization_1	[(None, 186)]	744	input[0][0]
batch normalization_2	[(None, 186)]	744	input[0][0]
dense (Dense)	(None, 256)	47,872	batch_normalization [0][0]
dense (Dense)_4	(None, 512)	95,744	batch_normalization_1 [0][0]
dense (Dense)_8	(None, 1024)	191,488	batch_normalization_2 [0][0]
dropout (Dropout)	(None, 256)	0	Dense [0][0]
dropout_3 (Dropout)	(None, 512)	0	Dense_4 [0][0]
dropout_6 (Dropout)	(None, 1024)	0	Dense_8 [0][0]
dense_1 (Dense)	(None, 256)	65,792	dropout[0][0]
dense_5 (Dense)	(None, 512)	262,656	Dropout_3 [0][0]
dense_9 (Dense)	(None, 1024)	1,049,600	Dropout_6 [0][0]
dropout_1 (Dropout)	(None, 256)	0	dense_1 [0][0]
dropout_4 (Dropout)	(None, 512)	0	dense_5 [0][0]
dropout_7 (Dropout)	(None, 1024)	0	dense_9 [0][0]
dense_2 (Dense)	(None, 256)	65,792	dropout_1 [0][0]
dense_6 (Dense)	(None, 512)	262,656	dropout_4 [0][0]
dense_10 (Dense)	(None, 1024)	1,049,600	dropout_7 [0][0]
dropout_2 (Dropout)	(None, 256)	0	dense_2 [0][0]
dropout_5 (Dropout)	(None, 512)	0	dense_6 [0][0]
dropout_8 (Dropout)	(None, 1024)	0	dense_10 [0][0]
dense_3 (Dense)	(None, 186)	47,802	dropout_2 [0][0]
dense_7 (Dense)	(None, 186)	95,418	dropout_5 [0][0]
dense_11 (Dense)	(None, 186)	190,650	dropout_8 0][0]
concatenate (Concatenate)	(None, 558)	0	dense_3 [0][0] dense_7 [0][0] dense_11 [0][0]
dense_12 (Dense)	(None, 128)	71,552	Concatenate [0][0]
output (Dense)	(None, 1)	129	Dense_12 [0][0]

Table 2.
CNN-based Denseblock structure.

Forget Gate, Input gate and finally, the output gate. The cell state helps the information to flow through the units. The forget gate decides which information from the previous cell state should be forgotten for which it uses a sigmoid function as represented in **Figure 3**. The input gate controls the information flow to the current cell state using a point-wise multiplication operation of "sigmoid" and "tanh," respectively. Finally, the output gate decides which information should be passed on to the next hidden.

The implemented BiLSTM is described by the following pseudo code algorithm:

Pseudo_code of proposed BiLSTM Algorithm: Model A

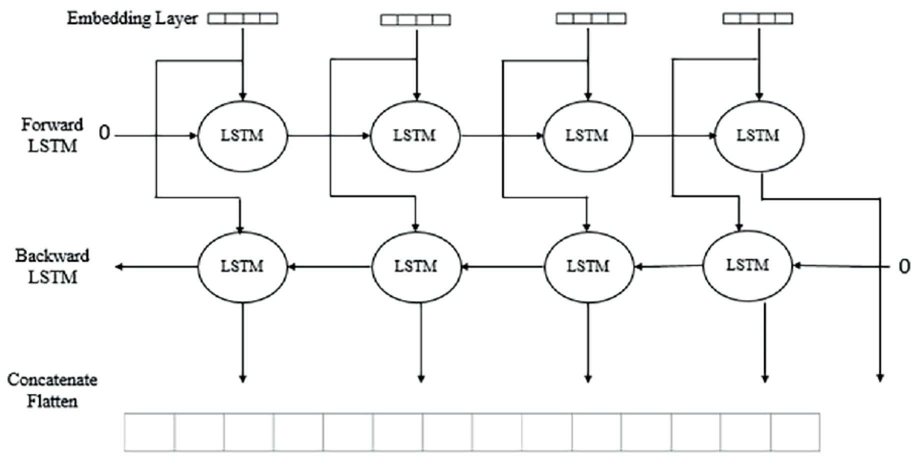


Figure 2.
BiLSTM graphical structure.

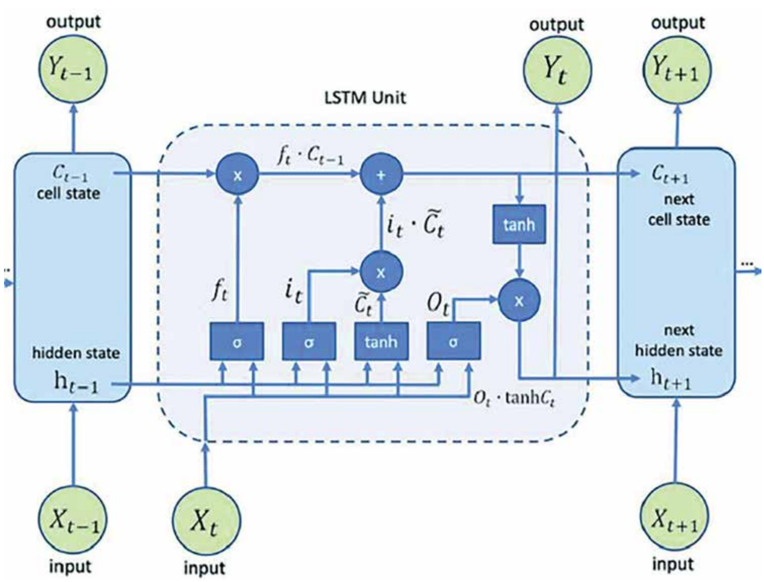


Figure 3.
LSTM graphical structure. X_t: numerical representation of the input object.

1. Input : data , data train , data test

2. Model: BiLSTM

```
dense = Dense(units=32, activation='relu', name='dense')(inputs_lstm)
lstm = layers. Bidirectional(LSTM(units=128, name='lstm'))(dense)
dropout = Dropout(0.3)(lstm)
batch_normalization
dense_1
dropout_2 = Dropout(0.3, name='dropout_2')(dense_1)
```

1. Save best weight model

2. Model Fit

model1.fit (data_train, label_train,epochs =10, batch_size=32, validation_data=(data_val,label_val)

1. Resave best weight model

2. y_pred=model1.predict(data_test)

3. Results

3.1 CNN-based Denseblock performance results

To implement our proposed algorithms, Python, Keras, and the tensor flow library were used with 4 GBRAM and an i7 processor. The MI-Bih ECG dataset has been split into three sets: 70%for the train, 25% for the validation, and 25% for the test, using a batch size equal to 32. Best performance are as depicted in **Table 3** in terms of training, validation and test accuracies have validated with 98.82, 98.56 and 98.23% respectively. F1–1 score reached more than 98% and the error shows too close to zero. These parameters are illustrated by the following equations, where TP represents the True classified signals, TN True Negative (TN), False Positive (FP) and False Negative (FN). The achieved results were explained by the deep proposed denseblock architecture, representing a huge number of parameters, representative of extracted feature weights.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1\,Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

Performances	Train	Validation	Test
Accuracy	98.82%	98.56%	98.23%
F1 Score	98.78%	98.36	98.48%
Loss	0.0004	0.0005	0.0004

Table 3.
CNN-based Denseblock performance results.

3.2 BiLSTM performance results

Best performance in terms of training, validation, and test accuracies have been reached due to BiLSTM implementation on Mit-Bih ECG original database through train, validation, and testing with 100, 100, and 100% respectively. The F1–1 score reached more than 99.9% and the error shows too close to zero (**Table 4**).

3.3 Best learning curves by BiLSTM implementation

As presented in **Figure 4**, the learning curves during training and validation process described BiLstm behavior, both accuracy and loss presented an excellent gap between the training and validation scores. In addition, accuracy result has been increased during the training from 0.42 to 100% through 10 epochs and 250 iterations/epoch. Simultaneously, validation accuracy curve shows close to the training. Moreover, losses curves represented a small error rate for both training and validation, explained by lightweight proposed LSTM architecture.

3.4 Arrhythmia ECG signals detection

Through this work, we have succeeded in detecting normal and abnormal Mit-Bih signal samples. As depicted in the following figures, normal ECG signal has been well shown, such as 730 Mit-Bih where waves P, Complex QRS, ST segment, and T wave, have normal behavior as depicted in **Figure 5**. The P wave corresponds to the atrial activation, whereas the complex QRS reflects the activation of ventricles. ST segment gives an idea if there exists an ischemic U wave that comes after the T wave. Moreover, abnormal ECG signals were well detected, as presented in 956 Mit-Bih sample. These results reflected the robustness of the proposed CNN-based denseblock and LSTM algorithms.

Performance	Train	Validation	Test
Accuracy	100%	100%	100%
F1 Score	99.99	99.99%	99.99
Loss	0.0000736	0.000005	0.0000003

Table 4.
BiLSTM Performance results.

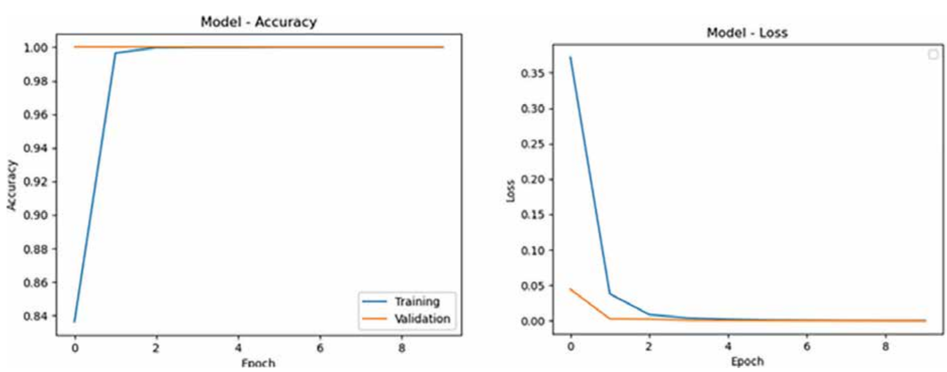


Figure 4.
Learning Curves: (a) Accuracy curves, and (b) loss curves.

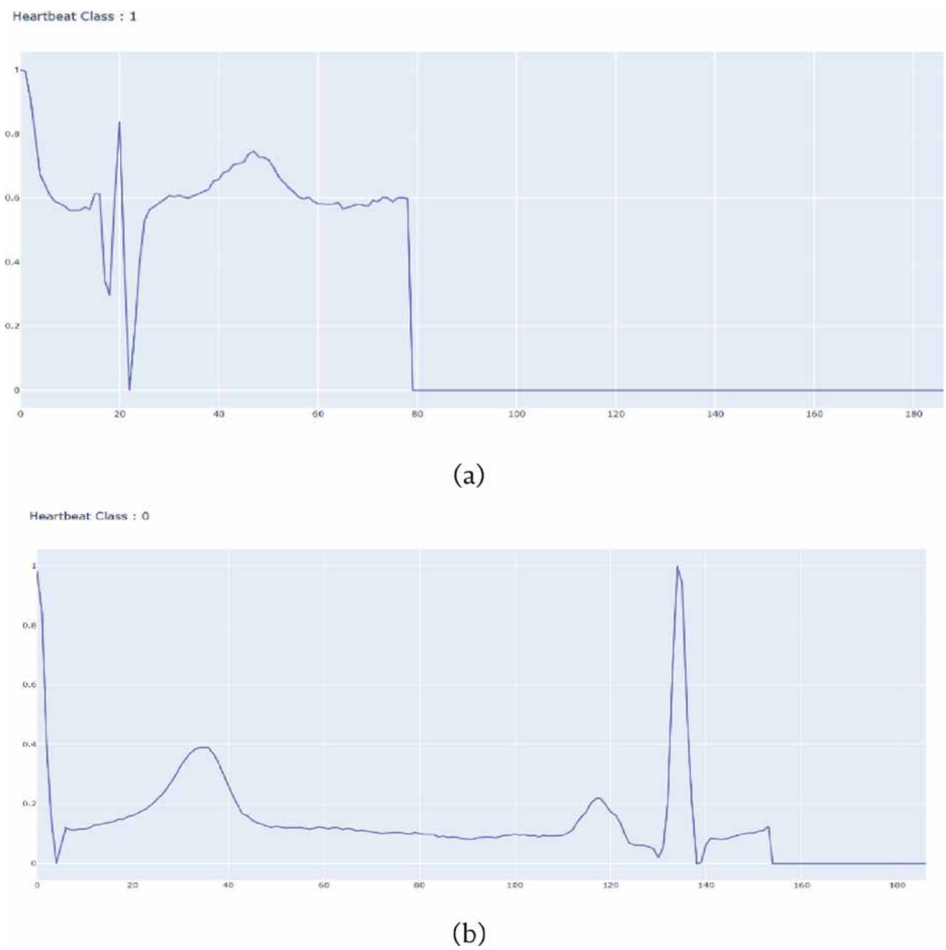


Figure 5.
ECG class detection: (a) Abnormal (1), and (b) Normal (0).

Neural Networks	Dataset	Accuracy (%)
FCN + CNN [17]	ECG Dataset	86
Dense Net [18]	ECG Dataset	89
ANN [19]	Mit-bih ECG Dataset	93
Our Work (CNN)	Mit-Bih ECG Dataset	100
Our work (LSTM)	Mit-Bih ECG Dataset	100

Table 5.
Comparative study with State-of-the-Art.

4. Discussions

The proposed two deep learning algorithms have reached accurate performance results. BiLSTM has presented lightweight architecture with less complexity. However, CNN-based dense block modules have shown more complexity because of the deep architecture where all layers were connected to each other, thus more parameters and more computational time. Achieved results have reflected the effectiveness of proposed DL algorithms, presenting superior results which have been achieved by previous works [13, 14]. Reached accuracies results show superior to state of the art works in terms of training, validation, and test accuracies with a gap of 5% [9], a gap of 3% [16], and a gap of 2% [15]. Thus, the effectiveness of our proposed method for automatic detection of arrhythmia ECG signals. Moreover, we have outperformed [17, 18] with 7% accuracy as depicted in **Table 5**, which explains the robustness of our model.

5. Conclusion

Through this work, two DL model approaches have been released and conducted to very competitive accuracies. The BiLSTM reached 100% test accuracy, showing superior performance than the first proposed one, thanks to its sample architecture, which is lightweight and has less complexity. Therefore, it is predicted that work will be focused on BiLSTM implementation on FPGA. Then, Denseblock complexity reduction will be done through CNN model compression.

Author details

Marwa Fradi[1]*, Wafa Gtifa[2], Ayoub Mhaouch[1] and Mohsen Machhout[1]

1 Laboratory of Electronics and Micro-electronics, Physic Department of Faculty of Sciences of Monastir, Monastir University, Tunisia

2 Laboratory of Automation and Electrical Systems and Environment, Monastir National School of Engineering, Monsatir University, Tunisia

*Address all correspondence to: marwa.fradi@gmail.com

References

[1] Saadatnejad S, Oveisi M, Hashemi M. LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE Journal of Biomedical and Health Informatics. 2019;**24**(2):515-523

[2] Hou B, Yang J, Wang P, Yan R. LSTM-based auto-encoder model for ECG arrhythmias classification. IEEE Transactions on Instrumentation and Measurement. 2019;**69**:1232-1240

[3] Pham B-T et al. Electrocardiogram heartbeat classification for arrhythmias and myocardial infarction. Sensors. 2023;**23**(6):2993

[4] Cui KX, Xia XJ. ECG signal anomaly detection algorithm based on CNN-BiLSTM. In: 2022 11th International Conference of Information and Communication Technology (ICTech). IEEE; 2022. pp. 193-197

[5] Gao J, Zhang H, Lu P, Wang Z. An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. Journal of Healthcare Engineering. 2019;**2019**(1):632065

[6] Zhou T et al. Dense convolutional network and its application in medical image analysis. Bio Med Research International. 2022;**2022**(1):2384830

[7] Ismail AR et al. ECG classification using an optimal temporal convolutional network for remote health monitoring. Sensors. 2023;**23**(3):1697

[8] Surbhi B et al. Classification of electrocardiogram signals based on hybrid deep learning models. Sustainability. 2022;**14**(24):16572

[9] Fradi M, Khriji L, Machhout M. Real-time arrhythmia heart disease detection system using CNN architecture based various optimizers-networks. Multimedia Tools and Applications. 2021;**81**(29):41711-41732

[10] Bui TH, Pham MT. Automatic varied-length ECG classification using a lightweight dense net model. Biomedical Signal Processing and Control. 2023;**82**:104529

[11] Salem M, Taheri S, Yuan J-S. ECG arrhythmia classification using transfer learning from 2-dimensional deep CNN features. In: 2018 IEEE Biomedical Circuits and Systems Conference (Bio CAS). IEEE; 2018. pp. 1-4

[12] Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine. 2018;**96**:189-202

[13] Cheng J, Zou Q, Zhao Y. ECG signal classification based on deep CNN and BiLSTM. BMC Medical Informatics and Decision Making. 2021;**21**:1-12

[14] Ma S et al. Deep learning-based data augmentation and model fusion for automatic arrhythmia identification and classification algorithms. Computational Intelligence and Neuroscience. 2022;**2022**(1):1577778

[15] Li P et al. High-performance personalized heartbeat classification model for long-term ecg signal. IEEE Transactions on Biomedical Engineering. 2017;**64**(1):78-86

[16] Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY, Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, arXiv preprint arXiv: 1707.01836v1, 2017

[17] Pyakillya B, Kazachenko N, Mikhailovsky N. Deep learning for ECG classification. Journal of Physics Conference Series. 2017;**913**(1):012004

[18] Guo L, Sim G, Matuszewski B. Inter-patient ECG classification with convolutional and recurrent neural networks. Biocybernetics and Biomedical Engineering. 2019;**39**(3):868-879

[19] Khriji L, Fradi M, Machhout M, Hossen A. Deep learning-based approach for atrial fibrillation detection. In: Springer Book Chapter in the Impact of Digital Technologies on Public Health in Developed and Developing Countries, COST 2020, Hammamet, Tunisia, June 24-26, 2020, Proceedings 18. Springer International Publishing; 2020. pp. 100-113

Section 3

AI, Big Data, and New Trends

Chapter 5

An Innovative Process Engineering Framework: Pipes and Puddles

Jose Costas and Rafael Pastor

Abstract

Quality Control is a key requisite in industry, and business standards are expected to follow process-centric and customer-focused principles. These requirements entail very difficult and challenging problems in terms of choreography (connections between processes—the overall process map of the company contains processes, subprocesses, and meta-processes, which involve the challenging task of designing process interconnections) and orchestration (roles of actors, material flow, data flow, decision-making, exception handling, and so forth). The pipes and puddles framework is a new trend in business organization: through a hydraulic schema, each process is built by three pipes: the green pipe handles the intended process function, the red pipe deals with troubleshooting, and the blue pipe provides channels to convey new knowledge to improve the red and green pipes. The blue pipe conveys jidoka (the integration of humans and machines) and yokoten (the transfer of knowledge gained through solving problems, increasingly by using Artificial Intelligence, Machine Learning, Big Data, etc.). Thus, this framework views the organization as a collection of intelligent agents, both human and artificial, that are continuously learning and improving.

Keywords: pipes and puddles, goal-oriented process engineering, multi-layered standards, distributed collaborative documentary management, fast diagnosis reactive procedures, multichannel improvement, holism

1. Introduction

With the increasingly intensive presence of big data [1] and the 2010 emergence of Artificial Intelligence after a period of dormancy [2], many aspects of business and their operational methods face new challenges in capturing the opportunities brought by these new trends, of which we now have tangible evidence [3] of the changes they are causing. However, there are certain invariants we need to understand and manage within business operating systems. For example, in the Lean Manufacturing philosophy, technology must always be subordinate to the process. We also cannot afford the absence of standards or to operate a business management system with poor, non-robust standards when faced with common disturbances that affect every productive system. Standards are the simplest possible way to minimize the variety of problems that may impact a process. Any reader familiar with Control Theory, and especially Ashby's Law of Requisite Variety [4], (this law here can be understood in the following terms: the quality control system must have a repertoire of responses that has at least as much variety as the diversity of problems that the manufacturing system is exposed to) will quickly understand the negative implications of a vague, poorly defined operational standard.

Beyond Control Theory, simple common sense shows us that the variety of situations that may arise when the actors of a process (people and machines) collaborate to achieve the "intended process function" depends on the variety of tasks, controls, communications, records, and other activities involved in the choreography of the process, as well as its orchestration with other processes with which it interacts. To this are added the sources of "common noise," such as disturbances due to geometry (special), disturbances due to wear-out, and disturbances due to the "gearing" (intricacy) of work elements that carry out the various primary and auxiliary functions for the process.

The authors of this chapter, with their extensive and varied industrial experience, have developed a framework to tackle the growing challenge faced by any organization that seeks, because it knows it is much more profitable and satisfying for its customers, to operate effectively, efficiently, and adaptably under the principles, mandatory under the Lean Manufacturing paradigm, of customer focus and process-centric. There is plenty of literature concerning how to develop the process-centric principle in industry, for instance [5], but throughout this chapter, we will discover some general shortcomings, which end up with a Quality Control system mostly dependent on inspection and reactive problem analysis for common, known, modes of failure in processes.

This is an engineering requirement to shape a business operating system (BOS) oriented toward the ultimate goal (theory of constraints [6]) of any business: making money today and tomorrow.

We are all aware of the enormous role played by the great gurus of Quality (Deming [7], Shewhart [8], Crosby, Juran, Feigenbaum, Ishikawa, etc.) in contributing specific insights, directives to guide thought, and in general, developing a powerful machinery that has led to drastic improvements toward business excellence by focusing on Quality.

However, complacency [9] ("One important aspect of automation misuse is reflected in an inappropriate monitoring or checking of automated functions, a phenomenon that commonly has been referred to as complacency.") is the worst enemy of continuous improvement. We have made great strides in productive systems, and now we face both new challenges and new opportunities.

The rest of this chapter is dedicated to explaining an innovation in organizational engineering: the Pipes and Puddles framework [10]. This framework is inspired by a hydraulic idea: to view processes as a network of pipes, which present challenges of connection and leakage. The mission of this framework is to make it affordable to operate with highly volatile standards, as this is intrinsic to continuous improvement, which naturally involves very frequent changes in standards. This presents a huge challenge in terms of requirements engineering for standardizing processes and their connections, as well as managing troubleshooting in an orderly manner, which is natural given the physical laws that govern productive processes. To achieve this, pipes and puddles deeply develop a sense of ownership, focusing on the process by reacting against departmental silos and poor teamwork dynamics. It resolves document management through multi-layered, decentralized standards, and the presence of more abstract processes, the Management-Oriented Processes (MOPs) which, from an abstract framework, through inheritance and polymorphism, extend into all Customer-Oriented Processes (COP) + Support-Oriented Processes (SOP).

In summary, with the Pipes and Puddles framework, we have a model for effective, robust, and adaptable governance of the standardized work required by each business process, with distributed ownership, but forming a cohesive and connected fabric, while remaining flexible and malleable to naturally embrace continuous improvement; that is, the frequent, day-to-day changes in business standards. And in this continuous improvement, Artificial Intelligence and Big Data techniques are playing an increasingly prominent role.

2. The need

According to Goal-Oriented Requirements Engineering [11] (GORE), one must first understand the needs before proceeding with solutions (**Figure 1**).

What we can see in the figure above (authors' elaboration) is a general idea of the major challenges, and obstacles that arise in order to create the framework that supports a defined way to run the business.

The organizational model we are all familiar with is the organization chart. This organizational scheme addresses how we define different job positions, functions, main tasks and roles, relationships, and competencies (soft and hard skills). In other words, it solves issues such as how to recruit, determine the size of the workforce, hierarchical relationships, job evaluation, training plan design, career development plans, etc. The organization chart gives us a departmental view of the company.

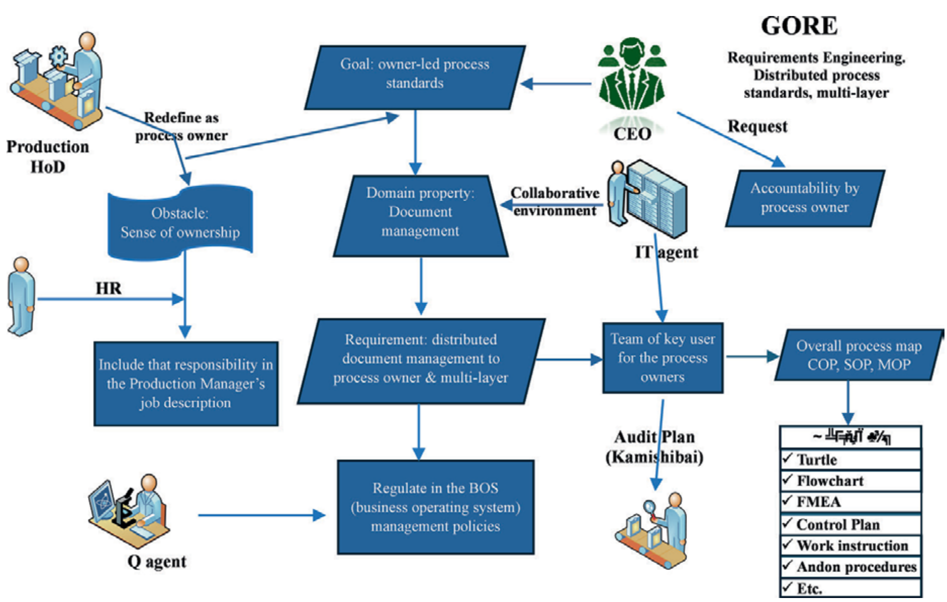


Figure 1.
Requirements engineering.

However, a company makes money [6] ("Theory of constraints views a company as a set of interdependent processes working in harmony to achieve the profit goal of the company, and thus it emphasizes total system performance over localized measures to guide operational decisions. T") by delivering products (and/or services) to its customers. And all of this does not happen through departments. Departments provide actors for the various processes. Value reaches customers through processes!

Therefore, beyond the organization chart, which solves needs in the hierarchical order, we have to address the needs linked to business flows (material, information, money, ...), first by creating a map of the overall processes [12], and, for each process, its operational standard. In turn, we need a standard, a detailed way of defining how we want the process to function, because we cannot bear the burden of problems that would arise if we limited ourselves to explaining the "what" (results-oriented) to the process agents, but do not specify the "HOW" (detailed procedure in the work instructions).

Thus, we quickly identify the range of needs:

Just as a department has a head of department (HOD), a process needs a process owner (PO). There needs to be a position defined in the organizational chart that is accountable for the process. This should be reflected in their job description. Because a process does not operate through the hierarchy of the organizational chart. However, we need the PO to have matrix authority over the actors from the various departments involved in a process.

Each process owner is accountable for having defined standards for their process down to the detail of the work instructions. This entails (1) requirements for document management, (2) training requirements for the standards to the applied by process actors, and (3) audit requirements to ensure that the standards are known and applied.

Each process requires a dashboard that measures: (1) customer satisfaction with the process (internal or external), (2) efficiency, and (3) flexibility/adaptability to common changes and disruptions.

Following requirements engineering, where we understand the needs and obstacles to overcome, we come to formulate a solution that, in this case, is structured as follows: (1) a collective of process owners defined in job descriptions, to formalize the company's process orientation, (2) a collective of process owners trained in process mapping, (3) reliable data collection systems to feed the dashboard, (4) a common document management system, and (5) defined process connectors (PCOM: process communications).

Once we understand the needs involved in process-based management and the obstacles the organization faces, we know that we want to arrive at a general process map that offers an effective solution to the challenges of orchestration (interactions between processes) and the challenges of choreography (collaborations between actors, task flow, information flow, material flow, etc.).

