

**Integration intelligenter Klassifikatoren und
Prüfmethodiken in die automatische optische Inspektion
von Automotive-Produkten in der SMT-Fertigung**

Dissertation
zur Erlangung des Doktorgrades der Ingenieurwissenschaften

vorgelegt von
Patricia Carlotta Matthies, M.Sc.
aus Gießen

genehmigt von der Fakultät für Mathematik/Informatik und
Maschinenbau der Technischen Universität Clausthal

Tag der mündlichen Prüfung
24. März 2023

Dekan
Prof. Dr. rer. nat. Jörg P. Müller

Dekan: Prof. Dr. rer. nat. Jörg P. Müller

Vorsitzender der Promotionskommission: Prof. Dr.-Ing. Armin Lohrengel

Betreuer: Prof. Dr.-Ing. Volker Wesling

Gutachter: Prof. Dr. rer. nat. Andreas Rausch

Gutachter: Prof. Dr.-Ing. Karl-Heinz Spitzer

D104

Dissertation Technische Universität Clausthal

2023

– There is a way out of every box, a solution to every puzzle;
it's just a matter of finding it. –

Jean-Luc Picard – Star Trek

Abstract

The continuously increasing miniaturisation of SMD components as well as the increasing integration density of electronic components on the PCBs of automotive products are challenging both the manufacturing processes and their quality assurance. In the mass production of SMT products, automated optical and X-ray inspection prove to be efficient and robust inspection methods.

The present work deals with the synthesis of modern methods of machine and deep learning and AOI technology. In two main research areas, on the one hand, an active modification of the AOI process is carried out with the help of modern deep and machine learning methods, while on the other hand, a post-AOI verification process contributes to the identification of the component condition in case of conspicuous AOI inspections.

An acquisition procedure of numerical, class-annotated AOI process information is developed and presented under consideration of data quality requirements, in particular the correctness, consistency as well as completeness of the information.

Conventionally designed AOI test plans are improved in their ability to classify solder joints and component states by adjusting the parameters of the rule-based classification integrated into the AOI process using decision tree-based algorithms (Anchor, Decision Tree, RuleFit, SkopeRules) and the heuristic ripper method.

Furthermore, machine learning methods (Support Vector Machine, variants of Random Forest), neural network architectures (TabNet, Multilayer Perceptron) and evolution methods (GANN, MOEFS) are able to process tabular AOI manufacturing data and generate efficient classifiers.

In addition, this research work contains recommendations to ensure the models are up-to-date and implementation suggestions in SMT manufacturing.

Zusammenfassung

Die kontinuierlich zunehmende Miniaturisierung von SMD-Bauelementen sowie die ansteigende Integrationsdichte elektronischer Bauteile auf den Leiterplatten von Automotive-Produkten stellen sowohl für die Fertigungsprozesse als auch deren Qualitätssicherung eine Herausforderung dar. In der Massenproduktion von SMT-Produkten erweisen sich die automatische optische und die röntgentechnische Inspektion als effiziente und robuste Prüfverfahren.

Die vorliegende Arbeit befasst sich mit der Synthese von modernen Verfahren aus den Bereichen Machine- und Deep-Learning und der AOI-Technologie. In zwei Hauptforschungsgebieten erfolgt zum einen eine aktive Änderung des AOI-Prozesses mithilfe moderner Deep- und Machine-Learning-Methoden während zum anderen ein Post-AOI-Verifikationsprozess zur Identifikation des Bauteilzustands bei auffälligen AOI-Prüfungen beiträgt.

Ein Akquisitionsverfahrens von numerischen, klassenannotierten AOI-Informationen wird unter Berücksichtigung von Anforderungen an die Datenqualität, im Besonderen die Korrektheit, Konsistenz sowie Vollständigkeit der Informationen, entwickelt und vorgestellt.

Konventionell angelegte AOI-Prüfpläne erfahren durch eine Parameteranpassung der in den AOI-Prozess integrierten regelbasierten Klassifikation mithilfe von auf Entscheidungsbäumen basierenden Algorithmen (Anchor, Decision Tree, RuleFit, SkopeRules) und der heuristischen Ripper-Methode Verbesserungen des Klassifikationsvermögens von Lötstellen und Bauteilzuständen.

Des Weiteren sind Verfahren des Machine-Learning (Support Vector Machine, Varianten des Random Forest), neuronale Netzwerkarchitekturen (TabNet, Multilayer-Perzeptron) sowie Evolutionsmethoden (GANN, MOEFS) in der Lage, tabellarische AOI-Daten zu verarbeiten und effiziente Klassifikatoren zu generieren. Darüber hinaus enthält diese Forschungsarbeit Empfehlungen zur Sicherstellung der Aktualität der Modelle und Implementierungsvorschläge in der SMT-Fertigung.

Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als Doktorandin in dem Bereich Industrial Engineering der Continental Automotive GmbH in Karben durch Kooperation mit dem Institut für Schweißtechnik und Trennende Fertigungsverfahren der Technischen Universität Clausthal.

Ich bedanke mich bei meinem Doktorvater Prof. Dr.-Ing. Volker Wesling für die Realisierung und die Bereitschaft, eine externe Promotion an seinem Lehrstuhl zu begleiten. Seine exzellente fachliche Betreuung, das mir entgegengebrachte Vertrauen sowie seine Erfahrungen und Ratschläge ermöglichten es mir, jede Klippe erfolgreich zu umschiffen.

Darüber hinaus bedanke ich mich bei meinen Gutachtern Herrn Prof. Dr. rer. nat. Andreas Rausch und Herrn Prof. Dr.-Ing. Karl-Heinz Spitzer für das Interesse an meiner Arbeit und die Annahme des Koreferats.

Ein großer Dank geht an die Continental Automotive GmbH in Karben und Regensburg, welche meinen Wunsch der Anfertigung einer Industriepromotion stets unterstützten und umgesetzt haben. Im Besonderen danke ich meinem fachlichen Industriebetreuer Herrn Dr. oec. Thomas Kleinert, der mir in den Bereichen Data Mining, Machine-/Deep-Learning und Gestaltung einer wissenschaftlichen Forschungsarbeit jederzeit mit Rat und Tat zur Seite stand, und Herrn Klaus Schauburger, dessen Unterstützung in der Thematik der Löt- und SMT-Fertigungsprozesse unerlässlich waren. Ebenfalls möchte ich mich an dieser Stelle bei meinem dritten Industriebetreuer Herrn Daniel Wichmann bedanken. Seine Organisation und Förderung führte zum erfolgreichen Abschluss meines Promotionsbestrebens.

Des Weiteren bin ich dankbar für das Engagement und die Unterstützung der gesamten Belegschaft der Abteilung Front-End und der SMT-Fertigung der Continental Automotive GmbH in Karben. Im Besonderen sind Herr Martin Dietrich †, Herr Axel Heise und Herr Heiko Müller zu nennen.

Ich bedanke mich auch bei meiner Familie und meinen Freunden, die mich während meiner Promotion stets unterstützten, ein offenes Ohr hatten sowie zu jeder Tages- und Nachtzeit motivierende Worte an mich richteten.

Inhaltsverzeichnis

Abbildungsverzeichnis	V
Tabellenverzeichnis	IX
Abkürzungsverzeichnis	XI
1. Exposition	1
1.1. Ausgangssituation	1
1.2. Problemstellung und Motivation	2
1.3. Zielsetzung	3
2. Forschungssynthese	5
2.1. Produktionstechnischer Bezugsrahmen	5
2.1.1. Fügendes Fertigungsverfahren des Wiederaufschmelzlötens	5
2.1.1.1. Oberflächenmontagetechnologie in der Praxis	5
2.1.1.2. Lötwerkstoffe und Leiterplatten-Material	10
2.1.1.3. Fehlerarten und Einflussfaktoren	11
2.1.2. Automatische Inspektion in der SMT-Produktion	19
2.1.2.1. Funktionsweise und Prozessablauf AOI	19
2.1.2.2. Analysealgorithmen AOI	22
2.1.2.3. Verifikation AOI-Ergebnis	24
2.1.2.4. Einflussanalyse AOI-Prozess	25
2.2. Informationstechnischer Bezugsrahmen	28
2.2.1. Maschinelles Lernen	28
2.2.1.1. Überwachtes Lernen	29
2.2.1.2. Vorstellung Algorithmen des überwachten Lernens	31
2.2.1.3. Interpretierbares maschinelles Lernen	35
2.2.1.4. Feature-Selektionsverfahren	44
2.3. Systematisches Literaturreview	50
2.3.1. Literatursuche	50
2.3.2. Literaturlauswertung	54
2.3.3. Literaturlausanalyse und Präsentation	55
2.4. Grundkonzeptionierung der Forschungstheorie	60

2.5. Vorgehensmodell und Methodik der Forschung	62
3. Datenakquise	65
3.1. Identifikation des Untersuchungsgegenstands	65
3.2. Datenflussplan und Beschreibung der Prozessstufen	68
3.3. Datenstruktur und Inhalte der AOI-Informationen	72
4. Statistische Datenanalyse	75
4.1. Maße der zentralen Tendenz und Datenvisualisierung	75
4.1.1. Binäre Hauptuntersuchung	75
4.1.2. Multikategorielle Nebenuntersuchung	82
4.2. Maße des Zusammenhangs und Feature-Ranking	89
4.2.1. Binäre Hauptuntersuchung	89
4.2.2. Multikategorielle Nebenuntersuchung	92
4.3. Anpassungstests	96
4.3.1. Binäre Hauptuntersuchung	96
4.3.2. Multikategorielle Nebenuntersuchung	100
4.4. Dependenzanalyse	102
4.4.1. Binäre Hauptuntersuchung	102
4.4.2. Multikategorielle Nebenuntersuchung	103
4.5. Zuordnung relevanter Featurewerte zu Lötstellencharakteristika	105
4.6. Zwischenfazit statistische Datenanalyse	106
5. Maschinelles und tiefes Lernen	109
5.1. Anwendung von Verfahren des erklärbaren Lernens	109
5.1.1. Aufstellung Methodenplan erklärbares Lernen	110
5.1.2. Regelextraktion und Anpassung Prüfmethodik	111
5.1.3. Klassifizierungsgüte der generierten Prüfpläne in Fertigungsum- gebung	114
5.2. Anwendung von Verfahren des überwachten Lernens	122
5.2.1. Aufstellung Methodenplan überwachtetes Lernen	122
5.2.2. Validierung binärer und multikategoriemer Klassifikatoren	123
5.2.3. Gewährleistung der Modellgültigkeit	128
5.2.4. Zwischenfazit maschinelles und tiefes Lernen	130
6. Diskussion und kritische Würdigung	133
6.1. Ergebnisübertragung auf die Forschungshypothesen	133
6.2. Ableitung ingenieurwissenschaftlicher Lehrsätze	135
7. Zusammenfassung	137

8. Ausblick	139
Literaturverzeichnis	XIII
A. Anhang	XXI
A.1. Multikategorielle Datenakquise	XXI
A.2. Statistik	XXVI
A.2.1. Zentrale Tendenz und Dispersion – Thema 1	XXVI
A.2.2. Zentrale Tendenz und Dispersion – Thema 2	XXXIII
A.2.3. Anpassungstest – Thema 1	XXXIV
A.2.4. Anpassungstest – Thema 2	XXXVI
A.2.5. Korrelationsanalyse – Thema 2	XXXVIII
A.2.6. Feature-Ranking – Thema 2	XXXVIII
A.2.7. Dependenzanalyse – Thema 1	XXXIX
A.2.8. Dependenzanalyse – Thema 2	XLIII
A.2.9. Zuordnung Featurewert – AOI-Messgröße	XLVIII
A.3. Maschinelles und tiefes Lernen	L
Lebenslauf	LI

Abbildungsverzeichnis

2.1.	Aufbauschema SMT-Produktionslinie	6
2.2.	Beispielsystem Reflow-Konvektionsofen	8
2.3.	Reflow-Lötanlage mit PCBs	9
2.4.	Relative Fehlerverteilung Analyse 1: Gesamte SMT-Produktion	13
2.5.	Relative Fehlerverteilung Analyse 2: Fünf Haupt-BG SMT-Produktion	14
2.6.	Versetztes Bauteil nach IPC-A-610 G	15
2.7.	Zulässiger Seitenüberhang nach IPC-A-610 G	15
2.8.	Zulässige Lötstellenbreite nach IPC-A-610 G	16
2.9.	Aufgerichtetes Bauteil (Grabstein) nach IPC-A-610	17
2.10.	Schema AOI-Bildakquisition	21
2.11.	SI Debugbetrieb: Ergebnisdarstellung	23
2.12.	Schema AOI-Streifenprojektion	24
2.13.	Gemessene Eigenschaften Planaritätsprüfung 3D-Analyse	24
2.14.	Parameterdiagramm AOI-Prozess	26
2.15.	Übersicht Machine-Learning-Verfahren	29
2.16.	Prozessablaufplan Supervised Machine-Learning	30
2.17.	Margin lineare SVM	33
2.18.	Soft-Margin-Funktion lineare SVM	34
2.19.	Beispiel Visualisierung DT	36
2.20.	Beispiel Regelextraktion aus DT	37
2.21.	Skizze Geltungsbereich Anchor-Algorithmus 2D-Daten	41
2.22.	Programmablaufplan Anchor-Rules	42
2.23.	Programmablaufplan Ripper-Algorithmus	43
2.24.	Programmablaufplan Boruta-Algorithmus	48
2.25.	SOAP-Algorithmus: Projektion Feature 3, 23, 24 und Labeländerungen	49
2.26.	Ablaufdiagramm Literaturreview	51
2.27.	Morphologischer Kasten Thema 1	56
2.28.	Morphologischer Kasten Thema 2	60
2.29.	Ablaufplanung der Forschungsinhalte	63
3.1.	Beispielbilder Prüfmuster	67
3.2.	Ablaufplanung der Datenakquisition	71

3.3. Relative Häufigkeitsverteilung – Berücksichtigung sämtlicher Fehlercodes	74
4.1. Kardinalität – binär – Analyse-C	77
4.2. Zentrale Tendenz – binär/ternär – Analyse-C	79
4.3. Boxplot-Whisker-Diagramme – binär/ternär – Analyse-C	80
4.4. Histogramme – binär/ternär – Analyse-C	81
4.5. Histogramme – binär/ternär – Analyse-C	81
4.6. Streudiagramme – binär/ternär – Analyse-C	82
4.7. Häufigkeitsverteilung nach Winkel und FC – Analyse-C	83
4.8. Kardinalität – multikategoriell – Analyse-C	84
4.9. Boxplot-Whisker-Diagramme – multikategoriell – Analyse-C	85
4.10. Histogramme nach Winkel – Analyse-C	86
4.11. Histogramme nach Fehlercode – Analyse-C	87
4.12. Zentrale Tendenz – multikategoriell – Analyse-C	88
4.13. Streudiagramme nach Winkel und FC – Analyse-C	88
4.14. Heatmap-Diagramm – binär – Analyse-C	90
4.15. Dendrogramm – binär – Analyse-C	90
4.16. Heatmap-Diagramm nach FC – Analyse-C	93
4.17. Heatmap-Diagramm nach FC – Analyse-C	93
4.18. Dendrogramm nach FC – Analyse-C	94
4.19. Dendrogramm nach FC – Analyse-C	94
4.20. Q-Q-Diagramme und P-P-Diagramme – binär – Analyse-C	99
4.21. Anpassungsdiagramm – binär – Analyse-C	99
4.22. Q-Q-Diagramme und P-P-Diagramme – multikategoriell – Analyse-C	101
5.1. Auszug Visualisierung Decision Tree	112
5.2. Ergebnisse Prüfplanerstellung – Analyse-A	116
5.3. Ergebnisse Prüfplanerstellung – Analyse-B	118
5.4. Ergebnisse Prüfplanerstellung – Analyse-C	120
5.5. Ergebnisse Prüfplanerstellung – SOT23 – Analysen D und E	121
5.6. Ergebnisse Klassifikatoren – binär – Analyse-A	124
5.7. Vergleich F1-Score – binär und multikategoriell – Analyse-A	125
5.8. Ergebnisse Klassifikatoren – binär – Analyse-B	126
5.9. Vergleich F1-Score binär und multikategoriell – Analyse-B	126
5.10. Ergebnisse Klassifikatoren – binär – Analyse-C	127
5.11. Vergleich F1-Score binär und multikategoriell – Analyse-C	128
5.12. Ergebnisse Klassifikatoren binär – SOT23 – Analysen D und E	129
5.13. Vergleich F1-Score Einzelmodell und Ensemble-Methode – binär	129
6.1. Wachstumsfaktor Original-PP und Ensemble-Methode	134

A.1. Zentrale Tendenz – binär – Analyse-A	XXVI
A.2. Zentrale Tendenz – binär – Analyse-B	XXVII
A.3. Zentrale Tendenz – binär – Analyse-D	XXVIII
A.4. Zentrale Tendenz – binär – Analyse-E	XXIX
A.5. Boxplot-Whisker-Diagramme – binär/ternär – Analyse-B	XXIX
A.6. Boxplot-Whisker-Diagramme – binär/ternär – Analyse-A	XXX
A.7. Streudiagramme – binär – Analyse-A	XXX
A.8. Streudiagramme – binär – Analyse-B	XXXI
A.9. Streudiagramme – binär – Analyse-D	XXXI
A.10. Streudiagramme – binär – Analyse-E	XXXII
A.11. Zentrale Tendenz – multikategoriell – Analyse-C	XXXIII
A.12. Q-Q-Diagramme und P-P-Diagramme – ternär – Analyse-C	XXXIV
A.13. Q-Q-Diagramme und P-P-Diagramme – multikategoriell – Analyse-C	XXXVII
A.14. Heatmap-Diagramm – multikategoriell – Analyse-C	XXXVIII
A.15. Binäre logistische Regression – Analyse-C – EF-/OK-Klasse	XLI
A.16. Multinominale logistische Regression – ternär – Analyse-C	XLII
A.17. Binäre logistische Regression – Analyse-C – OK-/Tombstone-Klasse	XLVIII

Tabellenverzeichnis

2.1.	Viscom-SI: AOI-Prüffamilien und Einsatzgebiet	21
2.2.	Auszug Variablen Parameterdiagramm AOI-Prozess	27
2.3.	Literatursuche Thema 1 – Zuordnung Durchlauf/Schlüsselbegriffe	52
2.4.	Ausbeute Literatursuche Thema 1	52
2.5.	Literatursuche Thema 2 – Zuordnung Durchlauf/Schlüsselbegriffe	54
2.6.	Ausbeute Literatursuche Thema 2	54
3.1.	Eigenschaften Prüfmuster	67
3.2.	Analysealgorithmen Prüfmuster	68
3.3.	Datensatz SMD-Testkomponenten	73
4.1.	Feature-Ranking – binär – Analyse-C	91
4.2.	Feature-Ranking – multikategoriell – Analyse-C	95
4.3.	KS-Test – binär – Analyse-C	98
4.4.	KS-Test – multikategoriell – Analyse-C	100
4.5.	Median-Test – binär – Analyse-C	102
4.6.	Median-Test – multikategoriell – Analyse-C	104
4.7.	Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-C	106
5.1.	Methodenplan: Entscheidungsbaum-unabhängige XAI-Algorithmen	110
5.2.	Methodenplan: Entscheidungsbaum-basierende XAI-Verfahren	111
5.3.	Auszug Ergebnisse Skope-Algorithmus	112
5.4.	Konfusionsmatrix – Originalprüfplan	115
5.5.	Konfusionsmatrix – RuleFit-Algorithmus	119
5.6.	Methodenplan AOI-Klassifikatoren – binär und multikategoriell	122
A.1.	Aufbau Datensatz C0402 – multikategoriell	XXI
A.2.	Aufbau Datensatz C0603 – multikategoriell	XXII
A.3.	KS-Test: Anteil [%] normalverteilte Featurewerte – Binär – EF-Klasse	XXXIV
A.4.	KS-Test: Anteil [%] normalverteilte Featurewerte – Binär – OK-Klasse	XXXIV
A.5.	KS-Test – multikategoriell – Analyse-C	XXXVI
A.6.	KS-Test: Anteil [%] normalverteilte Featurewerte – Multikategoriell – Tombstone-Klasse – Winkel 90°	XXXVII

A.7. KS-Test: Anteil [%] normalverteilte Featurewerte – Multikategoriell – OK-Klasse – Winkel 90°	XXXVII
A.8. Importance-Kennzahlen Winkelinformationen – multikategoriell mit FC 0 und 560	XXXVIII
A.9. U-Test von Mann-Whitney – binär – Analyse-C	XXXIX
A.10. Kruskal-Wallis-Test – ternär – Analyse-C	XL
A.11. Dunn-Bonferroni-Test – binär – Analyse-C	XLI
A.12. U-Test von Mann-Whitney – multikategoriell – Analyse-C	XLIII
A.13. Kruskal-Wallis-Test – multikategoriell – Analyse-C	XLIV
A.14. Dunn-Bonferroni-Test – multikategoriell – Analyse-C	XLV
A.15. Kruskal-Wallis-Test – multikategoriell – Analyse-C	XLVI
A.16. Dunn-Bonferroni-Test – multikategoriell – Analyse-C	XLVII
A.17. Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-A	XLVIII
A.18. Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-B	XLIX
A.19. Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-D & Analyse- E (nur SOT23)	XLIX
A.20. Steigerung F1-Score Einzelmodell und Ensemble-Methode – binär	L

Abkürzungsverzeichnis

AOI	Automated Optical Inspection
Alg.	Algorithmus
angl.	Angular
AXI	Automated X-Ray Inspection
BB	Bildbasis
BE	Bauelemente
BV	Bildverarbeitung
DL	Deep-Learning
dpmo	Defects per Million Opportunities
DT	Decision Tree
EF	Echtfehler
FC	Fehlercode
fcpmo	False Calls per Million Opportunities
FKT	Funktionstest
FPY	First Pass Yield
Gen.	Generalisiert
HARAN	Halbautomatische Reparaturplatzanlage
Hypergeom.	Hypergeometrisch
i.O.	In Ordnung (Bauteilzustand)
IC	Integrated Circuit
ICT	In-Circuit-Test
Imp.	Importance-Kennzahl
Inv.	Invers
IPC	Institute for Printed Circuits
JSON	JavaScript Object Notation
KDE	Kernel Density Estimation
KPI	Key Performance Indicator
KS-Test	Kolmogorov-Smirnov-Test
ML	Machine-Learning
n.i.O.	Nicht in Ordnung (Bauteilzustand)
OGP	Ordnungsgesichtspunkt

orth.	Orthogonal
PCB	Printed Circuit Board
PET	Polyethylenterephthalat
PF	Pseudofehler
PP	Prüfplan
ppm	Parts per Million
P-P	Probability-Probability
Q-Q	Quantile-Quantile
RES-Datei	Result-Datei
Rez.	Reziprok
RF	Random Forest
ROI	Region of Interest
SMD	Surface-Mounted Device
SMT	Surface-Mounted Technology
SnAgCu	Zinn-Silber-Kupfer
SPI	Solder Paste Inspection
SVM	Support Vector Machine
TFT-LCD	TFT: Thin-Film-Transistor, LCD: Liquid Crystal Display
VPC	Viscom Process Control
VOPS	Viscom-Online-Programmier-Station
WEEE	Waste of Electrical and Electronic Equipment
XAI	Explainable Artificial Intelligence

1. Exposition

Dieses Kapitel beinhaltet die Kurzbeschreibung der Ausgangssituation, siehe Kapitel 1.1 und ermöglicht die Offenlegung von Prozesseigenschaften in Kapitel 1.2, welche Verbesserungsmaßnahmen unterzogen werden können. Im Anschluss an die fachliche Einordnung des Forschungsfeldes wird die Zielsetzung der vorliegenden Forschungsarbeit abgeleitet und in Kapitel 1.3 vorgestellt.

