

# Sachbericht zum Verwendungsnachweis

## Verbundprojekt

Industrielles Reinforcement Learning zur Qualitätsregelung von  
Massivumformprozessen

IRLeQuM

## in der Fördermaßnahme

Pro-DL-A: Lernende Produktionstechnik - Einsatz künstlicher Intelligenz (KI) in der  
Produktion

## Autor(en)

IconPro GmbH, Dr.-Ing. Martin Peterek, Friedlandstraße 18, 52064 Aachen,  
martin.peterek@iconpro.com

**Projektlaufzeit:** 01.06.2021 – 31.10.2024

**Erstellungsdatum:** 25.02.2024

**Förderkennzeichen:** 02P20A075

## Projektpartner

**Mubea Tailor Rolled Blanks GmbH (*Mubea*)**

**Eichsfelder Schraubenwerk GmbH (*ESW*)**

**Schomäcker Federnwerk GmbH (*Schomäcker*)**

**IconPro GmbH (*IconPro*)**

**Quality Automation GmbH (*Quality Automation*)**

**RWTH Aachen:**

**Lehrstuhl für Fertigungsmesstechnik und Qualitätsmanagement (*WZL-MQ*) und**

**Lehrstuhl für Technologie der Fertigungsverfahren (*WZL-TF*) des  
Werkzeugmaschinenlabor WZL der RWTH Aachen**

*Assoziierte Partner:*

**MAWI GmbH (*MAWI*)**

**Schuler Pressen GmbH (*Schuler*)**

**Mubea**



„Dieses Forschungs- und Entwicklungsprojekt wurde durch das Bundesministerium für Bildung und Forschung (BMBF) im Programm „Zukunft der Wertschöpfung – Forschung zu Produktion, Dienstleistung und Arbeit“ (Förderkennzeichen 02P20A075) gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor Akshay Paranjape.

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

## Inhaltsverzeichnis

Sachbericht Teil I - Kurzfassung.....	4
I.1 Aufgabenstellung .....	4
I.2 Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens.....	4
I.3 Wesentliche Ergebnisse im Überblick.....	5
Sachbericht Teil II - Eingehende Darstellung.....	6
II.1 Motivation und Aufgabenstellung .....	6
II.2. Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens .....	6
II.3 Planung und Ablauf des Vorhabens .....	9
II.4 Erzielte Ergebnisse .....	15
II.5 Darstellung des während des Vorhabens bekannt gewordenen Fortschritts auf diesem Gebiet bei anderen Stellen .....	22
II.6 Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit der Ergebnisse 22	
II.7 Zusammenarbeit mit anderen Stellen außerhalb des Verbundprojektes....	23
II.8 Veröffentlichungen, Vorträge Referate, etc. ....	23
II.9 Literaturverzeichnis .....	23

# Sachbericht Teil I - Kurzfassung

## I.1 Aufgabenstellung

Das Ziel des Forschungsvorhabens IRLeQuM war die Steigerung der Prozess- und Produktqualität sowie die Reduktion des Ausschusses in der Massivumformung.

Dazu wurde ein intelligenter Regler entwickelt, der den Fertigungsprozess automatisch anpasst und optimiert. Ein Regler ist eine Steuerungseinheit, die Prozesswerte überwacht und gezielt Maschinenparameter anpasst, um eine gleichbleibende Qualität sicherzustellen. Der hier entwickelte Regler basiert auf Reinforcement Learning (RL), einer Form des maschinellen Lernens, die selbstständig optimale Einstellungen findet.

Der RL-Regler nutzt Qualitäts- und Prozessdaten (z. B. Maschinenparameter) und ermittelt daraus optimale Einstellungen, um Ausschuss zu reduzieren und die Qualität zu verbessern. Er greift dabei in den **Qualitätsregelkreis (QRK)** der Fertigung ein.

Um die Anlernzeit in der Produktion zu minimieren, wurde der RL-Regler vorab mit einem stochastischen Metamodell auf Basis historischer Daten und Simulationen trainiert. So konnte er bereits vor dem ersten Einsatz fundierte Entscheidungen treffen und schneller produktiv eingesetzt werden.

## I.2 Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens

Die Beherrschung von Massivumformprozessen stellt bis heute eine Herausforderung aufgrund hoher Prozessbelastungen und komplexer Bauteilgeometrien dar [1, 2]. Durch die Nutzung von QRK können gewonnene Informationen über Qualitätsmerkmale verarbeitet und Handlungen oder Handlungsempfehlungen zurück in den Prozess gespielt werden [3]. Eine Möglichkeit einen ganzheitlichen QRK zu entwickeln, ist die Integration von Künstlicher Intelligenz (KI) in den Produktionsprozess.

Reinforcement Learning ist ein Teilgebiet des Machine Learning (ML) und passt einen Agenten (Aktor) im QRK an den Prozess an. Das Ziel des Agenten ist das Erlernen des Zusammenhangs zwischen Zuständen, Aktionen und Nutzen (Bewertung), um darauf basierend die Kosten durch Anpassung der Stellgrößen zu minimieren. [5].

Beim *Transfer Learning* wird bestehendes (Domänen-)Wissen auf eine neue Aufgabe übertragen. Das Ziel von TL ist die Reduktion der Anlernzeit und -daten für den Algorithmus. Umgesetzt wurden die Ansätze bisher bei Strategiespielen [8] und Robotersteuerungen [6].

Das Vorhaben wurde in neun Arbeitspakete bearbeitet. Im Anschluss an die Anforderungsaufnahme seitens der Anwendungsunternehmen zu Beginn des Projektes wurden die IT-Infrastruktur und Schnittstellen definiert, das stochastische Metamodell aufgebaut, Transfer Learning Algorithmen sowie Reinforcement Learning identifiziert und implementiert. Die Einzelkomponenten wurden in einen übergeordneten Regler integriert und anhand der Prozesse der Anwendungsunternehmen validiert. Projektübergreifend wurden die Teilergebnisse

dokumentiert und der Öffentlichkeit zugänglich gemacht. Im Projektverlauf wurde festgestellt, dass das Reglertraining durch historische Daten zielführender ist und zusätzliche Kontrollinstanz durch Prozessexperten sinnvoll ist. Beide Aspekte wurden im Rahmen einer Vorhabensbeschreibungsänderung in den Projektfokus mit aufgenommen.

### **I.3 Wesentliche Ergebnisse im Überblick**

Das Projekt IRLeQuM hat bedeutende Fortschritte in der KI-gestützten Prozessoptimierung erzielt. In drei industriellen Anwendungsfällen wurden die entwickelten Methoden getestet und validiert:

#### **Anwendungsfälle und ihre Ergebnisse**

- **Mubea (Umformtechnik, Automobilzulieferer)**
  - Ziel: Ausschussreduktion durch KI-gestützte Prozessregelung.
  - Ergebnis: Der RL-Regler wurde erfolgreich in die IT-Datenpipeline integriert und zeigte erste Erfolge in der Prozessstabilisierung.
- **ESW (Schmiedetechnik, Simulation & Regelung)**
  - Ziel: Optimierung der Schmiedeprozesse durch Simulation und KI-Regelung.
  - Ergebnis: Die entwickelte Simufact-Simulation ermöglichte eine gezielte Prozessoptimierung mit verbesserten Werkzeugstandzeiten.
- **Schomäcker (Fertigungstechnik, Qualitätsmodellierung)**
  - Ziel: Erstellung eines prädiktiven Qualitätsmodells als Grundlage für die Prozessregelung.
  - Ergebnis: Ein expertenbasiertes Modell zur Qualitätsbewertung, trotz fehlender vollständiger Produktionsdaten.

#### **Machbarkeitsnachweis von RL-Modellen für die industrielle Prozessoptimierung**

Ein wesentliches Projektergebnis war der Nachweis, dass Reinforcement Learning erfolgreich mit prädiktiven Modellen kombiniert werden kann, um Prozessparameter automatisiert zu optimieren. Die theoretische Validierung zeigt, dass diese Methode in der Produktion zur Ausschussreduktion und Qualitätssteigerung einsetzbar ist.

Für IconPro sind die Ergebnisse von großer Bedeutung, da sie direkt in die Weiterentwicklung der ARES-Software zur KI-gestützten Prozessoptimierung einfließen und die Grundlage für neue industrielle Anwendungen bilden.