Many companies reach this point by implementing their process-based organization and redefining their accountability system so that it is based on processes and not departments. This dissipates much of the impedance stemming from departmental silos and allows for much more favorable interdepartmental collaboration.

But that is just the beginning. Now, once these basic needs are addressed, the organization recognizes other, much more challenging needs; just to mention some of them: (1) dynamic needs, which involve change management, continuous improvement, handling exception conditions, disaster recovery plans (degraded work, contingency plans, etc.), (2) the dashboards of the various processes need to form an integrated dashboard, derived from the strategic policy deployment, (3) the overall key performance indicators system should be supported by an embedded data collection system within the processes, (4) the document management of standards must be common, interoperable, and present a homogeneous appearance to all the interested parts, and (5) uniformity in the standards of the various processes is expected, despite the variety of process owners.

All these challenges posed by the standardization of different processes often leave certain aspects poorly resolved or inadequately addressed. Here, the authors have compiled a list according to their industrial experience: (1) we may find inconsistent dashboards, (2) dashboards that do not drive continuous improvement, (3) continuous improvement confused with incident management like a fire department, (4) severe deficiencies in the andon (exception handling) system, which takes on an improvised form rather than providing diagnosis and rapid response to common and known problems, (5) lack of a sense of ownership, leading to the dominance of departmental silos and weak teamwork, (6) risk management efforts, especially in Process Failure Mode and Effects Analysis (P-FMEA), which become mere paperwork with little (or no) business value, (7) artificial intelligence-based controls considered appendices, not integrated into the process, (8) big data solutions that end up feeding dashboards in an unorganized manner, (9) enormous effort to keep standards up to date and current, (10) process actors applying different versions of the standard because they are unaware of the current version, (11) actors untrained in the latest version of the standards, (12) process audits that do not identify significant deficiencies in the standard or even fail to notice that they are not being properly applied because each actor takes undue autonomy to deviate from the standard and adopt their own operational method, which multiplies the variety of problems and unnecessarily increases the complexity of problem analysis.

The Pipes and Puddles framework is an innovation in organizational engineering, proposing and developing mechanisms for the implementation of robust solutions for the adoption of Process-Based Organization, drastically promoting the "sense of ownership." Thus, it serves as a framework to build an effective Quality Control System in the organization.

3. Mathematical category theory

In this section, we will explain some very basic applications of mathematical category theory in terms of how can it be used to map business standards.

Category theory is a branch of mathematics that can be applied to improve many aspects of business. In the work of Baez et al. [13], we find the application of this mathematical categories approach to control systems. The underlying idea is that many properties of systems can be represented using arrows and objects. We find plenty of examples of this in business. Every time we use diagrams, we are employing these artifacts from category theory. However, we do not need a deep understanding of the rigor or formalism of this theory.

Let us consider some concrete examples.

3.1 The parameter diagram (p-diagram)

What is a parameter diagram? It is a tool commonly used in robust engineering that: (1) looks for a visual representation of a system, (2) aims to clarify the system and its boundaries, (3) identifies and categorizes input parameters, and (4) does the same for output parameters.

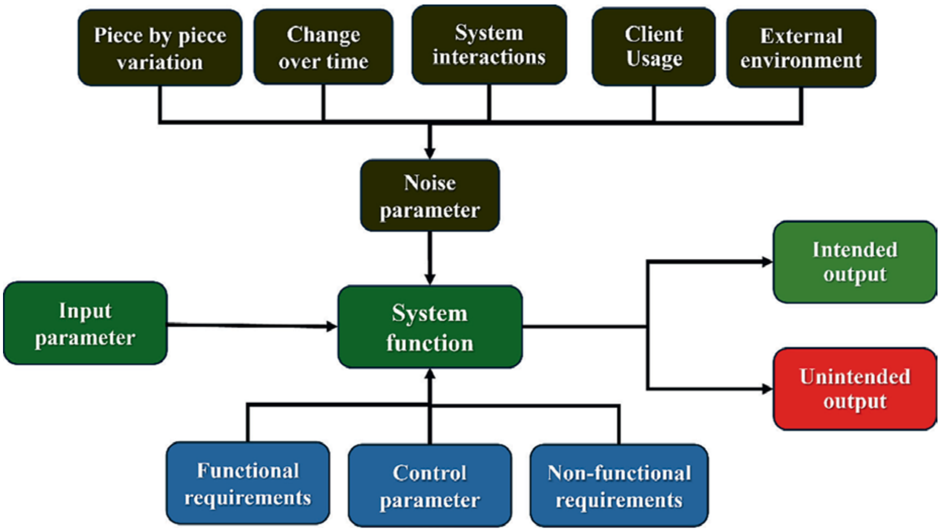


Figure 2.
p-diagram.

In the work of Goktas et al. [14], one can find specific applications of the p-diagram to reliability (for maintenance systems) (**Figure 2**).

Every time we plan an experimental design, the p-diagram is quite useful. Any process has an intended function, which can be seen as an arrow (or morphism, in the language of category theory) that transforms one object (such as a tuple of raw materials, order data, and other inputs) into another object (a tuple of value-added products, possibly with undesired outcomes and other data).

To design the experiment with the goal of learning more about the process behavior—for example, to increase process capability—we use our control space (other objects such as means of production: machines, parameters, operators, etc.) while being aware that the system will encounter common noise factors due to wear, object geometry, piping flow for heating, cooling, and other variables.

3.2 Ishikawa diagrams

The figure below (Ishikawa diagram) [5] is a well-known tool to show the knowledge the organization has about the different causes that may end up with a concrete failure, hence showing causal chains (here we have stopped just at the first level of branching for clarity). From the quality control perspective, this diagram is much more valuable when this knowledge is collected in a preventive way, and is used to provide fast diagnosis and reaction to already known problems in the system (**Figure 3**).

In the Pipes and Puddles framework, a collection of Ishikawa diagrams is generated by the FMEA team (of each process) to provide procedures for fast response to system failures (the red pipe). Any effect (failure) has one or more direct causes (failure modes) that the FMEA team (process risk management) has identified and prioritized.

Of course, we can mention many other examples of using diagrams within the company: Business Process Model and Notation (BPMN) flowcharts, work instructions, fault tree analysis, A3 thinking for communicating problem analysis, etc. The more mature an organization becomes, the more diagrams are used by people to solve various types of process-related problems.

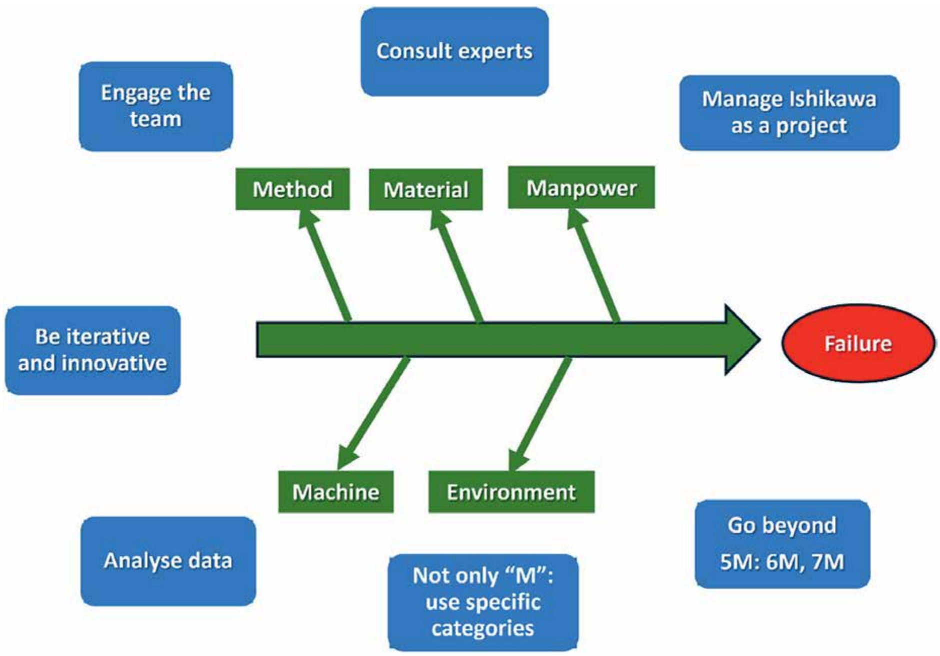


Figure 3.
The Ishikawa diagram.

We will soon see that Pipes and Puddles adopt this vision of arrows and objects to generate, maintain, and elevate process standards. The initial vision is straightforward: pipes and puddles are a categorical system that transforms the need for process choreography, the need to orchestrate the overall process map, the need to handle andon conditions, and opportunities to reduce all types of waste (continuous improvement), as well as the need to address risks, into standards that ensure the system remains effective, efficient, and adaptable to changes.

4. Orchestration

In this section, we first develop the general process map followed by a turtle diagram for each process.

4.1 The overall process map of a company

The overall process map of any industrial company will likely resemble this structure (**Figure 4**).

Pipes and Puddles adopt this common taxonomy of business processes.

In the central lane, we see the value stream. These are the customer-oriented processes (COP) and are the pipelined processes that generate business throughput (as defined in the theory of constraints [15]).

In the top lane, we find the meta-system, which consists of processes related to business policy (vision, mission), strategic plans linked to business intelligence, and the control function (supported by the audit system to ensure that reporting

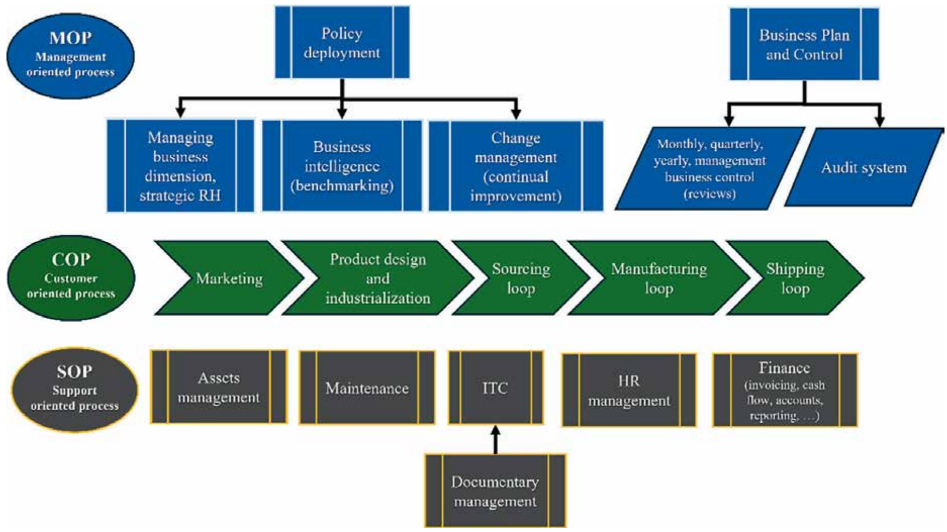


Figure 4.
Overall process map.

matches reality). These functions involve managing change, partly to adapt to the environment through intelligence insights and updates, and to drive continual improvement in all processes. They are known as management-oriented processes (MOP).

In the bottom lane, we see the support processes (SOP), which help customer and management processes fulfill their intended functions. These include managing middleware, resources, actors, technology communications, administrative tasks for compliance with norms and regulations, and so on.

The process-centric principle dictates that each process has a defined process owner in the organizational chart, and the control function of the business evaluates process performance. Processes have many interdependencies; therefore, orchestration must ensure well-defined process communication (sending signals between them in a highly organized manner). Technological solutions (including Artificial Intelligence, Big Data, and other new trends) play an important role in optimizing process communication.

Effectively integrating each tuple of interdependent processes means that the overall process orchestration can be mapped as a mathematical category, where the morphisms represent the different flows of entities (objects) such as data, materials, and money.

In fact, orchestration is about the composability of morphisms (each individual process function and its breakdown into sub-functions). Compositionality refers to the semantic view, meaning what happens in the business. Together, these aspects ensure proper syntax for achieving business goals. Simply put, business architects (process owners who shape their processes and their corresponding orchestration) build the business operating system, which we now refer to as the company standards.

Once we have this "gestalt" ("The Gestalt account of problem-solving tells us that the structural quality of our perception assists the solution process, and when we fail to solve problems, this amounts to a failure to perceive the structure of the problem situation. By contrast, information processing theory focuses on the mechanism of the problem-solving process. Both theories begin by looking at the ways in which people go through a problem-solving process, but they do it in different theoretical contexts, and so focus on different aspects of the situation" [16]), an overall perspective of how the company is organized, obtained through an overall process map, we are ready to distribute responsibilities among process owners to continue building standards by layers. This means we will engage in a highly distributed activity to define process standards (which also includes process communications, or orchestration), without falling into disarray.

From this point forward, we will realize that we use composability and recursion to develop each process choreography.

4.2 The turtle diagram

Having at sight the overall process map, for each process, its process owner is accountable for continuing this first layer of the standards by drawing a turtle diagram (**Figure 5**).

The turtle diagram is a sort of black-box representation of the process which expands the traditional analysis tool used to map and understand the key elements of a business process, from the five main components, suppliers, inputs, process, outputs, and customers (S-IPO-C), which was once the favorite, by adding branches to make visible what resources are needed for the process functions, who the stakeholders are (actors, stakeholders, customers, etc.), how the process operates (methods, instructions, guidelines, recipes, etc.), and the KPI's used to evaluate process performance.

Drawing this diagram for each process makes it easier for process owners, IT staff, HR personnel, and other stakeholders to discuss how to connect "this" process with others. For example, the production process needs material buffers, requires warehousing, operator training, machine uptime during scheduled production, and so on. While it is easy to understand these connections in the material or information flow, many other connections remain hidden until process owners discuss together who needs what, who supports what, with which service level agreements, and how signals should be designed to handle exceptions, among other details.

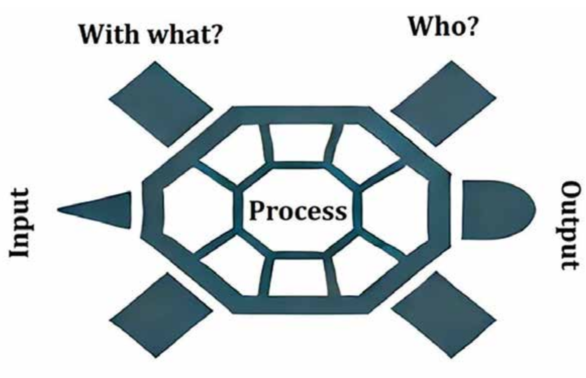


Figure 5.
Turtle diagram.

One point that deserves special mention here is that during Enterprise Resource Planning (ERP) implementation [17], a common mistake is for the consultant team and key users to produce a specifications document ("Technical specifications") to represent the information flow. This redundancy violates the laws of requisite variety (see the work of Ashby [4]) because the standards already include both information and material flow. Therefore, when implementing an ERP system, we must update our version of the standards. The correct approach is to use the current standards, which already cover material and information flows, and the ERP implementation must work with process owners to update business standards. In fact, the simplicity principle, which is key in Lean Manufacturing philosophy, is a way to effectively apply the laws of requisite variety. Having detailed, well-defined standards and adhering to them for process functions minimizes problem-solving efforts.

Another common mistake that can be easily avoided is allowing a variety of formats. We aim to keep the variety of formats in our standards at a minimum to address the requisite variety. The fact that each process owner might produce more variety than needed based on personal preferences is never an excuse to allow the proliferation of forms. The role of the document management process owner (an SOP process) is to act as a facilitator, ensuring that the variety of formats is driven by requisite variety, not personal preferences.

However, the most serious misunderstanding occurs when a process owner fails to see the value of this layer of standards. They may be asked to produce a turtle diagram, perhaps just to comply with ISO 9001: 2015, and they do so reluctantly. This is exactly the wrong approach. ISO 9001: 2015 recommends this practice to establish an effective quality system because experts understand that the first step is to comprehend exactly what each business process entails. Therefore, the value of creating a turtle diagram lies in the following:

The process's intended function must be properly understood. The "goal" of the process is to transform data (specific documents in the ERP or any supporting information system) and materials into information and value-added products. It is essential to describe rigorously what is being transformed and into what.

Inputs have a source, and outputs have a destination. These can be seen as the process's ports, linking upstream and downstream stations.

The different roles of the actors in the process must be clarified, as well as the rules they must follow. Typically, the "WHO?" branch of the turtle diagram is expanded in a chart, which represents the four key functions that each person or group can have in a specific process: Responsible, Accountable, Consulted, and Informed (RACI). Moreover, actors need the training to develop the necessary skills to follow standards within the cycle time, adhering to the "first time through" principle. This is detailed in job description sheets managed by the human resources process owner. The collective of process owners must be aware of these process interactions and provide definitions for these requirements.

The resources required for the process to function must be identified, as they are subject to consumption, wear, and so on. These needs (procurement, maintenance, etc.) are another source of process interactions with support processes that must be clarified in a "defined" way (i.e., clarifying transactions, process communications, signals, alarms, etc.).

The "HOW?" branch of the turtle diagram will be elaborated with flowcharts, which will recursively lead to work instructions for detailed tasks and controls. For risk management, especially in the production process, fault tree analysis is another important tool where the physics of the process can reveal hidden root causes that disrupt the process's intended function. Later in this chapter, we will explore an innovation within the Pipes and Puddles framework: we must provide "HOW?" methods not only for the "green" pipe but also for the "red" pipe. The "blue" pipe is a meta-process that will be discussed later.

The "results" branch is designed to evaluate process performance, ideally using well-known and defined business metrics that facilitate benchmarking. Data collection and processing should be cross-checked against the "HOW?" branch.

In the ongoing process of building process standards, we continually refine and cross-check the details of the turtle diagram's branches with the key points. We do not simply define standards—we are also accountable for the consistency across all layers. Multi-layered standards offer numerous advantages, particularly regarding continuous improvement and meeting requirements for training, auditing, and document management (versioning). The more standards we can embed within the ERP, the better. This is the key point of an ERP system: to simplify both inter-process orchestration and intra-process choreography.

5. Pipes and puddles: The conceptual framework

Once we realize that the need for process standards is not limited just to provide a solution for the process's intended function, we are ready to grasp the key ideas of the pipes and puddles framework (**Figure 6**).

Up to this point in the chapter, we have developed an awareness of common issues organizations face when dealing with standards. These issues manifest in terms of the cost of quality, particularly the cost of failure. This is often a consequence of poor integration of different requirements: the ERP, instead of being an instrument for the data flow in the standards, becomes a separate element; problem-solving is managed as "projects" involving research rather than having reactive procedures, which points to inefficiencies in risk management. The simplicity principle is also poorly understood.

We aim for simplicity in using tools because the tasks are repetitive, but we should instead focus on simplicity in the design phase, which happens once, and during this upstream phase is when the organization can provide a better Quality Control system, which avoids large future problems.

Moreover, applying *kaizen* (continuous improvement) to elevate standards often results in high documentary management costs and poor deployment of new versions of standards, which in turn generates more variety of special, assignable problems.

The framework Pipes and puddles is the result of extensive problem-solving in the industrial experience of the authors of this chapter, which has culminated in the framework we will now discuss.

First, a company's viability is seriously challenged when new components are continually added to the system, sometimes creating non-value-added paperwork and redundancies that generate more work and problems. It makes no sense to have the Enterprise Resource Planning (ERP) on one side and standards on another—conflicting with the ERP due to redundancies. It is equally nonsensical to separate risk management, *andon* and anti-error systems, and fast-response problem-solving. It is also unproductive to isolate 5S campaigns, Kanban initiatives, TPM, and other Quality Control foundations or quality basics from business process standards. All of this goes against the simplicity principle, or, as we prefer to call it, against the laws of requisite variety in goal-oriented requirement engineering. Piling up a mess of growing

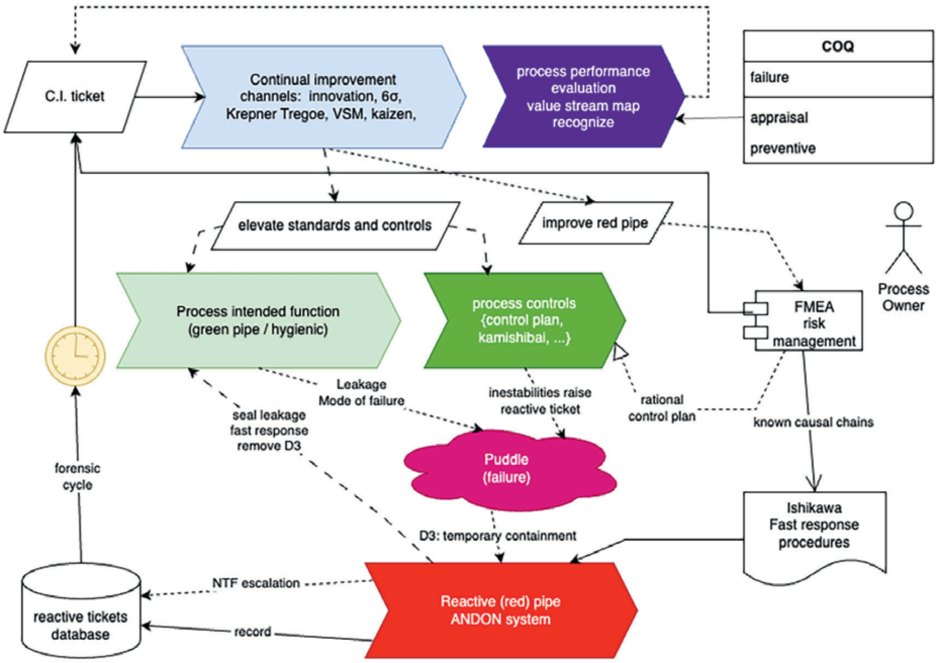


Figure 6.
The framework pipes and puddles.

requirements in a disconnected way eventually leaves management completely overwhelmed.

Second, organizations fail to leverage the advantages of a layered structure when standards are not distributed and/or multi-layered. When, as we have observed many times, the quality director is solely accountable for all business standards, this signals a poor allocation of human resources.

Third, organizations treat meta-processes, such as change management, continuous improvement, strategy deployment, and results evaluation, as ordinary processes, when, in fact, these processes must be embedded within all value stream and support processes.

The main contribution of Pipes and Puddles as a driver of organizational innovation toward effectiveness, adaptability, and simplicity is its view of each process as comprising three pipes:

The green pipe is what everyone naturally envisions: the one that carries out the intended process function.

The red pipe addresses the fact that most problems are common and can be identified and solved by applying the eight disciplines as potential problems through proper risk management. This approach ensures fast diagnosis and reaction procedures.

Additionally, prioritizing risks (in P-FMEA) creates a taxonomy that dictates the type of action to address each risk: "H" risks are not tolerable and should be reduced (or eliminated) before the start of production (SOP). "M" and "L" risks are expected to have verified diagnosis and reaction procedures for fast responses, which involve promptly removing the *andon* condition and implementing containment. "M" risks should have defined, verified, effective, and cost-efficient controls.

All risks in the green and red pipes are supported by lean foundations. We consider 5S, TPM, visual management, and other lean elements as ways to detect and/or prevent potential failure modes within the system.

The blue pipe is a transversal meta-process that defines channels to: (1) build the green and red pipes, ensuring stable and capable processes, and (2) drive continuous improvement, enhancing what already works well repeatedly.

The framework is complemented by forensic analysis to respond to new, unknown problems.

Ultimately, we view the organization as an intelligent machine, built with human and artificial intelligence, drawing on an understanding of the physics involved in each process, historical data, and experimentation and simulation techniques to generate new knowledge.

6. Conclusions

In the extensive experience of the authors, standards are often created in two layers. One is an overall synoptic map—usually bizarre—where the creator, typically the quality manager, attempts to draw the entire company's flow map on a single canvas. The other layer is a bulky document containing all low-level instructions. This approach treats standards merely as paperwork to satisfy mandatory external audits, ignoring their value in problem-solving by seeking simplicity and repeatability in achieving the intended function of each business process. Furthermore, changing methods is perceived as a nightmare, which discourages continuous improvement because it complicates keeping standards updated.

Decentralizing to process owners and adopting a multi-layered structure is a well-known solution to this problem. However, other problems arise when doing so. This approach often results in a limited view of standards, adding multiple requirements: the ERP becomes separate from the standards, FMEA is isolated, control plans are disconnected, and performance evaluation is done by departments rather than processes. Instead of aiming to improve overall system performance, the focus becomes hitting KPI's targets—this is not an objective but a constraint that signals when an organizational red flag should be raised.

Pipes and Puddles, like lean philosophy, operates under a holistic paradigm: it is the whole that matters, in stark contrast to the reductionism of mass production. Everything is interlinked and seamlessly connected: work instructions link to flowcharts, which connect to RACI charts and turtle diagrams. Turtle diagrams are used to define connection points with well-defined protocols, either synchronous or asynchronous, but robust enough to detect any message leakage or pending responses.

Technological solutions have always played a relevant role in running processes, though often in a scattered way. Today, with common, affordable solutions like Big Data, Artificial Intelligence, and Machine Learning, the need for a framework grounded in systems thinking becomes evident. Pipes and Puddles address this need, integrating all components—human, mechanical, and artificial—into the process, making the process-centric principle the cornerstone of the manufacturing system.

Author details

Jose Costas[1]* and Rafael Pastor[2]

1 Organizational Engineering Group, University of Oviedo, Spain

2 Iberomoldes Group, Iber-Oleff Pombal, Portugal

*Address all correspondence to: jose.costasgual@gmail.com

References

[1] Chen J, Chen Y, Du X, Li C, Lu J, Zhao S, et al. Big data challenge: A data management perspective. Frontiers of Computer Science. 2013;**7**:157-164

[2] Lu Y. Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics. 2019;**6**(1):1-29

[3] Chen Y, Zhang L. Data-intensive applications, challenges, techniques, and technologies: A survey on big data. Information Sciences. 2014;**275**:314-347

[4] Ashby WR. An Introduction to Cybernetics. London: Chapman & Hall; 1956

[5] Liliana L. A new model of Ishikawa diagram for quality assessment. In: Iop Conference Series: Materials Science and Engineering. Vol. 161. No. 1. United Kingdom: IOP Publishing; 2016. p. 012099

[6] Gupta M, Ko HJ, Min H. TOC-based performance measures and five focusing steps in a job-shop manufacturing environment. International Journal of Production Research. 2002;**40**(4):907-930

[7] Anderson JC, Rungtusanatham M, Schroeder RG. A theory of quality management underlying the Deming management method. Academy of Management Review. 1994;**19**(3):472-509

[8] Nelson LS. The Shewhart control chart—Tests for special causes. Journal of Quality Technology. 1984;**16**(4):237-239

[9] Manzey D, Bahner JE, Hueper AD. Misuse of automated aids in process control: Complacency, automation bias and possible training interventions. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 50, No. 3. Sage CA: Los Angeles, CA: Sage Publications; 2006. pp. 220-224

[10] Costas-Gual J, Pastor-Climent R, Puche-Regaliza J, Puente J. Pipes and puddles framework: Risk management in manufacturing processes to reduce the total cost of quality. Journal of Industrial Engineering and Management. 2024;**17**(1):35-62. DOI: 10.3926/jiem.6448

[11] Van Lamsweerde A. Goal-oriented requirements ENGINEERING: A guided tour. In: Proceedings Fifth IEEE International Symposium on Requirements Engineering. Toronto, Canada: IEEE; 2001. pp. 249-262

[12] Cobb CG. Enterprise Process Mapping. Milwaukee, Wisconsin, USA: ASQ, Quality Press; 2004

[13] Baez JC, Erbele J. Categories in control. arXiv. 2014;**30**:836-881

[14] Goktas Y, Hu Y, Yellamati DD. P-diagram driven robust PFMEA development. In: 2024 Annual Reliability and Maintainability Symposium (RAMS). Albuquerque, New México, USA: IEEE; 2024. pp. 1-7

[15] Balderstone SJ, Mabin VJ. A review of Goldratt's theory of constraints (TOC)–lessons from the international literature. In: Operations Research Society of New Zealand 33rd Annual Conference. Edinburgh, Scotland: Scottish Academic Press; 1998

[16] Laurillard D. Styles and approaches in problem-solving. The Experience of Learning. 1997;**2**:127

[17] Matende S, Ogao P. Enterprise resource planning (ERP) system implementation: A case for user participation. Procedia Technology. 2013;**9**:518-526

Chapter 6

Severity of Failures in Spur Gearboxes by Vibration Signal Analysis

Antonio Pérez-Torres, René-Vinicio Sánchez and Susana Barceló-Cerdá

Abstract

Gearboxes are a fundamental component in the operation of rotating machinery due to their efficiency in power transmission. Therefore, determining the severity level of a failure at an early stage not only avoids machine downtime but also unexpected maintenance activities. This chapter presents a methodology to determine the severity level of different failures in spur gearboxes by analysing the vibration signal, for which a data mining process is carried out using artificial intelligence (AI) techniques. The condition indicators (CIs) in the time, frequency, and time-frequency domains extract the vibration signal features. The CIs from all three domains are merged into a single database (DB), the dimensionality is reduced through a principal component analysis (PCA), and the main factors are used to determine the failure severity level through random forest (RF) and k-nearest neighbour (KNN) classification models, and factorial analysis of variance (ANOVA) tests are performed. Excellent results were obtained in classification accuracy and the area under the curve (AUC) of the receiver operating characteristic (ROC).