1.1. Ausgangssituation

Die Oberflächenmontage-Technologie (Abkürzung SMT; englisch Surface-Mounted Technology) stellt in der modernen Herstellung von Leistungselektroniken, insbesondere Automotive-Produkten, ein Verfahren dar, welches wirtschaftliche und fertigungstechnische Anforderungen in sich vereint. Aufgrund des Wegfalls von Verbindungsdrähten wird die Realisierung von hohen Packungsdichten und somit ein gesteigerter Miniaturisierungsgrad ermöglicht. Zusätzlich begünstigt die Automatisierbarkeit eine Massenfertigung von oberflächenbestückten Leiterplatten und erhöht die Wirtschaftlichkeit dieses Fertigungsverfahrens. Die SMT setzt sich aus drei Unterkategorien zusammen. Die Bauelemente und Leiterplatte entsprechen der ersten Kategorie, gefolgt von dem gewählten Bestückungssystem. Die Forschung dieser Arbeit widmet sich der dritten Kategorie: *Das Fertigungsverfahren und die Qualitätssicherung*. [Her17]

Ein effizientes Maschinenlötverfahren zur Herstellung der stoffschlüssigen Verbindung von oberflächenmontierten Bauteilen (Abkürzung SMD; englisch Surface-Mounted Device) und dem Basismaterial ist das Wiederaufschmelzlöten. Dies weist eine hohe Robustheit, Flexibilität und Gewährleistung kontinuierlicher Qualitätseigenschaften auf, sodass das Reflow-Löten das wichtigste Masselötverfahren in der SMT-Fertigung darstellt. [Bel05]

Im Anschluss an den Lötvorgang in Reflowöfen folgt die automatische Inspektion der generierten Lötstellen und oberflächenmontierten Bauteile. Abhängig von der Prüfbarkeit und den Charakteristika des Bestückungsplans finden optische (Abkürzung AOI; englisch Automated Optical Inspection) oder röntgentechnische (Abkürzung AXI; englisch Automated X-Ray Inspection) Prozesse Anwendung. Auffälligkeiten in der regelbasierten

Klassifikation der AOI-/AXI-Systeme führen zu einer manuellen Verifikation des Bauteilzustandes. Zu diesem Zweck werden die Ergebnisdaten in Form von Bilddaten der Region-of-Interest und Metainformationen, wie die Position auf Leiterplatte oder Identifikationsnummern des geprüften Mehrfachnutzens, an die Verifikationsstation im direkten Anschluss an die AOI-/AXI-Systeme weitergeleitet. Auf dem Monitor der Verifikationsstation können die Aufnahmen der Prüfung interpretiert werden. Zusätzlich bietet der Abgleich der Fertigungsdaten mit Gut-/Schlecht-Mustern eine Hilfestellung in der Entscheidungsfindung. [Vis22d]

1.2. Problemstellung und Motivation

Die Anforderungen an die Prüfpläne der AOI-/AXI-Systeme bestehen in der Realisierung einer maximierten Aufdeckungsrate von Echtfehlern mit gleichzeitiger Minimierung des Pseudofehlerrückkommens. Eine hohe Anzahl von Pseudofehlern bewirkt eine hohe Auslastung der Mitarbeiter an der Verifikationsstation, reduziert die Verfügbarkeit des Inspektionssystems und somit die Effektivität der gesamten Produktionslinie.

Die Parametrierung der Prüfpläne, einschließlich der Grenzwerte der regelbasierten Klassifikation, erfolgt manuell auf Basis archivierter Produktionsdaten unter Berücksichtigung der Erfahrungen des Mitarbeiters und Schwankungen/Veränderungen in der Produktion. Analog dazu wird die manuelle Klassifizierung des Bauteil-/Lötstellenzustandes an der Verifikationsstation durch die Erfahrung, Schulung und den Konzentrationszustand des Mitarbeiters beeinflusst.

Die Risikoanalyse identifiziert das Prüfsystem und den Mitarbeiter der Verifikationsstation als Einzelrisiken für das Post-Reflow-Inspektionsverfahren. Die Annotation eines Echtfehlers als einen zulässigen Bauteilzustand wird als technischer beziehungsweise menschlicher Schlupf definiert. Beide Risikoereignisse bewirken die Zuführung des fehlerhaften Produktes in nachfolgende Prozessschritte und eine Erhöhung der Wahrscheinlichkeit, dem Produkt mit unzureichender Funktionalität eine Kundenfreigabe zu erteilen. Die redundante Struktur des Qualitätsmanagements, welches eine Vielzahl von Prüfungen in der gesamten Prozesskette umfasst, wirkt diesem Szenario entgegen. Eine frühzeitige Erkennung von Löt-/Bauteilfehlern ist ebenso im Hinblick auf wirtschaftliche Aspekte zu realisieren.

Ferner wird in vergangenen Prüfungen von Testleiterplatten ein vermehrtes Aufkommen von Tombstone-Effekten bei SMD-Chipbauteilen und Bauteilgruppierungen mit gleichen Orientierungswinkeln auf der Leiterplatte nachgewiesen. Diese Wechselbeziehung stützt sich auf rein empirische Beobachtungen. Somit ist kein wissenschaftlicher Nachweis für die Existenz einer bestehenden Abhängigkeit zwischen der Begünstigung von Tombstone-

Defekten und dem Orientierungswinkel erbracht.

1.3. Zielsetzung

Aus der Prozess- und Risikoanalyse der automatischen optischen Inspektion von Erzeugnissen der Oberflächenmontagetechnologie sowie dem Wiederaufschmelzlöten entstehen zwei Forschungsschwerpunkte der vorliegenden Arbeit.

Eine Rahmenbedingung, die zu Beginn der Eingrenzung des Forschungsfeldes und Definition der Zielstellung feststeht, ist die Erhebung von numerischen, klassenannotierten AOI-Daten. Weitere Spezifizierungen der Aufgabenstellung, die Forschungsmethodik und die Auswahl der anzuwendenden Verfahren leiten sich aus dem Stand der Forschung und prozess- sowie informationstechnischen Machbarkeitsstudien der Fertigungssysteme ab.

Die erste zentrale Thematik widmet sich der Fragestellung, ob die Parameter von konventionell angelegten AOI-Prüfplänen durch Verfahren des maschinellen oder tiefen Lernens angepasst werden können, sodass die Klassifikationsgüte und Bewertungsmaßzahlen des AOI-Prozesses eine Verbesserung erfahren. Diese Aufgabenstellung sieht einen direkten Eingriff in die Prüfmethodik und den gesamten Inspektionsprozess vor. Die Prüfplanänderungen sollen dabei anwenderspezifisch erfolgen, ohne die Betriebssoftware und grundlegende Mechanismen der Prüfmethodik zu verändern. Implementierungen in der SMT-Produktion sind im Rahmen der herkömmlichen Prüfplangestaltung ohne Aktualisierungen der Software- oder Hardware durch den AOI-Systemhersteller zu realisieren.

Aufgrund der Einflussnahme des Menschen auf die Prozessgüte der Verifikationsstation wird die zweite Hauptthematik formuliert. Die Realisierbarkeit der Unterstützung des Verifikationsverfahrens durch Prognosen von modernen Klassifikationsmodellen aus den Bereichen des maschinellen und tiefen Lernens steht im Mittelpunkt dieses Forschungsthemas. Auf Basis der in dem AOI-Prozess erhobenen Inspektionsdaten sind Modelle zu generieren, welche dem Bediener der Verifikationsstation in der Erprobungsphase durch Vorhersage des Bauteilzustands eine Hilfestellung leisten. Es erfolgt kein aktiver Eingriff in den AOI-Prozess, sondern die Verifikationsstation liegt im Mittelpunkt der Untersuchungen.

Diese beiden Aufgabenstellungen werden in folgenden Kapiteln als das Hauptthema (Thema 1) betitelt.

Die Nebenuntersuchung (Thema 2) sieht eine wissenschaftliche Analyse der erhobenen AOI-Prozessdaten und Beurteilung der Einflussnahme der Bauteil-Orientierungswinkel auf die Entstehung von Tombstone-Fehlern vor.

2. Forschungssynthese

Dieses Kapitel stellt die systematische Erfassung und Zusammenfassung der wissenschaftlichen Erkenntnisse, welche für die Fragestellung dieser Promotionsschrift von Relevanz sind, vor. Abgrenzung und Definition des theoretischen Bezugsrahmens dienen als Grundlage dieser Forschungsarbeit. Bedingt durch die Aufgabenstellung erfolgt eine Unterteilung des Bezugsrahmens anhand produktionstechnischer, siehe Kapitel 2.1, sowie informationstechnischer, siehe Kapitel 2.2, Aspekte. Methodik und Ergebnisse des systematischen Literaturreviews werden in Kapitel 2.3 erläutert.

2.1. Produktionstechnischer Bezugsrahmen

In dem produktionstechnischen Bezugsrahmen werden in Abschnitt 2.1.1 zunächst die Definitionen und der Aufbau der relevanten Fertigungsprozesse, siehe Kapitel 2.1.1.1 vorgestellt. Eine Übersicht der verwendeten Lötmaterialien sowie der in der Fertigung auftretenden Lötfehlerarten erfolgt in Abschnitt 2.1.1.2 und 2.1.1.3. Gegenstand des Forschungsthemas dieser Arbeit ist die automatische optische Inspektion, sodass eine detailliertere Prozessbeschreibung und Vorstellung des Untersuchungsmodells in einem gesonderten Kapitel 2.1.2 erfolgt.

2.1.1. Fügendes Fertigungsverfahren des Wiederaufschmelzlötens

Dieses Unterkapitel umfasst die Definition und Eigenschaften des untersuchten Fügeverfahrens sowie eine Anwendungsbeschreibung in der automatisierten Produktion von oberflächenmontierten Automotive-Produkten. Des Weiteren wird die Entstehung von Lötfehlern und die Qualitätssicherung in einer modernen Fertigungsumgebung erörtert.

2.1.1.1. Oberflächenmontagetechnologie in der Praxis

Definition Reflow-Maschinenlöten Das Reflow-Maschinenlötverfahren, auch als Wiederaufschmelzlöten betitelt, wird durch die folgenden Normen charakterisiert und in

übergeordnete Prozesskategorien eingeteilt. Die Norm DIN 8580 ordnet der Hauptgruppe des Fügens die untergeordnete Klasse des *Fügens durch Löten* zu. [DIN03a]

Eine weitere Unterteilung wird in [DIN03b] anhand definierter Ordnungsgesichtspunkte (OGP) ermöglicht. Diese stellen in der nächstfolgenden Ebene zunächst die Löttemperatur dar, welche eine Aufspaltung in Verbindungs-, Weich- und Hartlöten erlaubt. Die Kategorisierung wird durch den Energieträger als OGP in der vierten Ebene finalisiert. [DIN03b] Neben der Prozessbeschreibung durch thermodynamische Charakteristika beinhaltet die Norm DIN ISO 857-2:2007-03 produktionsbezogene Eigenschaften des Reflow-Maschinenlötverfahrens. [DIN07]

Definition 1 (Löten mit deponiertem oder eingelegtem Lot [DIN07]):

Prozess, bei dem das Lot vor dem Erwärmen in der Fügezone angebracht und zusammen mit den zu lötenden Bauteilen (z. B. beim Weichlöten nur Reflowlöten) auf Löttemperatur gebracht wird

Um die wachsenden Anforderungen an Komplexität und Wirkungsweise von leistungselektronischen Produkten, wie höhere Packungsdichten, kleinere Bauelemente und vielfältigere Funktionalität, zu erfüllen, werden die Flachbauelemente auf der Leiterplattenoberfläche im Rahmen der Oberflächenmontagetechnologie befestigt. Neben der hohen Kosteneffizienz sind die guten thermischen, elektrischen sowie mechanischen Eigenschaften der stoffschlüssigen Materialverbindungen sowie die Prozessautomatisierbarkeit Hauptkriterien für die Anwendung des Wiederaufschmelzlötens in der Elektronikfertigung. Eine vollautomatisierte SMT-Fertigungslinie, siehe Abbildung 2.1, setzt sich aus verschiedenen Einzelprozessschritten und Handlingsmodulen zusammen. [Bel05]

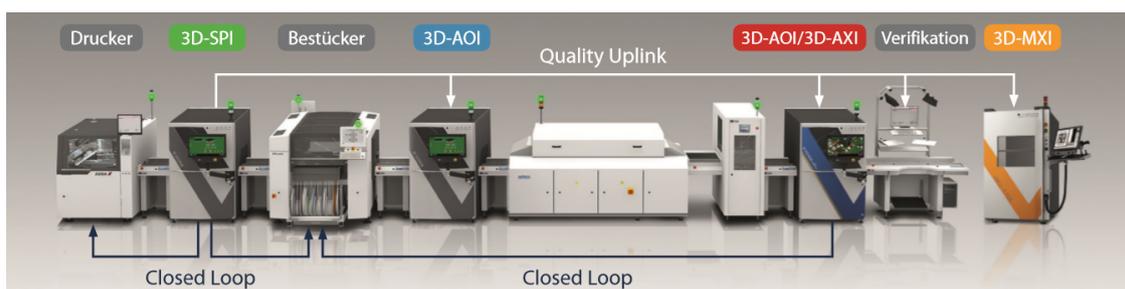


Abbildung 2.1.: Aufbaudchema SMT-Produktionslinie [Vis18]

Vorstellung gesamte SMT-Fertigungslinie Der gesamte Herstellungsprozess von Automotive-Produkten der Continental Automotive GmbH, beispielsweise Kombiinstrumente, moderne Langstrecken-Radarsensoren oder Head-up-Displays, wird in standortübergreifende Produktionssegmente unterteilt. Im Frontend-Segment erfolgt die nutzenbasierte Montage (Fügen durch Reflow-Löten) oberflächenmontierter Leiterplatten.

Die Einzelleiterplatten sind in diesem Produktionsschritt zu einem Verbund mehrerer Einheiten, dem Mehrfachnutzen, zusammengefasst. Der thematische Schwerpunkt dieser Forschungsarbeit liegt in der optischen Inspektion als Teil des Frontend-Segments. Die Einzelprozessschritte werden im Folgenden näher erläutert. Diese setzen sich aus dem Pastendruck, der Pastenkontrolle, den Bestückungsmodulen, dem Reflow-Maschinenlöten und der optischen Inspektion zusammen.

Zu Beginn der Prozesskette einer modernen, automatisierten SMT-Produktionslinie erfolgt der Lotpasten-Auftrag auf die Anschlussflächen der Leiterplatte im Rahmen des Schablonendrucks. Die hierfür verwendeten Metallschablonen weisen Öffnungen (Aperturen) auf, deren geometrische Abmessungen sowie Anordnung an das Leiterplattenlayout, die Anschlussflächen und die Löttaufgabe angepasst sind. Die viskose Lotpaste wird auf die Schablone aufgetragen. Unter Zuhilfenahme eines Rakels, welcher eine translatorische Bewegung entlang der Schablone ausführt, erfolgt das Füllen der Aperturen und somit die Benetzung des Substrats (Padflächen der Leiterplatte). Neben den Eigenschaften einschließlich des Layouts der Schablone beeinflussen sowohl die rheologischen und physikalischen Lotpastencharakteristika als auch die Druckparameter die Ergebnisse des Pastendrucks. Durch Auswahl geeigneter Leiterplatten-, Schablonenlayouts und Verfahrensparametrisierung kann die Gefahr einer Fehlerentstehung geschwächt und gegebenenfalls gänzlich beseitigt werden. Zu diesem Zweck gelten innerhalb der SMT-Fertigung von Continental firmeninterne Standards und Parametersollwerte, welche die effektive, reproduzierbare Realisierung IPC-konformer Leistungselektroniken ermöglichen. Dies beinhaltet eine präzise Abstimmung von Rakelkraft, Rakelwinkel sowie Druck-/Trenngeschwindigkeit unter Rücksichtnahme von Pasteneigenschaften (beispielsweise Scherrate und Korndurchmesser) und des Paddesigns. [Bel05]

Die Beurteilung des etablierten Pastendepots im Anschluss an den Druck erhöht die Ressourceneffizienz des Gesamtprozesses und erlaubt eine Markierung mangelhafter Bauteilbereiche. Zu diesem Zweck bestimmen In-Line-integrierte SPI-Systeme (englisch Solder Paste Inspection) in 3D-Methoden Volumen, Höhe, Position und Form gedruckter Pastendepots und sind somit in der Lage, Fehlplatzierungen bis hin zu Formabweichungen zu detektieren. Darüber hinaus ist eine Erkennung von Unregelmäßigkeiten innerhalb der Paste, wie Partikel, Kerben, Löcher und Ausfransungen möglich. [Bel05, GÖ21]

Im Anschluss an die erfolgreiche Lotpasteninspektion werden die Leiterplatten dem nächsten Verfahrensschritt, der Bestückung mit oberflächenmontierbaren, elektronischen Baugruppen, zugeführt. Die Aufnahme und Positionierung der Bauelemente erfolgt vollautomatisiert, wobei die SMD-Bauteile der Bauteilversorgungsstation in unterschiedlichen Transport- und Speicherformen, wie Blistergurte, Magazine oder Trays für ICs, übergeben werden. Blistergurte erfüllen die stets wachsenden Anforderungen nach einer kontinuierlichen, störungsfreien Bauteilzuführung und der Gewährleistung von konstanter,

verlässlicher Bauteilorientierung. [Vie21] Des Weiteren erfahren die einzeln verpackten SMD-Elemente Schutz vor Umwelteinflüssen und mechanischer Belastung bei Lagerung und Transport. Ein hoher Flexibilitätsgrad sowie eine Bestückungsleistung von bis zu 40.000 BE/h pro Bestückungsmodul wird in modernen Bestückungsautomaten durch schnelles Wechseln der Bestückprinzipien (Pick & Place, Collect & Place, MixedMode) erreicht. Optische Zentrierung, Leiterplattenfixierung, Verwendung von intelligenter Sensortechnik und Softwarelösungen unterstützen die Reduzierung der Lötfehlerrate und garantieren höchste Bestückqualität. [Her17, Fel13, Vie21]

Nach der exakten Bauteilpositionierung auf dem Lotpastendepot liegt eine temporäre stoffschlüssige Verbindung vor. Das Flussmittel und die viskosen Zusätze verleihen der Lotpaste adhäsive Eigenschaften, wodurch ein prozesssicheres Haften der Bauelemente auf der Leiterplatte realisiert wird. [Bel05] Da sowohl die Klebkraft als auch andere Verarbeitungseigenschaften einer Lotpaste in Abhängigkeit von der Verweildauer in normaler, sauerstoffhaltiger Atmosphäre abnehmen, wird ein effizienter, kontinuierlicher Produktfluss ohne Pufferzeiten mit unmittelbarer Weiterverarbeitung gefordert. [Bel05] Der Wiederaufschmelzprozess der Lotpaste und somit der eigentliche Lötvorgang findet in Reflowöfen, siehe Beispielsystem in Abbildung 2.2, im direkten Anschluss an die Bauteilbestückung statt. Auf Transportsystemen wie Ketten- oder Gitterförderbändern werden die Leiterplatten entlang der Heizstrecke des Reflowofens geführt und passieren dabei unterschiedliche Temperaturzonen, die unter Berücksichtigung der zeitlichen Komponente (Verfahrgeschwindigkeit, Verweildauer in Heizzonen) das Reflowprofil bilden. [Tec19, Fel13] Dieses wird durch unterschiedliche Temperatur- und Zeitintervalle charakterisiert. Eine Herausforderung in der Prozessparametrisierung liegt in der Bildung intermetallischer Phasen und Diffusionsschichten unter der Voraussetzung, dass die Liquidustemperatur der Lotlegierung ohne Aufschmelzung des Basismetalls der Bauteilanschlussfläche erreicht wird. [Bel05, Tec19, Fel13]



Abbildung 2.2.: Beispielsystem Reflow-Konvektionsofen [Reh22]

Der Wiederaufschmelzvorgang der Lotpaste wird im Rahmen des Reflowprofils haupt-

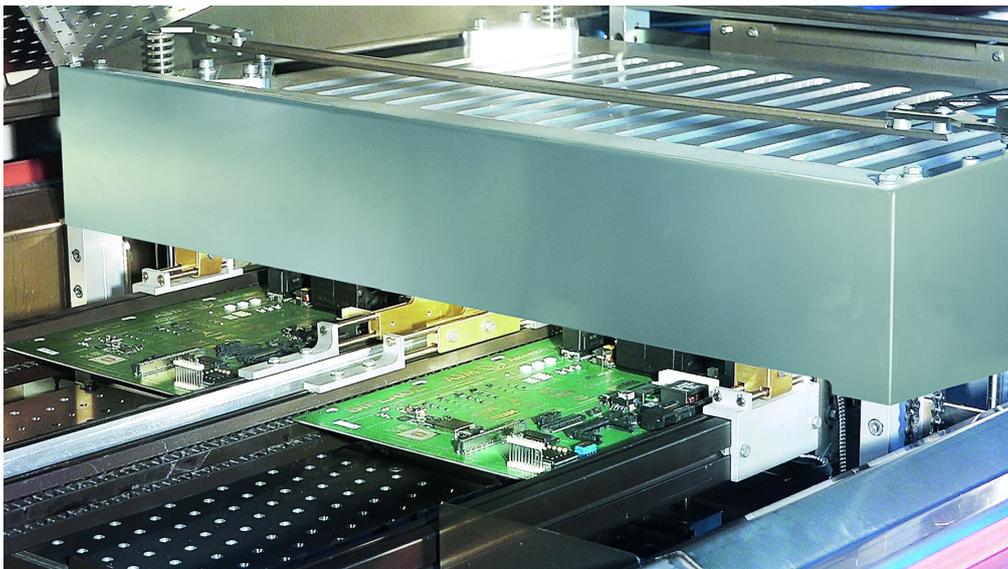


Abbildung 2.3.: Reflow-Lötanlage mit PCBs nach [Wil15]

sächlich durch zwei Kennwerte beschrieben: die Zeit über Liquidustemperatur und die Peaktemperatur. Das Umschmelzen aller Pastendepots, auch solcher in Kaltzonen, und somit die vollständige Benetzung der Bauteilmallisierungen stellen die Minimalanforderungen an das zeitliche Intervall dar. Jedoch ist die Ausprägung der Ablegerungseffekte durch zu lange Lötzeiten auf einem möglichst niedrigen Niveau zu halten. Die thermische Belastung der passiven und aktiven Bauelemente, insbesondere der integrierten Schaltungen, darf keinesfalls zu mechanischen oder funktionellen Schädigungen führen. [Bel05, Tec19, Fel13]

Qualitätssicherung in der SMT-Fertigung Die automatische Inspektion dient der optischen Qualitätsbeurteilung von leistungselektronischen Produkten und findet in der SMT-Fertigung in den meisten Anwendungsfällen nach dem Reflowprozess statt. Da es sich bei der automatischen Inspektion um ein berührungsloses Testverfahren handelt, ist dessen Einsatz in früheren Prozessschritten möglich, wie beispielsweise der Bare-Board-Inspektion, Lotpasteninspektion oder Überprüfung ungelöteter Paste vor dem Reflowprozess. [PCB21] In der automatischen SMT-Fertigung sind sämtliche Prüfsysteme (SPI, AOI, AXI) in der Lage, die geometrischen Informationen über Anzahl, Position, Drehlage sowie Bestückungsplan jeder Einzelleiterplatte zu erfassen und in ihren Fertigungsprozess einzubinden. Im Anschluss an die Erzeugung der stoffschlüssigen Verbindung und die optische Inspektion des Frontend-Bereiches werden die fehlerfreien Mehrfachnutzen dem Backend-Segment zugeführt. Dieses beinhaltet die Vereinzelung der Einzelleiterplatten aus dem Mehrfachnutzen sowie die Überprüfung der Funktionalität. Als Ergänzung zur optischen Inspektion im Rahmen der Frontend-Fertigung finden im Backend der In-Circuit-Test (Abkürzung ICT) und der elektrische Funktionstest (Abkürzung

FKT) Anwendung. Diese ermöglichen die Sicherstellung der Gesamt- und Teilfunktionen der Schaltung. [Alp20] Im Testsystem werden die Baugruppen durch Einspeisung von Testsignalen und die anschließenden Messung der Antwortsignale (Widerstands- und Kapazitätswerte, Entladeroutinen von Kondensatoren, Polaritätsprüfung oder Kurzschluss-tests) untersucht. Des Weiteren ist es möglich, durch Interpretation der Ausgangssignale Rückschlüsse auf Bestückung, Lötfehler oder Bauteilfehler zu ziehen. [Alp20] Danach erfolgt im Rahmen der Endmontage die Integration der Leiterplatte und weiterer Baugruppen, wie Sende-/Empfängerantennen bei Radarsensoren oder Abschirmkörpern, in die für das Produkt vorgesehenen Gehäuse.