# Sachbericht Teil II - Eingehende Darstellung

## II.1 Motivation und Aufgabenstellung

Das Ziel war die Entwicklung eines neuartigen Verfahrens zur Steigerung der Prozess- und Produktqualität sowie der Reduktion des Ausschusses im Bereich der Massivumformung. Hierfür wurde ein Regler entwickelt, der auf einem RL-Ansatz basiert und in einem geschlossenen **Qualitätsregelkreis (QRK)** verwendet werden kann (Abbildung 1). Dem RL-Regler wurden hierfür die geforderten Qualitätsmerkmale sowie Prozessdaten (u.a. Qualitätsmerkmalsausprägungen, Prozessparameter) übergeben. Der RL-Regler berechnet anschließend die optimalen Prozessparameter und führt diese zur automatischen Regelung zurück in den Massivumformprozess. Durch den Zugriff auf aktuelle Prozessdaten und der Schließung des QRK werden Instabilitäten (z.B. Umgebung, Rohstoff) vom RL-Regler adaptiert. Zur Bestimmung der optimalen Prozessparameter maximiert der RL-Regler seine Kostennutzenfunktion, welche die Prozessgüte basierend auf impliziertem Bedienerwissen (Domänenwissen) und den unternehmerischen Zielen quantifiziert. Um die Anlernzeit des RL-Reglers am realen Prozess zu minimieren, wird vorgelagert ein stochastisches Metamodel basierend auf historischen Daten und/oder, (Teil-)Prozesssimulation erstellt und der RL-Regler mit diesem über einen TL-Ansatz vortrainiert. Dies ermöglicht unter anderem die Anwendung in der Prototypenfertigung, bei der tendenziell weniger (Prozess-)Daten zur Verfügung stehen. Neben der Entwicklung des neuartigen RL-Reglers und dem TL-Ansatz zum Vortrainieren des Reglers über Historische Daten / Simulationsergebnisse, umfasste das Vorhaben dabei die Beschreibung der Umsetzung in Form von Leitfäden, den Aufbau von physischen Demonstratoren bei den Anwendungspartnern zur Erprobung und Validierung des entwickelten RL-QRK sowie die Entwicklung eines Python-Softwarepakets zur Unterstützung der Implementierung des Reglers bei den Unternehmen.

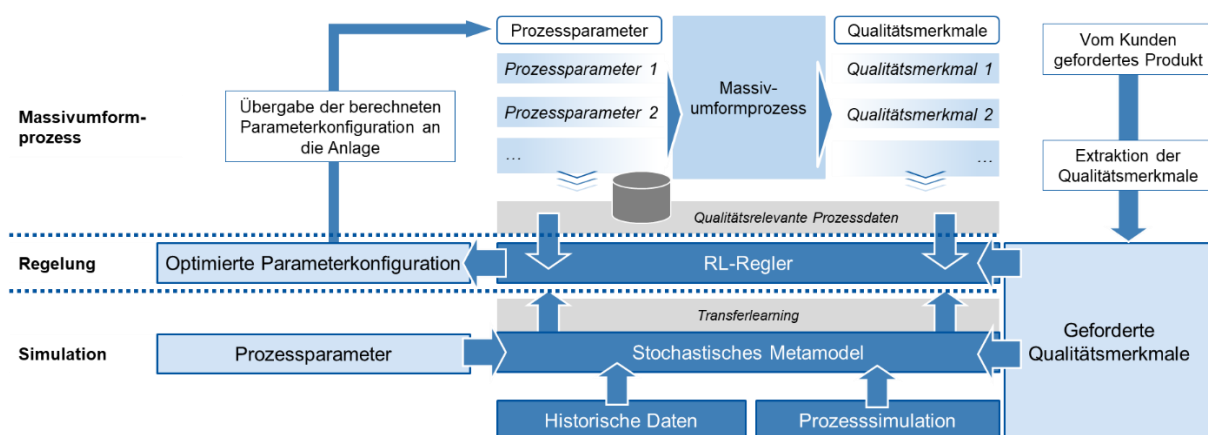


Abbildung 1: Qualitätsregelkreis zur Optimierung von Prozessparametern mittels Reinforcement Learning

## II.2. Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens

Nach DIN 8580 werden bei der *Massivumformung* Querschnittsveränderungen erzeugt, ohne dass Material abgetragen wird (Walzen, Biegen, Freiformschmieden,

Vorwärtsfließpressen) [1]. Die Beherrschung dieser Prozesse stellt bis heute eine Herausforderung aufgrund hoher Prozessbelastungen und komplexer Bauteilgeometrien dar [2]. Eine Möglichkeit zur Stabilisierung der Produktionsprozesse und der Optimierung der Produktqualität sind **QRK**. QRK nutzen gewonnene Informationen über Qualitätsmerkmale, verarbeiten diese und spielen Handlungen oder Handlungsempfehlungen zurück in den Prozess [3]. Hierbei sollen QRK alle qualitätsrelevanten Informationen berücksichtigen, in Echtzeit den Produktionsprozess regeln und ressourcenschonend (geringer Implementierungsaufwand, geringe Hard- und Softwarekosten) anwendbar sein. In Unternehmen insb. auch in Unternehmen der Massivumformung eingesetzte, maschinenexterne QRK beziehen sich häufig auf eine übergeordnete Managementebene (z. B. Deming-Zyklus) oder nutzen der Produktion nachgelagerte Informationen (z. B. Statistische Prozessregelung) [3]. Die statistische Prozessregelung (SPC) analysiert die gemessenen Qualitätsmerkmale und kann statistisch signifikante Auffälligkeiten detektieren. Die Analyse dieser Auffälligkeiten und die Anpassung möglicher Einstellparameter obliegt jedoch dem Menschen. Die maschinenexternen QRK haben somit keinen direkten Einfluss auf den aktuellen Produktionsprozess und regeln diesen über einen retrospektiven, manuellen Eingriff. Modellbasierte oder maschineninterne QRK sind oft statisch, können sich nicht automatisiert an veränderliche Umgebungsbedingungen (z. B. Rohstoffeigenschaften) anpassen oder beziehen nicht alle qualitätsrelevante Informationen (z. B. Messwerte von Qualitätsmerkmalen) ein. Die in der Massivumformung weit verbreiteten steuerungsinternen Regler ermöglichen es, einzelne Prozess- und Steuerungsparameter auf einen Sollwert zu regeln. Dadurch, dass diese Regler gezielt z. B. die Schwankungen eines bestimmten Prozessparameters minimiert, ist es selten möglich, eine übergreifende Regelung zu implementieren, welche alle qualitätsrelevanten Informationen mit einbezieht. Das hierzu durchgeführte Schwerpunktprogramm *Eigenschaftsgeregelte Umformprozesse* (SPP 2183) hat zum Ziel wissenschaftliche Grundlagen der prozessintegrierten Eigenschaftsregelung (Umformgrad, Härte, Kaltverfestigung), z. B. mit PID-Reglern für Bauteiltemperatur oder Prozesskräfte von Umformprozessen, zu erforschen. In dem Schwerpunktprogramm wird gezeigt, dass klassische Regler aus der Regelungstechnik zwar einzelne Prozessparameterschwankungen reduzieren können, durch die fehlende Korrelation mit den Qualitätsmerkmalen und weiteren Informationen (z. B. Umgebungsbedingungen) jedoch nicht zwingend der Ausschuss reduziert werden kann. Es besteht somit der Bedarf nach einem ganzheitlichen Konzept für einen QRK. Eine Möglichkeit einen ganzheitlichen QRK zu entwickeln, ist die Integration von Künstlicher Intelligenz (KI) in den Produktionsprozess. KI bietet die Möglichkeit automatisiert Verknüpfungen von verschiedensten Daten sowie Datenquellen zu erstellen und Handlungen daraus abzuleiten [4]. Vor allem in den Digitalisierungsabteilungen von großen Unternehmen kann man die ersten Ansätze der Integration von KI in die Produktion verfolgen. Beispiele hierfür sind Algorithmen des Predictive Maintenance, welche auf die Optimierung von Wartungszyklen abzielen, oder Algorithmen der automatisierten Fehlererkennung über Machine Vision. Diese Algorithmen bieten für Unternehmen bereits einige Vorteile, weil sie

Zusammenhänge erkennen, welche Menschen nicht zwingend erfassen können. Diese Algorithmen haben aber auch gemeinsam, dass sie nicht optimal für die Nutzung als QRK sind, da sie einen QRK nicht, oder zumindest nicht automatisiert schließen können. Den Algorithmen fehlt die entsprechende Feedback-Schleife in den Prozess. Algorithmen, welche diese Feedback-Schleife integriert haben, sind Algorithmen des RL.