Keywords: fault severity, machine learning, principal component analysis, spur gearboxes, vibration signal, condition indicators, feature extraction

1. Introduction

1.1 Gearboxes

In different industrial sectors, such as energy, automotive, railway, aerospace, manufacturing, agriculture, and mining, gearboxes play a fundamental role due to their efficiency in power transmission, low noise levels, and high load capacity. Given their operating conditions, wear impacts the durability of gear teeth [1]. For this reason, monitoring the condition of the equipment is essential to detect potential failures early, thereby preventing machine downtime or unexpected maintenance activities, which could ultimately result in significant economic losses [2].

Gearboxes are susceptible to several types of failures, such as wear, cracks, tooth fractures, and pitting. These failures can occur from the early stages of the gear's service life and may worsen either gradually or drastically, depending on operating conditions [3].

1.2 Vibration signal

A widely used signal for failure diagnosis is vibration, as it is easy to obtain using acceleration sensors that can be installed on the gearbox housing. However, this signal is complex, containing stationary and non-stationary components and resonant information. Therefore, it is necessary to select signal processing techniques that enable the extraction of its features to diagnose the condition of the equipment [4].

The vibration signal must be processed to extract its features. One feature extraction method involves calculating statistical parameters known as condition indicators (CIs). CIs can be calculated from the original signal acquired in the time domain (**Figure 1a**), in the frequency domain (**Figure 1b**) by transforming the signal using the fast Fourier transform (FFT), or in the time-frequency domain by transforming the signal using the wavelet package transform (WPT) [5, 6]. CIs are sensitive to different types and levels of failure severity, so it is necessary to combine them. However, this combination could lead to challenges, such as redundant information or, worse, contradictions that hinder detecting a failure or determining its severity [7].

1.2.1 Fast Fourier transform

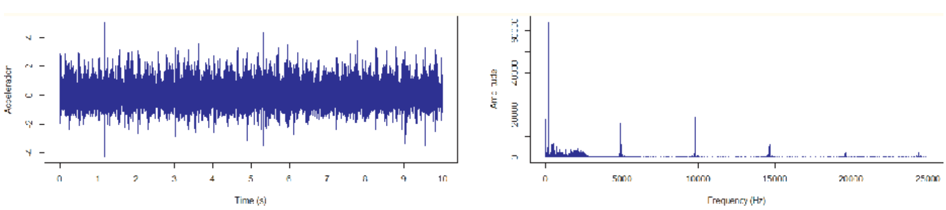
The fast Fourier transform (FFT) is an algorithm for computing the discrete Fourier transform (DFT). The DFT transforms a signal from the time domain into a frequency domain representation. The FFT significantly reduces the number of operations required, making the computation much faster, particularly for large datasets [8].

- The DFT is defined as:

$$\mathbf{X}[\mathbf{k}] = \sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N}-\mathbf{1}} \mathbf{x}[\mathbf{n}]\mathbf{e}^{-\mathbf{j}\frac{2\pi}{N}\mathbf{kn}}, \quad k = 0, 1, \ldots, N-1, \tag{1}$$

  Where:

  - $\mathbf{x}[\mathbf{n}]$ is the input signal in the time domain.
  - $\mathbf{X}[\mathbf{k}]$ represents the amplitudes in the frequency domain.



(a) (b)

Figure 1.
Gearbox vibration signal. (a) Time domain, and (b) Frequency domain.

- $\mathbf{N}$ is the total number of points in the signal.
- $\mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}\mathbf{kn}}$ is the complex rotational factor.

- The FFT algorithm exploits two important properties of the roots of unity:
  - Symmetry:

$$\mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}\mathbf{k}} = \mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}(\mathbf{k}+\mathbf{N})}. \tag{2}$$

  - Periodicity:

$$\mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}\mathbf{k}} = \mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}(\mathbf{k} \mod \mathbf{N})}. \tag{3}$$

- The most common FFT algorithm is the Cooley-Tukey algorithm, which divides the signal $\mathbf{x}[\mathbf{n}]$ into two parts:
  - Even indices: $\mathbf{x}_{\text{par}}[\mathbf{n}] = \mathbf{x}[\mathbf{2n}]$.
  - Odd indices: $\mathbf{x}_{\text{impar}}[\mathbf{n}] = \mathbf{x}[\mathbf{2n}+\mathbf{1}]$.

  The DFT of $\mathbf{x}[\mathbf{n}]$ can be expressed as:

$$\mathbf{X}[\mathbf{k}] = \mathbf{X}_{\text{par}}[\mathbf{k}] + \mathbf{e}^{-\mathbf{j}\frac{2\pi}{\mathbf{N}}\mathbf{k}} \cdot \mathbf{X}_{\text{impar}}\mathbf{k}. \tag{4}$$

- The recursive use of this division reduces the computational complexity:

$$\text{FFT complexity} : \mathbf{O}\,(\mathbf{N} \log \mathbf{N}), \tag{5}$$

  Compared to:

$$\text{Direct DFT Complexity} : \mathbf{O}\left(\mathbf{N}^2\right). \tag{6}$$

1.2.2 Wavelet packet transform

Starting from the discrete wavelet transform (DWT) represented in Eq. (7), wavelet packet transform is implemented with a low-pass filter $\mathbf{h}(\kappa)$ and a high-pass filter $\mathbf{g}(\kappa) = (-\mathbf{1})^{\kappa}\mathbf{h}(\mathbf{1}-\kappa)$, which depend on the wavelet function $\psi(\mathbf{t})$ and its scaling function $\phi(\mathbf{t})$, detailed in Eqs. (8) and (9), respectively.

$$\mathbf{DWT}(\mathbf{j},\kappa) = \frac{\mathbf{1}}{\sqrt{\mathbf{2}^{\mathbf{j}}}}\int \mathbf{x}(\mathbf{t})\psi^*\left(\frac{\mathbf{t}-\kappa\mathbf{2}^{\mathbf{j}}}{\mathbf{2}^{\mathbf{j}}}\right)\mathbf{dt} \tag{7}$$

$$\phi(\mathbf{t}) = \sqrt{\mathbf{2}}\sum_{\kappa}\mathbf{h}(\kappa)\phi(\mathbf{2t}-\kappa) \tag{8}$$

$$\psi(\mathbf{t}) = \sqrt{\mathbf{2}}\sum_{\mathbf{k}}\mathbf{g}(\kappa)\phi(\mathbf{2t}-\kappa) \tag{9}$$

The WPT is a wavelet decomposition generalisation that provides a signal's level-by-level transformation from the time domain to the frequency domain. It is computed using a recursive filtering operation that decreases resolution in the time domain while increasing resolution in the frequency domain [9]. Through WPT, the information is decomposed in greater detail by replacing the wavelet function $\psi(\mathbf{t})$

with $\mathbf{u_{2n}}(\mathbf{t})$ and the scaling function $\phi(\mathbf{t})$ with $\mathbf{u_{2n+1}}(\mathbf{t})$, resulting in Eqs. (10) and (11), respectively.

$$\mathbf{u_{2n}}(\mathbf{t}) = \sqrt{2}\sum_{\kappa}\mathbf{h}(\kappa)\mathbf{u_n}(2\mathbf{t} - \kappa) \tag{10}$$

$$\mathbf{u_{2n+1}}(\mathbf{t}) = \sqrt{2}\sum_{\kappa}\mathbf{g}(\kappa)\mathbf{u_n}(2\mathbf{t} - \kappa) \tag{11}$$

As a result, the signal is decomposed into a low-frequency approximation represented by $\mathbf{d_{j+1,2n}}$ in Eq. (12) and a high-frequency detail represented by $\mathbf{d_{j+1,2n+1}}$ in Eq. (13), for a decomposition level $\mathbf{j}$ with $\mathbf{2^j}$ subbands, where $\mathbf{m}$ represents the number of wavelet coefficients.

$$\mathbf{d_{j+1,2n}} = \sum_{\mathbf{m}}\mathbf{h}(\mathbf{m} - 2\kappa)\mathbf{d_{j,n}} \tag{12}$$

$$\mathbf{d_{j+1,2n+1}} = \sum_{\mathbf{m}}\mathbf{g}(\mathbf{m} - 2\kappa)\mathbf{d_{j,n}} \tag{13}$$

1.3 Dimensionality reduction

Given the need to combine condition indicators (CIs) extracted from signals in the time, frequency, and time-frequency domains, the matrix containing CIs becomes extensive. Therefore, dimensionality reduction is necessary. Principal component analysis (PCA) is a multivariate statistical technique to achieve this reduction. PCA analyses the data matrix, describing observations through several intercorrelated dependent variables. Its objective is to extract important information from the dataset by representing it as a set of new orthogonal variables called principal components while preserving the maximum variance [10].

Given a dataset represented by the matrix $\mathbf{X} \in \mathbb{R}^{\ltimes \times 1}$, where $\mathbf{n}$ is the number of observations and $\mathbf{p}$ is the number of variables, the process for performing PCA is described as follows:

- The data are centred by subtracting the mean of each variable:

$$\mathbf{X}_{\text{centered}} = \mathbf{X} - \text{mean}(\mathbf{X}), \tag{14}$$

Where $\text{mean}(\mathbf{X})$ is a column vector containing the means of the variables in $\mathbf{X}$.

- The covariance matrix of the centred data is calculated as:

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^{\top}_{\text{centered}}\mathbf{X}_{\text{centered}} \tag{15}$$

- The covariance matrix $\mathbf{C}$ is decomposed into its eigenvalues $(\lambda_{\mathbf{i}})$ and eigenvectors $(\mathbf{v_i})$:

$$\mathbf{C}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad i = 1, \dots, p, \tag{16}$$

Where:

- λ_i represents the variance explained by the ith principal component.
- $\mathbf{v}_i$ is the corresponding eigenvector, which defines the direction of the ith principal component.

- The original data are projected into the principal component space using the eigenvectors:

$$\mathbf{X}_{\text{projected}} = \mathbf{X}_{\text{centered}}\mathbf{V}, \tag{17}$$

Where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ contains the first $\mathbf{k}$ selected eigenvectors.

1.4 Classification models

Different artificial intelligence (AI) techniques, such as machine learning (ML), have become key elements in failure diagnosis and the estimation of severity levels [11]. These techniques use large datasets obtained from sensors to classify types of failures and evaluate their severity levels with high accuracy, depending on the training process of the selected model [12]. These AI systems can adapt to operating conditions, enhancing their robustness and making them valuable tools in condition-based maintenance (CBM) [13, 14]. Among the various ML algorithms used to determine the severity level of a failure in gearboxes, several studies employ random forest (RF) and k-nearest neighbour (KNN) algorithms, as detailed in Sanchez et al. [15].

1.4.1 Random Forest

RF is a classification model (Eq. 18) that follows a tree structure and consists of multiple classifiers. For each $\mathbf{j}$th tree, an independent random vector $(\mathbf{V_j})$ is generated. Each tree uses a training set and votes for the most popular category in the input vector $(\mathbf{x})$. The classification error in the model (Eq. 19) depends on the margin $(\mathbf{mg})$, which measures the average number of votes in the random vectors $\mathbf{X}$, $\mathbf{Y}$ that enable classification for the correct category, and $\mathbf{P_{X,Y}}$ represents the probability in the $\mathbf{X}$, $\mathbf{Y}$ space [16].

$$\mathbf{RF} = \mathbf{h}(\mathbf{x}, \mathbf{V_j})_{\mathbf{j=1}}^{\mathbf{N}} \quad j = 1,2,3\ldots \tag{18}$$

$$\mathbf{E} = \mathbf{P_{X,Y}}(\mathbf{mg}(\mathbf{X}, \mathbf{Y})) \tag{19}$$

1.4.2 k-nearest neighbours

KNN is a non-parametric classification model that uses a base training set to classify new observations in a test set. It applies nearest-distance weighting to all categories in the training set, and the category of the new observation is determined by a majority vote [17]. The training set $\mathbf{D}\left(\mathbf{x_j}, \mathbf{y_j}\right)_{\mathbf{j=1}}^{\mathbf{N}}$, where $\mathbf{x_j}$ is the training vector and $\mathbf{y_j}$ is its corresponding category, is used to determine the category of $\mathbf{y}'$ for a test vector $\mathbf{x}'$ in the test set $(\mathbf{x}', \mathbf{y}')$, as specified by (Eq. 20).

$$\mathbf{y}' = \underbrace{\text{arg max}}_{\psi} \sum_{(\mathbf{x}', \mathbf{y}') \in \mathbf{D}'} \mathbf{w_j} \delta\left(\psi, \mathbf{y_j}\right) \tag{20}$$

Where ψ is the category, $\mathbf{y_j}$ is the category of the $\mathbf{j}$th nearest neighbour, $\delta(\cdot)$ is a function that assigns a value of 1 if positive or 0 otherwise, and $\mathbf{w_j}$ is a weighting factor based on the distance $\mathbf{d}(\mathbf{x}', \mathbf{x_i})$ to the $\mathbf{j}$th nearest neighbour, which by default is the Euclidean distance.

There are a variety of metrics to evaluate the performance of classification models, including classification accuracy and the area under the curve (AUC) of the receiver operating characteristic (ROC), which represents the trade-off between sensitivity and specificity [18].

1.5 Resampling

Resampling techniques are fundamental in artificial intelligence algorithms, involving repeated sampling from a dataset and readjusting the model to gain insights into the fitted model. The main resampling methods include repeated hold-out, bootstrapping, leave-one-out cross-validation, and k-fold cross-validation [19].

The performance of the classification models was evaluated using the repeated hold-out resampling technique. In this method, the dataset is split into a percentage ($\mathbf{p}$) for training and a percentage ($\mathbf{1} - \mathbf{p}$) for testing. This approach allowed the variability of the results to be analysed in terms of both accuracy and AUC, and performance metrics were obtained across multiple repetitions.

As outlined, integrating advanced sensing technologies with artificial intelligence methods provides a significant framework for the early detection of failures and classifying their severity levels in gearboxes. By combining the analysis of signals obtained from sensors in gearboxes with ML algorithms, industrial sectors can enhance reliability and extend the lifespan of these mechanical systems.

The objectives of this study were:

- To measure the performance of the RF and KNN models in classifying the severity levels of different types of failures in spur gearboxes by fusing CIs from vibration signals in the time, frequency, and time-frequency domains, followed by dimensionality reduction using PCA.

- To determine whether the position of the acceleration sensor influences the feature extraction from vibration signals.

2. Materials, equipment, and method

2.1 Test bench

This study was conducted using data obtained from the experimental setup **(Figure 2)**, which includes a spur gearbox coupled to a 1.5 kW, 1200 rpm (revolutions per minute), three-phase motor with a 1.5 kW frequency inverter. Load conditions on the gearbox's output shaft were simulated using an 8.83 kW electromagnetic brake. Vibration signals were collected using four accelerometers (A1-A4) installed

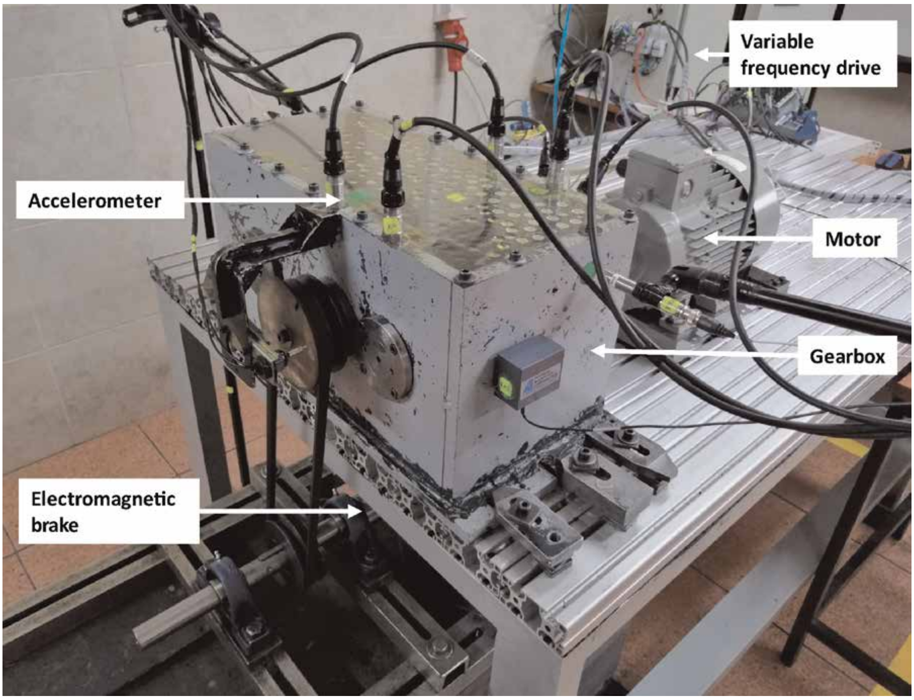


Figure 2.
Test bench for vibration signal acquisition.

vertically. Accelerometers A1 and A2 were placed on the driving shaft of the gearbox, while A3 and A4 were positioned on the output shaft.

Four types of failures were artificially simulated on a gear: the break (B) failure was simulated by tooth milling **(Figure 3a)**, while the crack (C) **(Figure 3b)**, pitting (P) **(Figure 3c)**, and scuffing (S) **(Figure 3d)** failures were simulated using electroerosion. For each type of failure, tests were performed under normal working conditions and nine severity levels. The motor speed was adjusted using the frequency inverter at F1 = 8 Hz, F2 = 14 Hz, and F3 = 20 Hz. At the same time, load conditions were modified using the electromagnetic brake at L1 = 0 V, L2 = 10 V, and L3 = 20 V. With 10 severity levels, varying motor speed and load conditions, and repeating the experiment 10 times, a database (DB) was generated containing 900 observations for each accelerometer and failure type.

2.2 Methodology

The methodological approach developed in this study is represented in **Figure 4** and consists of the following stages:

2.2.1 Data acquisition

The vibration signal in the time domain was acquired using the four accelerometers (A1-A4), each with a sampling capacity of 50 k samples/s. The signal was recorded for 10 s, resulting in 500 k acceleration data points in m/s^2. Subsequently, the time-domain signal was transformed into the frequency domain using the FFT and the time-frequency domain using the WPT.

Figure 3.
Simulated failure types in gear. (a) Break, (b) Crack, (c) Pitting, and (d) Scuffing.

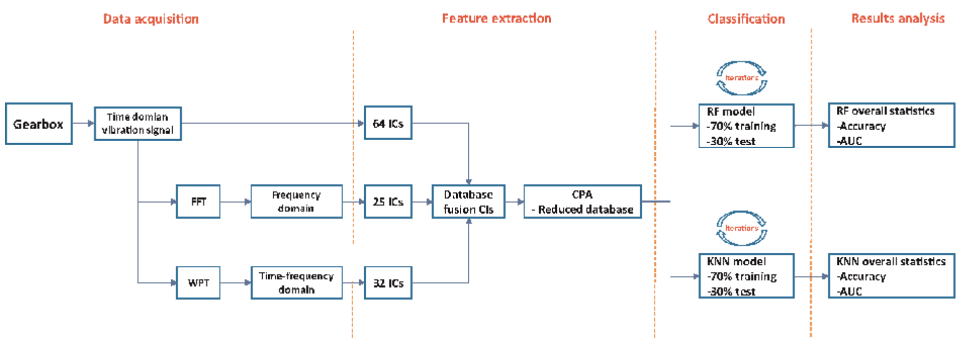


Figure 4.
Methodology.

2.2.2 Feature extraction

Feature extraction from the vibration signal for each accelerometer and failure type was performed by calculating 64 CIs in the time domain and 25 CIs in the frequency domain, with the formulas for these CIs detailed in Sánchez Loja [20]. For the time-frequency domain, the vibration signal was decomposed to the fourth level, using Lawton wavelets with filter 3 and Coiflets with filter 4. This process resulted in the extraction of 32 CIs, as detailed in Pérez-Torres et al. [21]. The CIs from the time, frequency, and time-frequency domains were combined into a single database (DB), and preprocessing was performed to eliminate highly correlated CIs and those with near-zero variance.

Subsequently, dimensionality reduction of the DBs was achieved using PCA by selecting the principal components, resulting in reduced DBs.

2.2.3 Classification

The severity level classification process was conducted with the reduced DBs obtained after PCA. Initially, the RF model was used, and subsequently, the KNN model was applied to compare the results.

	Break		Crack		Pitting		Scuffing	
A	**RF**	**KNN**	**RF**	**KNN**	**RF**	**KNN**	**RF**	**KNN**
A1	0.9049	0.8966	0.9306	0.9190	0.8995	0.9141	0.8937	0.9050
A2	0.8893	0.9047	0.9339	0.9427	0.9275	0.9070	0.9161	0.8978
A3	0.9153	0.9223	0.9530	0.9782	0.9331	0.9460	0.9341	0.9502
A4	0.9139	0.9428	0.9073	0.9049	0.9151	0.9039	0.9127	0.9035

Table 1.
Accuracy by accelerometer, failure, and classifier.

	Break		Crack		Pitting		Scuffing	
A	**RF**	**KNN**	**RF**	**KNN**	**RF**	**KNN**	**RF**	**KNN**
A1	0.9461	0.9507	0.9632	0.9539	0.9468	0.9574	0.9474	0.9408
A2	0.9352	0.9539	0.9555	0.9612	0.9660	0.9557	0.9585	0.9571
A3	0.9518	0.9584	0.9646	0.9828	0.9578	0.9682	0.9626	0.9766
A4	0.9576	0.9729	0.9404	0.9354	0.9550	0.9533	0.9588	0.9424

Table 2.
AUC multiclass by accelerometer, failure, and classifier.

Factor	Df	Sum square	Mean square	F-value	P-value
A	3	2.02	0.67	3499.1	<0.001
Failure	3	0.28	0.09	478.5	<0.001
Classifier	1	36.94	36.94	192312.7	<0.001
A:Failure	9	1.15	0.13	662.4	<0.001
Failure:Classifier	3	0.24	0.08	423.1	<0.001
A:Classifier	3	1.53	0.51	2654.6	<0.001

Df = Degrees of freedom.

Table 3.
ANOVA factorial test for accuracy.

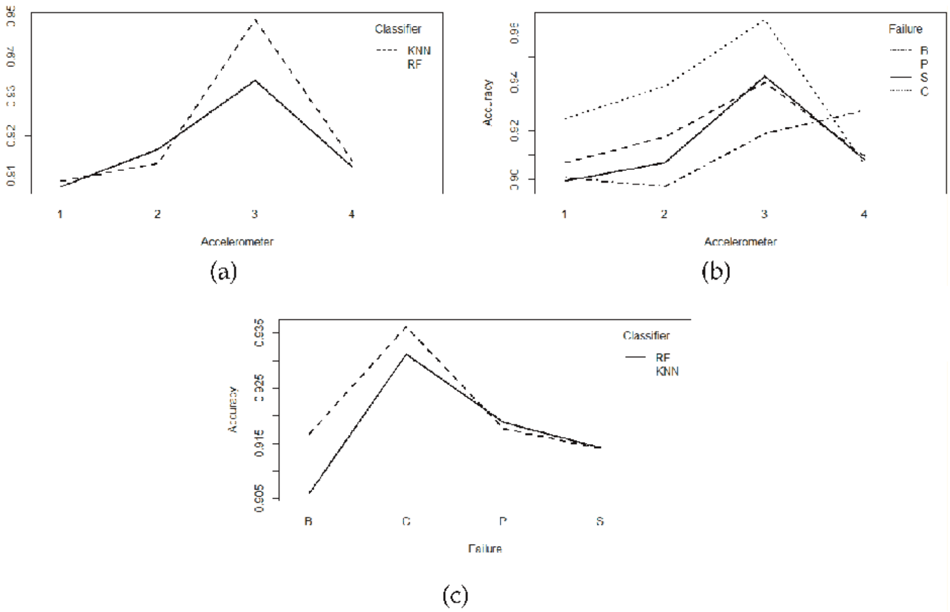


Figure 5.
Interaction factors for accuracy. (a) Accelerometer: Classifier, (b) Accelerometer: Failure, and (c) Failure: Classifier.

Factor	Df	Sum square	Mean square	F-value	P-value
A	3	0.516	0.172	1919.8	<0.001
Failure	3	0.044	0.015	163.8	<0.001
Classifier	1	11.441	11.441	127589.4	<0.001
A:Failure	9	0.490	0.054	607.6	<0.001
Failure:Classifier	3	0.033	0.011	121.7	<0.001
A:Classifier	3	0.495	0.165	1839.4	<0.001

Df = Degrees of freedom.

Table 4.
ANOVA factorial test for AUC.

The performance of the classification models was measured using the iterative repeated hold-out resampling process. For this purpose, the DBs were split into 70% for training and 30% for testing. This process generated two vectors of 100 observations each for every accelerometer, failure type, and classification model. The first vector corresponds to the accuracy values of the classification model, while the second to the area under the curve (AUC).

2.2.4 Results analysis

To conduct this analysis, the accuracy and AUC vectors obtained from the RF and KNN classification models were considered. Two factorial analyses of variance

(ANOVAs) and post-hoc Tukey tests were performed to determine whether accuracy and AUC are influenced by the position of the accelerometer, the type of failure, or the classification model. In other words, the analysis explored whether these factors affect the acquisition of the vibration signal.

The databases were processed and analysed usisng the R software [22].

3. Results

The first objective of this study was to measure the performance of the RF and KNN classification models in classifying the severity levels of different types of failures in spur gearboxes. This was achieved by fusing CIs from vibration signals in the time, frequency, and time-frequency domains, followed by dimensionality reduction using PCA. For this purpose, the average accuracy and AUC were calculated when classifying the severity levels of failures. The accuracy results for each accelerometer, failure type, and classifier exceeded 88% and are detailed in **Table 1**. In comparison, the AUC results exceeded 93% and are detailed in **Table 2**.

The second objective was to determine whether the position of the acceleration sensor influences the feature extraction from vibration signals. Therefore, factorial ANOVA and post hoc Tukey tests were performed to compare the mean values of accuracy and AUC in classification. The factors considered were the accelerometer, failure type, and classification model.