2.1.1.2. Lötwerkstoffe und Leiterplatten-Material

Der Eintrag von umweltgefährdenden Stoffen in die Umwelt wird seit Verabschiedung der EU-Richtlinie 2002/95/EG (Abkürzung RoHs1; englisch Restriction of Hazardous Substances) im Januar 2003 und der Umsetzung in nationales deutsches Recht durch Festlegung von Grenzwerten in bestimmten Erzeugnissen minimiert. Diese Reglementierung beinhaltet das für die Elektronikfertigung in Lotpasten verwendeten Legierungselement Blei mit einem Grenzwert von 1000 ppm (0,1 %). Im Jahr 2013 wurde RoHs1 von der EU-Richtlinie 2011/65/EU (RoHs2) abgelöst, die wiederum im Jahr 2019 geändert wurde. Jedoch sind die im Bezug auf die Verwendung sowie Verarbeitung von bleihaltigen Werkstoffen festgelegten Grenzwerte unverändert. [VDE19, Kru05]

Des Weiteren trat in Deutschland 2003 die EU-Richtlinie 2002/96/EG (Abkürzung WEEE; englisch Waste of Electrical and Electronic Equipment) in Kraft, welche einen gesetzlichen Rahmen schafft, um die Recycling- und Rücknahmepflichten von Elektro- und Elektronikernzeugnissen zu definieren und somit eine Abfallreduzierung zu erreichen. Die Neufassung und aktuelle Version der WEEE-Richtlinie ist in 2012/19/EU zu finden. [Hes12]

Aufgrund abweichender Werkstoffeigenschaften hatte die Substitution bleihaltiger Lotpasten durch moderne bleifreie Lotlegierungen Anpassungen der Anlagentechnik und Fertigungsparameter zur Folge. Da bleifreie Legierungen in der Regel einen höheren Zinnanteil von bis zu 99 % aufweisen, erhöht sich der Schmelzpunkt der Legierungen und somit die für den Umschmelzvorgang benötigte Wärmemenge um 20 % auf bis zu 230 °C. Zusätzlich sinkt die Fließgeschwindigkeit, wodurch längere Lötzeiten und höhere Löttemperaturen/-gradienten benötigt werden. Dies hat eine größere thermische Belastung der SMD-Komponenten sowie eine schnellere Verflüchtigung des Flussmittels und eine erschwerte Entfernung der Oxide zur Folge. Das Anlösen der Bauteilmattisierungen und des Kupfers der Landflächen auf Leiterplatten setzt früher als bei bleihaltigem Lot ein mit der Folge eines schmaleren Prozessfensters. In der Praxis haben sich Typ3 SnAgCu-Lotpasten als zuverlässige Lötlegierung bewährt. [Bel05, STA21]

Die SMT-Produktpalette der Continental Automotive GmbH Karben reicht von zahlreichen Leistungselektroniken bestehend aus starren Leiterplatten mit unterschiedlichen Oberflächenbeschichtungen bis hin zur Bauteilmontage auf flexiblen Leiterplatten aus Polyimidfolie. Im Rahmen dieser Arbeit werden ausschließlich Multilayer-Leiterplatten mit einer Endoberfläche aus chemisch Zinn in der Bearbeitung der Aufgabenstellung untersucht, da unter deren Verwendung von wichtigen Vorteilen Gebrauch gemacht werden kann. Dies sind beispielsweise geringere Kosten der Rohleiterplatten, kürzere Lötprozesszeiten (Verringerung der thermischen Bauteilbelastung), hohe Planarität des Ausgangsmaterials (somit geringere Fehleranfälligkeit, bessere Rahmenbedingung für Pastendruck und Bauteilbestückung) und geringe Abscheidungsneigung auf organischen Schichten. [Bru09]

Die angewendeten Auswahlkriterien, die Eingrenzung des Forschungsfeldes und Bestimmung der betrachteten Prozesse, Bauteilfamilien sowie Prozessparameter zwecks Akquirierung von Fertigungsdaten, welche die geforderten Qualitätskriterien erfüllen, werden in Kapitel 3 erläutert.

2.1.1.3. Fehlerarten und Einflussfaktoren

In der modernen Produktion von Leistungselektroniken werden durch die Vorteile der Oberflächenmontagetechnologie trotz kontinuierlich steigender Produktkomplexität, Miniaturisierung bei gleichzeitig vielfältigerer Funktionalität geringe Fehlerraten erreicht. Die Implementierung moderner Fertigungsprozesse sowohl auf der Hardware- als auch auf der Softwareebene erlaubt eine Annäherung an das übergeordnete Ziel der Null-Fehler-Produktion. Bei auftretenden Defekten und Unregelmäßigkeiten in der Produktion ist es erforderlich, deren Ursachen zu identifizieren und zu beseitigen. Zu diesem Zweck kann die Vereinigung der produktionstechnischen Gegebenheiten mit theoretischen Lösungsansätzen eine Prozessverbesserung herbeiführen. [Bel05]

Vorstellung der geltenden Industriestandards Die von der Continental Automotive GmbH in der Produktion berücksichtigten Industriestandards in der Elektronikfertigung werden von der weltweit tätigen Handels- und Standardisierungsorganisation IPC – Association Connecting Electronics Industries seit 1995 festgelegt sowie veröffentlicht und finden weltweit in zahlreichen Elektronikunternehmen Anwendung. Die Standards für die Elektronikfertigung umfassen sowohl die Definition der Abnahmekriterien sowie die Festlegung verschiedener Klassen, nach denen die Baugruppen und Bauteilzustände bewertet werden. Die in der SMT-Produktion hergestellten Leistungselektroniken, das Reflowlötverfahren sowie Beschreibungen von Chipgehäusen sind Bestandteile der Richtlinie, welche in dem Spezifikationsdokument IPC A 610 aufgeführt wird. Dort wird

zwischen drei unterschiedlichen Bauteilzuständen unterschieden. [Ind17]

In der Produktion ist der Idealzustand der Lötung für die Gewährleistung der Zuverlässigkeit der Komponente in ihrer Einsatzumgebung nicht zwingend erforderlich. Eine Abnahmefähigkeit und somit eine zulässige Lötverbindung liegt vor, wenn trotz geringer Abweichungen vom Idealzustand eine zuverlässige Funktionsweise garantiert werden kann. Ist die Ausprägung dieser Abweichungen jedoch zu groß und infolgedessen die Einsatzfähigkeit nicht mehr sichergestellt, wird der Bauteilzustand als unzulässig deklariert. [Ind17]

Die Automotive-Produkte der Continental Automotive GmbH unterliegen der strengsten Gruppierung – Klasse 3 Hochleistungselektronik – des vorgestellten Anforderungskataloges.

Im Folgenden werden die Ergebnisse zweier unabhängiger Fehleranalysen der Produktionsdaten, welche die Identifikation der relativen und absoluten Häufigkeitsverteilung in der SMT-Produktion ermöglichten, vorgestellt. Nachgewiesene Echtfehler werden in Defektklassen unterteilt. Diese Defektklassen entsprechen der Klassenannotation der automatischen Inspektion und anschließender Verifikation. Der Untersuchungszeitraum erstreckt sich bei beiden Untersuchungsreihen über 6 Monate.

Analyse 1: Gesamte SMT-Produktion In die Betrachtung fließen die Daten der gesamten SMT-Produktion inklusive aller Bauformen, Prüfpläne und Bauteilbibliotheken ein. Darüber hinaus erfolgt keine Filterung nach dem Produktionstyp, sodass sich der Datensatz aus Informationen der Serienfertigung mitsamt durchgeführter Muster- und Neuanläufe zusammensetzt. Die Bauteilklassifikation entspricht den Ergebnissen der automatischen optischen Inspektion und der manuellen Überprüfung an der Verifikationsstation.

Der ausgewertete Datensatz berücksichtigt 2,361 Mio. Einzel-PCBs und 1,24 Mrd. Bauteile mit 5,73 Mrd. Fehlermöglichkeiten und wurde zwischen dem 19.05.2021 und dem 19.11.2021 erhoben. Aus Datenschutzgründen wird die absolute Fehleranzahl nicht angegeben, sondern der relative Anteil der einzelnen Fehlerklassen der gesamten Fehlerzahl in diesem Datensatz. Die relative Häufigkeitsverteilung stellt Abbildung 2.4 dar. Die Fehler *Versatz* und *Brücke* mit einem Häufigkeitsanteil von 15 % beziehungsweise 13 % treten am häufigsten bei allen identifizierten Defekten auf. *Fehlende Bauteile* werden in 12 % der Fälle beobachtet. Die Fehlerklasse der *Benetzungsstörung*, sowie *zu wenig/kein Lot* weisen Häufigkeitsanteile von 11 % beziehungsweise 10 % auf.

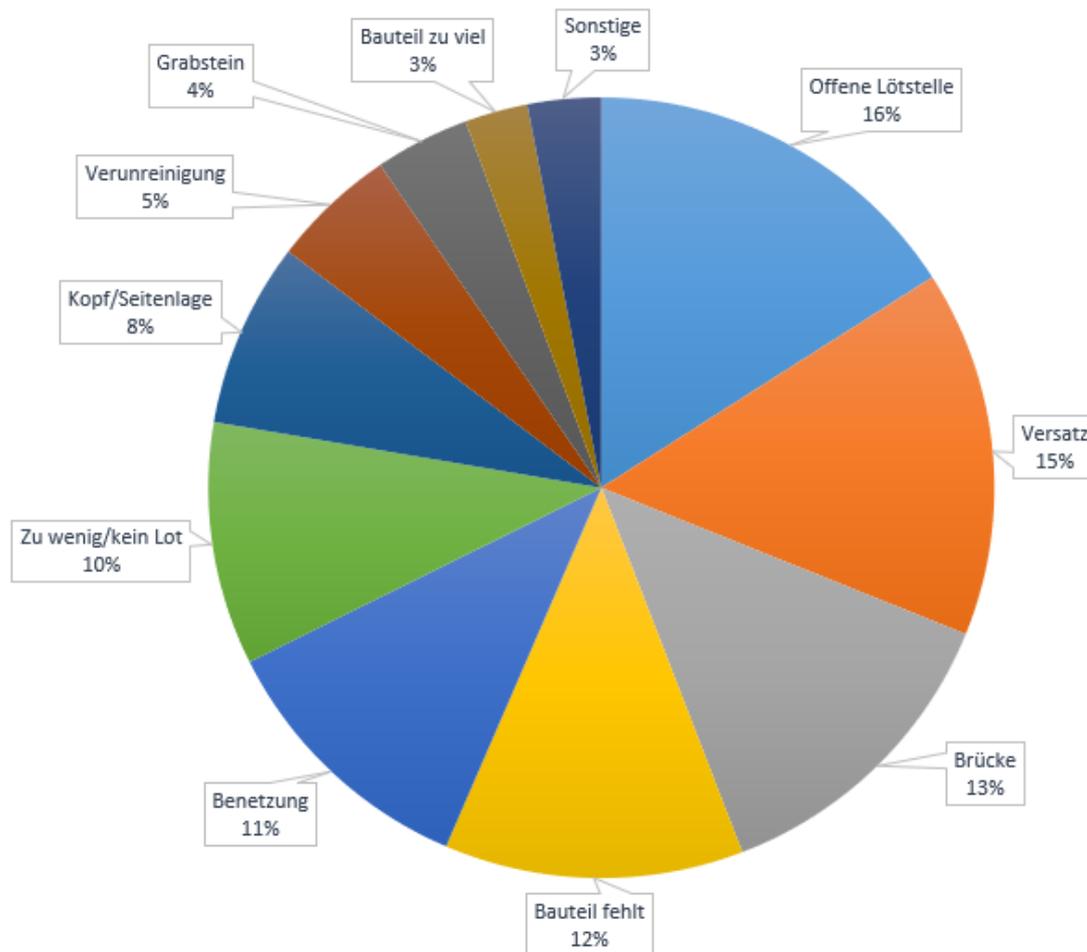


Abbildung 2.4.: Relative Fehlerverteilung Analyse 1: Gesamte SMT-Produktion

Analyse 2: Fünf Forschungs-Bauteile Eine eingeschränkte Untersuchung erfolgte in der zweiten Fehleranalysereihe. Gegenstand der Betrachtung sind fünf verschiedene Gehäuseformen von SMD-Bauteilen, vgl. Kapitel 3, die einer gemeinsamen Bauteilbibliothek entstammen. Dabei handelt es sich um vier Chip-Bauformen mit den Bezeichnungen C0402, C0603, R0402 und C0603 sowie der Transistor SOT23. Diese fünf Bauelemente sind Hauptbestandteil der Forschung dieser Arbeit. Analog zur Analyse 1 erfolgte die Datenakquise zwischen dem 19.05.2021 und dem 19.11.2021. Aufgrund des eingeschränkten Suchraums ergibt sich ein Datenumfang, der um den Faktor 6,3 reduziert ist. Abbildung 2.5 veranschaulicht die relative Fehlerverteilung für diesen Versuchsdurchgang.

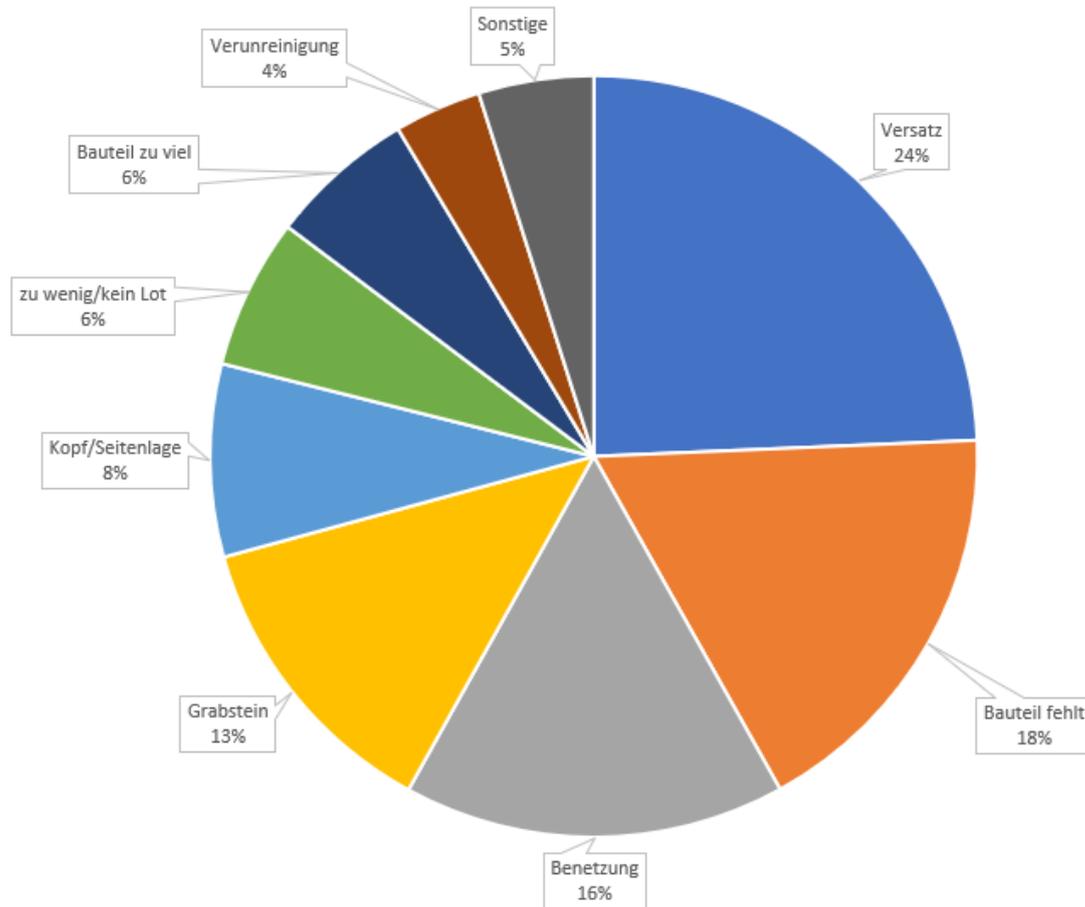


Abbildung 2.5.: Relative Fehlerverteilung Analyse 2: Fünf Haupt-BG SMT-Produktion

Auch hier ist der Fehler *Versatz* am häufigsten vertreten. In dieser Betrachtung sind Bauformen mit geringen Mitte-Mitte-Abständen der Lötkontakte ausgeschlossen, sodass die Fehlerklasse *Brücke* in der Verteilung nicht existiert. *Fehlende Bauteile* treten in 18 % der Beobachtungen auf, gefolgt von *Benetzungsstörungen* und dem für Chipbauteile typischen Defekt des *Grabsteins* (16 % und 13 %). Im folgenden Abschnitt werden die am häufigsten auftretenden Fehlerklassen und ihre Entstehungsursachen vorgestellt sowie die Einteilung in Echtfehler und Gutteil anhand des Industriestandards belegt.

Versatz Sitzt das Bauteil nicht in der vorgeschriebenen Toleranz auf den Pads, handelt es sich um den Defekt des Versatzes, Abbildung 2.6. [Ind17] Sowohl bei zweipoligen Gehäuseformen als auch bei solchen mit einer größeren Anzahl an Anschlüssen tritt dieser Defekt auf, sodass er in beiden Analysen vertreten ist. [Ind17]

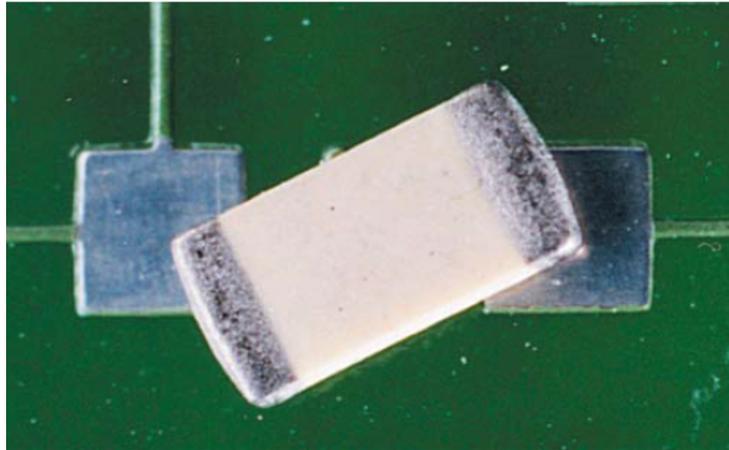


Abbildung 2.6.: Versetztes Bauteil nach IPC-A-610 G

Laut IPC-A-610 G ist für Klasse 3 der Hochleistungselektronik ein Seitenüberhang A , der kleiner/gleich 25 % der Bauteilanschlussbreite W oder Padflächenbreite P ist, hier ist die geringere Breite ausschlaggebend, zulässig, Abbildung 2.7. [Ind17]

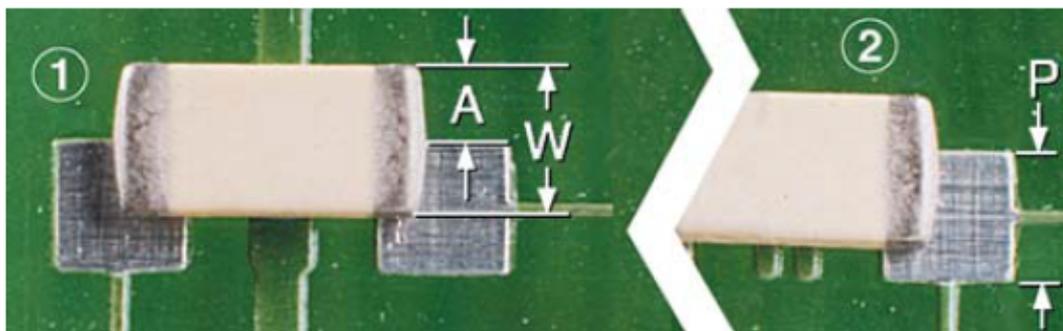


Abbildung 2.7.: Zulässiger Seitenüberhang nach IPC-A-610 G

Neben der relativen Position des Bauteils auf der Leiterplatte müssen ebenfalls Kennzahlen der Lötstellenbeschaffenheit festgelegte Anforderungen erfüllen. Hier darf beispielsweise die Breite der Lötstelle bei auftretendem Versatz nicht geringer als 50 % der Bauteilanschlussbreite W oder Padbreite P sein, siehe Abbildung 2.8. [Ind17]

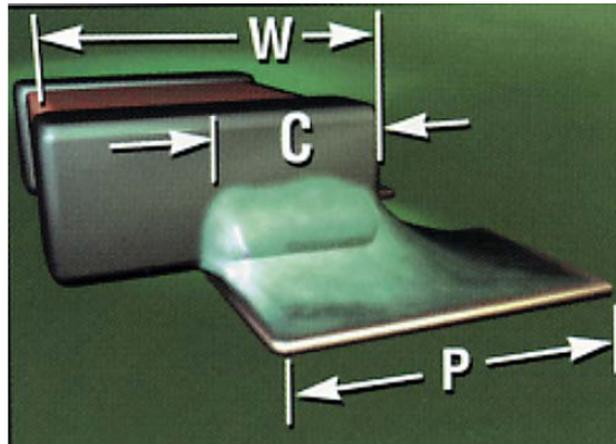


Abbildung 2.8.: Zulässige Lötstellenbreite nach IPC-A-610 G

Die Ursache für ein versetztes Bauelement lässt sich auf zwei grundlegende Prozessfehler zurückführen. Einerseits kann das Bauteil nach einem korrekten Bestückungsvorgang auf den Pads eine thermodynamische (während der Wiederaufschmelzphase beziehungsweise oberflächenspannungsinduziert) oder mechanische Krafteinwirkung erfahren. Andererseits wurde das Bauteil im Bestückungsvorgang nicht exakt in der vorgesehenen Position auf der Substratfläche platziert und zusätzlich die Selbstzentrierung im Ofen, auch als Selbstheilung betitelt, nicht erreicht. Aus diesen Gründen kann ebenfalls eine Positionsabweichung des Bauteils nach Vollendung des Fügeverfahrens auftreten. An diesem Vorgang sind einzelne oder gleichzeitig wirkende und sich beeinflussende Entstehungsmechanismen beteiligt. Ein ungeeignetes Layout mit zu großen Pads begünstigt einen Versatz ebenso wie schlechte Masseverteilungen. Erfolgt das Aufschmelzen des Lotes unregelmäßig und zeitlich versetzt, bewirkt der so entstehende Temperaturunterschied ein Verschieben der Bauteile. Ebenfalls ist ein Abgleiten von Bauteilen durch Unregelmäßigkeiten im Pastendruckprozess oder während der Bestückung möglich. Zu hohe Gasgeschwindigkeiten oder Schleppkräfte während der Gaseinleitung im Reflowofen sind ebenfalls mögliche Auslöser für eine Positionsabweichung des Bauteils. Zusätzlich fördern vibrierende Transportsysteme, zu hohe oder negative Beschleunigungen während des Bauteilhandlings die Entstehung und Ausprägung eines Bauteilversatzes. [Rah14]