**Reinforcement Learning** ist ein Teilgebiet des Machine Learning (ML) und beschreibt eine autonom durchgeführte Regelung. RL regelt über Optimierung einer Kostennutzenfunktion und passt mit Aktionen eines Agenten (Aktor) im QRK den Prozess an (Abbildung 2). Zum Zeitpunkt  $t$  wird durch den Agenten eine Aktion durchgeführt, indem eine Stellgröße (z.B. Vorschub  $a_t$ ) verändert wird. Dadurch ändert sich das physische System, sodass ein neuer Zustand (z.B. Blechdicke  $s_t$ ) erreicht wird. Der Zustand wird mit einer Kostennutzenfunktion bewertet. Das Ziel des Agenten ist das Erlernen des Zusammenhangs zwischen Zuständen, Aktionen und Nutzen (Bewertung), um darauf basierend die Kosten ( $r_t$ ) durch Anpassung der Stellgrößen zu minimieren. Für die Übertragung der Aktionen  $a_t$  und der Zustände  $s_t$  in den Massivumformprozess muss eine Schnittstelle zwischen Anlagensteuerung und RL-Hardware erarbeitet werden [5]. Die RL-Algorithmen bieten somit großes Potenzial für Unternehmen. Das Defizit von RL für QRK ist die nötige, große Datenbasis (lange Anlernphase) zum Lernen der Zusammenhänge und dem Erstellen der Kostenfunktion bei komplexen Prozessen, wie dem Massivumformen. Zudem reicht die IT-Infrastruktur in vielen Unternehmen nicht aus, solche Algorithmen zu implementieren und anzuwenden, da die IT-Infrastruktur entweder nicht alle Informationen abgreift und/oder aufgrund der Latenzen Entscheidungen und Regelanpassungen nicht in Echtzeit übermittelt werden können [6]. Dies sind jedoch Voraussetzungen, um RL-QRK zu schließen und erfolgreich anzuwenden. Somit führt der aufwendige Lernprozess und die nicht ausreichende IT-Infrastruktur dazu, dass RL in den Unternehmen bisher nicht eingesetzt wird. Um den Defiziten zu begegnen und die nötigen Daten zu reduzieren, soll existentes Domänenwissen über TL in das RL-Modell integriert werden.

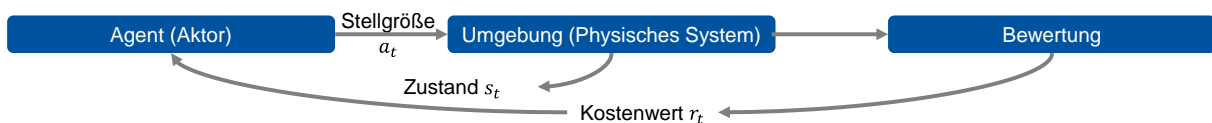


Abbildung 2: Vorgehen von Reinforcement Learning

Beim Transfer Learning wird bestehendes (Domänen-)Wissen auf eine neue Aufgabe übertragen. Das Ziel von TL ist die Reduktion der Anlernzeit und -daten für den Algorithmus. TL ist beim Supervised und Unsupervised Learning verbreitet, beim RL aber noch Neuland. Im Allgemeinen wird die Aufgabe, von der Wissen extrahiert werden soll, als Informationsquelle bezeichnet und die neue Aufgabe, auf die das Wissen angewandt wird, Zielaufgabe genannt. Die TL-Methoden können in verschiedene Verfahrensarten unterteilt werden. Hier wird unterschieden, ob die Variablen und Aktionen des Agenten bei der Informationsquelle und der Zielaufgabe

identisch oder unterschiedlich sind. Außerdem wird unterschieden, ob aus einer (Single-Task Learning) oder mehreren (Multi-Task Learning) Informationsquellen gelernt wird [7]. Umgesetzt wurden die Ansätze bisher bei Strategiespielen [8] und Robotersteuerungen [6]. Hier wurde Wissen über Simulationen auf die Zielaufgabe übertragen. Stochastische Unsicherheiten der Randbedingungen oder Simulationsparameter werden dabei nicht betrachtet. Dieser Ansatz eignet sich jedoch vor allem für Unternehmen in der Massivumformung, da hier häufig eine große Expertise im Bereich der Simulationen vorherrscht. Die Unternehmen nutzen die Simulationen, um Prozessverständnis aufzubauen, aber auch um Prozessanläufe zu verkürzen. Ein Beispiel hierfür ist DIGISIM von Schuler, welches die Erstellung einer Simulation vereinfachen soll. Die Herausforderung eine solche Simulation in eine Regelung zu überführen oder einen RL-Regler damit anzulernen, wurde bisher jedoch noch nicht gelöst.

### II.3 Planung und Ablauf des Vorhabens

Workshops wurden mit den Partnern (Mubea, ESW, Schomäcker) durchgeführt, um Prozessanforderungen zu definieren und ein gemeinsames Verständnis der Produktionsprozesse zu schaffen. Wichtige Parameter wurden identifiziert, und Prozessflussdiagramme wurden in Zusammenarbeit mit dem WZL auf einem Miro-Board erstellt.

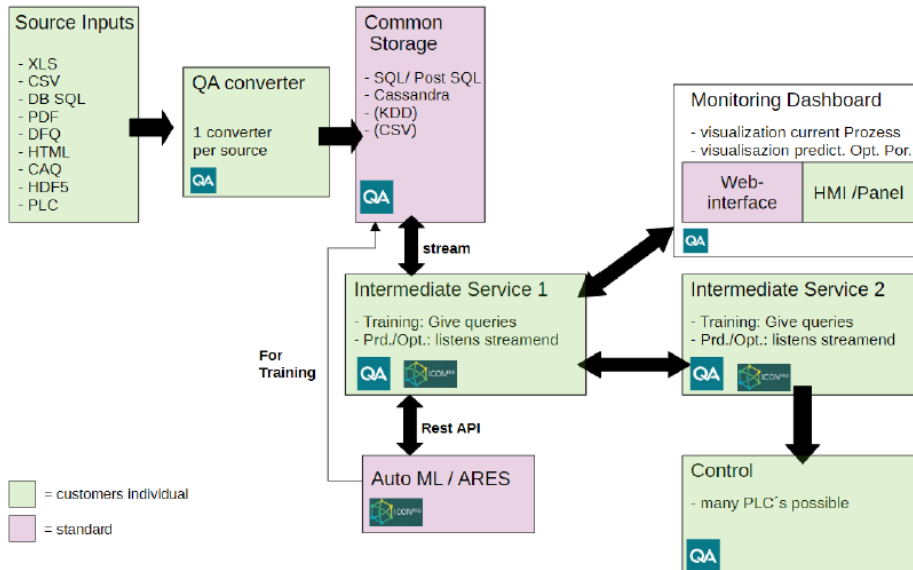


Abbildung 3: Vollständige Daten- und Machine-Learning-Operations-Pipeline.

Eine strukturierte IT-Infrastruktur wurde entwickelt, wobei der Schwerpunkt auf Datenbankkompatibilität, Datenkonvertierung und Speicheranforderungen lag. Die Integration verschiedener Datenbanken sowie die erforderliche Hardware und Software für KI-basierte Regler wurden analysiert. Die Module von QA und IconPro wurden in einem Flussdiagramm dokumentiert (Abbildung 3).

Im Rahmen dieser Arbeiten wurde in Zusammenarbeit mit Quality Automation (QA) eine übergeordnete Übersicht der zentralen Komponenten – von der Datenerfassung bis zur Parametersteuerung – in AP 2.1 erstellt. Dazu gehörte die Identifikation potenzieller Datenquellen, die von Unternehmen wie Mubea, ESW und Schomäcker genutzt werden, sowie deren Konvertierung in ein gemeinsames Speichersystem, das durch den Konverter von QA ermöglicht wurde. Nach der Speicherung wurden die Daten für das Training von KI-Modellen innerhalb der IconPro-ARES-Software genutzt. Die Komponenten wurden kategorisiert, um festzustellen, ob sie Standardlösungen darstellten oder eine individuelle Implementierung erforderten.

In AP 2.2 wurde die erforderliche Hardware und Software für ein KI-basiertes Steuerungssystem weiter analysiert, mit besonderem Fokus auf Datenformatkompatibilität und die von QA durchgeführten Konvertierungsprozesse. Abschließend behandelte AP 2.3 die allgemeinen Hardwareanforderungen für ARES- und RL-Regler, um eine nahtlose Integration in die gesamte IT-Infrastruktur sicherzustellen.

### **Metamodell-Erstellung**

Daten wurden extrahiert, bereinigt und verarbeitet, um prädiktive Machine-Learning-Modelle zu trainieren. Am WZL wurden Simulationen für den Ritzelherstellungsprozess von ESW entwickelt, wobei Faktoren wie Hub, Kraft und Temperatur unter Verwendung von Simufact berücksichtigt wurden. Dies führte zu 600 Datenpunkten, die verschiedene Parameterbereiche für Temperatur, Hub und die anfängliche Dicke der Spule repräsentieren. Das auf dieser Basis aufgebaute prädiktive Modell wird im nächsten Abschnitt näher erläutert.

Die Datenbank von Schomäcker (der letzten 10 Jahre) wurde analysiert, um 32 finale Qualitätsparameter für Blattfedern vorherzusagen. Ein umfangreicher Datensatz mit über 50.000 Stichproben war verfügbar. Diese Stichproben mussten nach verschiedenen Artikelnummern aufgeteilt werden. Allerdings war für die Analyse von Schomäcker der Qualitätsparameter nicht direkt verfügbar. Da dieser jedoch eine wesentliche Voraussetzung für den Aufbau prädiktiver Modelle darstellt, wurde er aus einer Regel abgeleitet, die von Schomäcker auf Basis des Verhaltens der Endpositionsmessungen demonstriert wurde. Abbildung 4 zeigt die Verteilung der verschiedenen Register für Blattfedern.