When performing the ANOVA test using the accuracy vector (**Table 3**) for classification, it was determined that there are significant differences (p-value <0.01) among the four accelerometers, the four failure types, and the two classification models. Additionally, it was found that there are interactions (p-value <0.01) between the factors, such as accelerometer-classifier (**Figure 5a**), accelerometer-failure (**Figure 5b**), and failure-classifier (**Figure 5c**).

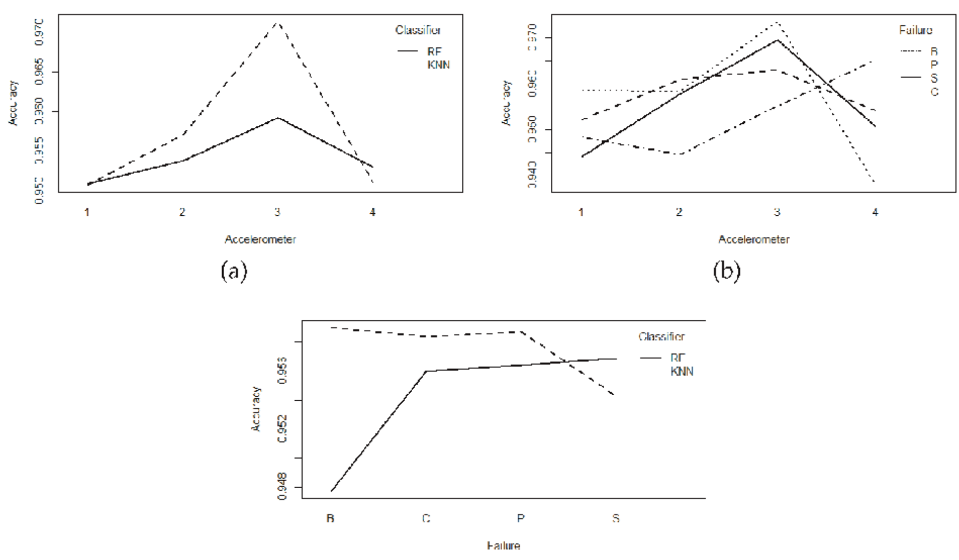


Figure 6.
Interaction factors for AUC. (a) Accelerometer: Classifier, (b) Accelerometer: Failure, and (c) Failure: Classifier.

Regarding the AUC vector, performing the ANOVA test (**Table 4**) for classification revealed significant differences (p-value <0.01) among the four accelerometers, the four failure types, and the two classification models. Furthermore, it was determined that there is interaction (p-value <0.01) between the factors, such as accelerometer-classifier (**Figure 6a**), accelerometer-failure (**Figure 6b**), and failure-classifier (**Figure 6c**).

Regarding the post hoc Tukey tests, graphical representations are provided in **Figures** 7 and **8**. These figures detail the significant differences and interactions observed between the factors. As shown, the differences between factor level pairs are almost all significant, with a few exceptions, notably in terms of accuracy (A4-A2) and (S-B), and in terms of AUC (A4-A1), (P-C), and (S-C), which do not show significant differences. However, these differences are not practically relevant as they

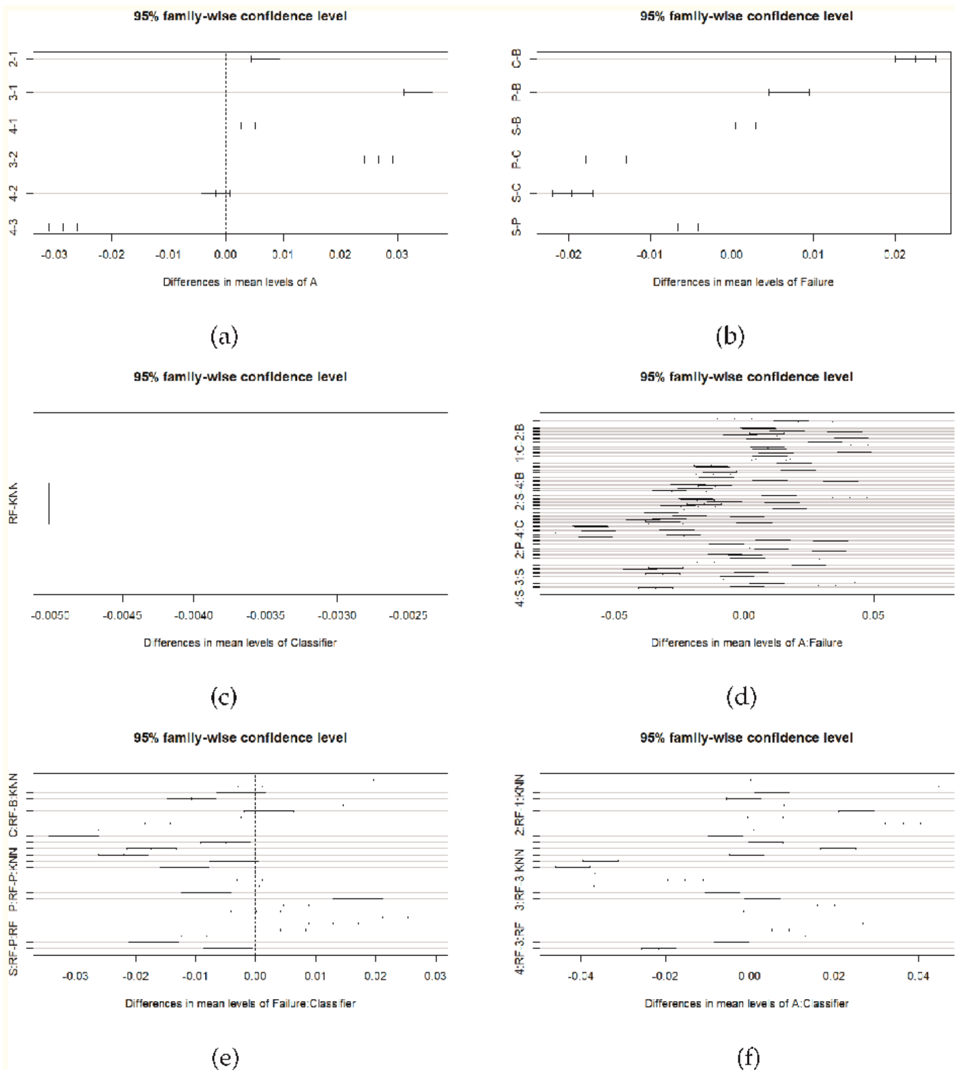


Figure 7.
Post hoc, factor, and interaction factors for accuracy. (a) Accelerometers, (b) Failure, (c) Classifier, (d) Interaction A: Failure, (e) Failure: Classier, and (f) Interaction A: Classifier.

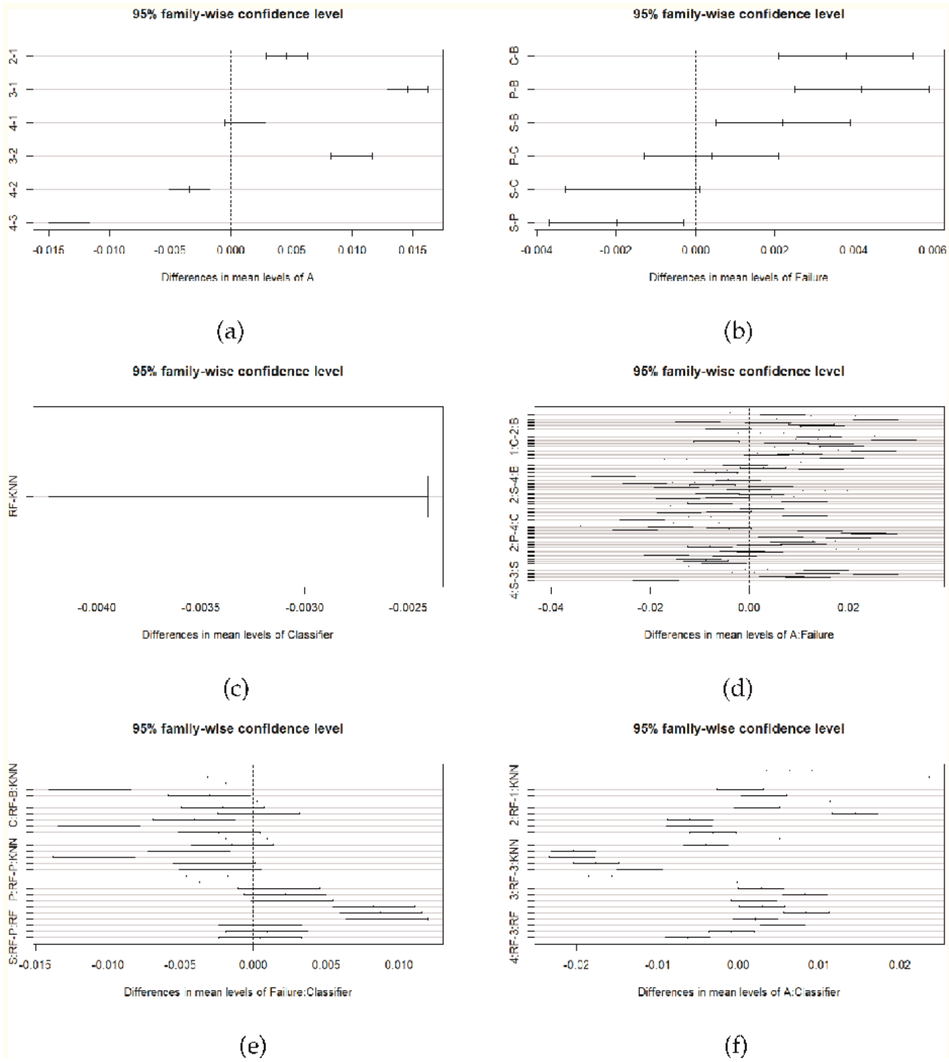


Figure 8.
Post hoc, factor, and interaction factors for AUC. (a) Accelerometers, (b) Failure, (c) Classifier, (d) Interaction A: Failure, (e) Failure: Classier, and (f) Interaction A: Classifier.

are minimal, as detailed in **Table 1** for accuracy and **Table 2** for AUC across the four accelerometers, failure types, and classification models. This is important because, in practice, the physical availability to position a sensor on a gearbox is often limited due to the equipment and attachments to which they are coupled.

4. Conclusions

This study extracted features from the vibration signal in the time, frequency, and time-frequency domains using CIs. After fusing them into a single DB, dimensionality was reduced using PCA. Subsequently, the severity levels of different types of failures were classified using RF and KNN models. Excellent results were obtained, with average accuracy values exceeding 88% and AUC values exceeding 93% across all four accelerometers and failure types. Therefore, the developed process is optimal for determining the severity levels of different types of failures in spur gearboxes.

When comparing the accuracy and AUC values obtained from the classification models using factorial ANOVA and post hoc Tukey tests, it was determined that significant differences exist based on sensor position, failure type, and classification model, with a few exceptions. Additionally, interactions were observed between these factors. However, these differences are minimal and lack practical significance. Thus, installing a sensor on a gearbox, whether positioned along the driving shaft or the output shaft, minimally affects the information obtained from the vibration signal. All these findings confirm that the methodology developed in this study is suitable for condition monitoring of spur gearboxes.

Acknowledgements

This work was sponsored by the Universidad Politécnica Salesiana through the research project entitled “Evaluación de la severidad de fallos en engranajes rectos y helicoidales mediante señales de vibración, corriente y emisión acústica” No. 013-003-2019-2105-22 of the (GIDTEC) “Grupo de Investigación y Desarrollo en Tecnologías Industriales”.

Abbreviations

CIs	Condition indicators
FFT	Fast Fourier transform
WPT	Wavelet package transform
DFT	Discrete fourier transform
DWT	Discrete wavelet transform
PCA	Principal component analysis
AI	Artificial intelligence
ML	Machine learning
CBM	Condition-based maintenance
RF	Random forest
KNN	K-nearest neighbour
AUC	Area under the curve
ROC	Reciever operating characteristic
A	Accelerometer
B	Break
C	Crack
P	Pitting
S	Scuffing
DB	Database
ANOVA	Analysis of variance

Author details

Antonio Pérez-Torres[1,2*†], René-Vinicio Sánchez[2†] and Susana Barceló-Cerdá[1]

1 Universitat Politècnica de València, Valencia, Spain

2 Universidad Politécnica Salesiana, Cuenca, Ecuador

*Address all correspondence to: jperezt@ups.edu.ec

† These authors contributed equally.

References

[1] Feng K, Ji JC, Ni Q, Beer M. A review of vibration-based gear wear monitoring and prediction techniques. Mechanical Systems and Signal Processing. 2023;**182**: 109605

[2] Alonso NF, Jos R, Fern PF. Determination of the bending strength geometry factor J for the calculation of stress in parallel-axis gears as a function of the reference correction and slope angle of the gear tooth. Journal of Mechanical Engineering and Automation. 2016;**3**(1):27-33

[3] Cabrera D, Sancho F, Li C, Cerrada M, Sánchez RV, Pacheco F, et al. Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation. Applied Soft Computing. 2017;**58**:53-64

[4] Halim EB, Choudhury MS, Shah SL, Zuo MJ. Time domain averaging across all scales: A novel method for detection of gearbox faults. Mechanical Systems and Signal Processing. 2008;**22**(2): 261-278

[5] Sharma V, Parey A. A review of gear fault diagnosis using various condition indicators. Procedia Engineering. 2016; **144**:253-263

[6] Pacheco F, de Oliveira JV, Sánchez RV, Cerrada M, Cabrera D, Li C, et al. A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions. Neurocomputing. 2016;**194**:192-206

[7] Lei Y, Zuo MJ. Gear crack level identification based on weighted K nearest neighbor classification algorithm. Mechanical Systems and Signal Processing. 2009;**23**(5): 1535-1547

[8] Brigham EO, Morrow RE. The fast Fourier transform. IEEE Spectrum. 1967; **4**(12):63-70

[9] Gokhale MY, Khanduja DK. Time domain signal analysis using wavelet packet decomposition approach. International Journal of Communications, Network and System Sciences. 2010;**3**(3):321

[10] Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;**2**(4): 433-459

[11] Bachar L, Matania O, Cohen R, Klein R, Lipsett MG, Bortman J. A novel hybrid physical AI-based strategy for fault severity estimation in spur gears with zero-shot learning. Mechanical Systems and Signal Processing. 2023; **204**:110748

[12] Resendiz-Ochoa E, Saucedo-Dorantes JJ, Benitez-Rangel JP, Osornio-Rios RA, Morales-Hernandez LA. Novel methodology for condition monitoring of gear wear using supervised learning and infrared thermography. Applied Sciences. 2020;**10**(2):506

[13] Li Z, Jiang Y, Hu C, Peng Z. Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. Measurement. 2016;**90**:4-19

[14] Aherwar A. An investigation on gearbox fault detection using vibration analysis techniques: A review. Aust. Journal of Mechanical Engineering. 2012; **10**(2):169-183.7

[15] Sanchez RV, Lucero P, Vasquez RE, Cerrada M, Macancela JC, Cabrera D. Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN. Journal of Intelligent Fuzzy Systems. 2018;**34**(6):3463-3473

[16] Breiman L. Random forests. Machine Learning. 2001;**45**:5-32

[17] Dudani SA. The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics. 1976;**4**:325-327

[18] Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997;**30**(7):1145-1159

[19] Raschka S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. arXiv:1811.12808v3 [Preprint]; 2018

[20] Sánchez Loja RV. Diagnóstico de fallos en cajas de engranajes con base en la fusión de datos de señales de vibración, corriente y emisión acústica; 2018

[21] Pérez-Torres A, Sánchez RV, Barceló-Cerdá S. Selection of the level of vibration signal decomposition and mother wavelets to determine the level of failure severity in spur gearboxes. Quality and Reliability Engineering International. 2004;**40**(6):3439-3451

[22] R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2024

Chapter 7

Perspective Chapter: QoS-Aware IoT Framework for Performance Control and Resource Management

Alem Čolaković, Bakir Karahodža and Adisa Hasković Džubur

Abstract

The Internet of Things (IoT) seeks to interconnect people and devices within integrated systems that support a broad spectrum of applications, ranging from smart cities to industrial automation. This diversity of use cases introduces substantial challenges in maintaining Quality of Service (QoS), which is essential for the efficient operation of IoT ecosystems. This chapter presents a framework for evaluating QoS performance by proposing a layered QoS architecture specifically designed for IoT systems. The proposed architecture provides a conceptual basis for implementing advanced Artificial Intelligence (AI)-based mechanisms that facilitate efficient management of QoS parameters across all system layers. As IoT systems increasingly converge with infrastructural technologies such as cloud computing, edge computing, fog computing, Multi-access Edge Computing (MEC), and Mobile Cloud Computing (MCC), overall system performance becomes highly dependent on efficient resource allocation and the underlying network characteristics of data transmission. While cloud platforms facilitate large-scale data processing due to their high computational capacity, they often incur increased network latency. In contrast, localized infrastructures mitigate latency but are limited by processing power and memory constraints. This chapter highlights the importance of emerging methodologies for optimizing service distribution and resource utilization within complex IoT environments. Special emphasis is placed on formulating guidelines for the design of intelligent mechanisms for QoS monitoring and control, enabling optimal decision-making and improving the overall efficiency of IoT systems.

Keywords: IoT architecture, Quality of Service (QoS), performance evaluation, QoS control, resource allocation

1. Introduction

The intensive development of Internet of Things (IoT) applications and their diversity pose significant challenges in ensuring Quality of Service (QoS). The integration of IoT architecture with various computing systems such as cloud, fog, Mobile Cloud Computing (MCC), Mobile Edge Computing or Multi-access Edge Computing (MEC), cloudlet, and edge computing systems directly impacts system performance, including the choice of infrastructure for data processing, network characteristics for data transmission, technological requirements of applications, and more [1, 2]. Therefore, efficient strategies and mechanisms for resource control and management are needed to achieve the desired QoS performance through task allocation to appropriate infrastructure [3–5].

Achieving the desired level of QoS performance requires an appropriate resource allocation model across the layers of the system architecture to efficiently utilize available resources. Some approaches include the rewriting-based approach, which focuses on managing infrastructure reconfiguration [6, 7]. Other proposals advocate for the development of a “smart middleware layer” between IoT devices and other infrastructure (e.g., developing a layer to enable efficient task allocation aimed at optimizing computational resources) [8–10]. Additionally, various data caching techniques have been proposed to optimize performance [11, 12], as well as numerous algorithms for task allocation to optimize resources and performance [13]. For instance, the authors in [14] utilized the Particle Swarm Optimization (PSO) algorithm to consider multiple parameters, including response time, network throughput, energy consumption, and latency, while [15] proposed an enhanced “firefly” algorithm to address workflow scheduling challenges in cloud-edge environments. The study [16] suggests applying a distributed deep reinforcement learning algorithm to optimize resource allocation decisions. The study [17] provides an overview and classification of dynamic load-balancing techniques to prevent congestion in IoT networks and identifies key metrics for evaluating these techniques. These works provide a foundation for theoretical modeling of IoT systems and identifying key components, which is essential for predicting system performance in various scenarios.

However, numerous challenges are associated with the application of these task-scheduling algorithms, such as device heterogeneity, the need for energy-efficient scheduling, and real-time scheduling requirements [18–20]. Most described solutions rely on the implementation of specific infrastructure, including servers, base stations, and network devices. Other limitations include high computational complexity, the omission of certain parameters in problem formulation, unrealistic assumptions, and a lack of evaluation in real-world scenarios [21]. Thus, an integrated approach is missing—one that synergistically encompasses various aspects such as considering IoT application requirements, observing system components individually and how they affect overall QoS, examining how specific technologies perform in different integration models, and developing practical solutions that do not demand extensive resources, among others.

These challenges represent open questions that require the future attention of the research community. A QoS-aware architecture for IoT systems needs to incorporate factors such as specific application requirements, performance indicators at various architecture levels, integration models with different computing systems, and more. Therefore, the main objective of this study is to propose a layered architecture for QoS management, which provides a structural framework for developing intelligent mechanisms capable of efficiently achieving desired QoS parameters. The key research question is, “How can IoT system performance be optimized through a layered architecture and efficient resource management to achieve the desired level of QoS performance?” The proposed conceptual QoS framework (architecture) for performance control in IoT systems can assist in determining optimal service allocation strategies across the layers of the IoT system architecture.

2. Conceptual architecture of IoT systems

IoT devices are typically limited in terms of memory and processing capabilities. As a result, most IoT solutions rely on the support of computing systems to "offload" tasks from IoT devices, such as edge, fog, and cloud systems. Technologies based on the cloud computing concept enable the storage and processing of large amounts of data and provide numerous other services requiring significant processing and memory capacities. However, this solution brings challenges, such as dependency on Internet network availability, service execution time due to data transfer over the Internet, system reliability, data transfer costs, and more.

The allocation of services closer to the data source has introduced new possibilities and addressed some of these challenges. Such solutions include fog computing, edge computing, MEC, MCC, cloudlet, and others.

These systems enable faster service execution in certain application scenarios, more reliable systems in terms of independence from the availability and state of the Internet network, and other advantages. These paradigms involve the allocation of computing services across various layers of system architecture based on infrastructure primarily located near IoT devices or at the network's edge. The key differences among these concepts lie in the distance of IoT devices from these systems and the type of applied infrastructure. The role of IoT architecture is to enable the integration of IoT with various systems and technologies [22].

Integrating IoT devices with one of the aforementioned infrastructural systems can substantially enhance system performance. Various services—such as data storage and processing, device control, and others—can be distributed across different layers of the IoT system architecture. However, infrastructure located closer to the IoT device often lacks the processing power and memory capacity characteristic of cloud-based solutions. Consequently, executing complex data analyses, enabling access for a large number of users, and storing vast volumes of historical data may pose significant challenges for edge- or fog-based components—tasks that can be seamlessly managed within cloud environments [23].

This highlights the importance of selecting the appropriate system architecture and IoT integration model with other computing systems. Essentially, these concepts offer a new approach to designing and developing IoT systems and can be considered integral to the architecture of specific IoT systems. Consequently, the generic architecture of IoT systems (designed from the perspective of infrastructure, technologies, and functionalities) can be represented as a three-layer structure, encompassing various interaction models between devices and computing systems for resource and task allocation across different layers. Thus, when considering data processing and analysis services, the system architecture can be viewed as having three primary layers: IoT devices (perception layer), edge/fog layer, and cloud layer (**Figure 1**). IoT devices are connected to one or more computing systems, and all layers of the system architecture can interconnect depending on the integration model.

Potential IoT architectures or integration models each present distinct advantages and limitations that must be carefully analyzed to determine the most suitable option based on the required performance criteria and the specific characteristics of the application domain. Among the key challenges are the selection of an appropriate system architecture to meet targeted QoS levels, the capacity for large-scale data storage, system reliability, support for mobility, as well as ensuring security and data privacy [24, 25].

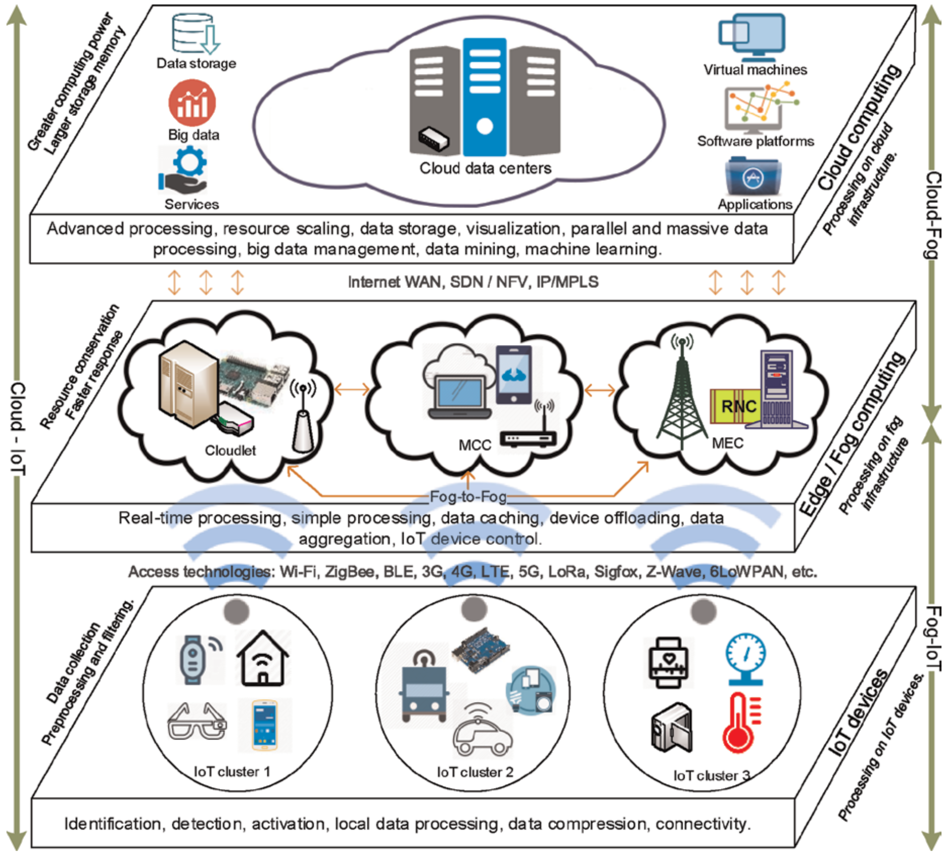


Figure 1.
Integration models of IoT and other computing systems (fog and cloud) [1].

3. QoS architecture of IoT system

Quality of Service (QoS) in an IoT system can be conceptualized as an expression of the system's performance, reflecting the extent to which application-specific requirements are satisfied. In traditional telecommunications services (e.g., voice services), standardized recommendations for the limits and optimal values of individual QoS parameters are well established—such as those found in the ITU-TG.1010 recommendation. However, due to the inherent heterogeneity and application diversity within IoT environments, such recommendations cannot be directly applied. Consequently, one of the primary challenges lies in identifying the most relevant indicators (metrics) for evaluating QoS performance in IoT systems.

For instance, the execution time of an IoT service may serve as a representative metric; it is influenced not only by network latency (data transmission delay) but also by the time required to process data generated by IoT devices and, in some cases, to transmit results to the target destination. Furthermore, determining appropriate boundary and threshold values for these indicators—aligned with the specific requirements of the application—remains an open and pressing research question.

An IoT system consists of many components, and it is necessary to implement a system with efficient resource distribution and service scheduling across the appropriate layers of the system architecture. According to the proposed architecture, the application layer directly responds to the user's request or the specific needs of the system application. The QoS achieved at the other layers reflects the overall system quality. The IoT application sets QoS requirements for the layers below, which must interact to implement the application according to the requirements. However, considering the different meanings and value ranges of individual parameters, it is necessary to normalize these parameters to evaluate the system's performance while taking all key indicators into account. Therefore, a systematic approach to performance assessment is required, which involves several stages (**Figure 2**).

These stages include selecting key performance indicators (KPIs) for each layer of the QoS architecture, P_{xi}, and then assessing them, $U(P_{xi})$. For further analysis, it is necessary to determine weight values, i.e., the significance of individual indicators, $P(w)_{xi}$. After that, the QoS performance at each layer is assessed, $f(QoS_x)$, providing insight into the impact of individual layers on the overall IoT system performance. To conduct such a layered QoS evaluation, which considers multiple performance indicators, it is essential to normalize KPIs to enable multi-criteria analysis. Based on the estimated QoS values at each layer, a QoS evaluation of the entire system, $IoT_{sys}(QoS)$, can be performed.

QoS must be maintained across all layers of the IoT system architecture to ensure that the overall service quality aligns with the specific requirements of the application. By the hierarchical approach to system architecture design, QoS requirements are propagated from the application layer down to the lower layers, while feedback flows upward from the lower layers to the application layer. Each layer within the IoT system incorporates distinct technologies and equipment, tailored to its specific functions and operational context.