Grabstein Der Grabsteineffekt tritt bei zweipoligen Bauteilen auf, wenn eine Lötseite keinen Kontakt zu dem Lotdepot aufweist und sich das Bauteil an der anderen Anschlusskante aufrichtet, Abbildung 2.9. [Rah14, Ind17] Verläuft die Benetzung der rechten und linken Anschlussfläche während des Reflowlötens zeitlich versetzt oder beginnt diese nicht synchron, entsteht ein Kraftgefälle. An dem zuerst aufschmelzenden Pastendepot bildet sich aufgrund der Benetzung und Oberflächenspannung ein Kraftvektor, welcher an der Außenseite der Anschlussfläche angreift, sodass eine Trennung des

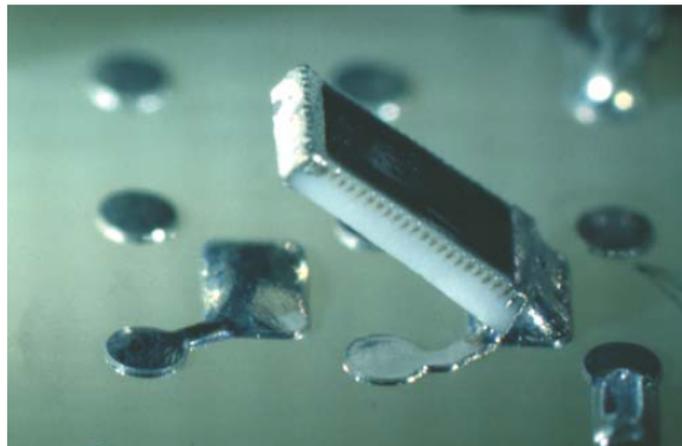


Abbildung 2.9.: Aufgerichtetes Bauteil (Grabstein) nach IPC-A-610

Bauteils und des Lotes der anderen Seite erfolgt. [Rah14] Die Fehlerbehebung bei auftretenden Grabsteineffekten wird durch die Vielfalt der Entstehungsursachen oftmals stark erschwert. Ein signifikanter Einfluss liegt in dem Versatz zwischen dem Pastendepot, der Bestückposition und der Padlage. Der Pastendruckversatz weist eine größere Einflussnahme als der Bestückversatz auf. Bei gegenläufigem Versatz erhöht sich die Anzahl an beobachteten Grabsteineffekten am stärksten. Die Versatzrichtung ist für die Anzahl und das Ausmaß der Aufrichtungen ebenfalls maßgebend. Während bei einem Querversatz nur ein geringer Anstieg in Versuchsreihen beobachtet werden konnte, bewirkt ein Längsversatz eine Erhöhung der Beobachtungen um den Faktor fünf. Weitere Ursachen liegen in der Verwendung von eutektischen Lotlegierungen, Löten unter Stickstoffatmosphäre und steilen Aufheizgradienten. [Rah14]

Benetzungsfehler Die Lötbarkeit eines Bauteils erlaubt eine Beurteilung über das Vermögen, eine stoffschlüssige Verbindung mit dem Lötpartner einzugehen, sodass die gestellten Anforderungen erfüllt werden. Hierfür wird der Begriff der Lötbarkeit in drei Kategorien unterteilt, welche von dem Werkstoff, der Fertigung und der Konstruktion abhängen. Die Benetzungsfähigkeit fällt unter die werkstoffabhängige Kategorie der Löt-eignung und beschreibt die physikalischen Eigenschaften. [DIN06]

Sowohl in der Literatur als auch in den Industriestandards stellt der Benetzungswinkel eines der wichtigsten Bewertungskriterien von Lötverbindungen dar. Der Benetzungswinkel entspricht dem Winkel zwischen der Tangente am Lotmeniskus und der Substratoberfläche (Padoberfläche). [Bel05] Einem kleineren Benetzungswinkel wird eine größere Benetzungsfähigkeit des Lotes gegenüber dem Fügepartner zugesprochen. Dieser Zusammenhang wird durch die folgende Young-Dupré-Gleichung erklärt, welche einen Zusammenhang zwischen den Spannungszuständen der Fügepartner in Abhängigkeit zum

Benetzungswinkel beschreibt. [Alm21, Bel05]

$$\sigma_{sv} = \sigma_{sl} + \sigma_{lv} \cdot \cos \Theta \quad (2.1)$$

Mit:

σ_{sv} : Oberflächenspannung des Festkörpers [mN/m] = Grenzflächenspannung zwischen Festkörper und Atmosphäre

σ_{sl} : Oberflächenspannung der Grenzfläche fest/flüssig [mN/m]

σ_{lv} : Oberflächenspannung der Flüssigkeit [mN/m] = Grenzflächenspannung zwischen Flüssigkeit und Atmosphäre

Θ : Randwinkel/Benetzungswinkel [$Grad$]

In [Ind17] wird ein Kontaktwinkel größer als 90° als unzulässig eingestuft. Weitere Abmessungskriterien für oberflächenmontierte Bauelemente entsprechen den geometrischen Abmessungen der Lötstelle in Bezug auf Bauteil- und Padflächenmaße (beispielsweise maximaler Seitenüberhang, maximale/minimale Länge beziehungsweise Höhe der Lötstelle). [Ind17]

Das Spektrum der Benetzungsstörungen weist eine Vielzahl von Erscheinungsformen und Entstehungsmechanismen auf. Eine *Entnetzung* liegt vor, wenn sich bei hohen Benetzungswinkeln nach dem Lötprozess Metalloberflächen ausbilden, die zwar benetzt wurden, jedoch keine Lötstelle formen. Durch die Entstehung von partiellen Lotanhäufungen werden die Lötqualität und Funktionalität der Lötung negativ beeinflusst. Mögliche Ursachen für Entnetzungsvorgänge sind Verunreinigungen der Grundmetallisierungen, unzulässige Lagerung, Verpackung oder mangelhafte Temperaturparameter/-verläufe während des Reflowlötens. Während bei der Entnetzung eine temporäre Verbindung zwischen dem Lot und der Anschlussmetallisierung, welche sich im Verlauf des Lötvorganges zurückbildet, zustande kommt, ist diese bei einer *Nicht-Benetzung* zu keinem Zeitpunkt des Reflowprozesses existent. Die Etablierung einer Diffusionsschicht wird somit vollständig verhindert. Eine vorliegende Nicht-Benetzung ist anhand hoher Benetzungswinkel über 90° und im Extremfall die Bildung einer kugelförmigen Lotanhäufung identifizierbar. [Rah14, Tec21]

Eine weitere Erscheinungsform der Benetzungsstörung ist die *schlechte Benetzung*. Diese häufig auftretende Defektart weist zwar die Entstehung einer Benetzung des Lotes mit der Metallisierung des Bauelements und der Anschlussfläche der Leiterplatte auf, jedoch können nicht alle Anforderungskriterien des Industriestandards erfüllt werden. Häufig bildet sich ein steiler Winkel des Lots zum Basismetall aus. Die Ursachenforschung bei auftretenden Benetzungsstörungen gestaltet sich aufgrund der Vielzahl der Einflussfaktoren als äußerst aufwendig. Zunächst ist die Lötbarkeit der Bauelemente und Leiterplatten im Hinblick auf die Rahmenbedingungen der Lötaufgabe zu überprüfen. Dies gilt auch für die

Wahl des Flussmittels. Die Lagerbedingungen sowie die thermische Beanspruchung der Bauelemente im Reflowofen üben ebenfalls Einfluss auf die Oberflächenbeschaffenheit, die Bildung von Verunreinigungen (beispielsweise Oxidierung) und chemische Reaktionsprozesse aus. [Rah14]

2.1.2. Automatische Inspektion in der SMT-Produktion

Die gesamte Produktpalette der Continental Automotive GmbH unterliegt höchsten Sicherheitsstandards sowie zwischen Kunde und Lieferanten definierten Abnahmekriterien. Eine kontinuierliche hohe Leistungsfähigkeit und -bereitstellung auf Abruf, auch unter ungewöhnlich rauen Einsatzumgebungen, ist unverzichtbar. Da ein Leistungsausfall in der Produkthanwendung innerhalb kritischer Einsatzgebiete lebensbedrohliche Folgen haben kann, ist dieser nicht zu tolerieren und muss durch die Produktentwicklung, den Herstellungsprozess und dessen Qualitätsprüfung ausgeschlossen werden. [Ind17, Wö12]

Zwecks Realisierung jener strengen Anforderungen ist die zuverlässige Identifikation fehlerhafter Bauelemente und Montage von hoher Bedeutung. Dies wird durch die Kombination von berührungslosen Inspektionsverfahren (AOI, AXI, SPI) und Prüfverfahren (ICT, FKT) mithilfe an das Produkt angepasster Prüfadapter erreicht.

Die optische Inspektion besitzt im Vergleich zu den kontaktherstellenden Testverfahren eine Reihe von Vorteilen. Optische und röntgentechnische Inspektionsverfahren sind nicht an die Vollendung des Montageprozesses gebunden, sodass die Integration und Positionierung der berührungslosen Systeme innerhalb der Fertigungslinie flexibel gestaltet und an die Produkt-/Herstellungsanforderungen angepasst werden kann. Die frühe Erkennung von Auffälligkeiten und Bauteildefekten erlaubt eine prozessnahe Aufdeckung von systematischen Prozessfehlern und negativen Einflussfaktoren. Defektarten, die im ICT- oder Funktionstest aufgrund ihrer Beschaffenheit und Charakteristika nicht detektierbar sind, werden durch eine optische Beurteilung aufgefangen. Eine Prüftiefe von 100 % und Abdeckung aller möglichen Fehlerarten wird durch die Abstimmung der Prüfstrategien aller am Prozess beteiligten Qualitätssicherungsprozesse erreicht. [Wö12]

2.1.2.1. Funktionsweise und Prozessablauf AOI

In dieser Arbeit werden ausschließlich optische Systeme der Viscom AG in die Forschung integriert. Die verwendeten Produktionsdaten entstammen der automatischen optischen Inspektion im Rahmen des Post-Reflow-Verfahrens. Direkt im Anschluss an den Wiederaufschmelzlötvorgang im Reflowofen erfolgt die Qualitätsbeurteilung des Produktes durch ein AOI-System. Auf eine genauen Angabe von in der Produktion der Continental Automotive GmbH etablierten Prüfeinstellungen wird aufgrund der Geheimhaltungsver-

einbarung verzichtet. Die Bewertung der Ergebnisse dieser Forschungsarbeit wird dadurch nicht beeinträchtigt.

Mechanische Eigenschaften AOI-System Die Systeme der automatischen optischen Inspektionstechnologie sind als Portalsysteme mit automatischen Bauteilbe- und -entladesystemen dank effizienter Implementierung von Kommunikationsschnittstellen ausgeführt. Der Mehrfachnutzen/Prüfling wird im AOI-System positioniert und durch eine pneumatische Leiterplattenklemmung fixiert. Zwecks Vermeidung von Krümmungserscheinungen des Leiterplattenverbands aufgrund großer PCB-Breiten sowie hoher Packungsdichten der Einzelleiterplatten und somit stärker auf die PCB wirkende Gewichtskräfte (maximaler Einfluss in der Mehrfachnutzenmitte) ist eine optionale Verwendung einer pneumatischen Leiterplatten-Mittenunterstützung möglich. [Wö12, Vis22c]

Bildaufnahme AOI-System Streng dem Prüfplan folgend, welcher speziell an den Bestückungsplan und die Eigenschaften eines Produkts und dessen Varianten angepasst wird, werden im bildgebenden AOI-Teilprozess optische Bildaufnahmen der Region-of-Interest (Abkürzung ROI) durch ein (auf der x- und y-Achse) translatorisches Mehrfachkameranystem akquiriert. Die simultane Nutzung von bis zu acht schrägen Kameras, einer orthogonalen (relativ zur Leiterplatten-/Kalibrationsebene positionierten) Kamera und ein durch den Prüfalgorithmus optimierter Verfahrensweg des Sensorkopfes ermöglichen die Durchführung eines effizienten In-Line-Prozesses, siehe Abbildung 2.10. Die oberflächenmontierten Bauteile samt ihrer Lötstellen werden innerhalb des Prüfplans in Prüffenster aufgeteilt und beinhalten anhand ihrer eindeutigen ID und Positionsbestimmungen die notwendigen Informationen für das bildgebende Aufnahmeverfahren. [Wö12, Vis22a] Aufgrund der Verwendung von leistungsfähigen Softwarekomponenten und -lösungen, wie modernen Framegrabbern und intelligenter Steuerungssoftware, im Zusammenspiel mit präzisen elektromechanischen Komponenten (Synchron-LinearMotoren als Verfahrenseinheit) werden hohe Inspektionsgeschwindigkeiten und optimale Prüfergebnisse generiert (mit einer Bilddatenrate von 3,6 Gigapixel pro Sekunde und Prüfgeschwindigkeiten bis zu 65 cm²/s). [Vis22a, Mar19]

Prozessablauf und Prüffamilie Die akquirierten Bildaufnahmen dienen als Eingangsgröße für den nächsten Teilprozess in der automatischen optischen Inspektion, die digitale Bildverarbeitung. Abhängig von den Metainformationen des zu untersuchenden Bildausschnittes erfolgt eine Zuordnung der SMD-Komponenten/Lötstellen zu den Analysealgorithmen und der damit verbundenen digitalen Bildverarbeitung. Diese sind in der Lage, eine Vielzahl von Löt-, Bestück- und Bedruckungsfehlern nach den Abnahmekriterien des geltenden Industriestandards zu detektieren.



Abbildung 2.10.: Schema AOI-Bildakquisition: Orthogonale und schräge Bildaufnahme von ROI [Mai18]

Zunächst teilen sich die Bildverarbeitungsmethoden in Prüffamilien, die sich in ihren Haupteinsatzgebieten voneinander unterscheiden, auf. Innerhalb einer Prüffamilie erfolgt erneut eine Untergliederung in unterschiedliche Algorithmen. Eine tabellarische Auflistung der Prüffamilienbezeichnungen und Beschreibungen ihrer Anwendungsbereiche erfolgt in Tabelle 2.1. [Vis20]

Tabelle 2.1.: Viscom-SI AOI-Prüffamilien und Einsatzgebiet – alphabetische Sortierung [Vis20]

Bezeichnung	Beschreibung
BOND	Prüfung von Bond-Drähten, ist nur mit spezieller Kamerasensorik möglich
BRDG	Lötbrückenprüfung mit der orthogonalen Kamera
GENR	Allgemeine Funktionen zur Analyse von Referenzmarken, Bauteilanwesenheit, Bauteilecken
LAND	Überprüfung der Pads (Landings) auf Anwesenheit von Paste, Zinn
MENI	Lötmeniskusprüfung mit der orthogonalen Kamera
POSC	Positionsprüfung von Bauteilen mit der orthogonalen Kamera
XRAY	Röntgenprüfung, ist nur mit spezieller Kamerasensorik möglich

Ein wesentlicher Bestandteil der AOI-Prüfstrategie ist die Aktivierung und Parametrierung von Analyseverfahren als Bestandteil der oben genannten Prüffamilien für jedes

Prüffenster der auf dem Produkt montierten SMD-Komponentenbauform. Das Prüffenster deckt hierbei das gesamte Bauteil oder lediglich einzelne Lötstellen ab, sodass für jede Bauform mehrere aktive Prüffenster existieren. [Vis20]

Ergebnis- und Qualitätsbeurteilung Der Berechnung von Schlüsselkennwerten (Abkürzung KPI; englisch Key Performance Index) zwecks Überwachung der Fertigungseffektivität und -qualität liegen somit mehrere Ausfallmöglichkeiten (abhängig von der Anzahl Prüffenster) pro Bauteil zugrunde. In den Inspektionsverfahren der SMT-Fertigung (SPI, AOI, AXI) werden fcpmo- und dpmo-Kennzahlen (englisch False Calls per Million Opportunities und Defects per Million Opportunities) sowie der First-Pass-Yield als Vertreter der produktionsbezogenen KPI-Kennwerte zur Qualitätsbeurteilung verwendet.

Um den Anforderungen an Prüftiefe und untersuchte Fehlermerkmale der geltenden Industriestandards gerecht zu werden, ist es möglich, eine Vielzahl von Analysealgorithmen je nach Anwendungsfall für die Bauformen und deren Prüffenster innerhalb des AOI-Prozesses zu vereinen. Neben der unabhängigen sequentiellen Durchführung und Klassifikation der Bauteil-/Lötstellenzustände stellen Linkfunktionen die Möglichkeit dar, die Prüfergebnisse einzelner Algorithmen und Prüffenster innerhalb eines gleichen Prüfmusters zu kombinieren. Beispielsweise ist es möglich, den in einer MENI-Prüfung gemessenen Bauteilversatz mithilfe des Linkbefehls `si_link_posm` an eine folgende LAND-Prüfung zu transferieren und die Position der Prüffenster anhand dieser Informationen zu aktualisieren. Ein für diese Forschungsarbeit wesentlicher Linkbefehl bewirkt die Weitergabe von Bildinformationen an nachfolgende Analysealgorithmen. Im Zuge dessen werden die Erweiterung des Analysealgorithmus durch die Bereitstellung von bis zu acht Eingangsbildern und somit Verbesserungen der Klassifikationsergebnisse durch mannigfaltige Subprüfungsmöglichkeiten ermöglicht. [Vis20]

Ungeachtet der unterschiedlichen Fehlerabdeckung der Prüffamilien und ihrer Subalgorithmen folgen diese einer allgemeinen gemeinsamen Struktur, welche die Definition der Prüfstrategie, die Prüfplanerstellung sowie die Überwachung des AOI-Prozesses im Fertigungsbetrieb vereinfacht. [Vis20]

In der Ergebnisdarstellung sind die gemessenen Featurewerte der Prüfung, die festgelegten Grenzwerte (für die regelbasierte Klassifikation), linguistische Beschreibungen der Messgrößen und die Klassifikationsergebnisse enthalten, siehe Abbildung 2.11. [Vis20]

2.1.2.2. Analysealgorithmen AOI

Im Folgenden werden die Grundstruktur und Funktionsweise von Analysealgorithmen des AOI-Prozesses erläutert.

Feat	Value	Thr	Description
1	231		Gesamtmatchwert
2	227		linker Matchwert
3	235		rechter Matchwert
5	117		X-Verschiebung
6	-117		Y-Verschiebung
7	220		Kontrast
8	227		kleinster seitlicher Matchwert
24	2		Klassifikation MENI

Result: good

Abbildung 2.11.: SI Debugbetrieb: Ergebnisdarstellung; Grenzwerte geschwärzt [Vis20]

Prüfstruktur Eine erhebliche Effizienz- und Qualitätssteigerung optischer und röntgentechnischer Inspektionsverfahren (AOI und AXI) konnte durch die Anwendung von globalen Prüfbibliotheken erreicht werden. Zu diesem Zweck werden Produkte, welche identische Leiterplattenmaterialien, Padgeometrien und somit Reflexionseigenschaften durch ihre Farbgestaltung (beispielsweise Lötstopplack) und Werkstoffcharakteristika aufweisen, in Gruppen zusammengefasst. Jede dieser Gruppen besitzt eine individuelle Bauteilbibliothek, in der für jede SMD-Gehäuseform und Prüfmuster, speziell angepasste Prüfstrategien hinterlegt sind. Eine Prüfstrategie ist eine Kombination einzelner Prüfalgorithmen, die durch die Weitergabe von Informationen miteinander verknüpft sind oder unabhängig voneinander ausgeführt werden. In den Prüfplänen eines jeden Produktes werden durch Pfadangaben zu der entsprechenden globalen Bauteilbibliothek Zuordnungen der oberflächenmontierten Bauteile und den standardisierten Prüfstrategien erstellt. Dies ermöglicht eine linienübergreifende Vereinheitlichung der Inspektionsverfahren. Weisen Produkte aufgrund ihres Layouts oder Bestückungsplans Auffälligkeiten auf, wie beispielsweise Abschattungen durch große Bauteile in der optischen Prüfung, ist es möglich, eine individuelle Ergänzung beziehungsweise Änderung des Prüfplans global oder lokal auf dem Einzel-AOI-System vorzunehmen. [Vis20]

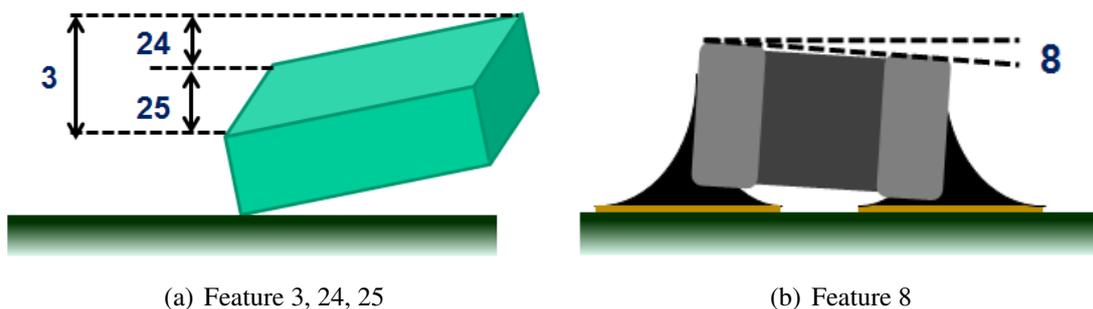
3D-Analyse Die Etablierung der modernen 3D-AOI-Technologie zur Anwesenheits- und Positionskontrolle begünstigt die Qualität der Klassifikationsergebnisse des gesamten Inspektionsprozesses maßgeblich. Unter Zuhilfenahme des Phasenverschiebeprozesses sowie der Implementierung eines Streifenprojektors in das AOI-System wird die Generierung von Höheninformationen der Oberflächen geprüfter SMD-Komponenten ermöglicht, siehe Abbildung 2.12. In dem Ergebnisbild entspricht jeder Pixelgrauwert einem



Abbildung 2.12.: Schema AOI-Streifenprojektion – vier geneigte und eine orthogonale Kamera [Vis22c]

konstruierten Höhenwert bezüglich einer ermittelten Referenzebene. Höhe, Koplanarität und Ausrichtung des Bauteils werden somit in der mehrschrittigen Analyse messbar. Das dreidimensionale Höhengrauwertbild wird zunächst Bildverarbeitungsfiltern unterzogen. Dadurch werden störungsfreie, geglättete Aufnahme-Bilder erzeugt. Die Bestimmung der Referenzebene ermittelt im Anschluss an einen Signaloptimierungs-Prozess die x -/ y -/ z -Ausrichtung des Bauteils.

Eine schematische Darstellung der Featurewerte 3, 24 und 25 des 3D-Analysealgorithmus ist in Abbildung 2.13a sowie Featurewert 8 in Abbildung 2.13b aufgeführt. Feature 24 und 25 geben den Höhenunterschied an der langen beziehungsweise kurzen Bauteilseite an. Der gesamte Höhenunterschied wird durch Feature 3 und die vertikale Bauteilneigung durch Feature 8 wiedergegeben. [Vis20]



(a) Feature 3, 24, 25

(b) Feature 8

Abbildung 2.13.: Gemessene Eigenschaften Planaritätsprüfung 3D-Analyse [Vis20]

2.1.2.3. Verifikation AOI-Ergebnis

Verletzungen der regelbasierten Klassifikation in der automatischen optischen Inspektion sind nicht automatisch als Bauteil- oder Lötstellenfehler einzustufen. Die finale Entschei-

dung über den Bauteilzustand wird manuell an einem Verifikationsplatz, der direkt im Anschluss an das AOI-System in der Fertigungslinie implementiert ist, getroffen.