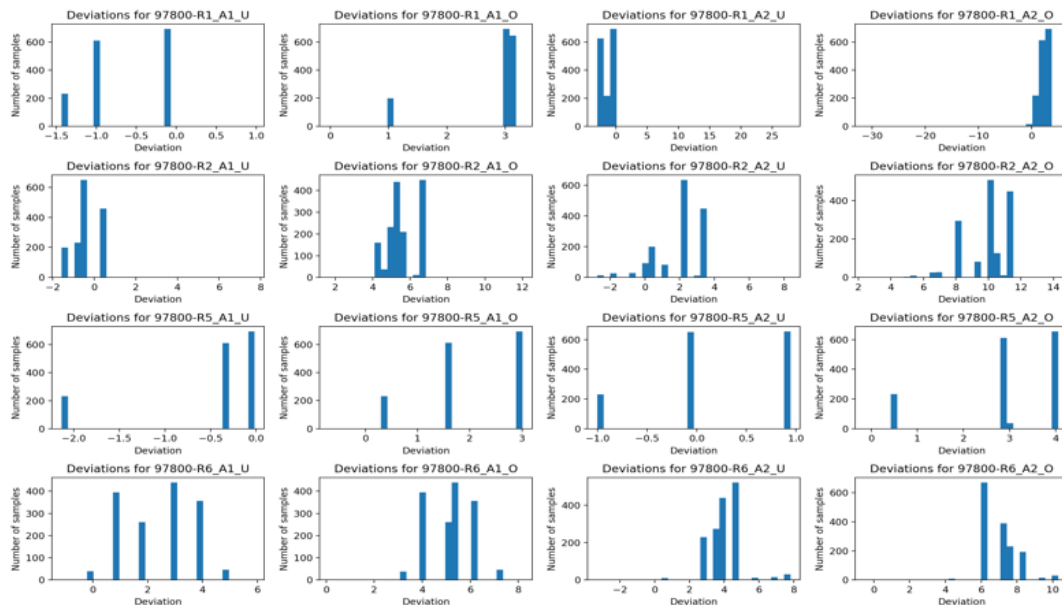


Abbildung 4: Verteilung der Endpositionswerte für verschiedene Motoren auf Registern in der Blattfederfertigung (Schomäcker).

## Reinforcement Learning

Im Arbeitspaket 5 des IRLeQuM-Projekts wurden bedeutende Fortschritte bei der Auswahl und Implementierung von RL-Algorithmen zur Optimierung industrieller Umformprozesse erzielt. Die ersten Arbeiten konzentrierten sich auf die Evaluierung modernster RL-Methoden, darunter Q-Learning, Deep Q-Networks (DQN) und Policy-Optimierungstechniken wie Maximum a Posteriori Policy Optimization (MPO). Aufgrund der multiobjektiven Natur industrieller Prozesse wurde die Forschung auf Multi-Objective MPO (MO-MPO) ausgeweitet, um eine Optimierung über mehrere Qualitätsziele hinweg zu ermöglichen, insbesondere in Bezug auf Radius und Dicke bei Umformanwendungen. Die Einteilung verschiedener RL-Algorithmen ist in Abbildung 5 dargestellt.

Zur Unterstützung der RL-basierten Entscheidungsfindung wurden Metamodelle entwickelt, die auf historischen und simulierten Daten trainiert wurden. Die Validierung erfolgte anhand von Datensätzen der Anwendungspartner Mubea, Schomäcker und ESW. Die Kosten-Nutzen-Funktion, ein entscheidender Bestandteil der RL-Implementierung, wurde iterativ verfeinert, um die Prozessstabilität und Leistung auszubalancieren und eine Echtzeit-Anwendbarkeit in Fertigungsumgebungen zu gewährleisten.

Eine REST-API wurde entwickelt, um eine nahtlose Interaktion mit den trainierten RL-Modellen zu ermöglichen und deren Integration in bestehende industrielle Workflows zu erleichtern. Dies wurde durch einen sicheren Authentifizierungsmechanismus mit JWT-Tokens unterstützt. Es wurden erhebliche Anstrengungen unternommen, um eine geeignete Belohnungsfunktion für den RL-Regler sowohl für Einzelziel- als auch Mehrzieloptimierung zu definieren wie in Gl. 1 dargestellt.

$$O_{\text{reg}}(\vec{s}) = \max(1 - |P_{\text{reg}}(\vec{s}) - \hat{y}|, 0) \in [0,1] \quad (1)$$

Diese Fortschritte tragen zum übergeordneten Ziel bei, KI-gesteuerte Optimierung in industrielle Produktionsumgebungen zu integrieren, um Variabilität zu reduzieren, die Qualitätskontrolle zu verbessern und die gesamte Prozesseffizienz zu steigern. Die Implementierung von RL-Reglern ermöglicht adaptive Entscheidungsfindung, wodurch kontinuierliche Prozessverbesserungen und eine Minimierung des Materialausschusses erreicht werden.

Um eine effektive Implementierung sicherzustellen, wurde die RL-Trainingspipeline systematisch strukturiert, indem Datenvorverarbeitung, Modellvalidierung und Hyperparameter-Tuning integriert wurden. Die Kosten-Nutzen-Funktion spielte eine entscheidende Rolle bei der Abwägung von Prozessstabilität, Produktionsgeschwindigkeit und Qualitätsmetriken, indem gewichtete Strafen für Abweichungen in zentralen Parametern berücksichtigt wurden. Das Reinforcement-Learning-Framework wurde so konzipiert, dass es sowohl diskrete als auch kontinuierliche Aktionsräume unterstützt, um den komplexen Steuerungsanforderungen industrieller Umformprozesse gerecht zu werden.

Zur weiteren Leistungssteigerung wurde eine Feature-Importance-Analyse der Metamodell-Ergebnisse durchgeführt, um sicherzustellen, dass die einflussreichsten Prozessparameter im Optimierungsprozess priorisiert werden. Fortgeschrittene Reward-Shaping-Techniken wurden angewendet, um den RL-Agenten gezielt zu optimalen Entscheidungen zu führen, wobei Domänenwissen von Industrieexperten einbezogen wurde.

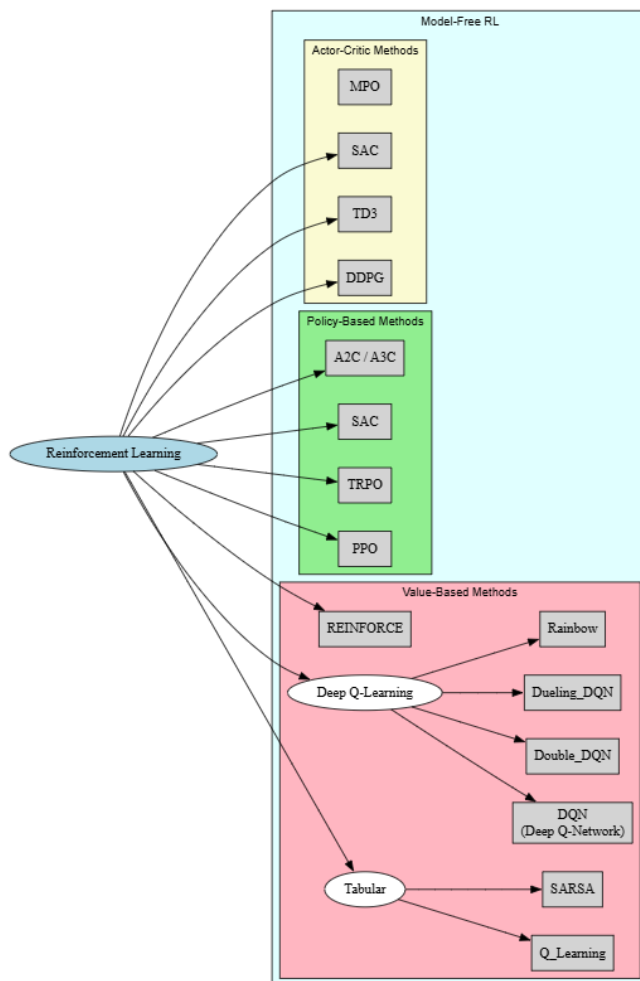


Abbildung 5: Klassifikation der Reinforcement-Learning-Algorithmen. Innerhalb der modellfreien Algorithmen gibt es hauptsächlich drei Kategorien: wertbasierte Methoden, Actor-Critic-Methoden und politikbasierte Methoden

## Transfer Learning

Im Arbeitspaket 4 des IRLeQuM-Projekts wurden verschiedene Transfer-Learning (TL)-Ansätze untersucht, um die Anpassungsfähigkeit von RL-Modellen zu verbessern, die auf simulierten Daten trainiert wurden, und sie für reale Anwendungen nutzbar zu machen. Die untersuchten Methoden umfassten Domänenanpassung (Domain Adaptation), Reward Shaping und Domänenrandomisierung (Domain Randomization), wobei jede Strategie unterschiedliche Ansätze zur Überbrückung der Lücke zwischen Simulation und realen Prozessdynamiken bot.