Furthermore, the application layer must remain cognizant of the operational status of subordinate layers and may need to exert control over their behavior. In response to these considerations, this chapter proposes a layered QoS architecture for IoT systems, comprising a QoS management object (QoS manager), a QoS broker, and a set of quality indicators defined per layer (see **Figure 3**). The proposed model establishes a

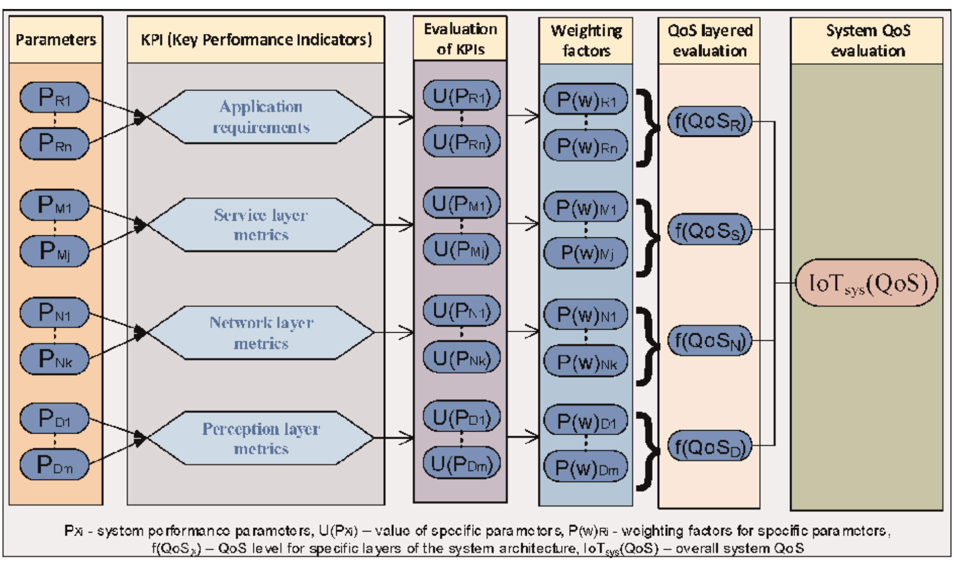


Figure 2.
Methodology for analyzing IoT system performance.

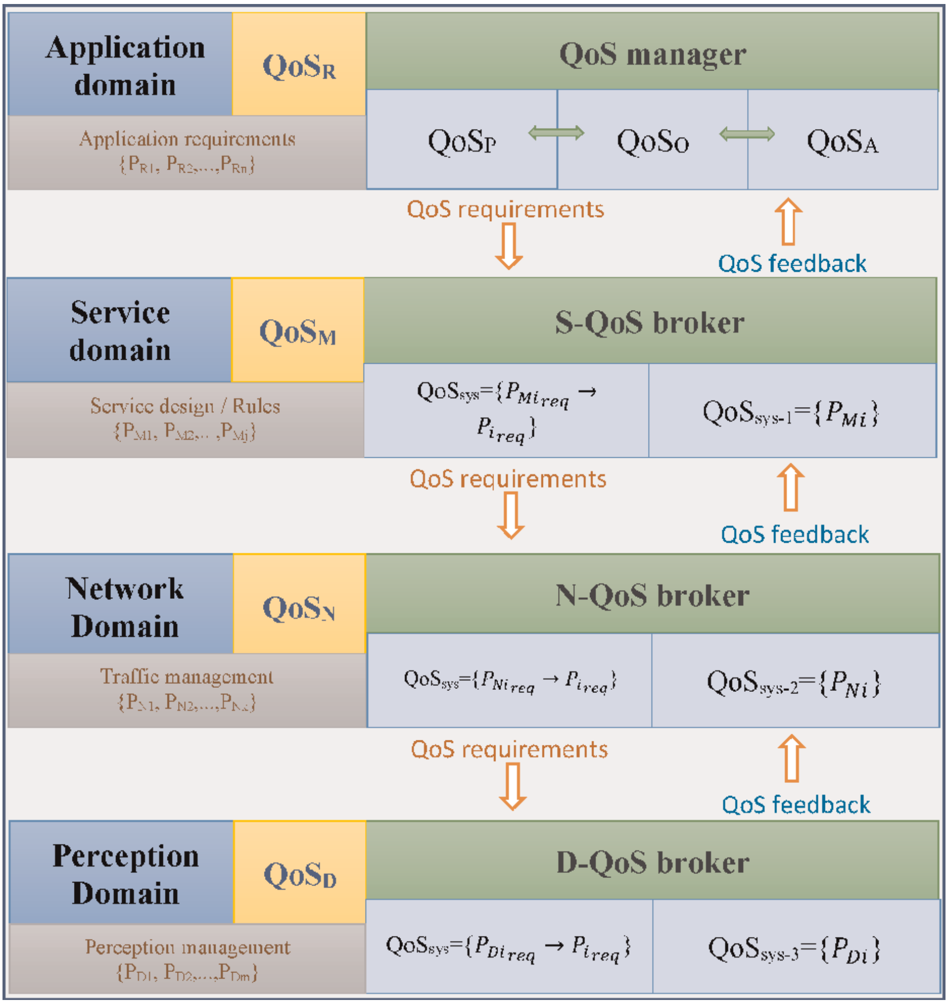


Figure 3.
QoS architecture in IoT systems.

conceptual framework for intelligent decision-making mechanisms, enabling selective data acquisition from devices, preliminary data filtering at the microcontroller or intermediary system level, and efficient forwarding of processed data to the appropriate destination.

Considering the number of devices acting as traffic generators, it is possible that, at different times, the IoT application will require different types and amounts of data. Therefore, a mechanism is required to use specific system resource state data and device metadata (e.g., location, load, energy consumption, etc.), which are necessary for optimization. This process is managed by the *QoS manager*, which addresses the specific requirements related to the context of the IoT application. Therefore, the IoT system can be context-aware and efficient when using available resources.

The *QoS manager* defines QoS rules (QoS_P) and manages operations regarding the adjustment of requirements and instructions sent to lower layers. Implemented at the application layer, it makes control decisions based on the established QoS rules. These rules are created based on the known application requirements ($QoS_R = P_{R1}, P_{R2}, \ldots, P_{Rn}$), system resources at each layer (QoS_P), and the analysis of system state (QoS_O) aimed at optimizing resource utilization and achieving the desired QoS level. The analysis of available resources is conducted based on feedback (*QoS feedback*) from lower layers (QoS_A), which is sent by QoS brokers.

The QoS manager, within its control decisions, creates system requirements for each application based on all key indicators (P_{ireq}) and sends them to all system components. At each layer, a mechanism (broker) is implemented that is responsible for addressing QoS requirements received from the QoS manager via the layer above. The QoS broker determines the impact of the layer it is implementing on the requested QoS performance and translates significant indicators of that layer into QoS requirements for the entire system. Based on this, it can perform necessary operations according to QoS manager instructions and forward the requests to the lower layer. In addition, the QoS broker measures the level of performance achieved at this layer and sends the results to the QoS manager, which can adjust the QoS rules and perform control operations based on optimization processes.

QoS brokers can be deployed across various components of the IoT ecosystem, including IoT devices themselves, service allocation platforms (e.g., fog and cloud infrastructures), and network gateways. Each architectural layer is responsible for executing operations that influence the overall QoS of the system. In the presented architecture, the parameters of individual layers are denoted as follows: QoS_M (parameters of the intermediary layer), QoS_N (parameters of the network layer), and QoS_D (parameters of the perception layer), with the possibility of a different number of parameters at each layer.

Effective optimization of available system resources in IoT environments requires continuous data collection. However, due to the inherent limitations in memory and processing capacity of IoT devices, it is essential to enable flexible and controlled mechanisms for data acquisition and transmission toward computing platforms, where data analysis, optimization, and policy generation are performed. Based on the derived insights, instructions are subsequently propagated to relevant system components.

For instance, if an application only requires periodic data, a predefined rule can be established to specify the timing and frequency at which the IoT device should transmit data. Filtering mechanisms can be implemented through algorithms within the QoS manager or distributed brokers across different layers of the architecture, taking into account the overarching application requirements. By analyzing incoming data, the QoS manager gains awareness of the characteristics of all accessible data sources—such as their geographic location, data collection frequency, and contextual relevance—and can make informed control decisions, which are then communicated to the respective brokers. This approach facilitates controlled and adaptive data collection aligned with application requirements while simultaneously supporting dynamic service allocation across system components to enhance QoS and optimize the utilization of available resources.

4. Key QoS indicators for IoT applications

Due to the complexity and diversity of IoT applications, creating QoS (Quality of Service) profiles for these applications requires extensive testing and performance analysis in various application scenarios and operating conditions (e.g., under different traffic loads). In real IoT systems, it is not possible to efficiently cover all service quality parameters, so it is necessary to identify key performance indicators (KPIs) and QoS metrics and determine their significance for specific application cases. The performance of IoT systems directly depends on meeting the application requirements across all system components and their combined effect, which requires understanding the interconnected indicators across all layers of the architecture.

Quality of Service (QoS) in IoT applications represents the cumulative service level achieved through the execution of all underlying system services. Consequently, Key Performance Indicators (KPIs) defined at the application layer serve as aggregate reflections of the overall QoS performance realized across the various layers of the system architecture. One of the core challenges in IoT system design lies in identifying appropriate KPIs and metrics, as well as in defining their corresponding reference or boundary values. Among the most critical QoS indicators in IoT systems are

- Service execution time (latency): This metric denotes the total time required to execute all relevant processes within the system architecture. It encompasses the system's response time to a given query or task and is typically divided into three main components: data collection, data processing, and result interpretation. Latency is particularly crucial in real-time applications such as monitoring, control, and management, where timely system responsiveness is essential.

- Network QoS parameters: These include parameters such as throughput, latency, packet loss, jitter (variation in latency), network availability, etc. Throughput reflects the amount of data transmitted across the network over time, which is crucial for data-intensive applications. Latency is the time taken for data transmission from source to destination. Low latency is critical for applications like smart transportation systems and medical devices. Packet loss refers to the percentage of data lost during transmission, which affects the transmission quality, particularly in applications like video surveillance or voice communications. Jitter refers to the variation in the arrival time of data, which can cause disruptions in latency-sensitive applications like video conferencing.

- Energy efficiency: Represents the amount of energy consumed by the system during its operation. Efficient energy management extends battery life and reduces operational costs. System uptime defines the period during which the system can operate without replacing or upgrading components. This is especially true for battery-powered systems, where energy efficiency is crucial.

- Capacity of connected devices: Refers to the number of devices the system can support. IoT systems must enable the connection of a large number of devices that exchange data with one another, and capacity limitations can directly affect system efficiency. This depends on the processing and memory capabilities of the system components (e.g., servers) as well as network characteristics.

- Reliability and availability: Refers to the system's ability to continuously provide services without unexpected interruptions. High availability ensures system stability even under increased load or hardware failures.

- Security and privacy: Involves protecting data from unauthorized access, manipulation, or theft, and controlling the collected data. These aspects are particularly important for applications that process sensitive information.

- Interoperability: The ability of different devices, platforms, and systems to communicate with each other, regardless of differences in technologies, protocols, or manufacturers. This is a key characteristic of the IoT ecosystem.

Each of these indicators and network QoS parameters plays a key role in evaluating and optimizing IoT systems. Proper understanding of them enables achieving optimal functionality and aligning performance with the requirements of specific applications.

In most literature, time and energy efficiency are identified as key performance indicators. The execution time of IoT services (T_{exe}), which must be less than the threshold value required by the IoT application (T_{app}). Two approaches can be used for service allocation to available resources to achieve the stated goal: partial or complete distribution of certain services to available resources, which can be formally written as follows in Eq. (1):

$$T_{exe} = \alpha_0 \bullet t_L + arg \max_{i=1\ldots k} \{\alpha_i \bullet t_{Fi}\} + arg \max_{j=1\ldots c} \{\alpha_j \bullet t_{Cj}\} + \sum_{i=1}^{N} (\beta_i \bullet T_{cmi}) \leq T_{app} \quad (1)$$

According to the three-tier model, T_{cp} consists of local processing time (t_L), remote execution time including processing time at the fog layer (t_F), and processing time at the cloud computing system (t_C). The computation time for multiple fog layers is the maximum time required to execute tasks at a fog node (t_{Fi}). It is necessary to determine where to allocate certain services and what percentage of processing will be performed on the device (α_0) and what percentage of tasks will be offloaded to offloading systems, including fog infrastructure (α_i) and cloud (α_j). In the case of multiple systems on the same layer of the system architecture (i.e., fog systems and j cloud servers), the maximum processing time required by any system on the same layer to complete tasks is taken into account when estimating processing time at that layer. This approach assumes that the data being processed arrived simultaneously at the systems of the same layer, which is a realistic assumption, considering that these systems are generally at approximately equal distances from the data source or are located at relatively small distances from each other with fast linking connections. This means that the data transmission time to each of these systems is approximately the same. However, if certain processes are offloaded to computational offloading systems, it is necessary to account for the time required to transfer a certain percentage of the data (β_i) to these systems (across different layers of the system architecture), with the data passing through N network segments.

The analogy from the expression for application execution time can be used to set the following task for energy consumption in Eq. (2):

$$E_{dev} = \alpha_0 (P_L \bullet t_L) + P_{id} \bullet arg \max_{j=1\ldots k} \left\{\alpha_j \bullet t_{off_j}\right\} + P_{tx} \bullet \sum_{i=1}^{n} (\beta_i \bullet t_{txi}) \leq E_{app} \quad (2)$$

E_{dev} represents the energy consumed by the IoT device during the execution of all required services (computational and communication processes). P_L denotes the power needed for local processing, i.e., executing services directly on the IoT device. P_{id} refers to the power consumed during the idle state of the device, while P_{tx} represents the power required to transmit data over the network to n systems used for offloading the IoT device.

The objective is to allocate services across various system components to minimize energy consumption. These challenges can be formulated as an optimization problem to determine the values of α_0, α_j, β_i depending on the number of offloading subsystems involved.

5. Distribution of IoT services

IoT architectures based on integration with cloud systems raise questions related to performance, cost-effectiveness, administrative challenges, and more. The performance of IoT systems depends on the computing infrastructure used, network characteristics for data transmission, employed technologies, application requirements, and other factors. Cloud systems have significantly better processing capabilities, enabling faster data processing. However, transferring raw data to the cloud introduces additional network latency due to transmission over the Internet. On the other hand, infrastructure located closer to the data source (IoT devices) eliminates network latency over the WAN (Wide Area Network) link but processes data more slowly due to weaker processing capabilities compared to the cloud. This highlights the issue of selecting an appropriate integration model and optimizing the use of available resources. Based on existing analyses, certain characteristics of systems (Edge, Fog, MCC, MEC, Cloud) that allocate services in IoT systems can be identified (**Table 1**). These characteristics help determine when to allocate services to the cloud or other layers of the system architecture to offload IoT devices.

Based on strategies for process allocation, suitable integration models can be proposed, emphasizing simplicity, interoperability, accuracy, and real-world applicability. The task allocation process within the system architecture involves addressing key strategic questions:

- Which services and resources should be allocated outside IoT devices?
- When and under what circumstances should IoT devices be offloaded?
- How should the allocation of services and resources be performed?
- How and where should services and resources be distributed within the system architecture?

The allocation of IoT services across various layers of the system architecture can be based on preconfigured instructions at each layer (static decision-making) or dynamically enabled based on real-time system resource states and application requirements. Static decision-making can be implemented on resource-constrained systems where system load can be estimated in advance. Dynamic decision-making requires higher system performance and can be executed on gateways or dedicated infrastructures in larger systems, determining where data processing will occur. Specific tasks can be distributed across one or more layers of the system architecture.

Static decision-making requires developing a matrix to select optimal solutions based on prior testing. In contrast, dynamic decision-making requires implementing components for automatic performance evaluation. QoS performance evaluation involves updating the matrix as system conditions change (e.g., system load variations). Dynamic decision-making necessitates complex algorithms capable of detecting repeating patterns through continuous system state monitoring. Using

Parameter	Edge, Fog, Cloudlet, MCC, MEC	Cloud system
Data storage	Limited memory resources for short-term data storage.	Large memory resources for long-term storage of vast amounts of data.
Data processing and analysis	Limited processing capabilities for simpler tasks involving small, significant data.	Big data analysis, data mining, and other complex processing tasks.
Data transfer time	Significantly shorter due to proximity of fog infrastructure.	Data must traverse the Internet, causing greater network latency.
Data processing time	Longer due to limited hardware capabilities.	High processing speeds are enabled by resource sharing across numerous servers.
Data transfer costs	Significantly lower due to reduced data transmission over WAN networks.	Higher costs due to transferring raw data over the Internet to cloud infrastructure.
Internet connectivity	Supports devices with limited hardware and software capabilities without requiring Internet connectivity, which can reduce energy consumption.	Devices with limited resources must have network interfaces for cloud connectivity, increasing energy consumption.
Network load	Reduced traffic through hierarchical processing and data filtering.	All data must be transmitted from IoT devices to the cloud, causing significant network load.
Data loss	Reduced or eliminated losses due to fewer hops and reduced network congestion.	Higher data loss rates due to transmission over the Internet.
Jitter (delay variations)	Minimal jitter due to fewer packet hops and low latency.	Higher jitter due to Internet transmission causing variable packet paths.
Infrastructure location	Centralized or distributed architecture closer to IoT devices.	Centralized architecture primarily consists of large data centers located worldwide.
Number of services and Management	Fewer services with adaptable management due to closer infrastructure access.	A greater number of services are managed depending on the cloud model (IaaS, PaaS, SaaS).
Programmability	Challenges remain due to early-stage concept development.	Extensive capabilities with numerous tested services.
Security and privacy	Reduced threats due to smaller data volumes transmitted. Functions as a proxy for the cloud and offers additional functionalities like status monitoring and threat detection.	Higher likelihood of data attacks due to communication channel vulnerabilities but better protection for stored data.
Location-based services and mobility	Enables real-time location-based services and full mobility support depending on communication technologies.	Limited real-time location-based services and constrained mobility support due to latency and connectivity issues.
System reliability	Supports autonomous operations without Internet connectivity, allowing uninterrupted services.	Requires constant Internet connectivity for uninterrupted services, challenging in areas with limited network coverage.
Energy consumption	Can significantly reduce IoT device energy consumption through energy-efficient components and networks.	Potentially energy-inefficient due to traditional Internet protocols and networks not optimized for low power usage.

Table 1.
Comparison of Edge, Fog, Cloudlet, MCC, MEC, and cloud characteristics in IoT Systems.

machine learning tools, it becomes possible to distinguish system conditions, recognize available resources, and assess performance based on these factors. Performance assessments of selected KPIs depend on system load, available resources, and weighted values of individual indicators.

The challenge arises when considering multiple QoS performance indicators simultaneously. This creates a need for a method that incorporates multiple metrics to determine acceptable implementations for each layer of the system architecture. Achieving desired performance levels for individual IoT applications should be viewed as the cumulative impact of the service, network, and perception layers on the application. Thus, there is a dependency between the QoS performance levels of IoT applications and the characteristics of the applied infrastructure and technologies. In prior research [1], a multi-criteria decision-making method was proposed as an effective tool enabling the application of different metrics within a normalized range, independent of measurement units. This method offers significant advantages, including independence from specific KPIs, simplified selection of reference values, and adaptability for achieving desired results through utility function adjustments.

The QoS-aware IoT architecture proposed in this study, combined with the previously mentioned method, can serve as a framework for developing decision-making modules capable of optimizing task allocation across available resources in the IoT system architecture. Future research should focus on identifying key indicators for specific groups of IoT applications, mapping their weighted values across all architectural layers, and developing energy-efficient mechanisms for devices with limited processing and memory capabilities.

6. Conclusion

Optimal allocation of resources across IoT architecture layers is crucial for achieving the desired QoS performance. This chapter presents the key aspects of QoS performance control and management, proposing a layered architecture that enables efficient resource management. The proposed QoS-aware architecture provides a flexible and structured framework for developing intelligent decision-making mechanisms, allowing service allocation to adapt to different infrastructures and layers of the IoT architecture. Its versatility makes it applicable to various scenarios, including fixed device locations and mobility cases.

For successful implementation in real-world environments, several key prerequisites must be fulfilled. First, it is essential to identify key performance indicators (KPIs) and metrics that define the QoS level, taking into account specific requirements and threshold values for individual services in different IoT applications. Future research should focus on developing QoS profiles for specific applications and determining the weighting factors for individual parameters to establish their relative importance in various application contexts.

One of the challenges is translating the QoS requirements of the application layer into requirements for the lower layers of the proposed architecture. This requires synergy between different system layers to ensure optimal performance and satisfactory service quality that meets user expectations. Effectively translating these requirements provides the foundation for planning, implementing, controlling, and managing IoT systems.

Identifying key metrics and developing appropriate QoS profiles for IoT applications enable the practical implementation of the proposed architecture. This opens the way for the development and application of intelligent decision-making mechanisms for resource allocation, optimizing their use, and achieving the desired level of QoS performance. Such an approach contributes to the creation of scalable and flexible IoT systems that meet the demands of modern applications and user expectations.

Conflict of interest

The authors declare no conflict of interest.

Author details

Alem Čolaković*, Bakir Karahodža and Adisa Hasković Džubur
University of Sarajevo, Bosnia and Herzegovina

*Address all correspondence to: alem.colakovic@fsk.unsa.ba

References

[1] Čolaković A. IoT systems modeling and performance evaluation. Computer Science Review. 2023;**50**:1-26. DOI: 10.1016/j.cosrev.2023.100598

[2] Abdulazeez DH, Askar SK. Offloading mechanisms based on reinforcement learning and deep learning algorithms in the fog computing environment. IEEE Access. 2023;**11**:12554-12585. DOI: 10.1109/ACCESS.2023.3241881

[3] Zahoor S, Mir RN. Resource management in pervasive internet of things: A survey. Journal of King Saud University-Computer and Information Sciences. 2021;**33**(8):921-935. DOI: 10.1016/j.jksuci.2018.08.014

[4] Alonso RS, Sittón-Candanedo I, García Ó, Prieto J, Rodríguez-González S. An intelligent edge-IoT platform for monitoring livestock and crops in a dairy farming scenario. Ad Hoc Networks. 2020;**98**(3):102047. DOI: 10.1016/j.adhoc.2019.102047

[5] Badidi E, Ragmani A. An architecture for QoS-aware fog service provisioning. Procedia Computer Science. 2020;**170**: 411-418. DOI: 10.1016/j.procs.2020.03.083

[6] Haidri RA, Katti CP, Saxena PC. Cost effective deadline aware scheduling strategy for workflow applications on virtual machines in cloud computing. Journal of King Saud University-Computer and Information Sciences. 2020;**32**(6):666-683. DOI: 10.1016/j.jksuci.2017.10.009

[7] Khebbeb K, Hameurlain N, Belala F. A Maude-based rewriting approach to model and verify cloud/fog self-adaptation and orchestration. Journal of Systems Architecture. 2020;**110**:10182. DOI: 10.1016/j.sysarc.2020.101821

[8] Zhang J, Ma M, He W, Wang P. On-demand deployment for IoT applications. Journal of Systems Architecture. 2020;**111**:10179. DOI: 10.1016/j.sysarc.2020.101794

[9] Xavier TCS, Santos IL, Delicato FC, Pires PF, Alves MP, Calmon TS, et al. Collaborative resource allocation for cloud of things systems. Journal of Network and Computer Applications. 2020;**159**(10259):1-18. DOI: 10.1016/j.jnca.2020.102592

[10] Tuli S, Mahmud R, Tuli S, Buyya R. FogBus: A blockchain-based lightweight framework for edge and fog computing. Journal of Systems and Software. 2019; **154**:22-36. DOI: 10.1016/j.jss.2019.04.050

[11] Safavat S, Sapavath NN, Rawat DB. Recent advances in mobile edge computing and content caching. Digital Communications and Networks. 2020; **6**(2):189-194. DOI: 10.1016/j.dcan.2019.08.004

[12] Wei H, Luo H, Sun Y, Obaidat MS. Cache-aware computation offloading in IoT systems. IEEE Systems Journal. 2020;**14**(1):61-72. DOI: 10.1109/JSYST.2019.2903293

[13] Kumari N, Yadav A, Jana PK. Task offloading in fog computing: A survey of algorithms and optimization techniques. Computer Networks. 2022;**214**:1-24. DOI: 10.1016/j.comnet.2022.109137

[14] Gill SS, Garraghan P, Buyya R. ROUTER: Fog enabled cloud based intelligent resource management approach for smart home IoT devices. Journal of Systems and Software. 2019; **154**:125-138. DOI: 10.1016/j.jss.2019.04.058

[15] Bacanin N, Zivkovic M, Bezdan T, Venkatachalam K, Abouhawwash M. Modified firefly algorithm for workflow scheduling in cloud-edge environment. Neural Computing and Applications. 2022;**34**:9043-9068. DOI: 10.1007/s00521-022-06925-y

[16] Wang S, Song X, Xu H, Song T, Zhang G, Yang Y. Joint offloading decision and resource allocation in vehicular edge computing networks. Digital Communications and Networks. 2023;**11**(1):71-82. DOI: 10.1016/j.dcan.2023.03.006

[17] Kanellopoulos D, Sharma VK. Dynamic load balancing techniques in the IoT: A review. Symmetry (Basel). 2022;**14**(12):1-60. DOI: 10.3390/sym14122554

[18] Bu T, Huang Z, Zhang K, Wang Y, Song H, Zhou J, et al. Task scheduling in the internet of things: Challenges, solutions, and future trends. Cluster Computing. 2023;**17**:1017-1046. DOI: 10.1007/s10586-023-03991-2

[19] Zhou S, Jadoon W, Khan IA. Computing offloading strategy in mobile edge computing environment: A comparison between adopted frameworks, challenges, and future directions. Electronics. 2023;**12**(11):1-30. DOI: 10.3390/electronics12112452

[20] Maray M, Shuja J. Computation offloading in mobile cloud computing and mobile edge computing: Survey, taxonomy, and open issues. Mobile Information Systems. 2022;**2022**:1-17. DOI: 10.1155/2022/1121822

[21] Gasmi K, Dilek S, Tosun S, Ozdemir S. A survey on computation offloading and service placement in fog computing-based IoT. The Journal of Supercomputing. 2022;**78**:1983-2014. DOI: 10.1007/s11227-021-03941-y

[22] Čolaković A, Hasković Džubur A, Karahodža B. Wireless communication technologies for the internet of things. Science, Engineering and Technology. 2021;**1**(1):1-14. DOI: 10.54327/set2021/v1.i1.3

[23] Babovic ZB, Protic J, Milutinovic V. Web performance evaluation for internet of things applications. IEEE Access. 2016;**4**:6974-6992. DOI: 10.1109/ACCESS.2016.2615181

[24] Madsen H, Burtschy B, Albeanu G, Popentiu-Vladicescu F. Reliability in the utility computing era: Towards reliable Fog computing. 20th International Conference Systems, Signals and Image Processing (IWSSIP); 10 October 2013. Bucharest, Romania: IEEE; 2013; pp. 43-46. DOI: 10.1109/IWSSIP.2013.6623445

[25] Chiang M, Zhang T. Fog and IoT: An overview of research opportunities. IEEE Internet of Things Journal. 2016;**3**(6):854-864. DOI: 10.1109/JIOT.2016.2584538

Chapter 8

An Application of Machine Learning, Big Data, and IoT of Enterprise Architecture: Challenges, Solutions and Open Issues

Abu Sarwar Zamani, Md. Mobin Akhtar and Nayyar Ahmed Khan

Abstract

Big data analytics has emerged as a tool to support clinical decision-making across a range of clinical domains because clinical judgments are robust and grounded in evidence. One of the most effective IoT and intelligent automation tools in enterprise architecture (EA) is machine learning (ML). Despite its potential, ML applications for IoT services and systems require specialized research. Analyzing IoT-generated data and IoT-based networks is essential given the importance of the Internet of Things. Numerous research has looked into how ML may address particular IoT-related problems. IoT applications may use sensor data to improve ML models, thanks to these mutually supporting technologies, which improves IoT operations and procedures. Additionally, ML approaches enable the detection of suspicious activity in smart objects and systems and provide knowledge to IoT systems. This study thoroughly examines the function of machine learning in Internet of Things applications. In the context of IoT, it offers a thorough examination of the most advanced machine learning techniques, stressing their advantages, disadvantages, and possible uses, and also, a big data analytics model based on machine learning is created for multi-disease prediction. The average weight function between the fuzzy classifier, KNN, and NN outputs yields the final prediction output. Finally, by contrasting the outcomes of the newly created prediction techniques, the efficacy calculated for the created multi-disease prediction framework is promised.