Prozessbeschreibung Verifikation Die Metainformationen des ausgefallenen Prüfensters dienen während der automatischen Inspektion der Akquisition von weiteren Bildern der ROI, in denen das betroffene Bauteil beziehungsweise die Lötstelle auf dem Prüfling montiert sind. Diese Nachlaufbilder stellen die ROI aus unterschiedlichen Aufnahmewinkeln (in der Regel acht geneigte und eine orthogonale Ansicht) dar. Diese werden der Bedienungssoftware des Reparaturplatzes HARAN (Abkürzung Halbautomatische Reparaturplatzanlage) samt der von der AOI-Prüfung generierten Ergebnis-Dateien zur Verfügung gestellt. Weitere Informationen, welche dem Bediener der Verifikationsstation bereitgestellt werden, sind der Barcode des Prüflings und die Position der aktuell betrachteten ROI auf der Leiterplatte. [Vis20] Die manuelle Einordnung des Bauteilzustandes erfolgt durch Beurteilung der Bildaufnahmen. Der Bauteilzustand wird in zwei mögliche Kategorien eingeteilt: Pseudofehler (Bauteilzustand akzeptabel) und Echtfehler. Im Falle eines Echtfehlers wird die vorliegende Fehlerklasse durch Auswahl aus einer vordefinierten Defektliste ausgewählt und dem Klassifikationsergebnis dadurch ein Fehlercode zugewiesen. Die Entscheidung des Bedieners und die aktualisierten Metainformationen werden in den Prozessdateien gespeichert und umgehend an das Datenmanagementsystem transferiert. [Vis20]

Integrierte Verifikation Ein weiterer Vorteil des Verifikationsprozesses und der Erhebung der daraus resultierenden Klassifikationsdateien liegt in der Realisierungsmöglichkeit einer integrierten Verifikation. Der Aufbau einer Bilddatenbasis mit korrespondierenden Metainformationen der Prüfung und Verifikation ermöglicht Offline-Optimierungen der Prüfpläne (Parametrierung und Schwellwerte). Zu diesem Zweck wird eine Simulationssoftware, welche einer exakten Abbildung des realen AOI-Prozesses samt der Prüfsoftware entspricht, auf einem lokalen Computer etabliert. Mithilfe dieser Nachbildung eines AOI-Systems der Serienfertigung werden, ohne dabei Prozessstörungen in der Produktion zu verursachen, neue Prüfpläne erstellt oder bestehende angepasst. Als Validierungsdatensatz dienen die akquirierten Bilddaten und Ergebnisdateien der Fertigung. Eine manuelle Änderung der Klassenlabel zwecks Beseitigung von Klassifikationsfehlern ist ebenfalls möglich. [Vis22b, Vis20]

2.1.2.4. Einflussanalyse AOI-Prozess

Eine Untersuchung der Einflussfaktoren des AOI-Prozesses ermöglicht die Eingrenzung des Forschungsfeldes und die Definition der Aufgabenstellung. Daraus lassen sich Forschungsfragen ableiten und im Anschluss ein Forschungsplan generieren.

Ziel der Einflussanalyse ist die Generierung einer Systemantwort, welche eine zuverlässige Verwendbarkeit und Gültigkeit der Daten gewährleistet. Zu diesem Zweck ist eine präzise Aufdeckung der Eingangssignale, Störfaktoren und Kontrollfaktoren hinsichtlich des AOI-Prozesses unerlässlich. Erst durch Kenntnisnahme der Wechselbeziehungen dieser Einflussgrößen und des gesamten Systemverhaltens werden Ausgangssignale geschaffen, die bestimmte Kriterien erfüllen und eine Informationsweiterverarbeitung im Rahmen dieser Forschungsarbeit ermöglichen. Eine graphische Beschreibung dieser Zusammenhänge erleichtert die Entwicklungsarbeit und bietet eine Übersicht des Prozesses und dessen Eingriffs- und Steuerungsmöglichkeiten. Die Anfertigung und Interpretation eines Parameterdiagramms eignet sich für dieses Anforderungsprofil, siehe Abbildung 2.14. [Fri06]

Die multivariate Systemfunktion $f(x)$ beinhaltet eine Verarbeitung des Signalinputs M_i , der Lenkungsgrößen C_i und der Störgrößen N_i . Auf eine detaillierte Aufzählung sämtlicher Einflussfaktoren wird verzichtet und stattdessen ein Auszug in Tabelle 2.2 dargestellt.

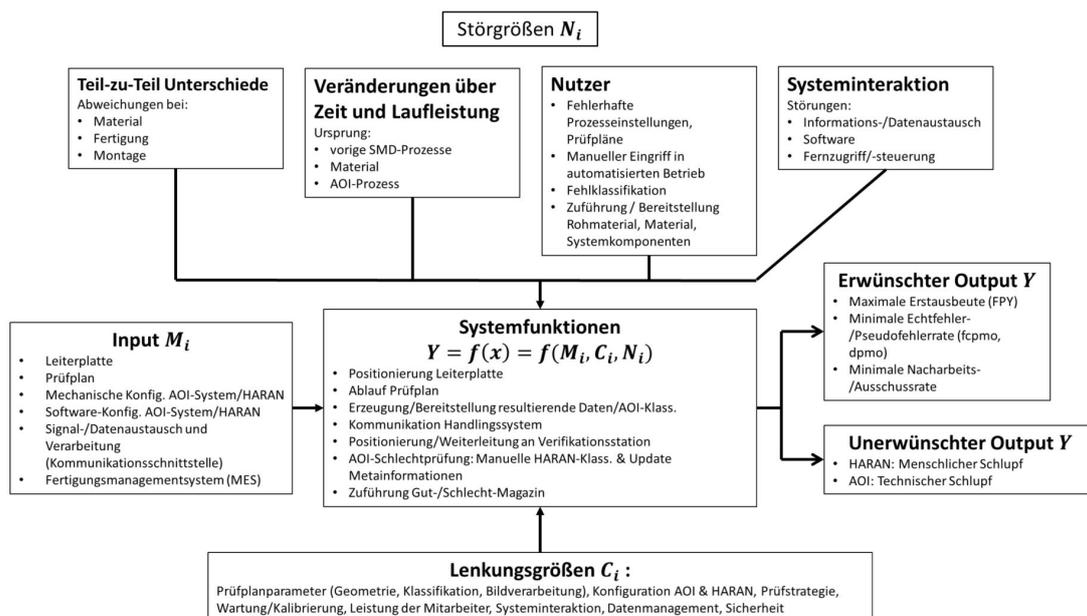


Abbildung 2.14.: Parameterdiagramm AOI-Prozess
Parameterdiagramm AOI-Prozess nach [Fri06]

Tabelle 2.2.: Auszug Variablen Parameterdiagramm AOI-Prozess

Variable	Einflussfaktor	Beispiel
Input	Leiterplatte	Bauteilbibliothek, Abmessungen, Sichtbarkeit, Material Basisplatte, Lötmaterial, Beschichtung, Layout (Packungsdichte, Abschattungen durch Bauteile, Prüfbarkeit)
	Prüfplan	2D/3D, Variante, Bauteilbibliothek, lokale PP (Ausnahmen, Alternativen), Verlinkungen mehrerer Analysen, Parameter-/Bildübergabe, mehrstufige Analysen, Prüfumfang, Prüfparameter (Geometrie, Klassifikation, Bildverarbeitung, Konfiguration)
Lenkungsgröße	Prüfplan	Geometrie: Größe, Position, Anzahl Analysefenster, Gradientenposition, Meniskus, Gap; Klassifikation: Matchschwelle, Klassifikationsschwelle, max. zulässige Grenzwerte, Klassifizierungsrichtung; Bildverarbeitung: Binarisierungen, Elimination, Schrittweiten bei mehrstufigen Analysen, Skalierung, Filterungsmethoden, Mediangröße
	Datenmanagement	Erfassung/Speicherung Produktionsdaten, statistische Auswertungen zwecks Identifikation von produkt-/system-/fehlerbezogenen Abweichungen im Produktionsprozess, Barcodesteuerung, Berechnung KPIs, Produktionsüberwachung in Echtzeit oder zeitversetzt durch Datenmanagement
Störgröße	Teil-zu-Teil-Unterschied	Material: Rohmaterial-/Chargenstreuung, Komponenten, Oxidation, Verunreinigung; Montage: Fehlerhaftes Bestückprogramm, Schwankungen Bestückungsparameter
	Nutzer	Fehlklassifikation: Ermüdung, Stress, fehlende Interpretierbarkeit optischer Bilder; Parametrierung: Fehler bei Erstellung Prüfpläne (Konzentration, falsche Daten), erschwerte Umsetzbarkeit aufgrund Layout/Bestückungsplan

2.2. Informationstechnischer Bezugsrahmen

Der informationstechnische Bezugsrahmen beinhaltet die Definition des maschinellen Lernens, welches in dieser Forschungsarbeit Anwendung findet. Des Weiteren werden die Unterteilungsebenen des maschinellen Lernens erörtert sowie Algorithmen unterschiedlicher Anwendungsbereiche vorgestellt.

2.2.1. Maschinelles Lernen

Das maschinelle Lernen (Abkürzung ML; englisch Machine-Learning) hat sich in den letzten Jahrzehnten als Teildisziplin des Forschungsgebietes der künstlichen Intelligenz herausgebildet. Dieses befasst sich mit dem Erlernen von Erkenntnissen und Regeln aus gegebenen Datenstrukturen mit dem Ziel, Vorhersagen zu treffen sowie vorhandene Strukturen beziehungsweise Wechselbeziehungen aufzudecken. Zu diesem Zweck werden seitdem zahlreiche selbstlernende Algorithmen entwickelt und stets optimiert. Folgende Definition des Begriffs Machine-Learning des Informatikers Tom Mitchell fasst die grundlegende Motivation und Funktionsweise dieser Methoden zusammen. [Mir18]

Definition 2 (Machine-Learning: Tom Mitchell [Mit97]):

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

Die Generierung eines Modells, welches Zusammenhänge und Situationen der realen Welt auf eine technisch-mathematische Ebene abstrahiert, ist eng mit dessen kontinuierlicher Bewertung und Überwachung unter Zuhilfenahme speziell auf für diese Aufgabe entwickelter Metriken verknüpft. Nur im Falle einer maximalen Anpassung an die Grundwahrheit ist eine Anwendbarkeit sowie Gültigkeit dieses theoretischen Konstrukts gegeben. Zu Beginn jeder Modellierung erfolgt stets die Festlegung der Zielsetzung, welche durch das Modell erreicht werden soll. Abhängig von der Zielsetzung ist die anschließende Definition der durch die Konzeptionierung abstrahierten Eigenschaften. Die nachstehende Aufzählung stellt mögliche Anforderungen an Modelle dar. [Mir18, Dau11]

- *Funktionalität:* Die Modelle sind in der Lage, bestimmte Funktionen des realen Prozesses/Objektes auszuführen und zu erfüllen.
- *Simulation:* Die Komplexität/Handhabbarkeit/Zugriffsmöglichkeit der Realität erschwert eine Untersuchung, sodass mithilfe von Modellen Operationen durchgeführt und getestet werden.
- *Erklärung:* Datenstrukturen, Phänomene, Wechselbeziehungen und Verhaltensweisen werden durch das Objekt erklärt.

- *Vorhersage*: Die Vorhersage ist essentiell für Supervised Machine-Learning. Klassifikationen und zukünftiges Verhalten des Originals werden generiert.

2.2.1.1. Überwachtes Lernen

Folgende Definition des überwachten Lernens (englisch Supervised-Learning) fasst dessen Motivation präzise zusammen.

Definition 3 (Supervised-Learning: Sotiris Kotsiantis [Kot07]):

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances.

Das Supervised-Learning bildet mit dem Unsupervised-Learning und dem Reinforcement-Learning die drei Hauptgebiete des maschinellen Lernens. Diese unterscheiden sich in den Anforderungen an die Datenstruktur, den Informationsgehalt sowie die Funktionsweise und Ausgangssignale des trainierten Machine-Learning-Modells. Abbildung 2.15 stellt eine Übersicht der Teildisziplinen samt ihrer Einstufungskriterien und Zielsetzungen dar. Die Implementierung von Machine-Learning-Modellen in reale Prozesse hängt wesentlich von der korrekten Methodenauswahl sowie der Datenakquisition ab. [Mir18, Kot07] In dieser Forschungsarbeit stehen Algorithmen des überwachten Lernens, insbesondere

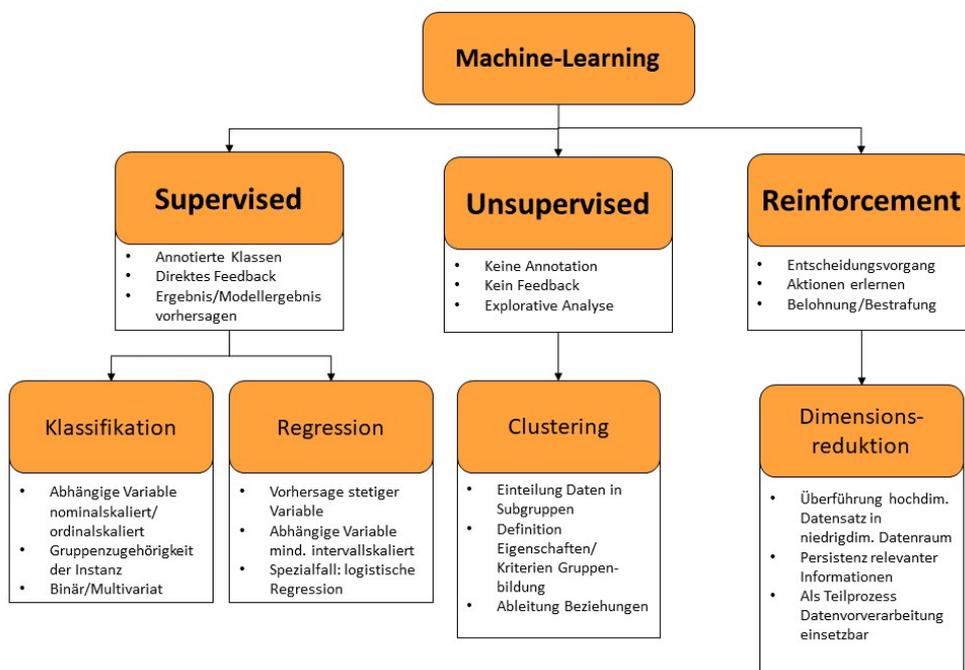


Abbildung 2.15.: Übersicht Machine-Learning-Verfahren (eigene Erstellung in Anlehnung an [Van09, Mir18, Fay17, Kot07])

die Verwendung von Klassifikationsverfahren, im Vordergrund. Zwar unterscheiden sich

die vielzähligen Methoden des überwachten Lernens in ihrer mathematischen Modellbildung und der Funktionsweise des erstellten Modells voneinander, jedoch vereint diese die Generierung einer Funktion, welche die Eingangssignale in gewünschte Ausgangssignale durch Approximation/Nachbildung des Verhaltens des realen Objektes transferiert. [Nas17] Der Prozess der Realisierung von Machine-Learning-Anwendungen ist in Abbildung 2.16 in Form eines Programmablaufplans, in Anlehnung an [Kot07], dargestellt. Dieser beinhaltet die Datenerhebung, Modellplanung und Implementierung des Modells.

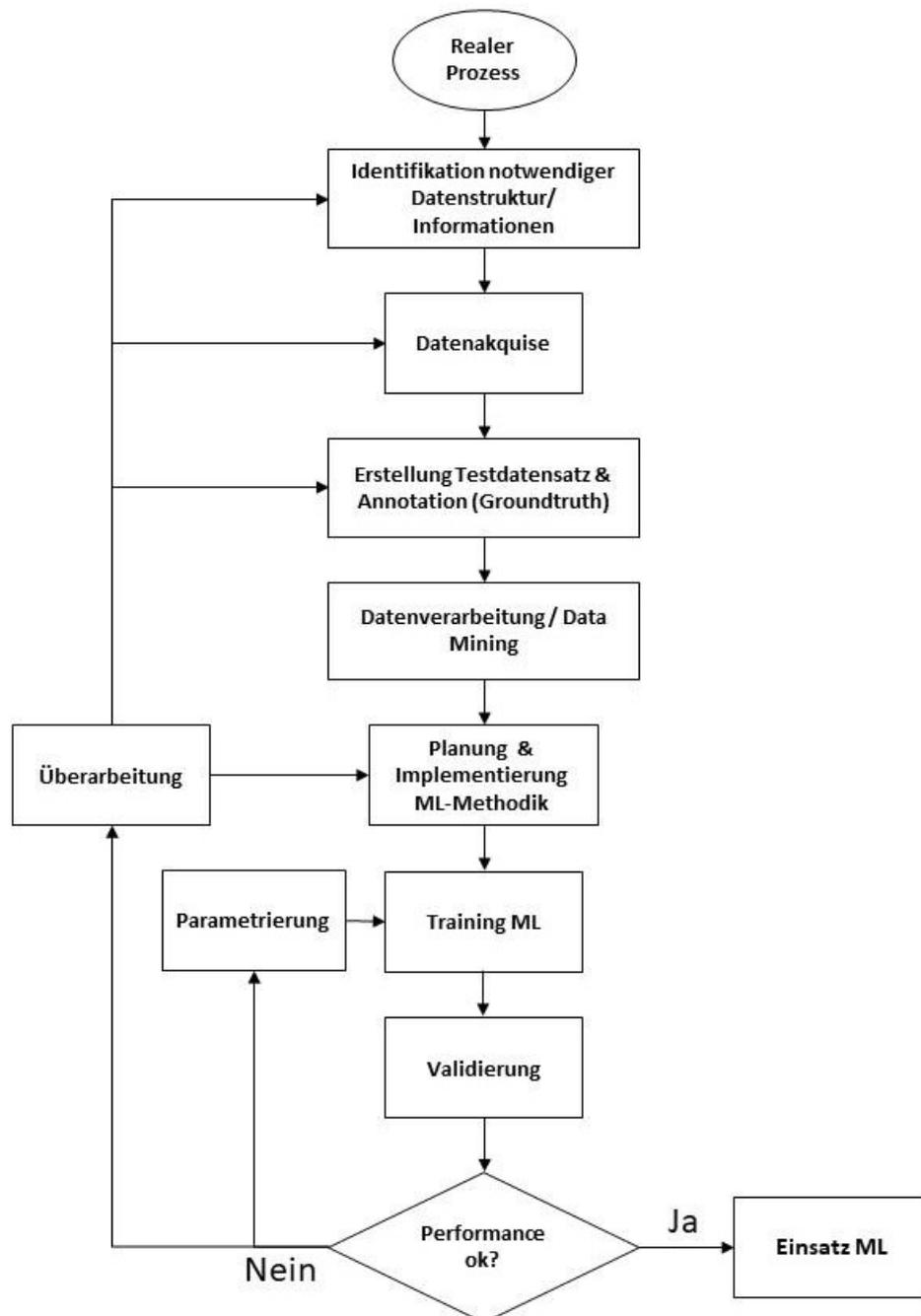


Abbildung 2.16.: Prozessablaufplan Supervised Machine-Learning (eigene Erstellung in Anlehnung an [Kot07])

2.2.1.2. Vorstellung Algorithmen des überwachten Lernens

Entscheidungsbaum Die Klassifikationsmethodik des Entscheidungsbaums (Abkürzung DT; englisch Decision Tree) beruht auf der Anwendung einer rekursiven Partitionsfunktion des Instanzraumes der Trainingsmenge. Entscheidungsbäume haben den Vorteil, dass sie einen hohen Grad an Modellinterpretierbarkeit zulassen. Somit räumen sie dem Anwender die Möglichkeit ein, Zusammenhänge und Informationen der Eingangsdaten zu identifizieren. [Mai05, Mir18]

Die Aufbaustruktur von Entscheidungsbaummodellen ähnelt der von Baumstrukturen in der Informatik. Diese setzen sich aus Knoten, welche einen verwurzelten Baum bilden, zusammen. [Mai05] Die Wurzel dieses Baums entspricht der ersten Ebene dieser Struktur. Ausgehend von der Wurzel entsteht ein gerichteter Baum, dessen Knoten jeweils nur einen Eingang aufweisen. Jeder Knoten der Baumstruktur teilt den Instanzenraum in mindestens zwei (entspricht binärem DT) oder mehr Subräume mittels einer definierten Funktion auf. In den meisten Anwendungsfällen erfolgt jede Unterteilung an dem jeweiligen Knoten anhand eines einzelnen Attributes des Trainingsdatensatzes univariat. Bei numerischen Attributen geschieht die Aufteilung durch Intervallbildung. Jedem daraus resultierenden Blatt wird eine Klasse beziehungsweise ein Wahrscheinlichkeitswert der Klassenzugehörigkeit zugeordnet. Die Voraussetzungen für eine Klassifizierungsmöglichkeit des Datensatzes werden durch Bildung eines Pfades, beginnend bei der Baumwurzel bis zum letzten Knoten/Blatt, formuliert. [Mai05]

Für die univariate Aufteilung der Instanzen und die Identifikation des informativsten Merkmals ist die Bildung einer Zielfunktion, im Rahmen eines Optimierungsproblems, erforderlich. Hierfür eignet sich die Maximierung des Informationsgewinns an jedem Knoten, sodass eine Definition eines Unreinheitsmaßes (englisch Impurity) gefordert wird. Die Unreinheitsmessung entspricht einer Funktion $\phi : [0, 1]^k \rightarrow R$ und folgt mit einer gegebenen Zufallsvariable x mit k diskreten Werten und der Verteilung $P = (p_1, p_2, \dots, p_k)$ den nachstehenden Konditionen. [Lig13, Mir18, Mai05]

- $\phi(P) \geq 0$
- $\phi(P)$ ist Minimum, falls \exists_i , sodass Komponente $p_i = 1$ ist
- $\phi(P)$ ist Maximum, falls $\forall_i, 1 \leq i \leq k, p_i = 1/k$
- $\phi(P)$ ist symmetrisch in Hinsicht auf P
- $\phi(P)$ ist stetig

Mit dem Trainingsset S wird in [Mai05] der Wahrscheinlichkeitsvektor der Ergebnisklas-

se y wie folgt definiert:

$$P_y(S) = \left(\frac{|\sigma_{y=c_1}S|}{|S|}, \dots, \frac{|\sigma_{y=c_{|dom(y)|}}S|}{|S|} \right) \quad (2.2)$$

Die Güte der Knotenaufteilung bezüglich eines diskreten Attributes a_i ist in [Mai05] als die Reduktion der Unreinheit des Zielattributes nach der Partition des Sets S hinsichtlich $v_{i,j} \in dom(a_i)$ definiert.