Die Domänenanpassung zielte darauf ab, die Merkmalsverteilungen zwischen der Quell- (simulierten) und Ziel-Domäne (realen) anzugleichen, sodass die gelernten Repräsentationen in beiden Umgebungen effektiv bleiben. Reward Shaping konzentrierte sich darauf, die Belohnungsfunktion so zu modifizieren, dass der RL-Agent effizienter zu optimalen realen Verhaltensweisen geleitet wird. Beide Ansätze stellten jedoch Herausforderungen hinsichtlich der Rechenkomplexität und des umfangreichen domänenspezifischen Kalibrierungsaufwands dar.

Nach eingehender Evaluierung wurde Domänen Randomisierung als die am besten geeignete Technik identifiziert, insbesondere für das Training von KI-Modellen mit simulationsbasierten Daten. Diese Methode führt kontrollierte Variationen in simulierten Trainingsumgebungen ein—z. B. Schwankungen in Materialeigenschaften, Maschinenparametern und äußeren Bedingungen—, um sicherzustellen, dass das RL-Modell sich gut an reale Unsicherheiten anpassen kann. Durch das Training des RL-Agents über eine Vielzahl von zufällig variierenden simulierten Szenarien lernt das Modell robuste Steuerungsstrategien, die weniger anfällig für Unterschiede zwischen Simulation und Realität sind.

Zur Umsetzung dieser Methode wurden in AP 3 gezielt Variationen in Schlüsselprozessparametern in die entwickelten Metamodelle integriert. Diese Variationen halfen dem RL-Agenten, adaptive Steuerungsstrategien zu erlernen, die unter verschiedenen Produktionsbedingungen zuverlässig funktionieren. Die trainierten Modelle wurden mit historischen Daten der Industriepartner validiert, um eine Übereinstimmung mit realen Prozessverhalten sicherzustellen.

Die Integration der Domänen Randomisierung hat die Fähigkeit des Modells erheblich verbessert, von rein simulationsbasiertem Lernen in reale Fertigungsumgebungen überzugehen. Dadurch wird der Bedarf an umfangreichen realen Trainingsdaten reduziert und die Produktionsausfallzeiten während der Modellanpassung minimiert. Für den ESW-Anwendungsfall wurde ebenfalls Transfer Learning angewendet, um das gelernte prädiktive Modell mit Hilfe neuer Daten zu übertragen, die aus realen Tests bei ESW gewonnen wurden.

## **Integration und Infrastrukturentwicklung**

Im Rahmen der Integration und Infrastrukturentwicklung (AP 6) wurden bedeutende Fortschritte erzielt, um die nahtlose Bereitstellung und Anwendung von RL-Controllern in industriellen Umgebungen zu gewährleisten. Die Entscheidung, ein Kubernetes-Cluster für das Training von Metamodellen und RL-Controllern zu nutzen, wurde erfolgreich umgesetzt. Die Trainingspipeline wurde erweitert, um sowohl eine automatisierte Bereitstellung als auch ein kontinuierliches Training und die Validierung der RL-Modelle zu unterstützen. Besonderes Augenmerk wurde auf Skalierbarkeit und Echtzeitfähigkeit gelegt, um eine reibungslose Integration in bestehende Produktionsumgebungen sicherzustellen.

Zur Interaktion mit den trainierten RL-Modellen wurde eine REST-API auf Basis der OpenAPI-Spezifikationen entwickelt. In enger Abstimmung mit den Projektpartnern wurden Spezifikationen für Ein- und Ausgabedaten definiert und verschiedene API-Endpunkte implementiert. Diese Endpunkte ermöglichen nicht nur das Training und die Bereitstellung von Modellen, sondern auch die Echtzeitvorhersage optimaler Steuerungsparameter. Darüber hinaus wurde eine benutzerfreundliche grafische Oberfläche entwickelt, die es Ingenieuren ermöglicht, RL-Controller zu konfigurieren und deren Auswirkungen zu analysieren. Die intuitive Benutzererfahrung erhöht die Zugänglichkeit und erleichtert die industrielle Einführung der entwickelten RL-Modelle.

Zur Erhöhung der Sicherheit wurde ein JSON-Web-Token (JWT)-Authentifizierungssystem implementiert. Jeder Benutzer, der mit dem RL-Controller interagiert, benötigt ein eindeutiges JWT-Token, wodurch sichergestellt wird, dass nur autorisiertes Personal Zugriff auf die Steuerungsparameter und Modelldaten hat. Dies schützt das System vor unbefugtem Zugriff und gewährleistet die Integrität der Prozessoptimierung. Zusammen mit Quality Automation (QA) wurde eine effiziente Datenpipeline konzipiert und implementiert. Der QA-Webserver kommuniziert mit dem RL-Controller von IconPro (ARES) und ermöglicht einen bidirektionalen Datenaustausch. Eingehende Prozessparameter werden vom trainierten RL-Modell verarbeitet, das anschließend optimierte Steuerparameter für die Maschinen zurückerliefert. Diese Interaktion erfolgt über die REST-API, die als standardisierte Schnittstelle dient und eine flexible Integration in bestehende industrielle Umgebungen ermöglicht.

Um die industrielle Anwendbarkeit weiter zu verbessern, wurde in enger Zusammenarbeit zwischen QA und IconPro eine Webanwendung entwickelt, die den spezifischen Anforderungen der Partnerunternehmen gerecht wird. Die Web-App vereinfacht den Zugriff auf RL-Modelle, visualisiert Modellvorhersagen und ermöglicht eine einfache Anpassung der Steuerparameter. Vor der finalen Implementierung wurde eine umfassende Kompatibilitätsprüfung zwischen der Web-App und der ARES-Software durchgeführt, um eine nahtlose Integration sicherzustellen. Die entwickelte Infrastruktur bildet somit die Grundlage für eine effiziente und sichere Bereitstellung von RL-Controllern in der industriellen Produktion.

### **Validierung und Erstbereitstellung**

Die Tests des RL-Controllers wurden bei Mubea und ESW durchgeführt, wo eine reale Validierung erfolgte, um seine Wirksamkeit bei der Optimierung von Umformprozessen zu bewerten. Bei Schomäcker wurde neben physischen Tests auch eine virtuelle Validierung durch eine Remote-Evaluierung durchgeführt, wobei Simulationsdaten und Expertenfeedback genutzt wurden, um die Modellgenauigkeit und Anwendbarkeit sicherzustellen.

---

## **II.4 Erzielte Ergebnisse**

Der Inhalt dieses Abschnitts ist je nach Anwendungsfall in drei Typen unterteilt: Mubea, ESW und Schomäcker. Verschiedene Analyseergebnisse zeigen die Entwicklung eines prädiktiven Modells, eines Optimierungsmodells mit RL sowie eine Analyse der Merkmalswichtigkeit.

Die gemeinsame Motivation der drei Anwendungspartner Mubea, Schomäcker und ESW fußte auf der Reduktion von Ausschuss der Steigerung der Prozessqualität. Durch die Einführung von RL-QRK, die eine prädiktive Steuerung der Anlagenparameter ermöglichen, sollten qualitätsrelevante Prozessfähigkeitskennwerte verbessert werden. Zum besseren Verständnis der von

Icon Pro und nachfolgend dargestellten Ergebnisse, werden die Anwendungsfälle der Anwendungspartner im Folgenden dargestellt:

### **Beschreibung des Mubea Anwendungsfalls**

Zur Herstellung von Karosserieblechteilen werden durch periodisches Anpassen des Walzspalts Dickenprofile in Stahlbänder mit konstanter Einlaufdicke gewalzt (flexibles Walzen). Die Einstellung der optimalen Walzparameter erfolgt auf Basis eines Regelkreises zwischen diesen Parametern und einer anschließenden Dickenmessung des Blechprofils. Im Schnitt müssen etwa drei Platinen aufgrund von abweichenden Toleranzen verschrottet werden, bis die korrekten Walzparameter eingestellt sind. Durch die Integration des neuartigen RL-QRK und der Berücksichtigung weiterer Prozesseinflussgrößen sollte der Einregelprozess verkürzt werden. Eine Ausschussreduktion von ca. 50 %, bzw. eine Reduktion auf eine Platine beim Anfahren der Anlage wurde erwartet. (Fokus: Reduzierung von Ausschuss im Einregelprozesses (Anlauf) beim flexiblen Walzen).

### **Ergebnisse des Mubea Anwendungsfalls**

Für den Mubea-Anwendungsfall wurde ein Datensatz über ein Jahr hinweg gesammelt und auf verschiedene Weise analysiert. Zunächst wurde ein prädiktives Modell entwickelt, um den Fehler zwischen dem tatsächlichen Ausgangsdickenwert vorherzusagen. Die Streudiagramme des Test-Splits sind in Abbildung 6 dargestellt.