Keywords: big data, machine learning techniques, Internet of Things, accuracy, healthcare

1. Introduction

Massive internet connectivity combined with abundant bandwidth has accelerated the development of the Internet of Things during the last few years. The IoT establishes its vital position for new electronics sector growth through remote control and automation capabilities, which engineers can use to operate their devices. IoT has executed a total transformation of medical practices among other domains. IoT technology enables the medical field to regenerate different healthcare applications while adding value to their development process [1]. The delivery of multiple medical services stands possible through IoT-based healthcare systems that enable both data transmission as well as information exchange and machine-to-machine communication while ensuring system interoperability [2]. Patients can obtain healthcare data remotely because of IoT-based wireless devices that use wearables and implants. The medical application of the Internet of Things is known as IoMT (Internet of Medical Things). The IoMT system receives vital patient symptoms from sensors that employ IoT technology in wireless remote monitoring [3, 4]. The information gathered through IoT-based applications allows healthcare professionals to develop suitable treatment plans and maybe require doctor intervention or webcam consultation in real time. Predictive analytics results from the gathered health patterns through IoT-based sensors to help doctors determine diseases with accuracy while processing data at high speeds [5]. The information equipment provides unique decision-making information that enables meaningful support in healthcare. The wide variety of data that need processing stem from the excessive noise contained within the data [6]. Multiple vital privacy and security risks appear during the processing of IoMT information because medical devices record and transfer data between practical spaces without standard data protocols or data ownership compliance systems [7].

The efficient management of various types of large datasets faces hurdles in information retrieval as well as visualization and processing and storage operations. Big data analytics faces its biggest challenge when seeking efficient methods to acquire information needed by different types of users according to Ref. [8]. Healthcare analytics makes use of highly valuable patient medical history data that result from continuous collection of various healthcare sources from clinical and non-clinical settings [9]. A distributed data system remains the only solution to address these following technical limitations. The main difficulties associated with information collection through shared locations begin with excessive heterogeneous data production [10]. The analysis of extensive mixed-dataset information produces storage problems [11]. A big data system operates most efficiently with an effective storage solution. Analysts need to overcome the difficulty of implementing big data analytics for working with large datasets across real-time platforms that require optimization, prediction, visualization, and model building [12]. Data management systems currently experience low efficiency when handling practical along with heterogeneous information, which requires a new processing approach for creating an innovative data management system [13].

The parallel processing approach known as MapReduce analyzes large, distributed datasets stored within commodity cluster environments. The framework includes a combination of Map as well as Reduce operations. The MapReduce framework solves complex problems and training complexities that emerge when working with big data [14]. Development of Hadoop began as a batch-wise processing architecture through MapReduce framework deployment for big data shared storage along with analysis purposes [15]. Machine learning functions as a scientific subdomain under computer science discipline. Machine learning represents an artificial intelligence subfield that makes machines learn independently for specific programming applications [16, 17]. Machine learning exists because of its origins in pattern discovery and computational learning theory [18]. The substantial rise in data volume together with various data types has enabled powerful tools for data storage analysis, which provides important insights for decision support [19].

Machine learning algorithms applied to healthcare enable disease forecasting, which allows patients to obtain beneficial medical actions when necessary. The research foundation originates from the points outlined below. The current health information system demonstrates substandard performance regarding data threats together with transmission delay errors although it achieves limited success. The system generates incorrect diagnoses for bigger datasets during its operation. This research develops a healthcare data prediction model through integrating different machine-learning techniques. The developed model efficiently manages high amounts of data through an accurate feature extraction process, which produces vital information. The result analysis has proven effective through validation testing.

2. Literature review

Kumar et al. [20] developed a genetic-based fuzzy C-means data clustering protocol to detect states of over-edge affected elements. The clustering method generated accurate data, which went through genetic modifications to prevent getting stuck at local optima. The classification process was conducted using a deep learning method as its conclusion. The system underwent validation procedures, while researchers checked how model execution results matched conventional techniques for time-related aspects. The data clustering strategy reduced processing times substantially while working with realistic data streams according to the conducted analysis.

The clinical data system developed by Tan et al. [21] received assistance through big data analytics before its implementation. To test and train the applied model, the patients' information was randomly sectioned off. This model used electronic health record processing to develop five end results through synthetic minority oversampling procedures. The assessment of the developed model established its superiority against conventional models in performance evaluation results.

Safa and Pandian and Zamani et al. [22, 23] developed a big data-based system to detect cardiac diseases through IoT devices using fuzzy rules in their implementation. The fuzzy rule produced characteristics that helped diagnose cardiac diseases before their employment in model training. The optimized recurrent neural network (RNN) performed the detection process. The classification happened after information processing through the RNN. The proposed model achieved a high performance rate through experimental evaluations that compared against traditional techniques.

Ed-daoudy et al. [24] created a practical health status prediction architecture and built an analytics system through big data approach support. Health status predictions became possible for the system by implementing selected input features while simultaneously sending alert notifications to care providers. The distributed database keeps this data for performing real-time stream reporting and health data analytics. A performance analysis took place by assessing execution pace and operational capacity. The developed model outperformed traditional models in terms of providing observational outcomes.

Manogaran et al. [25] built a "Bayesian Hidden Markov Model (HMM) along with the support of the Clustering technique" for creating a system to predict genomic DNA copy number variations. The created clustering technique established its links with conventional techniques through its relationship to the segment neighborhood and binary segmentation techniques. The developed model demonstrates its effectiveness according to analysis reports for detecting DNA number changes with high precision.

Nibareke and Laassiri [26] developed an ensemble machine technique dedicated to diabetes disease prediction. An analysis of performance took place using flight delays data. The developers used machine learning frameworks together with big data tools to perform the overview. A combination of evaluation metrics assessed the prediction accuracy. A test-analysis on diabetes prediction took place. Multiple conventional models underwent comparison with the produced results. This applied model demonstrated improved performance when compared to traditional frameworks according to the results.

Ashiku et al. [27] investigated Apache Spark's open-source capabilities to analyze massive data over clusters in order to evaluate big data and included technologies to support decision support systems in healthcare settings. The decision-making process for organ distribution used Apache Spark through developed machine learning frameworks, which selected kidneys for suitable candidates and increased donor effectiveness by localizing speedy recipients. The implemented framework successfully determines qualified waiting list patient candidates who accept kidneys that other systems ignore.

The combination of big data analytics and Deep Belief Network (DBN) forms the basis of a disease diagnostic framework with feature selection recommendations according to Pustokhin et al. [28]. The "Link-based Quasi Oppositional Binary Particle Swarm Optimization Algorithm" functioned to select optimal features by reducing feature numbers through a process that helps prevent dimensionality issues. The DBN performed disease classification with help from the dataset whose features were decreased. Experimental analysis confirms that the developed model delivers superior performance through its advanced configuration and stands superior to traditional disease detection procedures.

Marcin et al. [29] proposed an IoT model that combined Bi-LSTM along with decision trees with data balance strategies for building an automated diagnosis support system. An optimal solution requires the use of the data pre-processing model for training the effective network model. The framework goes through testing with several metrics, thus delivering effective precision, recall results, and accurate solutions. The 6G healthcare system utilizing artificial intelligence was developed through IoT by Vara Siddardha Reddy et al. [30]. The research establishes artificial intelligence and IoT implementation within medical infrastructure across specific clinical medical areas. The diverse analysis has demonstrated how the developed model performs effectively based on comparisons with other state-of-the-art procedures.

3. Different aspects of machine learning, IoT, and big data

3.1 Characteristics of machine learning

In the field of technology, machine learning has gained a lot of popularity in recent years. The features of machine learning in the data-rich world of today are echoed by several examples. Companies should choose machine learning over other technologies because of these seven important features.

3.1.1 The capacity to visualize data automatically

Businesses and everyday individuals regularly generate enormous amounts of data. Businesses can increase confidence and make better decisions by showing significant correlations in data. Numerous technologies for machine learning give rich data snippets that may be used with both structured and unstructured data. Businesses can gain a multitude of fresh insights to try to boost process productivity with the aid of machine learning's user-friendly automated data visualization tools.

3.1.2 The best of automation

The potential of machine learning to automate repetitive processes and hence boost productivity is one of its most notable features. Email and paperwork automation driven by machine learning is already being used by a vast number of enterprises. For instance, a great deal of predictable, data-intensive, and repetitive work is required in the banking industry. As a result, this industry makes extensive use of various machine learning systems. They improve the speed, accuracy, and insight of accounting tasks. Machine learning has already covered a number of areas, such as monitoring spending, automating bank reconciliations, making forecasts, handling financial inquiries using chatbots, and streamlining invoicing.

3.1.3 Engaging customers like never before

Starting meaningful conversations with its target customer base is one of the most important strategies for any organization to increase engagement, foster brand loyalty, and create enduring customer connections. In order for brands and companies to engage customers in more meaningful dialogs, machine learning is essential. Certain words, phrases, sentences, idioms, and material styles that appeal to particular audience members are analyzed by the technology. One example is Pinterest, which effectively uses machine learning to give its consumers personalized recommendations. Based on items users have already pinned, it leverages technology to find information that would pique their interest.

3.1.4 The potential to increase efficiency when combined with IoT

Machine learning has become increasingly popular. Many businesses have identified the Internet of Things as a strategically important sector. In order to assess the potential of IoT in relation to company operations, numerous others have started pilot projects. However, obtaining financial gains from IoT is not simple. Businesses that provide IoT platforms and consulting services must identify the areas that will change as a result of IoT strategy implementation in order to succeed. A lot of these companies have not dealt with it. The most effective technique in this situation for achieving more efficiency is most likely machine learning. Businesses may increase the efficiency of all of their production processes by combining machine learning with the Internet of Things.

3.1.5 The power to alter the mortgage industry

It is a well-known fact that many consumers require time, discipline, and extensive financial preparation in order to cultivate a positive credit score. One of the most important indicators of trustworthiness for lenders is the consumer credit score, which takes into account a variety of variables such as payment history, total debt, length of credit history, and so on. However, would it not be fantastic if there were a better and more straightforward measure? Machine learning has made it possible for lenders to have a more complete picture of their customers. They can now determine the customer's spending tipping point and determine if they are a high or low spender. Financial institutions are applying the same strategies to other consumer credit categories in addition to home lending.

3.1.6 Precise data analysis

Trial and error methods have long been a part of data analysis, but when dealing with vast and diverse datasets, this strategy becomes unfeasible. The best answer to all of these problems is machine learning, which provides efficient substitutes for the analysis of enormous amounts of data. Machine learning produces correct analysis and results by creating quick and effective algorithms and data-driven models for real-time data processing.

3.1.7 The best of business intelligence

When combined with big data analysis, machine learning features can produce extremely high levels of business information, which are being used by a number of industries to launch strategic initiatives. Machine learning has already emerged as one of the most powerful tools for improving corporate operations across a wide range of industries, including retail, finance, healthcare, and many more.

As stated in Section 3.1, despite the recent major advancements in machine learning brought about by the rise of big data, much more work needs to be done to address the numerous important issues raised by big data. As illustrated in **Figure 1**, we discuss the key concerns of machine learning approaches for big data from five distinct angles in this section. These include learning for large data scales, learning for various data types, learning for high streaming data speeds, learning for uncertain and incomplete data, and learning for extracting useful information from vast amounts of data.

3.2 Characteristics of IoT

IoT, machine learning, and artificial intelligence are all related or overlap in a variety of ways. Machine learning and artificial intelligence devices may be found in an Internet of Things (IoT). A machine application that simulates intelligent traits is a common definition of artificial intelligence. Machine learning is a branch of artificial intelligence that allows a machine to learn from the data it has access to. Machine

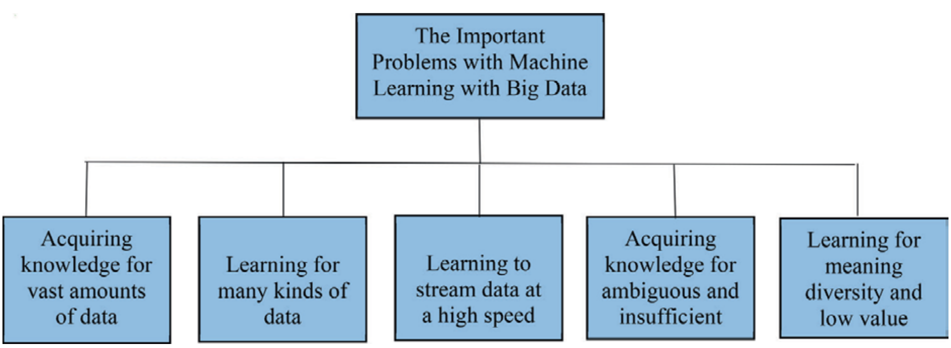


Figure 1.
The important problems with big data and machine learning.

learning makes cognitive decisions by applying supervised learning approaches to historical data. The algorithm's ability to make decisions improves with the amount of historical data. Because of this attitude, IoT is the perfect use case for machine learning because the devices typically generate a lot of data. Here are some such situations where machine learning and IoT collaborate to facilitate business optimizations.

3.2.1 Observation of anomalies

Anomalies in time series data, or data feeds transmitted by IoT devices that are consistently spaced in time, can be found using Azure machine learning. A machine learning algorithm that tracks the live stream of device data can identify anomalies such as spikes and dips, positive and negative trends.

3.2.2 Using predictive maintenance

One of the most widely used machine learning solutions is predictive maintenance, which has a direct impact on an organization's expenses. Machine learning algorithms' capacity to predict a device's likelihood of malfunctioning, its remaining life, and its causes can help businesses optimize operating costs by drastically cutting down on maintenance time.

3.2.3 Telemetry of vehicles

Machine learning solutions are a desirable technology to use for the transportation and logistics industries because of its capacity to absorb millions of events from vehicles to enhance their safety, dependability, and driving experience.

The cloud solution from Microsoft established itself first as the full IoT and machine learning system from a widely recognized cloud provider. Microsoft's offering includes a number of technologies that are provided as services to support various stages of the Internet of Things pipeline. Among these technologies are the following: Azure Stream Analytics, Azure Machine Learning, Azure IoTHub, Azure IoT suit, Azure IoTEdge, and Azure EventHub.

3.3 Characteristics of big data

Large and varied datasets that are enormous in number and fast expanding in size over time are referred to as big data. Predictive modeling, machine learning, and other advanced analytics employ big data to address business issues and make wise judgments. Big data is the term used to describe data that is too big or complicated to handle using conventional techniques. The five Vs that define it are volume, variety, velocity, veracity, and value.

Machine learning and big data are two sides of the same coin. The fuel that machine learning algorithms require to learn is big data. With the help of machine learning algorithms, we can glean insights from large amounts of data that would be impossible to uncover using conventional techniques.

Both machine learning and big data are effective tools, each with advantages and disadvantages of their own. While machine learning is better at drawing conclusions and predictions from data, big data is better at storing and analyzing massive quantities. Since they are complementary technologies, each one improves the other's functionality.

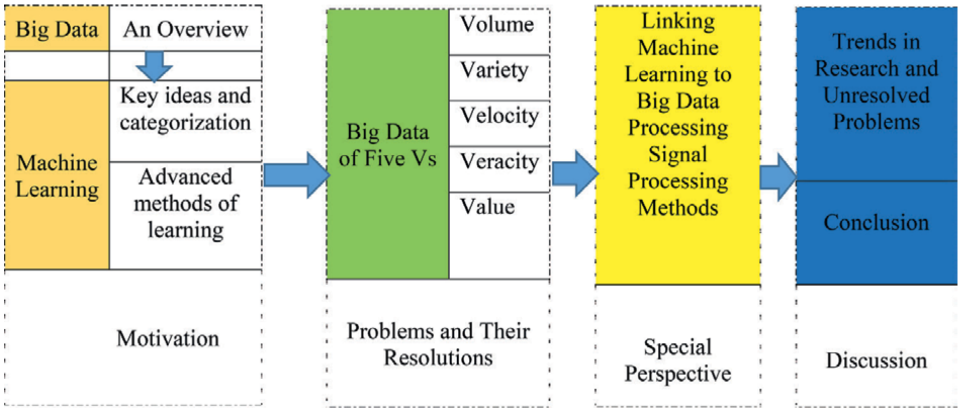


Figure 2.
Roadmap of the survey.

The roadmap in **Figure 2** illustrates how the rest of the paper is structured. After reviewing some key and pertinent machine learning ideas, we move on to several contemporary advanced learning approaches in Section 3.1. Machine learning offers a thorough analysis of the difficulties presented by massive data for machine learning, primarily from five perspectives.

4. Evaluation and challenges of machine learning, big data, and IoT

An established set of metrics, including accuracy, error rate, precision, recall, squared error, likelihood, posterior probability, information gain, K-L divergence, cost, utility, margin, optimization error, estimation error, approximation, mean, and worst outcome, are used in traditional machine learning to assess performance. The accuracy of ML predictions is the main emphasis of these measurements. Additionally, the focus of big data analytics is scalability, which is typically used to assess a parallel program. Data I/O performance, fault tolerance, real-time processing, memory utilization, data size supported, iterative task support, and throughput are some of the criteria that have been operationalized as scalability [4]. Analyzing large amounts of data, Machine Learning does not simply combines two kinds of metrics. It must handle intricate trade-offs between the measurements as well as trade-offs within each type of metric. For example, traditional performance trade-offs include accuracy and response speed, precision, and memory. Iterative task support conflicts with fault tolerance when it comes to scalability (MapReduce, for example, offers fault tolerance but not iteration). Furthermore, non-iterative methods (like the Nystrom approximation) perform somewhat worse than iterative ones (like Eigen decomposition), but they scale better.

The latter can be highly useful in reality, even though it is less theoretically interesting to build an algorithm that is slightly less accurate but has higher system dependability and ease of use [31]. As a result, creating useable machine learning on large data will help data scientists get trained (for example, in data preparation, process optimization, and parameter tuning), and ML will be widely used in practice. Charts, graphs, and other visual aids are commonly used to explain data because people are able to quickly make conclusions based on similarities and patterns. However, this strategy is becoming limited as the number of data increases [32]. Big data ML needs to be widely accepted in society in order to have an influence; therefore, data ethical concerns including data privacy, security, ownership, liability, and behavioral targeting must also be addressed [33, 34].

Analyzing IoT data could result in the discovery of important big data information in this respect. Machine learning techniques are also applicable in the field of large data analysis [35, 36]. These methods can also aid in lowering huge data storage space [37, 38]. IoT data analysis generally benefits from the use of machine learning techniques [39]. Big data is efficiently gathered and stored by IoT technology from IoT devices. Machine learning techniques can then be used to evaluate and extract valuable information from these data. Machine learning is frequently employed as a means of achieving artificial intelligence in place of deep learning. The ability to learn without explicit preparation was the definition of the phrase when it was first used by Arthur Samuel in 1959. The connection between IoT and machine learning is depicted in **Figure 3**.

A thorough evaluation is conducted on the published literature that deals with IoT-based big data and machine learning. Researchers may become more motivated to pursue this expanding field of study as a result of the examination of these articles, and new concepts may arise. To find publications, we scan digital scientific libraries like Scholar, Scopus, and the Association for Computing Machinery (ACM) [35]. The application of machine learning techniques has significantly improved Internet of Things effectiveness during the previous couple of years. Generally speaking, the future generation of artificial intelligence-based computing systems can be provided by three cutting-edge scientific technologies: machine learning, big data, and the Internet of Things. Machine learning-based big data analysis for the Internet of Things is acknowledged as a new, transdisciplinary discipline. Its expansion in both theoretical underpinnings and practical applications can be attributed to this characteristic as well as the transformative one.

We then examined the abstracts of the discovered papers to manually filter them. This allowed us to eliminate items that had nothing to do with the current area of study. **Table 1** indicates that, following the conclusion of the search process and in

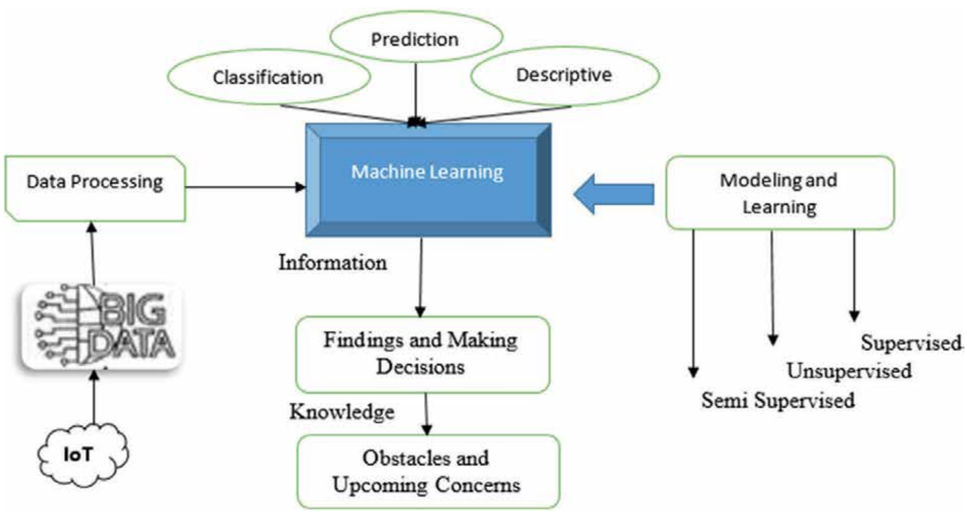


Figure 3.
Machine learning techniques and big data analysis for IoT.

Fields	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
IoT and big data	0	1	1	6	12	30	39	27	41	37	40	48	282
IoT and machine learning	1	1	1	0	1	3	7	4	7	8	11	26	70
Total	1	2	2	6	13	33	46	31	48	45	51	74	352

Table 1.
Details of articles found based on year of publication.

compliance with the established approach, 352 articles were located, of which 282 dealt with IoT and big data and 70 with the combination of IoT and machine learning [40]. The combination of Internet of Things and big data emerged for the first time in 2010, while experts started proposing it during 2011. Furthermore, it is evident that this area of study has expanded significantly since 2011.

The integration of IoT-based machine learning with IoT-based big data remains understudied because researchers consider it a promising research direction for the future.

5. Integration of IoT, big data platform, and ML

In addition to improving data processing capabilities, integrating a big data platform with IoT and machine learning enables real-time actions based on insightful insights. IoT and ML demonstrate a straightforward domain integration. Organizations may effectively handle the flood of data from various IoT sources by utilizing a big data platform, such as Kafka, which offers a strong data transport mechanism, guaranteeing data integrity and prompt delivery. When incorporated into this pipeline, machine learning models are able to instantly examine data for trends, patterns, and abnormalities. Applications that promote efficiency, cost savings, and innovation across industries are made possible by the combination of Kafka, IoT, and machine learning.

5.1 Big data-based healthcare data prediction using a map-reduce framework

MapReduce provides an approach to parallel programming, which distributes evaluation tasks across multiple computing devices for handling extensive datasets. The reduction phase operates after the mapping phase. The first stage requires the initialization of the map stage. The input data at this stage go through the map function to achieve analysis. The second operational phase generates particular intermediate outputs, which eventually create the final result. MapReduce model employs all processed data as key-value pair elements while functioning according to the key-value pair fundamental structure which serves as the data structure basis. The procedures of map phase and reduction phase follow the described pattern.

5.1.1 Map phase

In order to create autonomous blocks and distribute them across the block nodes, the header node participates in the segmentation of the input dataset. Additionally, the header node receives the solution after the block node processes the minor issues. The key-value pair is obtained as the input by the map function, which also generates a set of intermediate key-value pairs as outputs. The MapReduce library speeds up processing in the Reduce phase by combining all intermediate values associated with the relevant intermediate key before starting the Reduce phase. The MapReduce framework's Map phase is when methods for feature extraction are carried out. The statistical and Principal Component Analysis (PCA)-based approaches are used to carry out the feature extraction. More complexity is produced when using higher dimensional information for model training. Therefore, the PCA is used in this work to improve the interoperability of the healthcare data and minimize the feature-length [23].

5.1.2 Reduce phase

The header node unites the solutions from sub-problems before performing required combinations to generate the final output. The MapReduce library ensures transitional keys for the reduction phase, which produces end results for specific value-key combinations by considering their pair structures. The Reduce phase operates using the features derived from both PCA-based analysis as well as statistical feature extraction independently. The selected optimal features result from implementing the Hybrid Flower Pollination Bumblebees Optimization Algorithm (HFPBOA). The selection of the best prediction features for disease diagnosis in individual patients uses this HFPBOA (Hybrid Flower Pollination Bumblebees Optimization Algorithm) [23].

5.2 Modeling machine learning using neural networks (NN)

This ensemble-based multi-disease prediction model uses NN [41] and takes weighted data as input. The input units in this model pass the data function. The gadget manages different hidden layer modules and produces output-level results. An activation function makes up the complete unit [23].

5.3 A machine learning model that uses a fuzzy classification system

The fuzzy classifier is used for the multi-disease prediction [42]. Health-related diseases are identified using this methodology. The fuzzy classifier manipulates and expresses ambiguity and uncertainty by utilizing fuzzy set theory. Fuzzy logic is used to help construct the rules [23].

5.4 A K-nearest neighbor (KNN) machine learning model

This multi-disease prediction model uses KNN [43] to diagnose a variety of illnesses. The k closest training samples across the feature space are used as the input in this technique, which is regarded as a non-parametric approach. The output function is observed to be the class membership function [23].

6. Conclusion

Our everyday lives now revolve on the Internet of Things paradigm. Over the past 10 years, IoT technology advancements have resulted in the creation and gathering of data on a never-before-seen scale. Data management has become difficult due to the large amount of data produced by IoT devices. Machine learning has used its contemporary study approach to improve the resolution of real-life challenges including resource allocation, traffic engineering, secure systems, and routing procedures. In the context of the Internet of Things, this chapter emphasizes the significance of machine learning (ML) in facilitating independent, autonomous, and intelligent enterprise management. In today's businesses, integrating IoT technologies with either freshly developed or pre-existing application systems has become common procedure. ML has the ability to improve IoT-based enterprise systems by facilitating intelligent decision-making and automation. In comparison to the other models examined, the findings demonstrated the effectiveness and high accuracy of the suggested hybrid prediction model. However, the creation of adaptive, intelligent solutions using machine learning techniques is hampered by the limited computational and connectivity capabilities of IoT devices. Even though platform improvements and technological advancements open the door to a future with strong analytics of large volumes of IoT data, application deployment, and rapid IoT proliferation, we have maintained that it has proven challenging to integrate intelligent solutions from various domains. The authors have defended the use of ML in IoT in this chapter. The difficulties and potential applications were the main topics of discussion.