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|dom(a_i)|} \frac{|\sigma_{a_i=v_{i,j}}S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}}S)) \quad (2.3)$$

Der Informationsgewinn setzt sich aus der Differenz der Unreinheit des Elternknotens (Set S) und der Summe der Unreinheiten der Kindknoten zusammen. Je geringer die gesamten Unreinheiten der Kindknoten sind, desto größer falls der Informationsgewinn aus. Für die Anwendung von binären Entscheidungsbäumen existieren in den meisten Implementierungen drei gebräuchliche Maße für die Unreinheit, der Gini-Koeffizient, die Entropie und der Klassifizierungsfehler. [Lig13, Mir18]

Der Gini-Koeffizient, auch Ungleichverteilungskoeffizient, gibt an, inwieweit die Wahrscheinlichkeit einer Fehlerklassifizierung minimiert ist beziehungsweise wie stark ausgeprägt die Differenzen der Wahrscheinlichkeitsfunktionen der Aufteilungen sind. Ein Gini-Index von 0 entspricht einem reinen Knoten also genau dann, wenn alle Fälle eines Knotens einer gleichen Klasse entsprechen. Das Maximum des Gini-Koeffizienten ist erreicht, wenn jede Klasse in dem betrachteten Knoten mit der gleichen Wahrscheinlichkeit auftritt. [Lig13, Mir18, Mai05] In [Mai05] wird der Gini-Index wie folgt definiert:

$$Gini(y, S) = 1 - \sum_{c_j \in dom(y)} \left(\frac{|\sigma_{y=c_j}S|}{|S|} \right)^2 \quad (2.4)$$

Und somit der Informationsgewinn durch den Gini-Koeffizienten nach [Mai05]:

$$Ginigain(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}}S|}{|S|} \cdot Gini(y, \sigma_{a_i=v_{i,j}}S) \quad (2.5)$$

Das Baumwachstum wird von dem Algorithmus so lange fortgesetzt bis ein Abbruchkriterium erfüllt wird. Dies geschieht u. a. durch Angabe einer Minimalanzahl an Samples pro Klasse in einem Kindknoten, Definition einer Maximaltiefe des Baumes oder das beste Splittingkriterium liegt unterhalb eines festgelegten Grenzwertes (kein Informationsgewinn an dem betroffenen Knoten erreichbar). [Mai05]

Support Vector Machine Support Vektor Maschinen (Abkürzung SVM; englisch Support Vector Machine) sind in der Lage, Probleme von Klassifikations-, Regressions- und Anomalieerkennungsaufgaben zu lösen und konnten in den letzten Jahren im Bereich des maschinellen Lernen immer größere Popularität erlangen. Sie bestimmen die Modellparameter mithilfe eines konvexen Optimierungsverfahrens, sodass jede lokale Lösung ein globales Optimum darstellt. Zur Vereinfachung wird zunächst von einer binären Klassifizierung durch Generierung eines linearen Modells folgender Form ausgegangen. [Bis06, Mir18]

$$y(x) = w^T \phi(x) + b \quad (2.6)$$

Mit:

$\phi(x)$: Transformation des Merkmalsraumes

b : Bias Parameter

w^T : Optimierungsparameter

x : N-facher Inputvektor x_1, \dots, x_N ; Klassenvektor t_1, \dots, t_N und $t_n \in \{-1, 1\}$

Ziel von Support Vektor Maschinen ist die Etablierung eines zu maximierenden Trennungsbereiches, dem Margin, um die Klassengrenzen. Dieser Bereich ist als der minimale orthogonale Abstand zwischen der klassentrennenden Hyperebene und den Datenmengen im Merkmalsraum definiert. Die Position dieser Hyperebene (in Abbildung 2.17 rot hervorgehoben) wird durch eine Teilmenge der Datenpunkte des Inputvektors, die Stützvektoren (in Abbildung 2.17 durch Kreise gekennzeichnet), aufgespannt. [Bis06, Mir18]

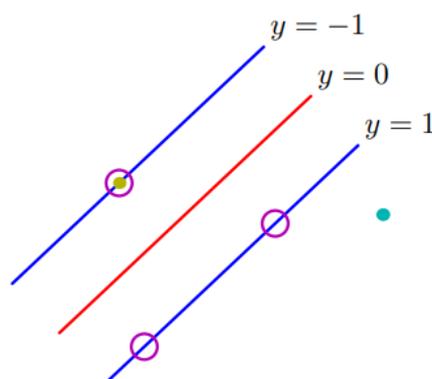


Abbildung 2.17.: Margin lineare SVM [Bis06]

Der Abstand der Datenpunkts x von der Hyperebene $y(x) = 0$ wird in Abbildung 2.17 durch den Term $|y(x)|/||w||$ angegeben. Somit gilt für den Abstand eines Punktes unter Berücksichtigung von Gleichung 2.6 und richtiger Klassifikationen des Modells, sodass

in [Bis06] $t_n y(x_n) > 0$ für alle n gilt:

$$\frac{t_n y(x_n)}{\|w\|} = \frac{t_n (w^T \phi(x_n) + b)}{\|w\|} \quad (2.7)$$

Daraus entsteht folgende Optimierungsaufgabe. [Bis06]

$$\operatorname{argmax}_{w,b} \left\{ \frac{1}{\|w\|} \min_n [t_n (w^T \phi(x_n) + b)] \right\} \quad (2.8)$$

Aufgrund der Komplexität und des benötigten Aufwandes der Lösung von Gleichung 2.8 wird diese unter Berücksichtigung der Nebenbedingung, dass die Datenpunkte durch die Entscheidungsgrenze bezüglich ihrer Klasse korrekt getrennt werden, durch folgendes äquivalentes quadratisches Optimierungsproblem ausgedrückt. [Bis06]

$$\operatorname{argmin}_{w,b} \frac{1}{2} \|w\|^2 \quad (2.9)$$

Unter der Bedingung in Form einer linearen Ungleichung. [Bis06]

$$t_n (w^T \phi(x_n) + b) \geq 1 \quad n = 1, \dots, N \quad (2.10)$$

Die darauf folgende Aufstellung und Lösung der Lagrangefunktion ist der Literatur in [Bis06, Mir18, Sch04] zu entnehmen. Ebenfalls ist zu erwähnen, dass für die Lösung von nicht linear trennbaren Daten und der daraus resultierenden nicht linearen Bedingung die Einführung einer Schlupfvariablen (Straffunktion bei Fehlklassifizierungen) im Rahmen der Soft-Margin-Klassifikationsmethode notwendig ist, siehe Abbildung 2.18. [Bis06, Mir18, Sch04]

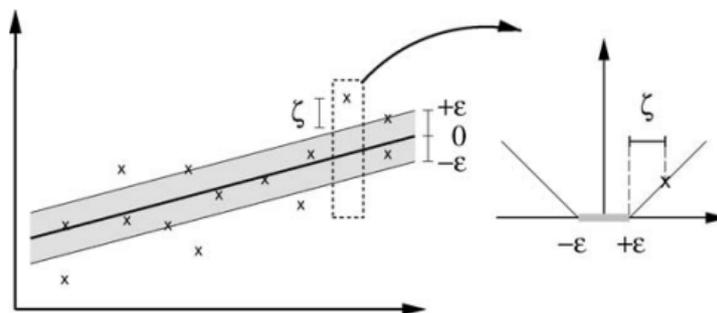


Abbildung 2.18.: Soft-Margin-Funktion lineare SVM [Sch04]

2.2.1.3. Interpretierbares maschinelles Lernen

Die Funktionsweise von generierten Modellen aus dem Bereich Machine-Learning (Abkürzung ML) und Deep-Learning (Abkürzung DL) ist in den meisten Fällen für den Nutzer nicht nachvollziehbar. Die anwendbaren Performance- und Kontrollmetriken werden für eine Vielzahl spezieller Anwendungsgebiete von ML-/DL-Verfahren kontinuierlich erweitert, sodass deren Bewertung und Freigabe mit hinreichender Prozesssicherheit gewährleistet wird. Jedoch ist aufgrund dieser Informationslage lediglich die Bewertung und Messung des äußeren Systemverhaltens beziehungsweise der geltenden Input/Output-Beziehungen möglich. [Pol14] Moderne Forschungszweige des maschinellen Lernens und des Deep-Learning befassen sich mit der Erweiterung eines zulässigen Systems zu einem interpretierbaren System. Ein besseres Verständnis dieses Bestrebens wird durch die Abgrenzung und Erläuterung der Begriffe *Interpretierbarkeit*, siehe Definition 4, sowie *Rechtfertigung*, siehe Definition 5, eines intelligenten Systems erlangt. [Cot17]

Definition 4 (Interpretability in ML [Cot17]):

Systems are interpretable if their operations can be understood by a human, either through introspection or through a produced explanation.

Definition 5 (Justification in ML [Cot17]):

A justification explains why a decision is a good one, but it may or may not do so by explaining exactly how it was made.

In den frühen Arbeiten dieses Forschungsgebiets lag der Fokus auf der Visualisierung von Modellvorhersagen zwecks Unterstützung der Anwender bei der Modellbewertung oder der graphischen Darstellung von Hiddenlayern in tiefen neuronalen Netzwerken. Aktuelle wissenschaftliche Untersuchungen konzentrieren sich auf weitere Formen der Modellinterpretierbarkeit: die Interpretierbarkeit der Modellvorhersage und die Interpretierbarkeit des gesamten Modells. [Cot17]. Zusammenfassend wird interpretierbares ML in [Abb19] als die „Extraktion von Wissen aus einem ML-Modell bezüglich der Beziehungen, die entweder in den Daten enthalten sind oder vom Modell gelernt wurden“ definiert.

Modelle, die auf Grundlage von Entscheidungsbäumen generiert werden, ermöglichen die Untersuchung von nicht-linearen Wechselbeziehungen zwischen den Merkmalen (Featurewerte) eines Datensatzes und der Modellantwort (Klassifikations-/Regressionsresultat). [Mol22] Die Aufteilung des Datensatzes in Subdatensätze pro Knoten anhand eines festgelegten Reinheitsmaßes und weitere mathematische Hintergründe während des Wachstums des Entscheidungsbaumes werden in Kapitel 2.2.1.1 erläutert.

Diese Eigenschaften werden in der vorliegenden Forschungsarbeit zur Aufstellung von Wenn-Dann-Regeln genutzt. Die graphische Interpretation und die daraus resultierende Ableitung von Regeln wird in Abbildung 2.19 anhand eines Beispiels verdeutlicht.

Der Entscheidungsbaum besteht aus zwei Knoten, wobei am Wurzelknoten eine Aufteilung des Datensatzes anhand des Merkmals 1 und des numerischen Grenzwertes 3 dieses Merkmals erfolgt. Die Datenpunkte, welche einen Merkmalswert 1 größer als 3 besitzen, folgen der rechten gerichteten Kante, hier findet keine weitere Unterteilung des Subdatensatzes statt. Dem linken gerichteten Pfad folgen die Datensätze mit Featurewerten 1 kleiner/gleich 3. In diesem Fall findet unter Berücksichtigung des Merkmals 2 eine weitere Unterteilung statt. Die Wenn-Dann-Regeln der einzelnen Knoten werden durch die linguistische Variable *UND* entlang eines Pfades (vom Wurzelknoten zum letzten Kindknoten) des Entscheidungsbaumes kombiniert. [Mol22]

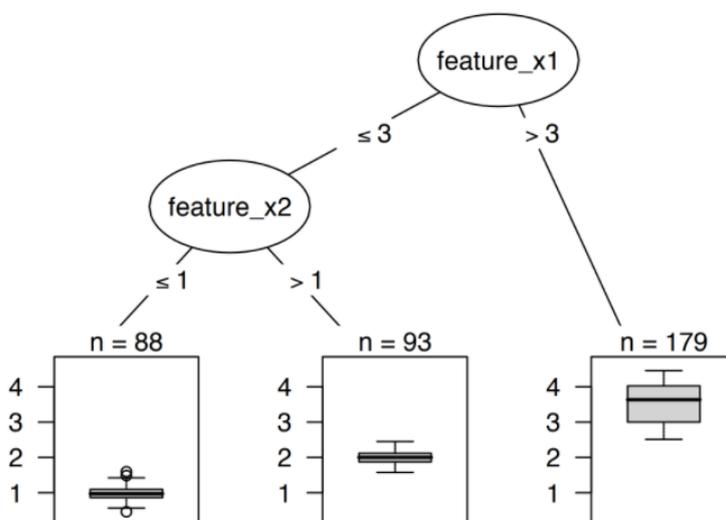


Abbildung 2.19.: Beispiel Visualisierung DT [Mol22]

Das Ergebnis der Klassifizierung beziehungsweise Regression \hat{y} durch einen Entscheidungsbaum wird in Formel 2.11 definiert. [Mol22]

$$\hat{y} = \hat{f}(x) = \sum_{m=1}^M c_m I\{x \in R_m\} \quad (2.11)$$

Mit:

R_m : Subdatensatz

c_m : Betrachtete Klasse des Subdatensatzes R_m

$I_{x \in R_m}$: Identitätsfunktion (1 wenn x in Subdatensatz R_m , sonst 0)

RuleFit Algorithmus Entscheidungsbäume oder gar Ensemblemethoden bestehend aus diesen dienen vielen Verfahren des interpretable Machine-Learning als Grundlage für die Extraktion von Klassifikations-/Regressionsregeln. [Mol22]

Der RuleFit-Algorithmus verbindet die Vorteile eines linearen spärlichen Modells mit den positiven Eigenschaften der Entscheidungsbäume (singular oder im Ensemble). Dieser Algorithmus weist einen großen Flexibilitätsgrad auf, da die Generierung der Entscheidungsbäume nicht von der Durchführung der RuleFit-Methode abhängt. [Mol22]

Die Grundidee besteht darin, ein spärliches lineares Modell aufzustellen, welches leicht interpretierbar ist, und dieses durch Hinzufügen der gelernten Regeln (extrahiert aus einem Entscheidungsbaum) zu erweitern. Abbildung 2.20 enthält eine beispielhafte Darstellung der vier Regeln, welche aus einer binären Baumstruktur mit drei Endknoten extrahiert werden. Regeln 1 und 2 enthalten jeweils nur ein Antezedens (linguistische Wenn-Teile der Wenn-Dann-Implikation) bezüglich Feature 2. Regel 3 und 4 bestehen jeweils aus 2 Antezedenzen. [Pop08, Mol22]

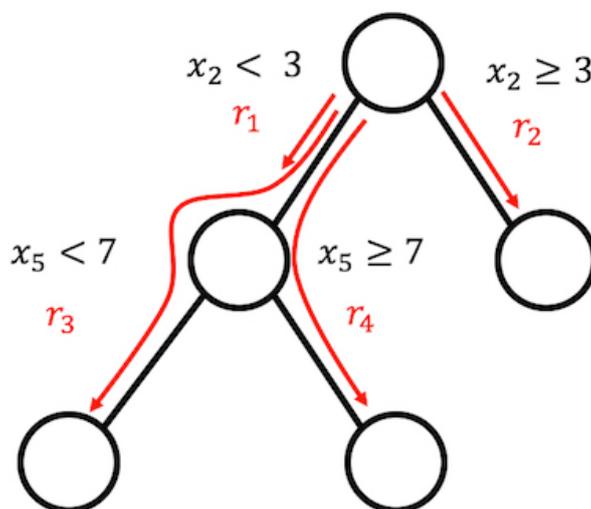


Abbildung 2.20.: Beispiel Regelextraktion aus DT [Mol22]

Durch die Kombination aus Originalmerkmalen und DT-Implikationen wird ein Kompromiss aus leichterer Interpretierbarkeit (lineares Modell) und Berücksichtigung der Korrelationen der Merkmale untereinander (Regeln aus DT) geschaffen. [Pop08] Die Anzahl der durch ein Ensemble aus M Entscheidungsbäumen und jeweils t_m Endknoten generierten Regeln wird in [Pop08] definiert als:

$$K = \sum_{m=1}^M 2(t_m - 1) \quad (2.12)$$

Aufgrund der starken Zunahme der Regelanzahl bei Verwendung von komplexen Baum- sowie Ensemblearchitekturen ist eine Begrenzung der im RuleFit-Modell verwendeten Regeln erforderlich. Diese Problematik wird durch Anwendung eines Regularisierungsverfahrens gelöst, indem durch Messung der Merkmalsrelevanz (originale Features des

Datensatzes und der hinzugefügten Regeln aus den DTs) und Einführung eines Bestrafungsterms im Rahmen einer LASSO-Regularisierung eine Schrumpfung des Merkmalsraums realisiert wird. [Pop08, KI-22]

RRule Algorithmus Der RRule-Algorithmus stellt eine Optimierung des Rules-Klassifizierers nach [Ask95] dar. Analog zu dem RuleFit-Verfahren wird die Aufstellung von relevanten Wenn-Dann-Implikationen zur Klassifizierung von kategorischen oder numerischen Datensätzen angestrebt. Jedoch wird im Rahmen einer RRule-Analyse keine Erstellung eines Modells mit singularen oder vielzähligen Baumarchitekturen gefordert, sondern gänzlich auf eine Modellgenerierung als Informationsgrundlage für die Regelextraktion verzichtet. Mithilfe iterativer Durchsuchung der Trainingsdaten nach Attribut-Wert-Paaren, welche die Datenpunkte eindeutig verschiedenen Klassen zuteilen, werden die Bildung von linguistischen, für den Menschen interpretierbaren Implikationen sowie Erkenntnisse über vorhandene Strukturen im Datensatz realisiert. [Pal21, Ask95]

Die *Attribute* eines Attribut-Werte-Paares entsprechen den Merkmalen oder Features im Datensatz und die *Werte* den im Datensatz vorkommenden Merkmalsausprägungen (numerisch oder kategorial). Die Gesamtanzahl n_a der Attribute im Datensatz limitiert die maximale Länge der Wenn-Teile der extrahierten Regeln ($\in \{1, \dots, n_a\}$), in denen jede Antezedens ein unterschiedliches Attribut-Werte-Paar enthält. Zu diesem Zweck wird ein Array erstellt, welches eine Zuordnung der Attribute und der Attributausprägungen enthält. [Pal21, Ask95]

In der ersten Iteration werden die Attribut-Werte-Paare, welche eine eindeutige Aufteilung der Datenpunkte in eine Klasse ermöglichen, identifiziert und als neue Regeln definiert. Diese Regeln werden auf den kompletten Datensatz angewendet und die Datenpunkte, welche anhand dieser Regeln nicht eindeutig klassifiziert werden, als Basis eines Subdatensatzes (vorkommende Attribute und Attributausprägungen) verwendet. Dieses Subarray ist die Eingangsgröße des darauffolgenden Iterationsschrittes. Es folgt eine Bildung aller möglichen Attribut-Werte Kombinationen dieses Subarrays und die Überprüfung dieser möglichen neuen Regeln anhand des gesamten Datensatzes T und des Subdatensatzes N . Wird durch diesen Regelkandidat eine eindeutige Klasseneinteilung ermöglicht, erfolgt eine Aufnahme dieser Regel in den bestehenden Regelkatalog und ein Update des zu untersuchenden Submerkmalsraumes N , indem die durch die Regel erklärten Datenpunkte aus diesem entfernt werden. [Pal21, Ask95]

Mit jeder Iteration steigen Anzahl sowie Länge der extrahierten Regeln. Gleichzeitig nimmt die Menge der nicht klassifizierten Datenpunkte ab, bis diese entweder leer ist oder das Abbruchkriterium erfüllt ist. Im letzteren Fall entspricht die Anzahl der Iterationen und somit die Anzahl der Bedingungsteile n_c in den Regeln der Gesamtanzahl an Merkmalen n_a im Datensatz. Die Funktionsweise des RRules-Verfahrens wird im Pseu-

docode 1 nach [Pal21] vorgestellt. [Pal21]

Algorithm 1 RRules nach [Pal21]

```

1: Input  $T$  ▷  $T$  ist Trainingsmenge
2:  $N \leftarrow T$  ▷ Initialisierung Menge  $N$  der nicht-klassifizierten Instanzen
3:  $Rules \leftarrow \emptyset$  ▷ Initialisierung Menge  $Rules$ 
4: for  $n_c \in (1, \dots, n_a)$  do ▷  $n_a$  ist die Anzahl Featurewerte
5:   Ermittlung aller möglichen Bedingungen  $Cond$  als Kombination der  $n_c$  Attribut-
     Wert-Paare in  $N$ 
6:   for all  $Cond$  do
7:      $M_T \leftarrow$  Instanzen in  $T$ , die  $Cond$  als Antezedens erfüllen
8:      $M_N \leftarrow$  Instanzen in  $N$ , die  $Cond$  als Antezedens erfüllen
9:     if  $|M_T| = 0$  then ▷ Kein Datenpunkt in  $T$  erfüllt Bedingung
10:      Löschung Bedingung und Betrachtung nächster
11:     end if
12:     if  $|M_N| = 0$  then ▷ Irrelevante Bedingung, Dopplung v. Aussagen möglich
13:      Löschung Bedingung und Betrachtung nächster
14:     end if
15:     if Alle Instanzen  $m_i \in M_T$  gehören derselben Klasse  $C$  an then
16:       Generierung Regel  $R$  aus  $Cond \rightarrow C$ 
17:        $Rules \leftarrow Rules + R$  ▷  $R$  ist neue gültige Regel/Update  $\{Rules\}$ 
18:        $N \leftarrow N - \{M_N\}$  ▷ Löschung klass. Datenpunkte aus  $N$ /Update  $\{N\}$ 
19:       if  $|N| = 0$  then ▷ Jeder Datenpunkt klass.
20:         return  $Rules$ 
21:       end if
22:       else if  $n_c = n_a$  then ▷ Inkonsistenz Datenpunkte: Gleiche Featurewerte,
         unterschiedliche Klassen
23:         Definition am meisten besetzter Klasse  $C^*$  in  $M_T$ 
24:         Generierung Regel  $R$  aus  $Cond \rightarrow C^*$ 
25:          $Rules \leftarrow Rules + R$  ▷  $R$  ist neue gültige Regel/Update  $\{Rules\}$ 
26:          $N \leftarrow N - \{M_N\}$  ▷ Löschung klass. Datenpunkte aus  $N$ /Update  $\{N\}$ 
27:         if  $|N| = 0$  then ▷ Jeder Datenpunkt klass.
28:           return  $Rules$ 
29:         end if
30:       end if
31:     end for
32:   end for
33: return  $Rules$ 

```

Anchor-Rules Während die RuleFit- und RRULE-Algorithmen intrinsisch evaluierbare Konstrukte zwecks Generierung von Regeln verwenden und somit den Verfahren, die interpretierbare Modelle hervorbringen, angehören, beschäftigt sich die Anchor-Rules-Methode mit der agnostischen Untersuchung eines Black-Box-Modells. Die Interpretierbarkeit der Vorhersage steht dabei im Vordergrund. Agnostische Verfahren erklären somit, in welcher Art und Weise ein trainiertes ML- oder DL-Modell die Einteilung von Daten in Klassen unter Berücksichtigung des Merkmalsraumes durchführt. [Dig21, Mol22]

Ein deutlicher Vorteil dieser Strategie zur Steigerung der Modellinterpretierbarkeit liegt in dem hohen Flexibilitätsgrad, da die Architektur und das verwendete ML-Verfahren völlig unabhängig von der Ausführung des Anchor-Rules-Algorithmus sind. Das zu untersuchende Modell $f : X \rightarrow Y$ und der betrachtete Datensatz X bestehend aus Einzelinstanzen $x \in X$ dienen als Eingangsgrößen des Anchor-Verfahrens. $f(X)$ entspricht einer Black-Box, deren inneres Systemverhalten für den Verwender komplett unverständlich und nicht von Relevanz ist, da lediglich die Input-Output-Wechselbeziehungen das Hauptuntersuchungsmerkmal sind. [Mol22, Kau13]

Ziel des Anchor-Rules-Verfahrens ist die Formulierung von simplen lokalen Wenn-Dann-Regeln durch Identifikation von Anchors (deutsch Ankern), die eine sichere Vorhersage ermöglichen. Ein solcher Wenn-Teil einer Regel bezüglich eines Merkmals gilt als sicher, also „verankert“, wenn lokal beschränkte Änderungen an anderen Merkmalen die Vorhersage/Klasseneinteilung nicht beeinflussen. Diese Regeln werden auf ihren Geltungsbereich untersucht, um für unbekannte Daten Vorhersagen zu treffen. Dies wird durch die Begriffseinführung und Optimierung des Abdeckungsbereiches (englisch coverage) realisiert. Für jede betrachtete Instanz aus dem Datensatz werden künstliche, benachbarte Datenpunkte erzeugt, diese bilden die Perturbationsverteilung \mathcal{D}_x des Datenpunktes x für eine geltende Menge A von Regeln, siehe Abbildung 2.21. [Mol22, Kau13, Gue18]