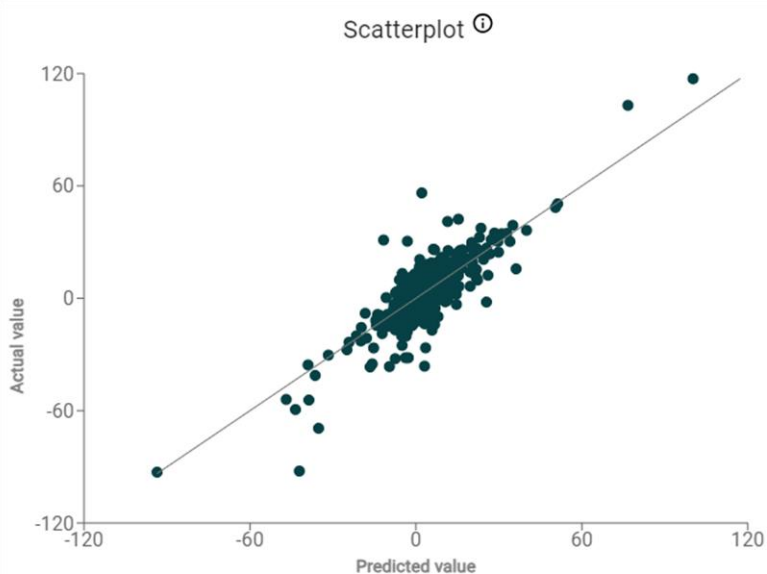


Abbildung 6: Streudiagramm, das den Fehler zwischen der vorhergesagten und der tatsächlichen Ausgangsdicke im Vergleich zur Nominaldicke für den flexiblen Walzprozess von Mubea zeigt.

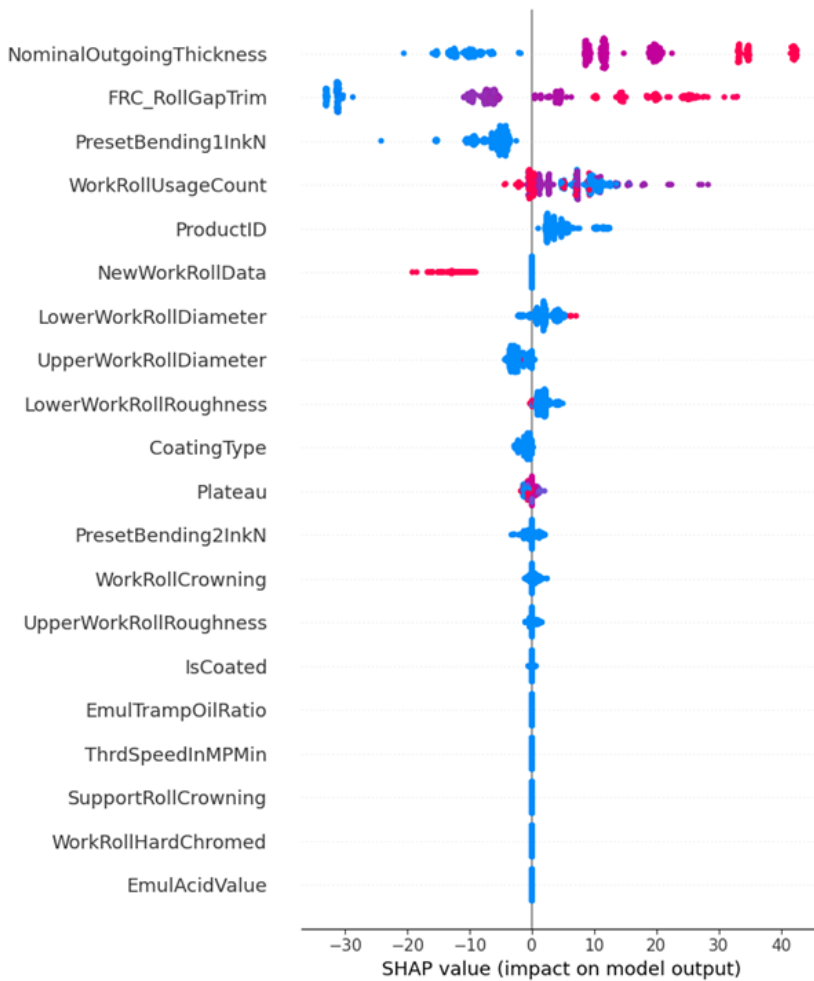


Abbildung 7: Diagramme zur Merkmalswichtigkeit für die Analyse des flexiblen Walzprozesses. Es ist deutlich zu erkennen, dass die FRC-Rollspaltenanpassung und die Produkt-ID eine entscheidende Rolle im Entscheidungsprozess spielen.

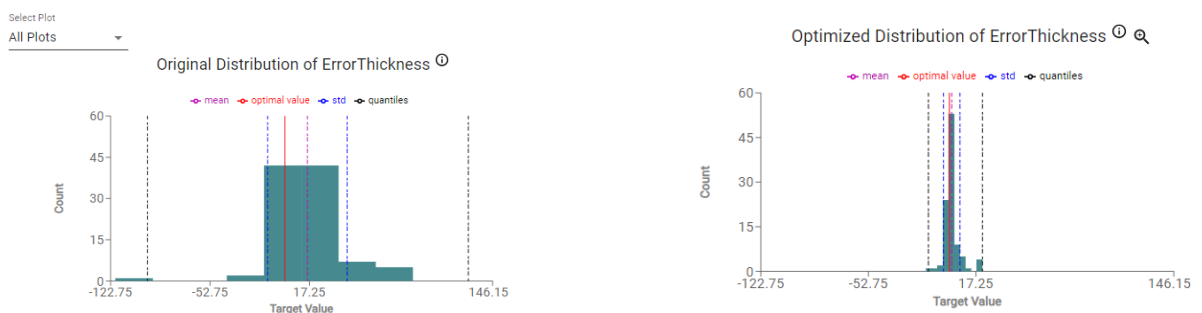


Abbildung 8: Histogramm, das die Optimierungsergebnisse vor und nach der Optimierung für den Fehler in der Ausgangsdicke und der nominalen Ausgangsdicke zeigt.

## **Beschreibung des ESW Anwendungsfalls**

Beim Vollvorwärtsfließpressen von Sicherheitsbauteilen für die Automobilindustrie regeln Bediener basierend auf Erfahrungswerten die Fließpressgeschwindigkeit, die Werkzeugposition und -zusammenstellung in den Umformpressen. Mit der Einführung eines RL-QRK sollen die Werkstückausformung durch eine prädiktive Regelung optimiert werden und den Bedienern als Hilfswerkzeug zur Verfügung gestellt werden. Hierdurch kann die Unabhängigkeit von Fachkräften gesteigert werden und eine Produktivitätssteigerung von ca. 25 % wird erwartet. (Fokus: Reduzierung von Ausschuss in der laufenden Fertigung)

## **Ergebnisse des ESW Anwendungsfalls**

Das WZL erstellte in Absprache mit ESW eine Simulation für den Kaltumformungs-Anwendungsfall der Ritzelfertigung in SimuFact. Die Simulation wurde genutzt, um 600 Punkte zu generieren, die den Parameterraum für die Ritzelfertigung abdecken. Eine Beschreibung der simulierten Parameter ist in Tabelle 1 dargestellt. Besonders wichtig war die Entscheidung, welche Parameter verwendet werden sollten, da Parameter, die während des Prozesses generiert werden, nicht für die Erstellung eines Prozessmodells genutzt werden sollten. Der Grund dafür ist, dass diese Parameter erst nach Eingabe der Steuerparameter generiert werden. Die Parameter wurden von IconPro umfassend für die Prozesspartner klassifiziert.

Zwei Metamodelle, eines für die Dicke des Ritzels und eines für den Radius des Ritzels, wurden mit einem Feedforward-Neuronalen Netz trainiert. Beide Modelle erreichten eine hohe Genauigkeit mit einem Score von 0,97 für die Dicke und 0,99 für den Radius. Die Merkmalswichtigkeiten für die entsprechenden Modelle sind in Abbildung 9 dargestellt. Gemeinsam zeigen sie den Einfluss verschiedener Prozessparameter auf die angestrebten Dicken- und Radiuswerte. Dies unterstützt auch die prozesstechnische Interpretation. Zum Beispiel sollte ein höherer Hubwert zu einem größeren Radius führen, was ebenfalls in Abbildung 11 ersichtlich ist.

Ein Multi-Agenten-Reinforcement-Learning-Modell wurde trainiert, um beide Zielgrößen – Radius und Dicke – gleichzeitig zu optimieren. Abbildung 10 zeigt das Histogramm der Dicke für eine Stichprobe vor und nach dem Optimierungsprozess. Eine theoretische Steigerung der Prozessfähigkeit wird für die Dicke von 0,33 auf 0,44 und für den Radius von 0,32 auf 0,38 berichtet.