Author details

Abu Sarwar Zamani[1]*, Md. Mobin Akhtar[2] and Nayyar Ahmed Khan[3]

1 Department of Computer and Self Development, Prince Sattam Bin Abdulaziz University, Al Kharj, Saudi Arabia

2 Department of Basic Sciences, Riyadh Elm University (REU), Riyadh, Saudi Arabia

3 Department of Computer Science, College of Computing and Information Technology, Shaqra University, Saudi Arabia

*Address all correspondence to: a.zamani@psau.edu.sa

References

[1] Yu W, Liu Y, Dillon T, Rahayu W, Mostafa F. An integrated framework for health state monitoring in a smart factory employing IoT and big data techniques. IEEE Internet of Things Journal. 2022;**9**(3):2443-2454

[2] Zheng M, Bai S. Implementation of universal health management and monitoring system in resource-constrained environment based on internet of things. IEEE Access. 2021;**9**:138744-138752

[3] Sivakumar NR, Karim FKD. An IoT based big data framework using equidistant heuristic and duplex deep neural network for diabetic disease prediction. Journal of Ambient Intelligence and Humanized Computing. 2021:1-11

[4] Singh D, Reddy CK. A survey on platforms for big data analytics. Journal of Big Data. 2014;**2**:1-20

[5] Subramaniyaswamy V, Manogaran G, Logesh R, Vijayakumar V, Chilamkurti N, Malathi D, et al. RETRACTED ARTICLE: An ontology-driven personalized food recommendation in IoT-based healthcare system. The Journal of Supercomputing. 2019;**75**:3184-3216

[6] Andrew Onesimu J, Karthikeyan J, Sei Y. An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services. Peer-to-Peer Networking and Applications. 2021;**14**:1629-1649

[7] Zhong Y et al. A systematic survey of data mining and big data analysis in internet of things. The Journal of Supercomputing. 2022;**78**(17):18405-18453

[8] Kumar A et al. Retracted article: Big data analytics to identify illegal activities on Bitcoin Blockchain for IoMT. Personal and Ubiquitous Computing. 2024;**28**(Suppl 1):5

[9] Khowaja SA et al. VIRFIM: An AI and Internet of Medical Things-driven framework for healthcare using smart sensors. Neural Computing and Applications. 2023;**35**(22):16175-16192

[10] Hashem M, Youssef AE. Teeth infection and fatigue prediction using optimized neural networks and big data analytic tool. Cluster Computing. 2020;**23**:1669-1682

[11] Moqurrab SA, Tariq N, Anjum A, Asheralieva A, Malik SUR, Malik H, et al. A deep learning-based privacy-preserving model for smart healthcare in internet of medical things using fog computing. Wireless Personal Communications. 2022;**126**:2379-2401

[12] Ji C, Yupeng H, Wang K, Zhan P, Li X, Zheng X. Identifiable temporal feature selection via horizontal visibility graph towards smart medical applications. Interdisciplinary Sciences: Computational LIfe Sciences. 2021;**13**:717-730

[13] Shen Y-C, Hsia T-C, Hsu C-H. Software optimization in ultrasound imaging technique using improved deep belief learning network on the internet of medical things platform. Wireless Personal Communications. 2022;**127**(3):2063-2081

[14] Lin J-W, Arul JM, Kao J-T. A bottom-up tree based storage approach for efficient IoT data analytics in cloud systems. Journal of Grid Computing. 2021;**19**:10

[15] Mohapatra AG, Talukdar J, Mishra TC, Anand S, Jaiswal A, Khanna A, et al. Fiber Bragg grating sensors driven structural health monitoring by using multimedia-enabled IoT and big data technology. Multimedia Tools and Applications. 2022;**81**:34573-34593

[16] Adi E, Anwar A, Baig Z, Zeadally S. Machine learning and data analytics for the IoT. Neural Computing and Applications. 2020;**32**:16205-16233

[17] Jasti VDP, Zamani AS, Arumugam K, Naved M, Pallathadka H, Sammy F, et al. Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. Security and Communication Networks. 2022;**2022**(1):1918379

[18] Himeur Y et al. AI-big data analytics for building automation and management systems: A survey, actual challenges and future perspectives. Artificial Intelligence Review. 2023;**56**(6):4929-5021

[19] Jamil H, Umer T, Ceken C, Al-Turjman F. Decision based model for real-time IoT analysis using big data and machine learning. Wireless Personal Communications. 2021;**121**:2947-2959

[20] Kumar A, Sharma K, Sharma A. Genetically optimized fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics. Applied Soft Computing. 2021;**109**:107525

[21] Tan T-H, Hsu C-C, Chen C-J, Hsu S-L, Liu T-L, Lin H-J, et al. Predicting outcomes in older ED patients with influenza in real time using a big data-driven and machine learning approach to the hospital information system. BMC Geriatrics. 2021;**21**:280

[22] Safa M, Pandian A. Retracted article: Intelligent big data analytics model for efficient cardiac disease prediction with IoT devices in WSN using fuzzy rules. Wireless Personal Communications. 2022;**127**(Suppl 1):23

[23] Zamani AS, Hashim AHA, Shatat ASA, Akhtar MM, Rizwanullah M, Mohamed SSI. Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics. Biomedical Signal Processing and Control. 2024;**94**:106247

[24] Ed-daoudy A, Maalmi K. A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. Journal of Big Data. 2019;**6**:104

[25] Manogaran G, Vijayakumar V, Varatharajan R, Kumar PM, Sundarasekar R, Hsu C-H. Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. Wireless Personal Communications. 2018;**102**:2099-2116

[26] Nibareke T, Laassiri J. Using big data-machine learning models for diabetes prediction and flight delays analytics. Journal of Big Data. 2020;**7**:78

[27] Ashiku L, Al-Amin M, Madria S, Dagli C. Machine learning models and big data tools for evaluating kidney acceptance. Procedia Computer Science. 2021;**185**:177-184

[28] Pustokhin DA, Pustokhina IV, Rani P, Kansal V, Elhosenye M, Joshi GP, et al. Optimal deep learning approaches and healthcare big data analytics for mobile

networks toward 5G. Computers and Electrical Engineering. 2021;**95**:107376

[29] Woźniak M, Wieczorek M, Siłka J. BiLSTM deep neural network model for imbalanced medical data of IoT systems. Future Generation Computer Systems. 2023;**141**:489-499

[30] Vara Siddardha Reddy M, Sai Prasad R, Sai Jagan R, Selvi M. Artificial intelligence for IoT-based healthcare system. In: 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE; 2023. pp. 1-5

[31] Tong S. Lessons learned developing a practical large scale machine learning system. In: Google Research Blog. Vol. 2016. 2010

[32] Armes T, Refern M. Using big data and predictive machine learning in aerospace test environments. In: 2013 IEEE AUTOTESTCON. IEEE; 2013. pp. 1-5

[33] Thuraisingham B. Big data security and privacy. In: Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, ACM Digital Library. 2015. pp. 279-280

[34] Nelson B, Olovsson T. Security and privacy for big data: A systematic literature review. In: 2016 IEEE International Conference on Big Data (Big Data). IEEE; 2016. pp. 3693-3702

[35] Amanullah MA, Habeeb RAA, Nasaruddin FH, Gani A, Ahmed E, Nainar ASM, et al. Deep learning and big data technologies for IoT security. Computer Communications. 2020;**151**:495-517. DOI: 10.1016/j.comcom.2020.01.016

[36] Irshad RR, Sohail SS, Hussain S, Madsen DØ, Zamani AS, Ahmed AAA, et al. Towards enhancing security of IoT-enabled healthcare system. Heliyon. 2023;**9**(11)

[37] Rana B, Singh Y, Singh PK. A systematic survey on internet of things: Energy efficiency and interoperability perspective. Transactions on Emerging Telecommunications Technologies. 2021;**32**(8):e4166. DOI: 10.1002/ett.4166

[38] Koot M, Mes MR, Iacob ME. A systematic literature review of supply chain decision making supported by the internet of things and big data analytics. Computers & Industrial Engineering. 2021;**154**:107076. DOI: 10.1016/j.cie.2020.107076

[39] Li T, Rezaeipanah A, El Din EMT. An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. Journal of King Saud University, Computer and Information Sciences. 2022;**36**(4):3828-3842

[40] Wang F, Wang H, Dehghan OR. Machine learning techniques and big data analysis for internet of things applications: A review study. Cybernetics and Systems. 2024;**55**(1):1-41

[41] Khelifi F, Jiang J. K -NN regression to improve statistical feature extraction for texture retrieval. IEEE Transactions on Image Processing. 2010;**20**(1):293-298

[42] de Medeiros Martins A, Neto ADD, de Melo JD. Neural networks applied to classification of data based on Mahalanobis metrics. In: Proceedings of the International Joint Conference on Neural Networks. 2003. Vol. 4. Portland, OR, USA: IEEE; 2003. pp. 3071-3076

[43] Angelov PP, Zhou X. Evolving fuzzy-rule-based classifiers from data streams. IEEE Transactions on Fuzzy Systems. 2008;**16**(6):1462-1475

Chapter 9

Perspective Chapter: Quality Control in New Conditions

Věra Pelantová and Jan Kamenický

Abstract

An organisation is a complex of parts and relationships. One of its parts is quality management. It consists of three parts: technical, personnel and soft elements. Common practice concerns quality control equipment and procedures. Human resources supervises the training of employees in basic quality management. However, the management of the third part and its relationships is often neglected. Due to economic, technical and socio-political aspects, quality management must adapt. This chapter describes the current external context and its effects on quality management. It is based on the basic terminology of quality management. It describes the issue of quality management in the form of the Control Plan and practical knowledge of quality management from organisations. It specifies relationships to promoted digitalisation and a neglected sustainability. It presents appropriate management tools for new conditions. The chapter also relies on the findings of its own research in organisations. It describes new conditions of the external context. It provides an overview of new concepts of quality management. This chapter describes how to combine the above areas to be effective and to improve quality control.

Keywords: quality control, new conditions, big data, artificial intelligence, organisation, process approach, customer

1. Introduction

The organisation is a living organism, as Tomáš Baťa said. This slogan has long been forgotten. Most people still think that organisational procedures, the market and the area of quality do not change over time. Therefore, they rely on the original professional knowledge from school when they were young and hope it lasts until retirement. Another mistake is understanding quality as exclusively the machine measurement of a component. Since 2019, this area has changed a lot due to external pressures. The primary reason was the Covid-19 pandemic. The subsequent shocks have affected the economies of many countries. Socio-political aspects increase market turbulence; this increases prices of all commodities. At the same time, regional and global markets are being pulled down. Organisations are modernising their technical equipment and hoping to bridge this problematic period. They do not see the broader context; they perceive risks narrowly. Managers are used to noticing marketing proclamations such as the application of digitalisation. On the other hand, environmental requirements resonate with the public. And the personalities of company employees are also changing, as they begin to balance work and personal life more. The management of the organisation should approach these facts appropriately. In fact, management often does not understand or perceive changes around them and therefore does not even work with them, let alone address them within the framework of quality management. It does not deal with the concept of quality; it considers it unchangeable.

This chapter deals with the area of quality management from a wider spectrum of stimuli. It helps the reader to navigate the complex situation of the external context of organisations. It builds on previous research on the situation of quality management in organisations. It draws appropriate recommendations for current organisational practice.

2. Economic, socio-political and technical aspects in the organisation

First, it is necessary to describe the current change in conditions of the surrounding environment of organisations. Risk management is based on economic, socio-political and technical aspects, which can be described as follows. Competition is growing in many areas; customers are demanding, and accordingly, there are changes in the market [1]. According to publications in Refs. [2, 3], the current period can be called 'mass customization' or 'customization' of products to customers. However, the text in Ref. [4] still talks about a large-scale production. The decline in sales of organisations and lower consumption in households affect production and logistics and the state of warehouses and transport. A raw material base is changing, and prices of materials and energy are rising. Securing resources of all kinds is becoming increasingly difficult for organisations. The turbulent environment leads to problematic forecasts and thus to difficulties with production planning and strategic planning in organisations in general. Organisations have had crisis scenarios since Covid-19. Disinfection and cleanliness in production are being paid more attention than before. Outsourcing of warehouses is also offered [5]. Organisations often have many subcontractors, and production is often affected by low added value of production, which reduces its competitiveness. Profits due to 'contracting' often do not stay in the country. Wages are relatively high. Inflation is high. Many types of subsidies are provided. For example, in the Czech Republic, the core of the economy is the automotive industry. Political will in this regard has been consistent for a long time. However, this does not correspond to the current economic and socio-political situation. Moreover, managerial work is needed differently than it was in the era of globalisation, as the text in Ref. [4] states, for example. All these areas, of course, also affect the area of quality control. There are problems with inputs for quality and speed in supply chains. There are not enough qualified employees for specific technical work, as the article in Ref. [6] states. Education in the field of quality is insufficient, as the article in Ref. [2] states and authors of this chapter confirm. Sensors of various types that replace human senses are developing rapidly [4], but their price is still high. Technical and organisational systems are becoming increasingly complex [3]. This entails the accumulation of secondary errors that affect the quality of production. Detection of nonconformities and their resolution are then diverse in individual organisations. Modern systems collect a huge amount of data across processes. This requires appropriate tools to process them for human needs, as publications in Refs.

[7, 8] write. The text in Ref. [9] deals with sources of fear of digitalisation and artificial intelligence in organisations. Not only in quality, but there are concerns about job losses, misunderstanding of digital technologies, insufficient explanation of the issue, general hustle and bustle, loss of privacy and cyber attacks. Added to this are large legislative requirements and imminent financial sanctions. Many employees are exhausted by these changes. The reality in many organisations is described in the article in Ref. [9], which states: 'Most organisations run their technologies on older systems'. This leads to difficulties with digital transformation in terms of cost, safety, complexity of setup, time and production interruptions but also quality control. The goal of most managers is simply to reduce costs and increase profits [10, 11]. On the other hand, some managers uncritically perceive artificial intelligence as the salvation of production and quality control. But what exactly specifies quality and its control?

3. Quality control and its starting points

The quality is defined in the standard ISO 9001 [12] as 'the degree to which customer needs are satisfied by a set of inherent characteristics of an object, which is any product with specified functions'. Quality management can be defined according to the publication in Ref. [12] as 'a procedure for determining and improving the quality of production and support activities in organisations'. Quality management is based on characteristics (properties) that characterise a product, as well as a process. Each character consists of a value and a unit. The organisation should also determine its optimal value so that it can use regulation to influence factors that affect this characteristic. The value of a character is determined by measuring quality characters or comparing quality characters. The quality employee must have a calibrated measuring instrument or a calibrated comparison set or template (e.g., a sample book, a document and an alternative calculation).

The reason for quality management is to ensure a quality product for the customer or for multiple stakeholders (e.g., the product of affected entities). At the same time, quality control supports the rationalisation of the production process and the standardisation of production, which facilitates subsequent maintenance of the product in terms of used spare parts and tools. The quality control also monitors compliance with legislative and safety requirements. What will be checked on the product is determined by the nature of the product and customer requirements. Legislative and safety restrictions are also related to this. From the point of view of efficiency, it is appropriate to choose critical safety values and values describing the added value of the product as basic characteristics. The total number of characters per product should be between five and seven characters due to sufficient characters and, at the same time, for effective quality control management. Otherwise, the control becomes very lengthy and can cause the product to become more expensive.

In the main production process, quality control can be carried out (counted from the beginning of the process) as input, interoperation and output. This division depends on the organisation. If the management system is appropriately designed and based on the process approach, it is sufficient to have input and output control in the organisation. Individual activities in the organisation then only need to be controlled by employees who perform their tasks conscientiously and in accordance with the product documentation. Alternatively, control will be ensured by correctly setting up equipment with a connected control system that will control certain characteristics directly on the manufactured product. It is advisable to integrate the quality control into production and not separate it within the organisational department of the organisation. Reasons are time, cost and space, and the resulting production may be complicated. Where to place the quality control within the layout (spatial arrangement) is a question of the organisation's spatial and financial capabilities. It is generally recommended to integrate it into its own production workshop or line. If this is not possible, measurements are carried out in a nearby control laboratory. This may also be related to media connections (for example, electricity, water, compressed air and so on). Another limit is the requirements of a measuring instrument or a comparison set for working conditions. They are usually set by the employee of a measuring instrument or a comparison set. Limits may consist of workplace temperature, humidity, avoidance of mechanical shocks, influence of magnetic fields, influence of noise and other factors [13–16].

When to measure product quality? It is related to the completion of a given production activity according to the workflow when the value of the character is decisive for further work tasks or already clearly affects the requirements and expectations of the customer. The stability of the characteristic value is important (e.g., after the product has cooled down after heat treatment). The number of measuring stations is not prescribed. It depends only on the organisation. Each measurement, of course, increases the price of the product. How many products to measure is a question of the size of the production batch, the requirements of the contract with the customer, and, if applicable, the corresponding type of statistical acceptance. It is also related to the type of measurement test. Most quality controls are non-destructive, which means that after the inspection, the product continues its journey to production or to the customer. In the case of destructive inspection, the product is physically destroyed, and its further journey through the production process to the customer does not take place. The product is measured by a designated quality employee. This is either given by the work procedure or determined on the basis of the relevant qualification examination. In addition to technical knowledge, this employee must also have knowledge in the field of quality. Furthermore, characteristics such as reliability, self-regulation, consistency, ethical behaviour, communication skills and so on are necessary.

From the point of view of documentation, the Control Plan is essential for quality management. It is based on the requirement for quality measurement according to the ISO 9001 standard [12]. This document describes the control procedure for ensuring product quality. It should be a part of the technical documentation of the product. As a rule, it takes the form of the Control Table. It contains all main data, such as: specification of the character, number of measured products, specification of the measuring instrument, place of measurement, time of measurement, who the employee in charge of the control is, recording form, date of measurement and what to do if a nonconformity is detected. Recording quality management records is the basis for product identification and traceability. The measured values cannot be recorded only in writing or digitally but must be further processed. This means analysing these values for quality management with regard to limits and effectiveness and then taking appropriate preventive or corrective actions. Of course, it should be mentioned that the Control Plan should not be confused with the often-mentioned Quality Plan. In the case of the Quality Plan, it is a standardised document according to the standard ISO 10005 [17]. The Quality Plan is de facto a Quality Management System Manual for one specific order, which differs in some parts from the previously implemented management system of the organisation. In essence, it supplements it with specific issues affecting the quality of production.

4. Quality management and appropriate methods

Many organisations are now engaged in the digitalisation of quality control management in connection with the digitalisation of the production process. The core, of course, is methods contained in these software solutions. The current turbulent period is methodologically unfavourable for lean production, which has problems with fluctuations in the basic production and market environment [2]. However, the opposite opinion is expressed in the article in Ref. [18]. Therefore, Production Planning and Control (PPC) systems, such as the Theory of Constraints (TOC) method, are coming to the fore because they solve bottlenecks in processes and support the process approach in the organisation during an unclear external context. The Value Stream Mapping (VSM) method tracks the flow of added value. The Process Decision Program Chart (PDPC) method and the 5x why method can solve the key causes of nonconformities detected during the quality control and help find appropriate measures. Organisational flexibility is supported by flexible heterarchical structures [2]. Benchmarking is suitable for quality management [2] because it allows comparing the status of the quality management system or the quality of products of competing organisations. The most followed metrics are Key Performance Indicators (KPI) and Overall Equipment Effectiveness (OEE) [7]. However, in organisational practice, quantitative (economic) characters are needed and sometimes neglected qualitative characters [10]. It is their processing that is difficult for many organisations. Therefore, the simplest solution to this problem is their conversion and processing using the semi-quantitative analysis.

5. Quality management and a new concept

Changing conditions of the external context of organisations require the application of a new concept of quality management. Existing concepts, such as the current version of the ISO 9001 standard [12], are already outdated. New concepts of quality management were identified in the literature review [19–35]. They are partly based on the ISO 9001 standard [12] or the total quality management (TQM) concept, but they also try to cover other aspects of new conditions that were described above in the text. These are the following concepts:

QM 2030—includes product added value, sustainable development, system perspectives, intelligent self-organisation models and ensuring stability during changes [19]. The advantage is the inclusion of change management. The disadvantage is the explicit non-inclusion of occupational safety and information security.

Quality 4.0 A—includes quality management procedures and components of Industry 4.0, IoT, big data, cloud and artificial intelligence (AI) [20]. The advantage is the abundance of technical information. The disadvantage is the lack of inclusion of sustainability, occupational safety and change management.

The triple bottom line—includes TQM, circular economy and social responsibility [21]. The advantage is the inclusion of the main concept of quality management, sustainability and social aspects. The disadvantage is the lack of inclusion of occupational safety (although it could be understood as a point of social responsibility), and technical information is not explicitly stated.

Quality definition model—includes meeting customer requirements, agreed deliveries, integration into the ecosystem and meeting company requirements [22]. The advantage is a broader concept of the quality management and the inclusion of the environmental aspect. The disadvantage is the lack of inclusion of soft skills, occupational safety and change management.

Quality 4.0 B—includes quality management practices, digitalisation basics, creativity, leadership, analytical thinking, ability to solve complex problems, decision-making ability, proactivity, ability to manage change and emotional intelligence [23]. The advantage is a strong emphasis on the soft skills of employees in new conditions and the inclusion of sustainability and occupational safety. The disadvantage is general technical information.

The Quality 5.0 model, which consists of social, economic and environmental sustainability, is still more of an academic concept [24]. The inclusion of economic management is taken for granted in business.

The issue of quality management is best represented by the triple bottom line concept because it is based on TQM, which is the most widespread quality concept in the world. The issues of new conditions are well met by concepts of QM 2030 in general and Quality 4.0 B, thanks to their strong focus on employee education in the field of quality. This will be decisive for the development of organisations. Companies now need, in particular, technically educated employees with critical thinking skills and soft skills, with an overlap with holistic. However, many managers have not yet accepted this idea and see the core of their future success in the digitalisation of the organisation.

6. Digitalisation and the quality control

Digitalisation is a rapidly developing field. It is intended to help organisations optimise management system processes, streamline production and improve control measures, especially the accuracy and speed of the quality control. Appropriate organisational adjustments could help with this [9]. But what does this digitalisation actually involve? Artificial intelligence (AI) is 'a set of statistical methods for data processing' according to the publication in Ref. [36]. Mainly, image or signal inputs are converted into two-state outputs [37]. A related term is big data. These are data files of tens of TB, which include text, calculations, images, databases and blogs about a certain object. They have different degrees of trustworthiness depending on the owner [38, 39]. ChatGPT is simpler and faster, but can sometimes return wrong answers, so-called 'hallucinations' [36, 37]. Copilot is a more sophisticated, 'intelligent search engine' [37]. ChatGPT requires a correct query (prompt), while Copilot 'gives context' that modifies the query. The most widespread application of artificial intelligence is large-scale language models (LLM) [37, 40, 41], which require a lot of data. The data are collected from sensors on devices in real time. Then, employees interpret it (e.g., using the Excel) [4, 7].

For example, by analysing the state of the system and the Internet of Things (IoT), 'tailor-made' medicines are produced for a specific person [11]. Generative artificial intelligence creates new things from data that a neural network has previously learnt from a teacher [4, 8]. Networks are divided into feedforward and recurrent. Feedforward networks study one pattern at each step. Recurrent networks can handle the entire sequence of patterns in one step. They produce images and texts and are not so concerned with process optimisation and control [8]. However, they can generate, for example, process diagrams based on management notes [1, 7]. Predictive artificial intelligence helps, for example, with predicting maintenance and monitoring the condition of equipment in the environment [4]. Machine vision is used to increase productivity [36, 41], for example, to check whether containers are overfilled with media or whether products are properly labelled, which is difficult to check. Visual inspection of the workplace using artificial intelligence is performed by rapid scanning regardless of lighting conditions [41]. According to the article in Ref. [8], organisations have a problem with optimisation, which is a task performed in a control loop according to the state of a process. Artificial intelligence helps with this, called Reinforcement Learning (RL). The key to the application of artificial intelligence is the correct formulation of the problem task [37]. An artificial intelligence system can help find weak points in the production process in terms of capacity, nonconformities or equipment failures, suggest measures and determine the causes of nonconformities [7]. The edge learning is based on examples of products with good and bad properties. The application of artificial intelligence in the field of quality and safety in the food industry is popular [40, 42]. Artificial intelligence can also connect logistics, production processes and quality control [18]. Training quality control staff and their subsequent control in accordance with work procedures is easy and fast [2, 36]. For example, a survey [43] from 2024 tracks the integration of AI into industrial practice. Benefits of AI applications include increased productivity and data processing. However, there are concerns that the development of AI could lead to a trend where quality is not a priority [3].

The Artificial Intelligence Act (AIA) contains rules for the development, implementation, and use of AI in practice in the EU. It is prohibited to insensitively collect biometric data and personal data of citizens and to monitor emotions during working hours [36]. However, the basic requirements are based on information security management according to the ISO/IEC 27000 standard [44].

7. Sustainability and the quality control

The basic standard for environmental management is the ISO 14001 standard [45]. On the one hand, it talks about sustainable resource management. It is important that organisations help to limit all types of media leakage [3], reduce carbon emissions and reduce energy consumption [9]. The publication in Ref. [2] presents some practical findings on the connection between artificial intelligence and sustainability. The most common applications are in meteorology, which helps to determine the impact of climate change on industrial production and the quality and durability of products, as well as on logistics routes [3]. It deals with eco-technical innovations (e.g., replacing batteries instead of charging). Other quality tools for environmental management are not so developed. They rely more on the connection between the environment and reliability. Production must technologically minimise environmental impacts. Waste from production should always be recycled. The text in Ref. [6] mentions the circular economy and the need to develop green technologies. On the other hand, in recent years, the environment has been neglected at the expense of economic indicators. Investments in environmental management are reduced.

In general, there is talk of sustainable development, but these are rather economic proclamations. True sustainability lies in the modesty of the individual, the organisation and society. Quality control can help with this, both in terms of removing low-quality products from production and in terms of specifying useful and recyclable products intended for production. The quality control, therefore, newly plays the role of arbiter of the ecological usefulness of products versus marketing extravagance. But are these all problems that quality control faces?

8. Problems of the quality control in new conditions

In connection with the development of society, several topics related to quality control have emerged that will need to be resolved. The reason is that the quality control can once again fulfil its role.

The problem is determining the characteristics of products and of the system and the characteristics of processes and setting their optimal values in the new conditions [22, 25, 33, 34]. Another problem is errors that occur during control and subsequently have an impact on the quality of production [26]. It is not easy to connect quality management, production and waste reduction [22, 23, 34]. In connection with digitalisation, many people do not know the terminology and do not have sufficient skills in this area [23, 26]. However, a part of the public has unrealistic expectations in this regard that 'AI will solve everything correctly'. Competencies of employees, not only qualities needed to cope with new market conditions [23, 24], are not specified. The quality control must resolve contradictions of the TQM and the circular economy, which resonates with the involvement of sustainability [21, 23, 35]. The quality control, in order to make the measurement results more accurate, comes up against the sensitivity of production parameters [29]. The effort to introduce artificial intelligence clashes with the need for a large amount of diverse data and a long training time on the one hand and with the lack of data standardisation and quality assurance on the other hand [20, 29, 33]. Article in Ref. [30] mentions the phenomenon of false positive discrepancies and the need to strengthen knowledge, for example, about the surface properties of materials for the needs of quality control (a similar problem is also addressed in the article in Ref. [32]). A separate ergonomic and emotionally unresolved chapter is the cooperation between an employee and a robot and the control of their work [29]. Despite digitalisation, unidentified or misidentified documents and objects (such as X-rays) still occur, which makes quality control difficult [31]. In addition, digital systems are still unable to 'complete without inventing' a missing part of the object in the image, and the laterality of the inspected object makes it difficult for them [31, 32]. A persistent problem in quality control is determining the causes of nonconformities [33]. There is a lack of determination of the characteristics of products and quality management and control tools that will be in line with the sustainability [35].