Es gilt $A(x) = 1$, falls alle Featurebedingungen der Regel eine richtige Einteilung der Instanz x und des Perturbationsraumes unter gleichen geltenden Bedingungen ermöglichen. Mit der bedingten Wahrscheinlichkeitsverteilung $\mathcal{D}(\cdot|A)$ wird die Regel A als gültig (Anker) definiert, wenn $A(x) = 1$ gilt und für $f(x)$ mit hoher Wahrscheinlichkeit ein Nachbarpunkt z aus dem Perturbationsraum $\mathcal{D}(z|A)$ zusätzlich $f(x) = f(z)$ gilt. Somit wird Regel A in [Kau13] wie folgt definiert. [Mol22, Kau13, Gue18]

$$\mathbb{E}_{\mathcal{D}(z|A)} = [\mathbb{1}_{f(x)=f(z)}] \geq \tau, A(x) = 1 \quad (2.13)$$

Mit:

x : Instanz aus Datensatz X

A : Regel/Anker = 1, wenn alle in A berücksichtigten Merkmale den Merkmalen

- von x entsprechen
- f : Black-Box-Modell
- $\mathcal{D}(\cdot|A)$: Nachbarschaft von Instanz x , die Regel A erfüllt
- τ : Grenzwert für lokale Genauigkeit mit $0 \leq \tau \leq 1$. Nur Regeln, die Grenzwert erreichen, werden als Anker bestätigt

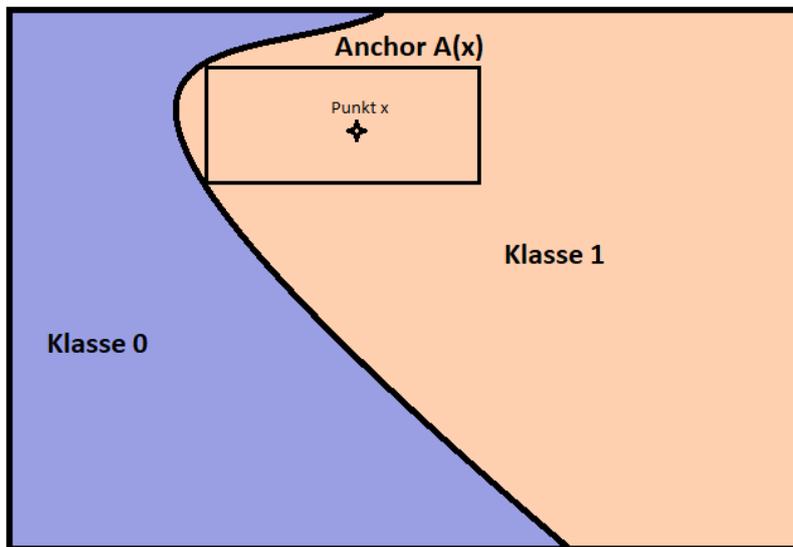


Abbildung 2.21.: Skizze Geltungsbereich Anchor-Algorithmus 2D-Daten (eigene Erstellung in Anlehnung an [Mol22])

In [Kau13] wird ein Präzisionsmaß $prec(A) \geq \tau$ eingeführt, welches für die Bestätigung einer Regel A erreicht werden muss, siehe Formel 2.14. Der erforderliche Rechenaufwand und somit die Kosten würden bei hochdimensionalen, numerischen Datensätzen derart stark ansteigen, dass die Anwendung von Anchor-Algorithmen nicht mehr praktikabel wäre. Aus diesem Grund wird durch Hinzufügen eines neuen Parameters $0 \geq \delta \geq 1$ eine probabilistische Betrachtungsweise ermöglicht. Die Wahrscheinlichkeit, mit der eine Regel einen hohen Präzisionswert erreicht, ist in Formel 2.15 nach [Kau13] definiert. [Mol22, Kau13]

$$prec(A) = \mathbb{E}_{\mathcal{D}(z|A)} = [\mathbb{1}_{f(x)=f(z)}] \tag{2.14}$$

$$P(prec(A) \geq \tau) \geq 1 - \delta \tag{2.15}$$

Der Abdeckungs-begriff findet Verwendung, wenn mehrere Regeln dieses Kriterium erfüllen. Nur die Regel, welche einen relativ zu anderen Regeln großen Perturbationsraum aufweist und somit die größte *coverage* besitzt, wird als gültiger Anker definiert. Daraus ergibt sich nach [Kau13] folgende und zu lösende Optimierungsaufgabe. [Kau13]

$$\max_{\text{A.s.t. } P(prec(A) \geq \tau) \geq 1 - \delta} cov(A), \quad cov(A) = \mathbb{E}_{\mathcal{D}[z]}[A(z)] \tag{2.16}$$

Zwei wichtige Prozessschritte des Anchor-Algorithmus sind die Generierung von zu untersuchenden Regeln (englisch candidate rules) und die Identifikation der besten Regeln, welche die Eingangsinstanzen mit großer *coverage* erklären. Diese beiden Verfahren werden im Rahmen einer modifizierten Strahlensuche kombiniert und ermöglichen somit eine effiziente Interpretation des Modells $f(x)$. Die Bewertung der Regelkandidaten erfolgt mithilfe eines Multi-Armed-Banditenverfahrens und stammt aus dem Forschungsgebiet des verstärkenden Lernens. [Mol22, Kau13] Ein vereinfachter Prozessablaufplan ist Abbildung 2.22 zu entnehmen.

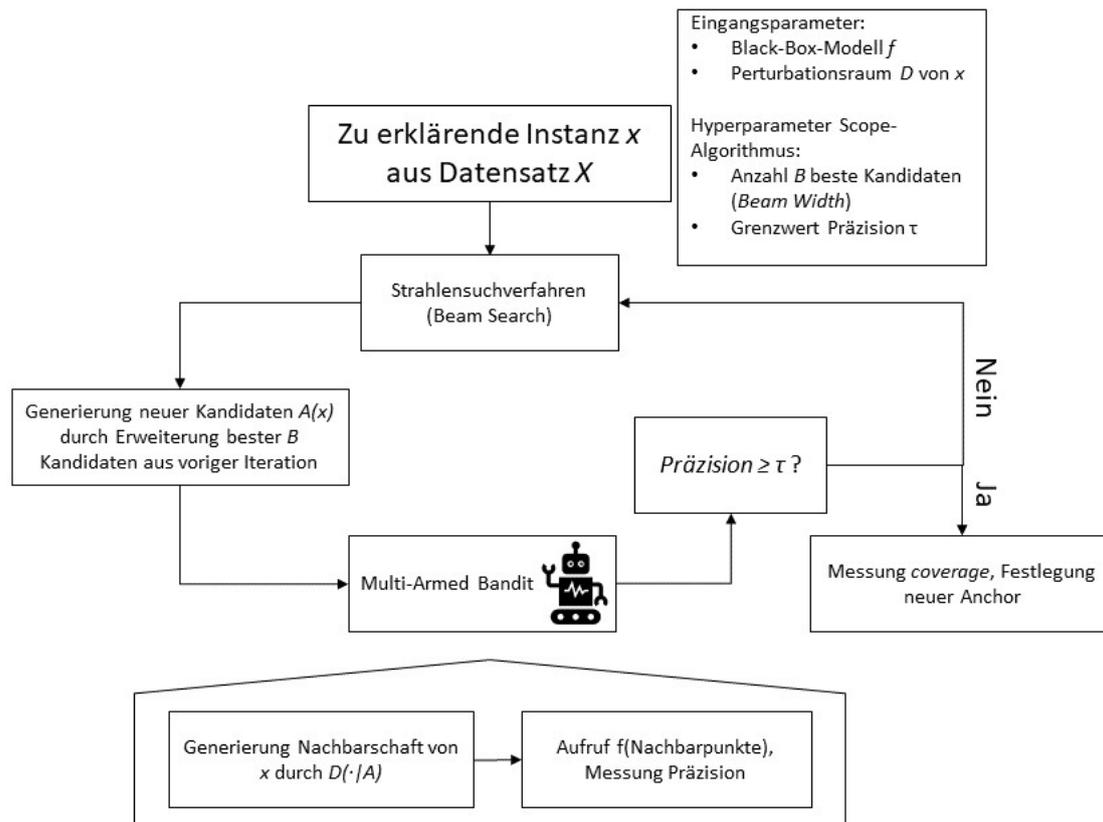


Abbildung 2.22.: Programmablaufplan Anchor-Rules (eigene Erstellung in Anlehnung an [Kau13])

Ripper-Algorithmus Der Ripper-Algorithmus (Abkürzung Repeated Incremental Pruning to Produce Error Reduction) wurde in [Coh95] als Verbesserung des IREP-Algorithmus vorgestellt und ist in der Lage, einen Regelsatz für verrauschte Informationen und große Datensätze zu generieren. Die erhobenen Regeln setzen sich aus Wenn-Dann-Implikationen zusammen und erlauben ein intuitives Verständnis der vorliegenden Datenstruktur. Der Ripper-Algorithmus lässt sich in vier Einzelschritte aufteilen, siehe Abbildung 2.23. [Mos19, Coh95]

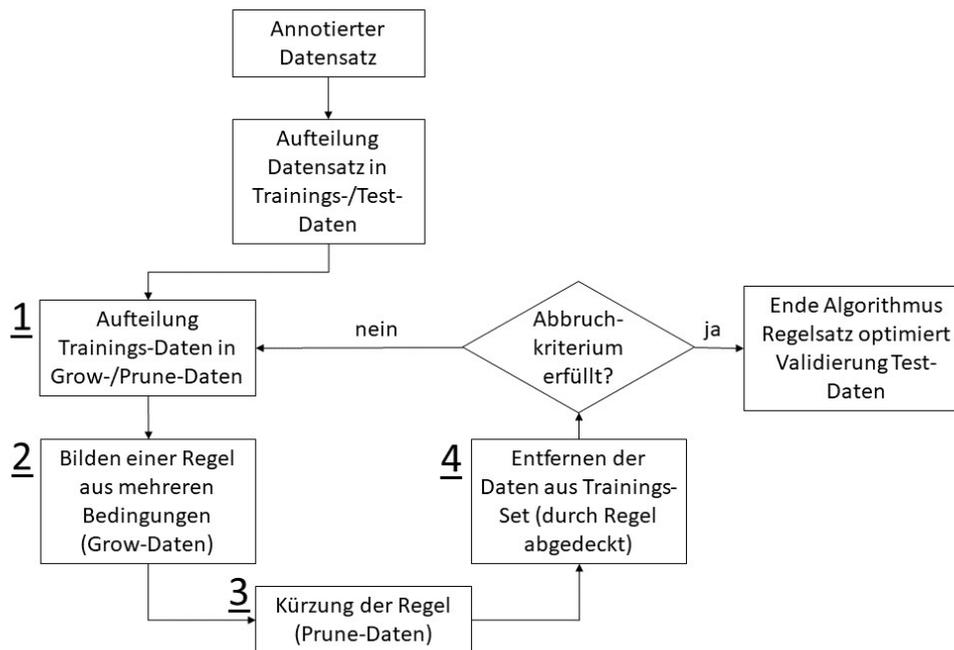


Abbildung 2.23.: Programmablaufplan Ripper-Algorithmus (eigene Erstellung in Anlehnung an [Coh95])

Nachdem die Eingangsdaten in Trainings- und Test-Daten aufgeteilt wurden (Schritt 1), bildet der Trainings-Datensatz die Eingangsgröße für den Ripper-Algorithmus und wird zur Regelextraktion verwendet. In dem ersten Prozessschritt erfolgt eine erneute Aufteilung des Trainings-Datensatzes in den Grow- und den Prune-Datensatz (nach [Coh95] in einem Verhältnis von 2/3 zu 1/3). Der Grow-Datensatz wird zur Generierung der ersten Regel genutzt. Diese beinhaltet zunächst eine Leermenge von Wenn-Teilen und wird im Anschluss um Bedingungen der Form $A_n = v$ mit $A_c \leq \theta$ oder $A_c \geq \theta$, mit A_n als nominales Attribut und dem gültigen Wert v sowie dem numerischen Attribut A_c und dem gültigen Grenzwert θ , erweitert. [Coh95]

Sukzessive wird die Regel um weitere Bedingungen, welche den FOIL-Informationsgewinn maximieren, erweitert (Schritt 2). Der FOIL-Informationsgewinn $FoilGain$ ist in Formel 2.17 definiert und enthält Tupel von verknüpften Variablen mit den jeweiligen gültigen Werten oder Grenzwerten. [Mos19]

$$FoilGain(R_0, R_1) = p_0 \left(\log_2 \left(\frac{p_1}{p_1 + n_1} \right) - \log_2 \left(\frac{p_0}{p_0 + n_0} \right) \right) \quad (2.17)$$

Mit:

R_0 : Ausgangsregel vor Hinzufügen einer neuen Bedingung

R_1 : Spezialisierung von R_0

p_0, n_0 : Anzahl positiver/negativer Tupel, die durch R_0 abgedeckt werden

p_1, n_1 : Anzahl positiver/negativer Tupel, die durch R_1 abgedeckt werden

Im direkten Anschluss an die Regelextraktion erfolgt der Prune-Prozess (Schritt 3), indem sukzessive Bedingungen aus Regel R (beginnend bei letzter Bedingung in R) entfernt werden und für jede Subregel ein zu maximierender Metrikkennwert v ermittelt wird. Dieser Prozess wiederholt sich, bis kein Prune-Vorgang eine Verbesserung der Fehlermetrik v zur Folge hat und somit die finale Regel der aktuellen Algorithmus-Iteration vollständig definiert ist. Die Fehlermetrik wird nach [Coh95] durch Formel 2.18 ermittelt. Im vierten Schritt werden alle positiven und negativen Beispiele, die durch die generierte Regel abgedeckt werden, aus dem Test-Datensatz entfernt. Im Falle einer Nichterfüllung des Abbruchkriteriums (beispielsweise Vorhandensein weiterer positiver Beispiele) wird eine neue Algorithmus-Iterationsschleife begonnen. [Mos19, Coh95]

$$v(R, PPos, PNeg) = \frac{p + (N - n)}{P + N} \quad (2.18)$$

Mit:

R : Zu kürzende Regel

$PPos$: Subdatensatz mit positiver Klassenannotation im Prune-Datensatz

$PNeg$: Subdatensatz mit negativer Klassenannotation im Prune-Datensatz

p, n : Anzahl positiver/negativer Daten, die durch R abgedeckt werden

P, N : Anzahl positiver/negativer Daten im Prune-Datensatz

Als Abbruchkriterium nutzt der Ripper-Algorithmus die Beschreibungslänge. Das Abbruchkriterium ist erfüllt, wenn die Beschreibungslänge des Regelsatzes einen festgelegten Grenzwert (in [Alp19] 64 Bit) überschreitet. Durch die steigende Länge der Einzelregeln und der Gesamtanzahl der Regeln im Regelsatz nimmt die Komplexität des gesamten Modells zu. Die Anzahl der nicht durch den Regelsatz beschriebenen Datenpunkte wird im Gegensatz dazu kontinuierlich reduziert. Die Beschreibungslänge erreicht eine Optimierung der Forderung nach Minimierung der Komplexität bei gleichzeitiger Minimierung der Anzahl nicht beschriebener Datenpunkte. [Mos19, Coh95] Das Prinzip der minimalen Beschreibungslänge wird in [Sch05] definiert.

Definition 6 (Prinzip der minimalen Beschreibungslänge [Sch05]):

Minimiere Summe aus Information der Hypothese und Information der Rekonstruktionsfehler bei gegebener Hypothese.

2.2.1.4. Feature-Selektionsverfahren

In den letzten Jahren wurde durch die Forschung im Bereich Machine-Learning beziehungsweise Deep-Learning und die Anwendung moderner Klassifikations- und Regressionsmethodiken die Erkenntnis darüber gewonnen, dass unter bestimmten vorliegen-

den Rahmenbedingungen die Modellperformance erheblich gesteigert werden kann, falls während des Modelltrainings ein optimierter Datensatz, welcher lediglich für die Klassifikation oder Regression relevante Informationen enthält, verwendet wird. Diese Feature-Selektionsverfahren reduzieren die Anzahl der verwendeten Merkmale (englisch Features) unter Berücksichtigung von zwei Hauptkriterien. Zum einen darf die Performance-Metrik des Modells durch die Verringerung des Merkmalsraumes nicht signifikant abnehmen und andererseits wird gefordert, dass die Unterschiede der Klassenverteilungen des originalen Datensatzes und des bereinigten Datensatzes minimiert werden. Neben der möglichen Effektivitätssteigerung der verwendeten Algorithmen ermöglichen Feature-Selektionsverfahren darüber hinaus die Untersuchung sowie Aufdeckung von im Datensatz vorliegenden Strukturen und Relevanzverteilungen. Dieses Wissen dient der kontinuierlichen Verbesserung von produktionstechnischen Abläufen und Prozessen. [Wit18, Di05]

Abhängig von den Zielen, die durch Anwendung von Feature-Selektionsverfahren erreicht werden sollen, existieren zwei Untergruppen dieses Forschungsbereiches: Erstellung von Subdatensätzen und Feature-Ranking-Verfahren. Ersteres verfolgt die Reduzierung des Merkmalsraumes zwecks Generierung effizienterer, intelligenter Anwendungen. Die zweite Unterkategorie beschäftigt sich mit der Relevanzbewertung und Untersuchung des Informationsgehaltes der einzelnen Merkmale in einem Datensatz. Im Rahmen dieser Forschungsarbeit werden Feature-Ranking-Methoden für die Bestimmung der Featurewerte, welche erheblichen Einfluss auf das Klassifikationsergebnis im AOI-Prozess ausüben, implementiert. [Wit18, Di05]

Definition 7 (Feature Ranking nach [Di05] – aus dem Englischen übersetzt):

Sei $A = \{a_1, a_2, \dots, a_m\}$ die Menge von m -Attributen und wird jedem Attribut $a \in A$ aus D eine Wertigkeit durch Funktion $r : A_D \rightarrow \mathbb{R}$ zugeordnet. So ist Feature-Ranking eine Funktion F , die jedem Attribut $a_i \in A$ einen Relevanzwert zuordnet und eine Liste von Attributen ($a^*_i \in A$) geordnet nach Relevanzkriterium und mit $i \in \{1, \dots, m\}$:

$$F(\{a_1, a_2, \dots, a_m\}) = \langle a^*_1, a^*_2, \dots, a^*_m \rangle \text{ und :}$$

$$r(a^*_1) \geq r(a^*_2) \geq \dots \geq r(a^*_m)$$

Eine weitere Unterteilung der Feature-Selektionsverfahren wird in [Bog15] formuliert. Darin wird zwischen Filter-, Wrapper-, Embedded- und Hybridmethoden unterschieden.

Filtermethoden sind unabhängig von dem verwendeten ML- oder DL-Modell und dienen in den meisten Anwendungsfällen als Vorstufe des eigentlichen Klassifikations- oder Regressionsalgorithmus, indem sie nur die relevanten Merkmale an das Modell transferieren. Einer Bewertung durch Filtermethoden werden einzelne Features oder zusammengefasste Featuresubsets unterzogen. Prominente Performancemetriken von Filterungsverfahren

sind Informationsgehalt, Distanzmaße, Konsistenz, Ähnlichkeit oder statistische Kennwerte. Zu dieser Algorithmuskategorie gehört der, im Folgenden näher erläuterte, SOAP-Algorithmus. [Bog15, Wit18]

Wrapper-Verfahren sind im Gegensatz zu Filter-Methoden auf die Anwendung eines Black-Box-Modells angewiesen. Diesem Black-Box-Modell werden in mehreren Iterationsschritten sukzessive Subdatensätze bestehend aus Featureteilmengen zugeführt und die Performance für jeden Einzeldurchlauf ermittelt. Je nach verwendetem ML-Modell erfolgt die Bewertung anhand speziell an die Aufgabe angepasster Metriken. Wrapper-Methoden für Klassifikationsmethoden greifen hierfür beispielsweise auf SVM oder Naive-Bayes-Algorithmen zurück. Die Erzeugung der Subdatensätze und die Featureauswahl für jede Iteration folgen der ausgewählten Suchstrategie. Aufgrund der Berücksichtigung von Featurekombinationen sind Wrapper-Algorithmen in der Lage, redundante Attribute zu identifizieren. Der BORUTA-Algorithmus als Bestandteil dieser Forschungsarbeit ist ein Vertreter der Wrapper-Methodik. [Wit18, Df05, Bog15]

Boruta-Algorithmus Entscheidungsbäume beziehungsweise Random Forests eignen sich nur bedingt für die Bestimmung von Relevanzkennzahlen einzelner oder gruppierter Merkmale innerhalb eines Datensatzes. Die Bildung des Entscheidungsbaumes beruht auf der Annahme, dass die Features untereinander unkorreliert vorliegen. Nicht alle Anwendungsfälle erfüllen diese Forderung, sodass eine Optimierung der Importance-Messung mithilfe von Entscheidungsbäumen erforderlich ist. Der Boruta-Algorithmus vereint die erforderlichen Anpassungen, sodass die Importance-Beurteilung eines stark korrelierten Datensatzes realisierbar ist. [Rud10]

Zu diesem Zweck werden dem Datensatz im Rahmen einer Boruta-Analyse Shadow-Features hinzugefügt. Dabei handelt es sich um randomisierte Kopien der einzelnen Featurewerte. Auf Grundlage dieses neuen, erweiterten Datensatzes wird in einem mehrstufigen Iterationsverfahren das Klassifikationsmodell (Random Forest) trainiert. Merkmale, die eine höhere Kennzahl in der Importance-Metrik als die maximale aller randomisierten Kopien aufweisen, werden als signifikant eingestuft und dem nächstfolgenden Iterationsschritt zugefügt. Features, welche dieses Kriterium nicht erfüllen, werden aus dem Datensatz entfernt. [Df05, Rud10]

Durch die Ausführung dieser Methode in einem Iterationsverfahren und die Erweiterung des Datensatzes um die Shadow-Features wird die Auftretenswahrscheinlichkeit von zufälligen Korrelationen vermindert. Der Z-Score dient hierbei als Wichtigkeitsmaß und wird in jedem Iterationsschritt für jedes Merkmal pro Einzelentscheidungsbaum ermittelt. Der gesamte Z-Wert eines Random Forests setzt sich aus dem Quotienten aus Mittelwert und Standardabweichung der Z-Werte aller Entscheidungsbäume des Random Forests zusammen. [Df05]

Im Detail wird jedem Merkmal des Datensatzes pro Iterationsschritt eine ermittelte Anzahl an Hits zugeordnet, siehe Diagramm 2.24. Ein Hit wird erreicht, wenn in einem Entscheidungsbaum der Z-Score des betrachteten Merkmals über dem maximalen Z-Score aller Shadow-Features liegt. Die Aufsummierung der Hits über alle Einzelbäume des Random Forests und anschließende Anwendung eines zweiseitigen Hypothesentests für jedes Merkmal ermöglicht die Einstufung in relevant oder nicht relevant und anschließende Beibehaltung beziehungsweise Entfernung dieses Merkmals aus dem Datensatz. [Dí05, Rud10] Als Abbruchkriterium ist entweder die vollständige Beurteilung aller Features oder das Erreichen einer maximalen Anzahl an Iterationsschritten definiert. Die finale Rangliste der Merkmalswichtigkeit wird durch Mittelwertbildung der Z-Scores aller Iterationen erstellt. [Dí05]