Tabelle 1: Beschreibung der mit der SimuFact-Simulation simulierten Parameter

Prozessparameter [Einheit]	Beschreibung
Hub	Der für den Umformprozess angewendete Hub
Werkzeugtemperatur	Temperatur des Werkzeugs während des Umformprozesses
Maschinensteifigkeit	Steifigkeit der Maschine bei Raumtemperatur
Flanschdicke	Anfangsdicke des Flanschs
Ausgabe Kräfte	12 Kraftwerte, die in gleichen Hubintervallen während der Simulation aufgezeichnet wurden
Ausgabe Flanschdicke	Finale Flanschdicke des Produkts
Ausgabe Flanschradius	Finaler Flanschradius des Produkts

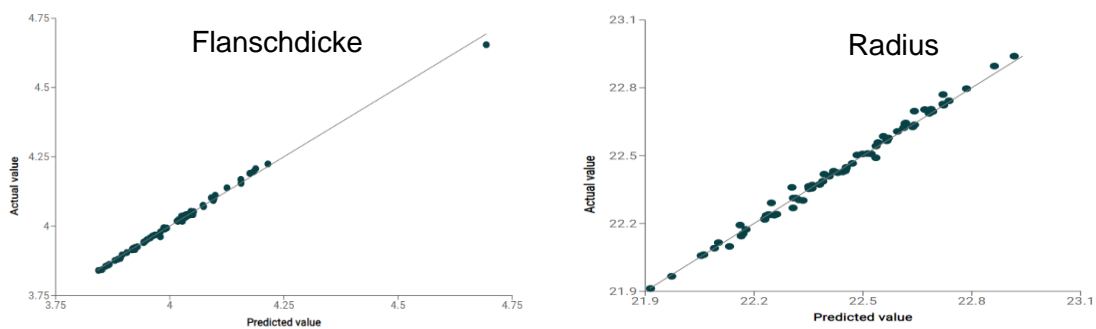


Abbildung 9: Streudiagramm für den tatsächlichen vs. vorhergesagten Wert der Dicke und des Radius des Ritzels

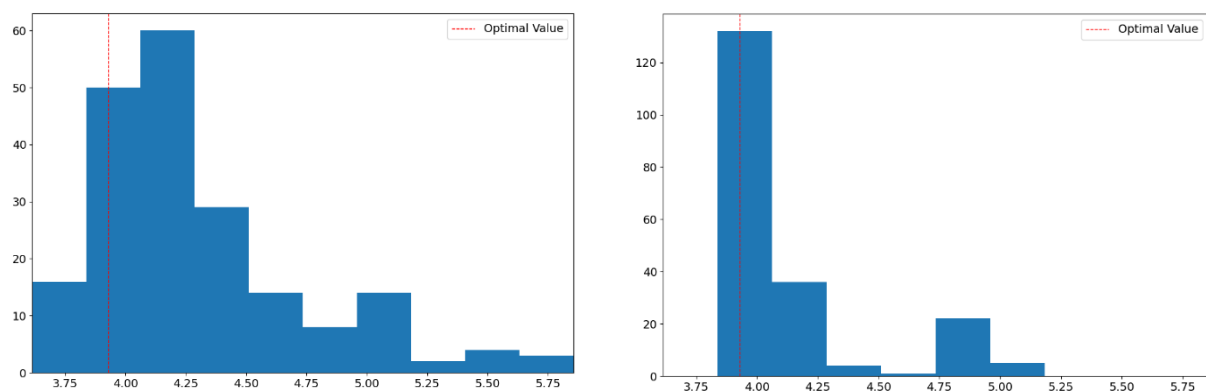


Abbildung 10: Histogramm der Dickenwertverteilung vor und nach der Optimierung. Die rote Linie zeigt den optimalen Wert. Es zeigt sich, dass sich die Prozessfähigkeit um 30 % verbessert.

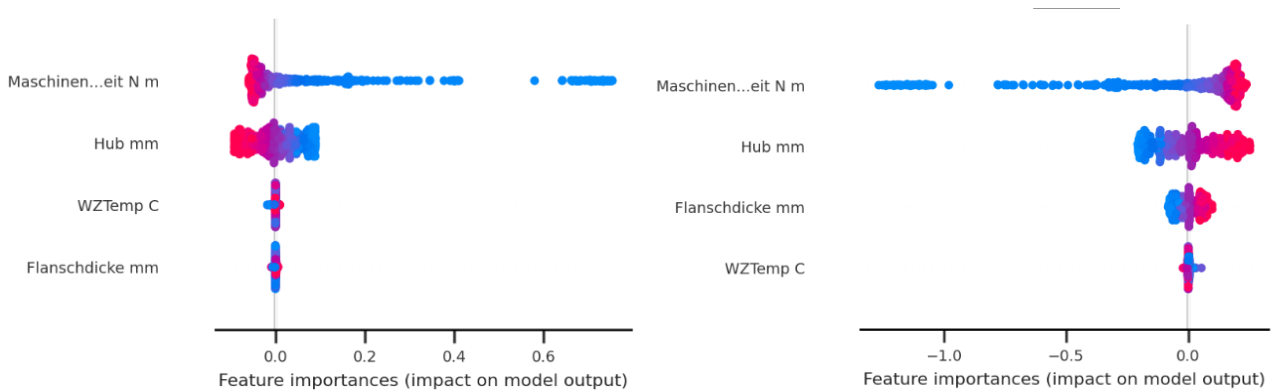


Abbildung 11: Diagramm zur Merkmalswichtigkeit, das den Einfluss verschiedener Merkmale zeigt (links: Dicke, rechts: Radius).

## Beschreibung des Schomäcker Anwendungsfalls

Die Erwärmung, Umformung und das nachfolgende Abschrecken der Werkstücke sind wichtige Prozessschritte bei der Herstellung von Blattfedern. Die zu erreichende Bauteilkontur und Bauteil- bzw. Mikrostruktureigenschaften werden durch die Temperatur- sowie Formgebungsparameter beeinflusst. Durch die Integration eines RL-QRK sollen Schwankungen bei den Bauteilabmessungen minimiert werden. Hierdurch wurde eine Steigerung des Prozessfähigkeitskennwert cpk von ca. 30 % erwartet. (Fokus: Reduzierung von Formabweichungen beim erwärmten Biegen mit anschließender Abschreckung als Folge von temperaturbedingten Mikrostruktur- bzw. Gefügeveränderungen)

## Ergebnisse des Schomäcker Anwendungsfalls

Für den Schomäcker-Anwendungsfall zur Qualitätsvorhersage von Blattfedern waren keine Qualitätsparameter verfügbar. In enger Abstimmung mit Experten von Schomäcker wurden diese Werte generiert (vgl. II.3). Basierend auf diesen Qualitätsparametern und mehr als 80 Produktionsparametern wurde ein prädiktives Qualitätsmodell entwickelt. Die Testergebnisse dieses Modells sind in Abbildung 12 dargestellt. Für eine Artikelnummer wurde eine hohe balancierte Genauigkeit von 95 % berichtet. Abbildung 13 zeigt die fünf wichtigsten Merkmale aus der Merkmalsanalyse, wobei ersichtlich wird, dass der Parameter Overmann\_SW einen großen Einfluss hatte. Dies stellte eine wertvolle Erkenntnis für die Prozessexperten von Schomäcker dar. Die Empfehlungen des RL-Controllers funktionierten jedoch für Schomäcker nicht. Ein möglicher Grund dafür ist das Fehlen präziser Qualitätsparameter. Obwohl die Qualitätsparameter anhand logischer Regeln generiert wurden, entsprechen sie nicht zwangsläufig den tatsächlichen Labels. Die Validierungstests wurden von ESW am in der Cloud bereitgestellten RL-Modell von IconPro durchgeführt.

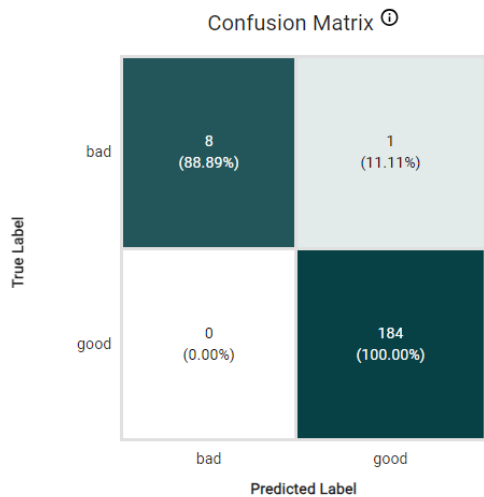


Abbildung 12: Konfusionsmatrix für die generierten Qualitätsparameter, die die endgültige Artikelqualität als gut oder schlecht klassifiziert.