Overall, it is clear that a new concept of quality management needs to be created that will correspond to the new conditions of the external context mentioned above [19–35].

9. Authors' view on the topic of quality control

The field of quality management is quite specific and quite conservative. This is due to the responsibility that this field bears for the quality of products and processes. Not all control activities can be automated. Products and processes are characterised by quantitative and qualitative characteristics that also need to be controlled, not ignored. It is often forgotten that the organisation is a soft system. It also includes employees who create and evaluate quality both technically and according to legislation and by their own judgement. The pressure for transparency of flows and clear decision-making according to the text in Ref. [3] combines not only the economic characteristics of management systems but also social characteristics. This, in turn, prevents the 'opening of the scissors' in the society. Thus, they prevent its destabilisation. Currently, in connection with the use of artificial intelligence not only for quality control, it happens that artificial intelligence delays the solution of tasks, simplifies it without the client's consent, adjusts task limitations, provides unethical advice and encounters basic ignorance in areas such as mathematics and quality. This can lead to secondary nonconformities within the quality management. Moreover, many organisations believe that the only way to win customers is through constant innovation. However, most customers do not want this constant product modification. This creates secondary pressure on education and a loss of routine, which leads to time savings in the production processes of enterprises. According to articles in Refs. [2, 9, 10], changes in the concept of the process approach crystallise around changes in the basic environment and internal context of organisations. They will include both digitalisation and a complex of changes in decision-making models and the need for qualified employees and organisational flexibility in a safe environment.

Quality management is an important part of organisational management. In a broader sense, it consists of three parts. The first part is the technical management of production quality, based on the processes of the management system. This includes technical equipment and measuring instruments (quality control), as well as the management of processes and product properties. The second part is the competences of employees, which consist of their qualifications, knowledge and skills. The third part is so-called 'soft elements', which consist of individual employees who have relationships with each other, with parts of the organisation, and with external objects. This also includes organisational structure, emotions, self-management, culture and so on. This is the internal context of the organisation's management system. The external context consists of economic, socio-political (including the environment) and technical aspects. If the organisation is perceived as a system, inputs (material, energy) and outputs (products and waste) are also visible, because the organisation is an open system. The core of quality management is where knowledge, skill and sustainability are tied into products for customers but also for the public. In order for the organisation to improve in all areas, it needs to educate its employees and understand all connections holistically because they lead to the creation of effective quality control in new conditions.

10. Research of the status quo of quality control in organisations

Although the production and the quality management have been organisationally linked to digitalisation for a long time, according to the publication in Ref. [46], this state needs to be investigated. For the needs of this topic, a survey of the state of quality management was conducted in 15 organisations in the Czech Republic. Of these, eight were large organisations and seven were small and medium-sized organisations (SMEs). They operate in the following sectors: automotive industry, mechanical engineering, electrical engineering, plastics processing, food industry, healthcare and services. The types of production are represented by piece and serial production. Seventeen areas of the topic were investigated. Their evaluation was carried out in the form of a yes/no answer or using a five-point response scale, where five points means the best.

The quality of the provided product from the customer's perspective is better in SMEs. Material savings are at an average level in both types of organisations. Occupational safety is better in SMEs. They also have much better communication with customers. Interpersonal consideration of employees towards each other and customers is better in SMEs. Despite a significant handicap in financial resources, these organisations are usually better at prices and lower production and quality costs. Large organisations are more successful in meeting deadlines. They also have a more sophisticated waste management system, which mostly helps them promote environmental care. Both types of organisations have a software system for recording data. Large organisations have hardware for checking the quality of materials and products. SMEs usually do it manually. Information is more easily transferred between departments in small and medium-sized enterprises. The Internet of Things (IoT) is used by more than half of large organisations. Expertise is used in both types of organisations for their activities and is usually linked to a specific production process. Large organisations often use cloud storage. Big data in the sense defined above is generated by approximately half of large organisations. Some forms of artificial intelligence are tried to be used by employees of most organisations. The reason was the visualisation of data for production innovations and prediction of the condition of objects. The use of modern digital devices is more common in large organisations, as it is a significant investment. Large organisations often have a more extensive quality management system (QMS). The assessment in terms of a better QMS corresponds to the organisation. This means that the more efficient and simpler the overall management in the organisation better. In this respect, some small and medium-sized enterprises are better. Control of input material and output product is carried out in all organisations studied. Product controls are also often carried out between production activities. There are many control activities, carried out manually and electronically, even directly by production employees. It depends on the financial capabilities of the organisation and how the measuring workplace is equipped. But even small and medium-sized enterprises often have quality control laboratories equipped at a high level. The state metrological control, which all organisations in the Czech Republic should undergo, is also obvious. In most organisations, all prescribed product quality controls are carried out. One-hundred percent control is carried out, especially if the customer is a large, significant organisation. In other cases, control by selecting several pieces according to the rules of the statistical acceptance is usually applied. The control of one piece is omitted. Most inspection tests are non-destructive. The creation of measurement records in quality control should also be easier and more consistent. Problems arise with input materials and semi-finished products, which are often of insufficient quality and contaminated. Sometimes they have incorrect parameters. The inability to manage production preparation is also to blame. Another reason is the indifference of employees to the quality of their own work. They lack concentration and self-control. This again indicates a lack of quality education from childhood.

Small organisations are better at managing soft skills. Interpersonal relations are usually better here because a small group of employees is involved. Large organisations are better at managing technical aspects and digitalisation because this entails significant financial costs. The emerging new situation is more evident in the state of QMS, which confirms the hypothesis of the authors of this chapter that quality management and maintenance management react more quickly to stimuli from the external context. Strengthening competences in the area of quality management in new conditions is more the domain of younger employees. It is not yet sufficiently supported in the education of organisations. It still corresponds to the classic basic quality training of employees. In-depth training in the area of quality control is completed only by quality managers and metrologists. Training in the field of occupational safety is guaranteed by the state. Information on sustainability is not common in most organisations. Education in the field of digitalisation and technology is also the domain of rather young employees.

The sample of organisations in the survey is small and territorially limited to the Czech Republic. The picture of the state of quality management in other aspects is less favourable. In many organisations, the quality control thus benefits from previous work habits of employees, regardless of the degree of digitalisation and other aspects. Soft factors in the quality management are usually not considered and not sufficiently worked with.

11. New conditions specification

The new conditions have not yet been satisfactorily described anywhere, although they concern everyone and will, of course, affect the quality control. Based on a survey of organisations and an analysis of the current external context under investigation, the following characteristics of new conditions can be deduced. The new conditions include a new geopolitical situation, increased aggression in the society, reduced safety, depletion of material and energy resources, high prices for these resources, high taxes and duties, complex legislation, migration and great cultural differences, many types and complexity of waste disposal, and reflection of environmental aspects and impacts. The Internet and mobile signal, as well as water supplies, electricity and so on, are taken for granted. Equipment and flows within organisations are often complex. Logistic routes are destroyed. There is general laziness, and distrust and intergenerational assistance is decreasing. Administration is increasing again. People believe in fraud and do not understand the difference between fiction and reality. They are inconsistent, unscrupulous, less physically fit and lack quality education. Digital technologies are developing rapidly in both hardware and software. In this regard, organisations are lagging behind, although their management usually claims otherwise. Devices and premises are equipped with many types of sensors, which provide a lot of data in different formats for different stakeholders. Sometimes data collection occurs without the knowledge of the owner. The owner of the facility does not always have control over stored data. On the other hand, a person outside the organisation can remotely change, for example, line settings and the accepted level of quality control. Thus, the global concept of the world is falling apart.

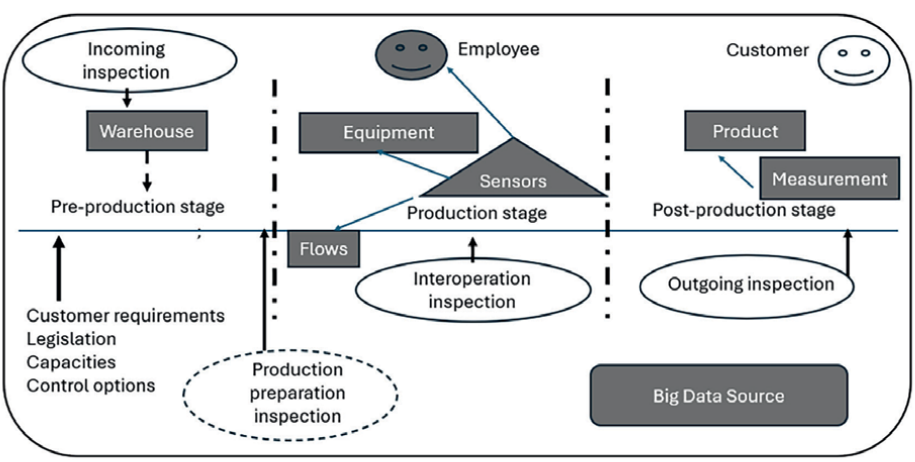


Figure 1.
Stages of the production process and detection of possible big data (shown in grey) (figure is own of authors).

The quality control will have to rely more on local, qualified employees, and it will be necessary to specify appropriate characters and tools for quality work in such a complex environment. The quality control will deal more with soft factors. On the contrary, the technical support of quality control will see minor changes (**Figure 1**).

12. Quality improvement

Quality improvement is a common goal of organisational management. It is usually understood as an increasing technical level and a possible increase in values of characters, such as the number of quality products per total batch. However, implementing quality improvement in an organisation's management system is more difficult. The comparative analysis conducted indicates the following trends in terms of foundations of organisational management systems in relation to, for example, publications [13, 16]. Given the geopolitical situation, it is appropriate to focus on one's own region or country and draw input resources and have customers there. The legislative component of the organisation also depends on geographical location and the type of product. The source should be product design and its production and subsequent effective use, but not at the expense of safety and environmental difficulties. Product characteristics should fully describe it and be in accordance with the characteristics of the organisation's processes, which are determined, for example, using the quality function deployment (QFD) diagram. It is necessary to include the issue of recycling in this design phase; otherwise, the circular economy in organisations will not start. This area should also be managed by quality employees. This idea corresponds to the process approach. And process control is de facto a system control loop. When it comes to added value to the product, the organisation should rely on its industry competencies and adhere to the technological discipline of production. Non-compliance with technological discipline and omissions in production preparation are frequent causes of poor-quality production and may not be detected by quality control. This technical part should be supplemented by a soft part. Occupational safety is required by law (at least in the Czech Republic); therefore, it is not necessary to include it in production and control as sustainability. The soft part includes creating a pleasant organisational culture, based on a flat organisational structure, and strengthening positive emotions in employees without pushing individuals out of the group. All employees in the organisation must be educated and trained in the field, in specialisation and in building awareness of the external context. It is necessary to methodologically educate employees in the field of data analysis and critical thinking. Attention should be paid to algorithms, which means using methods proven in the field. Given the complexity of systems and possible risks in organisations, it is recommended to start with a root cause solution (the 5x why method) and the holistic concept, where it is also necessary to understand the links between objects. The quality management model for the application of artificial intelligence must generally correspond to the process approach. This chapter, therefore, also presents the identification of production process tasks that artificial intelligence could support (**Figure 2**) to make the quality control more efficient. The absolute basis for all employees is to maintain self-control and concentration at work. Data analytics should be used to a greater extent than now to increase the accuracy of employees' decision-making. This will prevent waste, including energy, and reduce production and control costs. This should support customer and employee satisfaction and safety.

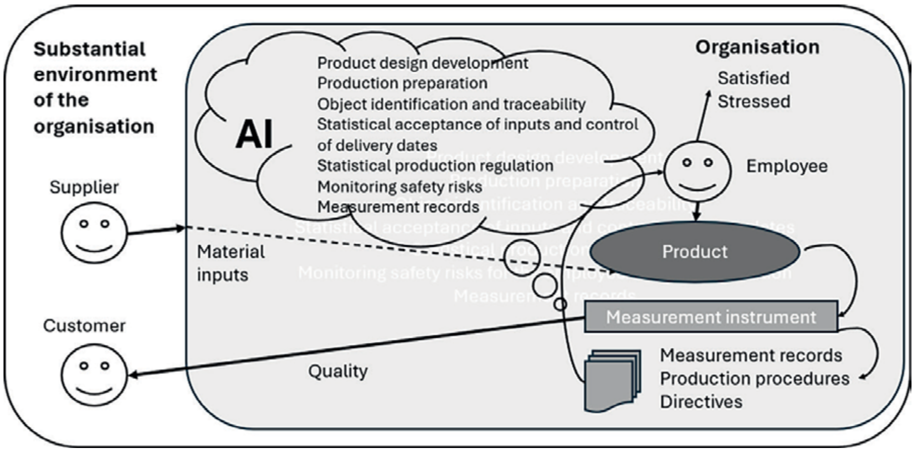


Figure 2.
Identifying production process tasks that could be supported by AI (figure is own of authors).

From the perspective of the above-mentioned new concepts of quality management, the authors' own approach to the issue is outlined here. It combines findings of the triple bottom line, QM 2030 and Quality 4.0 B and adds their own practical recommendations regarding organisational practice.

13. Conclusion

The quality management in new conditions brings a completely new situation for all organisations. The paradigm of the process approach was quite large. However, the current constellation changes the internal context and the external context of organisations at the same time. Only with time will parts of this change be transferred to the wording of the quality management standard.

This chapter describes current economic, socio-political and technical aspects that will change the quality management. Their reach is considerable. The following is a description of quality management from the point of view of basic terminology to recommendations related to the control procedure in the form of the Control Plan. Suitable methods for supporting quality management are presented. The essential part is the comparison of new quality management concepts, which is based on a literature search and their subsequent comparative analysis. In this regard, the triple bottom line, QM 2030 and Quality 4.0 B are significant candidates. Digitalisation and the use of artificial intelligence for quality management are described in terms of its possibilities and current problems. A brief overview of AI terminology is presented. Equally important is the description of the relationship between sustainability and quality management, although it is neglected in many organisations. Problems of quality management in new conditions are specified from the point of view of practical and theoretical foundations. Results of a survey of the status quo of the quality management in new conditions in a sample of organisations in the Czech Republic are presented. It brings interesting results that point out that large organisations have financial resources, but small and medium-sized organisations are usually better able to develop the quality through their own employees, their professional and human contributions. The second result is partial findings on the relationship between quality management and digitalisation. The latter is presented despite great proclamations rather than taking over the role of quality employees. Artificial intelligence in quality management must be perceived as an auxiliary tool, not as a necessary fashionable element. Its disadvantage is its rapid development. This chapter also contains a specification of the concept of quality management for new conditions. It is complex and places new demands, especially on quality employees and their education. Of course, new conditions of the external context are specified.

The view of the issue is formed from the position of quality management and reliability management, which the authors of this chapter have been dealing with for a long time. This is not a programming or an economic concept. The sample of organisations in the survey is small and geographically limited. Nevertheless, partial results can be considered indicative of further investigation. The trend is towards soft systems in the organisation, which is a new topic for the quality management and, of course, for the quality control. This topic should be further investigated, and appropriate quality tools for new conditions should be evaluated. The main task of quality management in turbulent market conditions and geopolitical changes, regardless of the possibility of digitalisation of production, is to produce quality products, not to control them.

Acknowledgements

This work was supported by the institutional support for long-term strategic development of the Ministry of Education, Young and Sports of the Czech Republic.

Author details

Věra Pelantová* and Jan Kamenický
Technical University of Liberec, Faculty of Mechatronic, Informatics and Interdisciplinary Studies, MTI, OSR, Liberec, The Czech Republic

*Address all correspondence to: vera.pelantova@tul.cz

References

[1] Jiras P. Generativní umělá inteligence v ERP. MM Průmyslové Spektrum. 2023;**1**(11):32. ISSN: 1212-2572

[2] Dvořák R, Witassek L. Inovace bez implementace je halucinace. MM Průmyslové Spektrum. 2024;**1**(9):8-12. ISSN: 1212-2572

[3] Rane J et al. A review on artificial intelligence, blockchain, big data analytics, cloud computing, and internet of things. In: Generative Artificial Intelligence in Agriculture, Education, and Business. Deep Science Publishing; 2024. Chapter 1. ISBN: 978-81-981271-7-4_1. pp. 1-26. DOI: 10.70593/978-81-981271-7-4_1

[4] Hloska J et al. Globální trendy v robotizaci. Automa. 2024;**30**(2):28. ISSN: 1210 9592

[5] Knaisl D et al. Současné změny na trhu a výzvy logistiky. MM Průmyslové Spektrum. 2023;**1**(10):28-29. ISSN: 1212-2572

[6] Schneiderová A et al. Veletrh inteligentní automatizace a robotiky. MM Průmyslové Spektrum. 2023;**1**(5):48. ISSN: 1212-2572

[7] Vitásek D. Dělejte to jinak. Zpracování a interpretace provozních dat. MM Průmyslové Spektrum. 2023;**1**(5):29. ISSN: 1212-2572

[8] Koutník J. Průmyslová umělá inteligence. Automatica. 2024;**30**(4):15-17. ISSN: 1210-9592

[9] Rane NL et al. Emerging trends and future research opportunities in artificial intelligence, machine learning, and deep learning. In: Artificial Intelligence and Industry in Society 5.0. Deep Science Publishing; 2024. pp. 95-118. Chapter 6. ISBN: 978-81 981271-1-2_6. DOI: 10.70593/978-81-981271-1-2_6

[10] Aderemi S et al. Big data analytics in the financial services industry: Trends, challenges, and future prospects: A review. International Journal of Science and Technology Research Archive. 2024;**06**(01):147-166. DOI: 10.53771/ijstra.2024.6.1.0036. Available from: https://sciresjournals.com/ijstra/content/big-data-analytics-financial-servicesindustry-trends-challenges-and-future-prospects-review

[11] Mehta A et al. Advancements in manufacturing technology for the biotechnology Industry: The role of artificial intelligence and emerging trends. International Journal of Chemistry, Mathematics and Physics (IJCMP). 2024;**8**(2):1-12. ISSN: 2456-866X. DOI: 10.22161/ijcmp.8.2.3

[12] Standard ISO 9001. Quality Management Systems—Requirements. Prague: UNMZ; 2016

[13] Veber J et al. Management Kvality, Environmentu a Bezpečnosti Práce. 1st ed. Prague: Management Press; 2006. ISBN: 80-7261-146-1

[14] Pelantová V, Havlíček J. Integrace a Systémy Managementu. Monografie. 1st ed. Liberec: TUL, FM; 2014. ISBN: 978-80-7494-164-1

[15] Legát V et al. Management a inženýrství údržby. 2nd ed. Příbram: Karel Mařík—Professional Publishing; 2016. ISBN: 978-80-7431-163-5

[16] Madzík P. Nástroje systematického riešenia problémov. 1st ed. Ružomberok: VERBUM; 2017. ISBN: 978-80-561-0478-1

[17] Standard ISO 10005. Quality Management—Guidelines for Quality Plans. Prague: UNMZ; 1997

[18] Štefan M. Inovace v automobilovém průmyslu. Část 9. Automatizace a digitalizace v logistice. MM Průmyslové Spektrum. 2024;**1**(7-8):20-21. ISSN: 1212 2572

[19] Fundin A, Lilja J, Lagrosen Y, Bergquist B. Quality 2030: Quality management for the future. Total Quality Management and Business Excellence. 2020;**36**(3-4):264-280. DOI: 10.1080/14783363.2020.1863778

[20] Escobar CA, Cantoral-Ceballos JA, Morales-Menendez R. Quality 4.0: Learning quality control, the evolution of SQC/SPC. Quality Engineering. 2024;**37**(1):92-117. DOI: 10.1080/08982112.2024.2356519

[21] Dezi L, Hysa X, Calabrese M, Mercuri F. Open total quality management in the circular economy age: A social enterprise perspective through the case of Patagonia. Total Quality Management and Business Excellence. 2022;**36**(3-4):309-323. DOI: 10.1080/14783363.2022.2051698

[22] Martin J, Elg M, Gremyr I. The many meanings of quality: Towards a definition in support of sustainable operations. Total Quality Management and Business Excellence. 2020;**36**(3-4):185-198. DOI: 10.1080/14783363.2020.1844564

[23] Santos G, Sá JC, Félix MJ, Barreto L, Carvalho F, Doiro M, et al. New needed quality management skills for quality managers 4.0. Sustainability. 2021;**13**(11):6149. DOI: 10.3390/su13116149

[24] Deleryd M, Fundin A. Towards societal satisfaction in a fifth generation of quality—The sustainability model. Total Quality Management and Business Excellence. 2020;**36**(3-4):292-308. DOI: 10.1080/14783363.2020.1864214

[25] Kolus A, Duffuaa S. Determining optimal process means in a multi-stage production system with inspection errors in 100% inspection. Quality Technology & Quantitative Management. 2024;**22**(1):105-130. DOI: 10.1080/16843703.2024.2312016

[26] Silva M, Alvelos H, Xambre AR. Quality problems reporting: Digitalization of the processes in an automotive company. Procedia Computer Science. 2025;**253**(1):737-744. ISSN: 1877-0509. DOI: 10.1016/j.procs.2025.01.135

[27] Závadská Z, Závadský J. Quality managers and their future technological expectations related to Industry 4.0. Total Quality Management and Business Excellence. 2018;**31**(7-8):717-741. DOI: 10.1080/14783363.2018.1444474

[28] Mahdikhani M, Karbasi A. Integration of TQM and circular economy practices in supply chain: Insights from industry perspective. Total Quality Management and Business Excellence. 2025:**36**(7-8):741-758. DOI: 10.1080/14783363.2025.2477168

[29] Papavasileiou A, Michalos G, Makris S. Quality control in manufacturing—Review and challenges on robotic applications. International Journal of Computer Integrated Manufacturing. 2024;**38**(1):79-115. DOI: 10.1080/0951192X.2024.2314789

[30] Chang C, Di Maio F, Bheemireddy R, Posthoorn P, Gebremariam AT, Rem P. Rapid quality control for recycled coarse aggregates (RCA) streams: Multi-sensor integration for advanced contaminant detection. Computers in Industry. 2025;**164**(1):104196. ISSN: 0166-3615. DOI: 10.1016/j.compind.2024.104196

[31] Hembroff G, Klochko C, Craig J, et al. Improved automated quality control of skeletal wrist radiographs using deep multitask learning. Journal of Imaging Informatics in Medicine. 2025;**38**:838-849. DOI: 10.1007/s10278-024-01220-9

[32] Fernández J, Valerieva D, Higuero L, et al. 3DWS: Reliable segmentation on intelligent welding systems with 3D convolutions. Journal of Intelligent Manufacturing. 2025;**36**:5-18. DOI: 10.1007/s10845-023-02230-0

[33] Chuenmee N, Phothi N, Chamniprasart K, Khaengkarn S, Srisertpol J. Machine learning for predicting resistance spot weld quality in automotive manufacturing. Results in Engineering. 2025;**25**:103570. ISSN: 2590-1230. DOI: 10.1016/j.rineng.2024.103570

[34] Attia PM, Moch E, Herring PK. Challenges and opportunities for high-quality battery production at scale. Nature Communications. 2025;**16**:611. DOI: 10.1038/s41467-025-55861-7

[35] Misztal A, Ratajszczak K. Possibilities of using contemporary quality management methods and tools for the sustainable development of the organization. Sustainability. 2025;**17**:617. DOI: 10.3390/su17020617

[36] Bartošík P. Využití umělé inteligence v průmyslové praxi. Automatica. 2024;**30**(7):35. ISSN: 1210-9592

[37] Bartošík P, Froněk T. Velké jazykové modely a generativní umělá inteligence v průmyslové praxi. Automatica. 2024;**30**(5):15-17. ISSN: 1210-9592

[38] Můčka J. Big Data Jsou víc než Jen Petabajty. Jejich Zpracovávání je Příležitost i Hrozba. Master.cz. 2020. Available from: https://www.master.cz/blog/bigdata-vysvetleni-a-zpracovani/

[39] Doskočilová V. Big Data: Jaká je Jejich Využitelnost a co to Znamená pro Vás? Měšec.cz. 2019. Available from: https://www.mesec.cz/clanky/big-data-jakaje-jejich-vyuzitelnost-a-co-to-znamena-pro-vas/

[40] Ogrizović M, Drašković D, Bojić D. Quality assurance strategies for machine learning applications in big data analytics: An overview. Journal of Big Data. 2024;**11**(156):1-48. DOI: 10.1186/s40537-024-01028-y. Available from: https://link.springer.com/article/10.1186/s40537-024-01028-y

[41] Bartošík P. Umělá inteligence rozšiřuje možnosti strojového vidění a usnadňuje jeho použití. Automatica. 2024;**30**(7):18-19. ISSN: 1210-9592

[42] Yi L et al. Recent advances of artificial intelligence in quantitative analysis of food quality and safety indicators: A review. TrAC Trends in Analytical Chemistry. 2024;**180**:117944. DOI: 10.1016/j.trac.2024.117944. ISSN: 0165-9936. Available from: https://www.sciencedirect.com/science/article/pii/S0165993624004278

[43] Janišová H. Jak správně (před) vybrat lidi. Umělá inteligence pracuje v každé druhé firmě. MM Průmyslové Spektrum. 2024;**1**(7-8):19. ISSN: 1212-2572

[44] Standard ISO/IEC 27000. Information Technology—Security Techniques—Information Security Management Systems—Overview and Specialist Dictionary. Prague: UNMZ; 2020

[45] Standard ISO 14001. Environmental Management Systems—Requirements and Instructions for Use. Prague: UNMZ; 2016

[46] Autors. Efektivní a zisková—Integrovaná chytrá továrna na výrobu elektroniky. Automatica. 2019;**25**(6):14-16. ISSN: 1210-9592

Edited by Sayyad Zahid Qamar, Nasr Al-Hinai, Sandeep Kumar, Shilpa Choudhary, Arpit Jain and Ankita Tiwari

Quality Control – Artificial Intelligence, Big Data, and New Trends provides a timely and comprehensive examination of how digital transformation is revolutionizing quality management across various industries. As artificial intelligence (AI), machine learning (ML), and Big Data analytics become central to engineering and manufacturing systems, this book examines their profound impact on quality assurance practices—from predictive maintenance and intelligent diagnostics to smart process control and dynamic optimization. The book begins by framing the transformation of traditional quality control into a data-driven, AI-enabled discipline, setting the stage for deeper technical explorations. Readers are then guided through a diverse set of case studies and innovations, including AI analysis of emotional development in fictional datasets, paradigm-shifting applications in software development, and comparative assessments of deep learning models for arrhythmia detection. Unique frameworks, such as "pipes and puddles" for process visualization and intelligent strategies for defect reduction in automotive paint shops, demonstrate how quality engineering is being reimagined. Vibration signal analysis for gearboxes, QoS-aware IoT frameworks, and enterprise architecture powered by ML and IoT further enrich the discussion with practical insights and future-ready methodologies. Later chapters critically assess the challenges and opportunities posed by remote operations, digital workflows, and adaptive manufacturing, providing a forward-looking perspective on quality control in evolving production ecosystems. Whether you are a researcher, quality engineer, industrial technologist, academic or student in the QA-QC area, this book provides essential knowledge and innovative perspectives on the intersection of quality control, AI, and Big Data. It is not just a reflection of current capabilities but a roadmap to future excellence in intelligent quality management.

Fausto Pedro Garcia Marquez,
Industrial Engineering and Management Series Editor

ISSN 3029-0511
ISBN 978-1-83634-537-4

Published in London, UK

IntechOpen