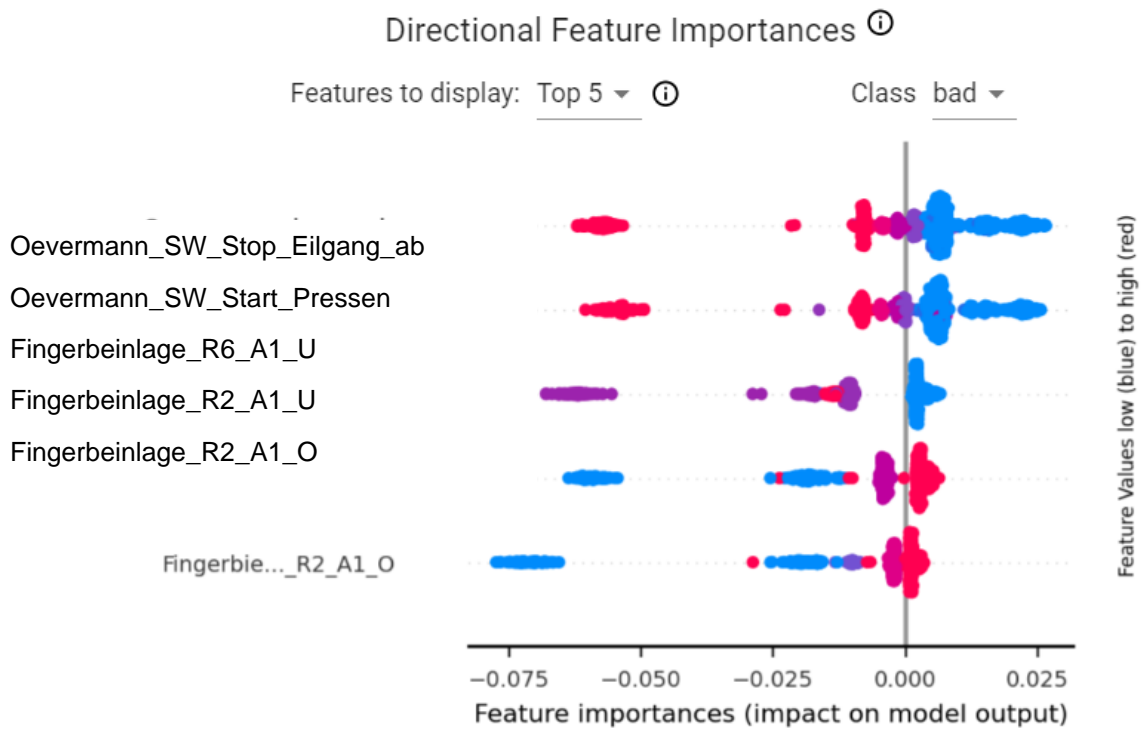


Abbildung 13: Diagramm zur Merkmalswichtigkeit für die Blattfederanalyse. Besonders bemerkenswert ist, dass Oevermann\_SW eine hohe Bedeutung aufweist.

## **Bedeutung der Projektergebnisse für IconPro und potenzielle Anwender**

Das wesentliche Projektergebnis für IconPro ist die Entwicklung und Validierung eines RL-gestützten Steuerungsmodells, das in Kombination mit prädiktiven Modellen zur automatischen Prozessoptimierung eingesetzt werden kann.

- ✓ **Technologischer Vorteil:** Die Kombination aus Predictive Modelling und Reinforcement Learning ermöglicht erstmals eine adaptive, selbstlernende Prozessregelung, die sich flexibel auf unterschiedliche Fertigungsszenarien übertragen lässt.
- ✓ **Direkte Verwertung:** Die entwickelten Algorithmen und Architekturansätze werden in die ARES-Software von IconPro integriert und dort als modulares KI-Optimierungsmodul angeboten.
- ✓ **Mehrwert für Unternehmen:** Produktionsbetriebe können diese KI-gestützte Prozessregelung nutzen, um Ausschuss zu reduzieren, Qualität zu steigern und Prozesse robuster zu gestalten – ohne tiefgehendes KI-Expertenwissen.
- ✓ **Industrielle Anwendung:** Die positive theoretische Validierung und die Tests in realen Fertigungsumgebungen zeigen, dass die KI-gestützte Prozessregelung praxistauglich ist und wirtschaftlichen Mehrwert bietet.

Mit diesen Ergebnissen positioniert sich IconPro als innovativer Partner für KI-gestützte Fertigungsoptimierung und bietet Unternehmen eine skalierbare, industrietaugliche Lösung zur intelligenten Prozessregelung.

### **II.5 Darstellung des während des Vorhabens bekannt gewordenen Fortschritts auf diesem Gebiet bei anderen Stellen**

Während des Vorhabens ist kein Fortschritt anderer Stellen auf diesem Gebiet bekannt geworden.

### **II.6 Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit der Ergebnisse**

Die Ergebnisse des IRLeQuM-Projekts zeigen den industriellen Mehrwert datengetriebener Prozessoptimierung und bestätigen die Wirksamkeit von KI-gestützten Steuerungsmodellen in realen Fertigungsumgebungen.

- **Für Mubea:** Die gewonnenen Erkenntnisse belegen die Machbarkeit KI-gestützter Prozessregelung zur Reduzierung von Ausschuss und Effizienzsteigerung im Umformprozess.
- **Für Schomäcker:** Die Merkmalsanalyse ermöglichte ein tieferes Verständnis der entscheidenden Prozessparameter, das zu präziseren Qualitätsvorhersagen beiträgt.
- **Für ESW:** Der Simulationsprozess legte die Grundlage für industrielle Tests und demonstrierte das Potenzial von RL-Optimierung für Multi-Objektiv-Prozesse.

**Ein zentrales Ergebnis für IconPro** ist die erfolgreiche Entwicklung und Validierung von Reinforcement Learning (RL)-Methoden in der Prozessoptimierung, insbesondere in Kombination mit prädiktiven Modellen.

#### **Wann und wie Unternehmen von den Ergebnissen profitieren können:**

- ✓ **Ab 2025:** Integration der entwickelten RL-Algorithmen in die ARES-Software von IconPro als neues Prozessoptimierungsmodul.
- ✓ **Für Bestands- und Neukunden:** Ermöglichung einer automatisierten, selbstlernenden Prozessregelung, um Ausschuss zu reduzieren, Prozessstabilität zu erhöhen und Qualitätsschwankungen zu minimieren.
- ✓ **Einfache Integration in bestehende Produktionsumgebungen:** Das neue Modul wird als Cloud- oder On-Premise-Lösung angeboten und kann an bestehende MES- und ERP-Systeme angebunden werden.

#### **Wirtschaftliche und wissenschaftliche Anschlussfähigkeit**

- Die entwickelten Methoden werden in zukünftige Industrieprojekte und F&E-Kooperationen eingebunden.
- IconPro plant die gezielte Weiterentwicklung der KI-Modelle für weitere Fertigungsbranchen, z. B. 3D-Druck, sensorbasierte Echtzeitregelung oder energieintensive Prozesse.
- Die Projektergebnisse fördern die breitere Anwendung KI-gestützter Optimierungsmethoden und positionieren IconPro als führenden Anbieter für KI-gestützte Prozessregelungslösungen in der Industrie.

Durch diese gezielte Verwertung und Skalierung wird sichergestellt, dass die Forschungsergebnisse langfristig in der industriellen Praxis genutzt werden, sowohl durch bestehende Kunden als auch durch neue Partnerschaften und wissenschaftliche Kooperationen.

#### **II.7 Zusammenarbeit mit anderen Stellen außerhalb des Verbundprojektes**

IconPro möchte das im Forschungsprojekt gewonnene Wissen nutzen, um weiter an bilateralen Industrieprojekten zu arbeiten. Aufbauend auf dem bestehenden Demonstrator plant IconPro, diesen weiterzuentwickeln.

#### **II.8 Veröffentlichungen, Vorträge Referate, etc.**

IconPro reicht eine wissenschaftliche Arbeit beim IJCIM Journal ein.

#### **II.9 Literaturverzeichnis**

- [1] Laska, R. u. Felsch, C.: Werkstoffkunde für Ingenieure. Viewegs Fachbücher der Technik. Wiesbaden: Vieweg+Teubner Verlag 1992
- [2] Klocke, F.: Fertigungsverfahren 4. VDI-Buch
- [3] Schmitt, R. u. Pfeifer, T.: Qualitätsmanagement. Strategien - Methoden - Techniken. 2015
- [4] Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow

- [5] Hguyen, H. u. La, H.: Review of deep reinforcement learning for robot manipulation. Third IEEE International Conference on Robotic Computing (IRC) 2019
- [6] Taylor, M. E. u. Stone, P.: Cross-domain transfer for reinforcement learning. Proceedings of the 24th international conference on Machine learning. New York, New York, USA: ACM Press 2007, S. 879–886
- [7] Taylor, M. E. u. Stone, P.: Transfer Learning for Reinforcement Learning Domains: A Survey. Journal of Machine Learning Research (2009)
- [8] Manu Sharma, Michael P. Holmes, Juan Carlos Santamaría, Arya Irani u. Ashwin Ram: Transfer Learning in Real-Time Strategy Games Using Hybrid CBR/RL. Hyderabad, India 2007, S. 1041–1046
- [9] Paranjape, A., Plettenberg, N., Ohlenforst, M., & Schmitt, R. H. (2023). Reinforcement learning for quality-oriented production process parameter optimization based on predictive models. In Industrial Engineering and Applications (Vol. 35, pp. 327–344). Advances in Transdisciplinary Engineering. IOS Press. <https://doi.org/10.3233/ATDE230059>