SOAP-Algorithmus Als ein Vertreter der Filtermethoden verzichtet der SOAP-Algorithmus (englisch Selection Of Attributes by Projection) auf die Anwendung von ML-Verfahren und erzielt somit erhebliche Effektivitätssteigerungen durch reduzierte Berechnungskosten und einen hohen Flexibilitätsgrad aufgrund der geringen Anforderungen an die Struktur des Eingangsdatensatzes. Sowohl kategoriale als auch numerische Eingangswerte sowie binäre und multikategoriale Klassifizierungsaufgaben können ohne erforderliche Transformationen verarbeitet und durchgeführt werden. [Agu02] Im Rahmen einer SOAP-Analyse erfolgt die Generierung von Projektionen bezüglich jedes Featurevektors. [Agu02, Wit18]

In Abbildung 2.25 werden die Datenpunktverteilungen der Featurewerte 23 beziehungsweise 24 hinsichtlich des Features 3 zweidimensional dargestellt. Im Falle des linken Graphs (Feat 3 zu Feat 23) werden zwei Projektionen realisiert. Dabei handelt es sich einerseits um die Projektion der Klassenannotationen auf die Abzisse (Feat 23) und andererseits auf die Ordinate (Feat 3), sodass entlang der Achsen eine Klassenverteilung entsteht und daraus die Labeländerungen ermittelt werden. In diesem Beispiel liegen bei der Ordinatenprojektion sieben und bei der Abszissenprojektion drei Änderungen der Klassenzugehörigkeit vor. Liegen die Datenpunkte gleicher Klassen in voneinander abgrenzbaren Clustern in einer Projektion vor, sinkt die Anzahl der Labeländerungen und somit wird den betrachteten Variablen ein größerer Einfluss auf die Klassenzuordnung zugeschrieben. [Agu02, Wit18]

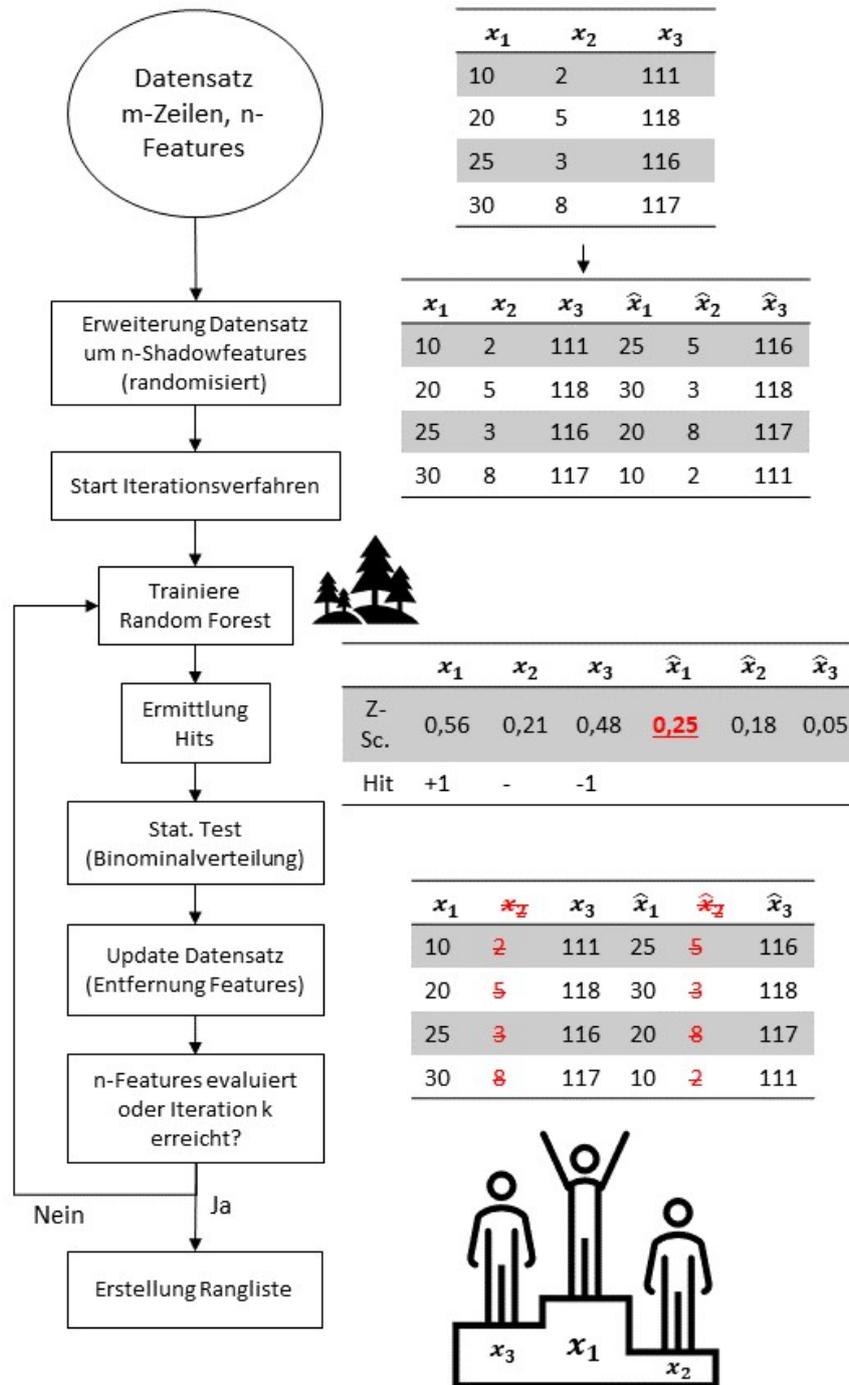


Abbildung 2.24.: Programmablaufplan Boruta-Algorithmus (eigene Erstellung in Anlehnung an [Wit18])

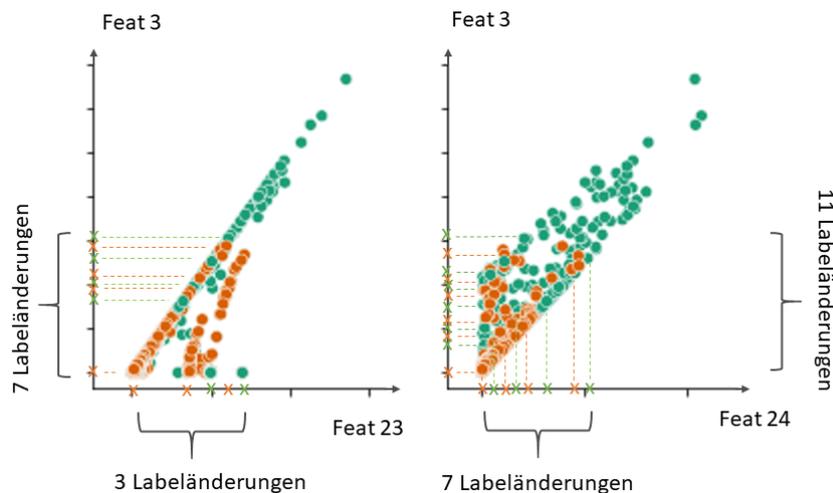


Abbildung 2.25.: SOAP-Algorithmus: Projektion Feature 3, 23, 24 und Labeländerungen (eigene Erstellung in Anlehnung an [Wit18])

Dieser Zusammenhang wird vom SOAP-Algorithmus zur Messung der Merkmalswichtigkeit genutzt. Iterativ wird für jedes Merkmal die Kennzahl der Klassenänderungen bestimmt und daraus die finale Rangliste abgeleitet, siehe Algorithmus 2. In der vollständigen Implementierung des SOAP-Algorithmus ist eine Strategie für den Spezialfall erforderlich, falls in einer Projektion die Mehrfachbesetzung eines Punktes auf der Projektionsachse mit unterschiedlichen Klassenannotationen vorliegt und somit die eindeutige Zuordnung der vorliegenden Klasse zu jedem Punkt auf der Projektionsachse verhindert wird. Diese Forschungsarbeit berücksichtigt das Worst-Case-Szenario nach [Agu02]. Dieses beinhaltet die Annahme der maximalen Anzahl an Labeländerungen in einem Punkt. [Agu02, Wit18]

Algorithm 2 SOAP

- 1: **Input** D \triangleright Datensatz $D \in \mathbb{R}^{m \times n}$, Klassenlabel y und Featurespalten X_1, \dots, X_n
 - 2: **for all** Feature X_i mit $i \in \{1, \dots, n\}$ **do**
 - 3: Generierung aller möglichen Projektionen für X_i
 - 4: Sortierung nach y bezüglich X_i
 - 5: Bestimmung Labelchanges (mit Worst-Case-Szenario)
 - 6: **end for**
 - 7: Generierung Rangliste anhand Labelchanges
 - 8: **return** Feature-Rangliste
-

2.3. Systematisches Literaturreview

Erst auf Grundlage der Bewertung des aktuellen Forschungsstandes wird die Formulierung und Gültigkeit der in Kapitel 2.4 aufgestellten Hypothesen ermöglicht sowie legitimiert. Zu diesem Zweck folgt dieses im Folgenden vorgestellte, systematische Literaturreview der Methodik nach [Fet06], siehe graue Ebenen im Ablaufdiagramm 2.26. Der Aufbau und die thematische Untergliederung der Forschungshypothesen in ein Haupt- und ein Nebenforschungsgebiet verlangen die Durchführung von zwei unabhängigen Literaturreviews. Diesen werden für eine bessere Übersichtlichkeit die Nummern 1 (Hauptthematik; Regelextraktion AOI-Prüfplan und ML/DL in AOI-Prozessen) und 2 (Nebenuntersuchung; Korrelation Montagewinkel/Tombstone in AOI-Daten) zugeordnet. Die erste Phase des systematischen Literaturreviews beinhaltet die Formulierung der Problemstellung, siehe oberste Ebene in Abbildung 2.26. Auf Grundlage dieser Anforderungen an das Literaturreview ist die Durchführung der Literatursuche in der zweiten Phase des Literaturreviews realisierbar, siehe zweite Ebene in Abbildung 2.26. Die generierten Suchergebnisse mehrerer Suchprozesse durch Nutzung verschiedener Datenbanken und Literaturquellen dienen als Eingangsgröße der Literaturlauswertung, in der ein mehrstufiges Ausschlussverfahren der wissenschaftlichen Veröffentlichungen anhand an die Forschungsaufgabe angepasster Kriterien durchgeführt wird. Hierzu wird im ersten Filterungsprozess unter Berücksichtigung der Metainformationen und Werktitel eine Vorauswahl getroffen. Die Kurzzusammenfassungen und Resultate der Publikationen, welche den Anforderungen entsprechen, werden in dem darauffolgenden Selektionsverfahren analysiert. Anhand der darin enthaltenen Informationen werden diese entweder von dem Literaturreview ausgeschlossen oder in dem nächsten Vorauswahldurchgang untersucht. In diesem wird eine Eignungsprüfung der wissenschaftlichen Arbeiten im Rahmen einer Volltextanalyse durchgeführt.

Die vierte Ebene des systematischen Literaturreviews nach [Fet06], die Literaturlauswertung und -interpretation, umfasst die Bewertung der verbliebenen Veröffentlichungen unter Berücksichtigung der Aufgabenstellung und Forschungshypothesen. Die letzte Phase des Literaturreviews stellt die Präsentation der Ergebnisse dar.

2.3.1. Literaturlauswertung

Die erste Phase des Literaturreviews umfasst zunächst die Bestimmung der zu verwendenden wissenschaftlichen Datenbanken. Daraufhin werden die Schlüsselbegriffe, welche auf der Problemstellung in Kapitel 1.2 basieren, festgelegt. Die Literaturlauswertung wurde zu Beginn der Forschungsarbeit 2019 und am Ende der Bearbeitung im Jahre 2022 durchgeführt, um eine Verletzung des Einzigartigkeitskriteriums dieser Dissertation gänzlich auszuschließen. Da während der Forschungszeit nicht zu vernachlässigende wissenschaftli-

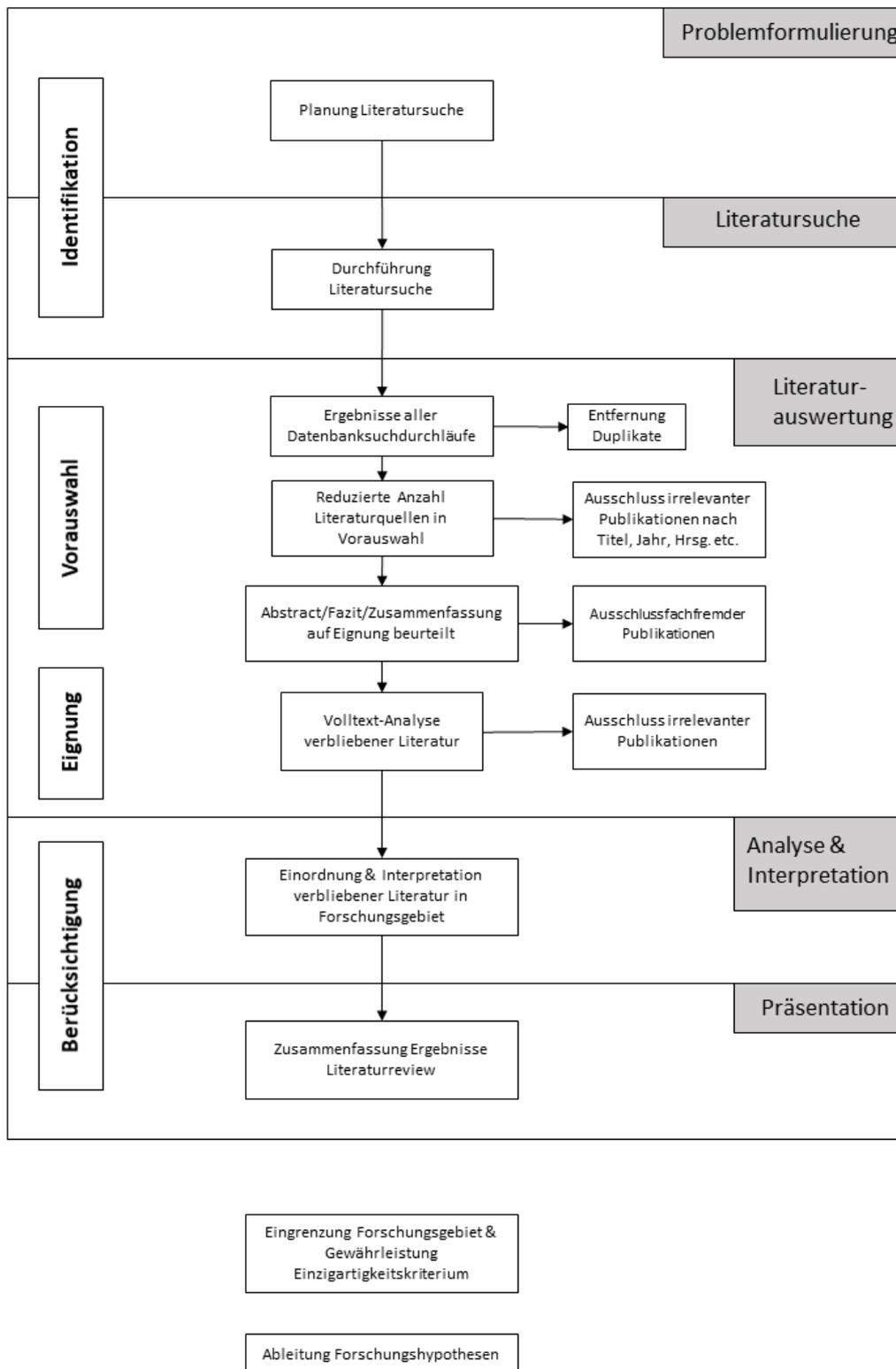


Abbildung 2.26.: Ablaufdiagramm Literaturreview (eigene Erstellung in Anlehnung an [Fet06] und PRISMA-Modell [PRI20])

che Arbeiten veröffentlicht wurden, werden die Resultate der aktuellsten Literatursuchen vorgestellt.

Tabelle 2.3 enthält die Zuordnung der Durchlaufnummerierungen zu den jeweils eingestellten Suchbegriffen. Diese werden als Phrase in die Suche integriert und durch die logische Und-Verknüpfung miteinander verbunden. Während der Parametrierung und Planung der Literatursuche erwiesen sich die wissenschaftlichen Fachdatenbanken Google Scholar, Springerlink, IEEE Xplore und ScienceDirect als effektive Umgebungen mit hoher Flexibilität an Einstellungsmöglichkeiten sowie umfassenden Datenbeständen. Eine Ergebniszusammenfassung der Literatursuche von Thema 1 ist in Tabelle 2.4 dargestellt.

Tabelle 2.3.: Literatursuche Thema 1 – Zuordnung Durchlauf/Schlüsselbegriffe

Durchlauf	Schlüsselbegriffe
1	"AOI" + "Decision Tree"
2	"AOI" + "Machine Learning"
3	"SMD" + "Machine Learning" + "PCB"
4	"Viscom" + "Machine Learning"
5	"Surface-Mounted" + "Feature" + "Machine Learning"
6	"Surface-Mounted device" + "Feature" + "Decision" + "Rule"
7	"AOI" + "SMD" + "Decision rule"
8	"SMD" + "AOI" + "Rule extraction"
9	"Visual inspection" + "SMD" + "PCB" + "Decision rule"
10	"SMD" + "AOI" + "Machine Learning" + "Threshold" + "Feature"

Tabelle 2.4.: Ausbeute Literatursuche Thema 1 – Zeitraum 14. Februar – 21. März 2022

Durchlauf Quelle	1	2	3	4	5	6	7	8	9	10
Google Scholar	1620	6470	613	68	1140	82	290	224	291	71
Springerlink	257	624	96	11	283	166	54	50	64	28
IEEE Xplore	4	60	2	0	1	0	0	0	0	0
ScienceDirect	152	353	34	4	152	71	14	7	19	3

Die Durchläufe 1 und 2 erreichten in Google Scholar 1620 beziehungsweise 6470 Treffer. In beiden Suchvorgängen setzen sich die Sucheinstellungen aus dem Begriff *AOI* gefolgt von einem Ausdruck aus dem Themenbereich *Machine-Learning* zusammen. Im Anschluss an eine Sortierung der Suchergebnisse nach Relevanz und Begutachtung der

jeweils ersten 1000 Suchergebnisse konnte eine Einschätzung der notwendigen Anpassungen nächstfolgender Suchprozesse erfolgen. Die in der Elektronikbranche und auch dieser Forschungsarbeit häufig verwendete Abkürzung *AOI* wird dem Prozess der automatischen optischen Inspektion zugeordnet. Jedoch existieren weitere Wort- und Begriffsgruppen mit der gleichen Abkürzung, sodass diese ebenfalls in die ersten beiden Suchdurchläufe eingeflossen sind und die Qualität der Suchergebnisse durch fachfremde, irrelevante Publikationen gestört wurde.

Die zahlreichsten Vertreter dieser Störungen sind Artikel und Veröffentlichungen mit den Schlüsselbegriffen *Age Of Information* (aus der Informatik), *Area Of Interest* und *Attribute Oriented Induction*. Bei dem Begriff *Attribute Oriented Induction* handelt es sich um eine Algorithmusfamilie, die sich mit der Wissensextrahierung relationaler Daten im Rahmen einer ML-Clusteranalyse beschäftigt.

Der Forschungsgegenstand des Themas 1 dieser Arbeit verbindet den Reflowprozess oberflächenmontierter Leistungselektroniken, deren optische Inspektion und moderne ML-Verfahren miteinander. Aus diesem Grund besteht die zugehörige Literatursuche aus der inhaltlichen Kombination dieser Fachgebiete. Dabei schließen die Suchdurchgänge mit den Nummern 1–5 die Untersuchung des aktuellen Forschungsstandes hinsichtlich der Erweiterung konventioneller Verbindungstechnologie (Reflowlötten und AOI-Prozesse) um moderne Industrie 4.0-Verfahren ein. Eine Spezialisierung auf die Extraktion von Regeln sowie Grenzwerte in der Prozessparametrierung erfolgt in den verbliebenen Suchdurchgängen (Nummern 6 – 10). In die nächstfolgende Phase des Literaturreviews, die Literaturauswertung, fließen sämtliche Ergebnisse aller Durchgänge (bei Durchlauf 1, 2 und 5 in Google Scholar die ersten 1000 Veröffentlichungen) ein.

Zunächst werden in Tabelle 2.5 den Durchlaufnummerierungen die verwendeten Suchphrasen bezüglich des Nebenthemas zugeordnet. Da bereits in der Datenbanksuche des Themas 1 mehrere Durchläufe bezüglich der Synthese von Lötverfahren und ML-Anwendungen durchgeführt wurden, entfällt diese Analyse in der aktuellen Planung. Es wird direkt eine Spezialisierung auf den geforderten Lötfehler (Tombstone) und die Einflussanalyse des Montagewinkels von SMD-Komponenten durchgeführt. In der Literatur werden für den Tombstone-Defekt häufig verwendete Synonyme wie Drawbridging und Manhattan-Defekt verwendet, welche ebenfalls in eigenen Suchprozessen (Nummern 1 und 7) Anwendung finden.

Die Literatursuche setzt sich aus drei Hauptthematiken zusammen. Zwei dieser Themen befassen sich mit dem Einfluss des Montagewinkels und des Leiterplatten-Designs auf die Entstehung von Tombstone-Defekten. Die dritte Thematik ist die Untersuchung des Forschungsstandes bezüglich des Selbstheilungseffektes, welcher eine eigenständige Bauteilzentrierung des SMD-Bauteils hervorruft, im Reflowofen. Durch den anfänglich höheren Spezialisierungsgrad der Aufgabenstellung und folglich der Literatursuche des Themas 2

sind die Trefferquoten, siehe Tabelle 2.6, reduziert.

Tabelle 2.5.: Literatursuche Thema 2 – Zuordnung Durchlauf/Schlüsselbegriffe

Durchlauf	Schlüsselbegriffe
1	"SMD" + "Drawbridging" + "Angle" + "Placement" + "Reflow"
2	"SMD" + "Tombstone" + "Angle" + "Placement" + "Reflow"
3	"SMD" + "Tombstone" + "Angle" + "Placement"
4	"SMD" + "Tombstone" + "Position" + "Component"
5	"SMD" + "Tombstone" + "Influence" + "Angle"
6	"Surface-Mounted" + "Drawbridging" + "Angle"
7	"SMD" + "Manhattan" + "Angle" + "PCB"
8	"SMD" + "PCB design" + "Tombstone"
9	"Surface-Mounted" + "Tombstone" + "Angle" + "Placement"
10	"SMD" + "Self-alignment" + "Chip"

Tabelle 2.6.: Ausbeute Literatursuche Thema 2 – Zeitraum 28. Februar – 14. März 2022

Durchlauf \ Quelle	1	2	3	4	5	6	7	8	9	10
Google Scholar	31	37	39	110	48	35	167	279	61	234
Springerlink	2	6	18	28	16	4	22	15	16	33
IEEE Xplore	0	0	0	1	0	0	0	2	0	1
ScienceDirect	3	16	18	28	12	45	6	24	30	23

2.3.2. Literaturoauswertung

Die Relevanzbeurteilung und fachliche Einordnung von Publikationen unter Berücksichtigung der Aufgabenstellung wird in mehreren Stufen und Textanalysetiefen durchgeführt. Zunächst erfolgt in der ersten Stufe die Literaturselektion anhand der Titel und Metainformationen jedes Suchergebnisses. Die Metainformations-Kriterien sind wie folgt definiert:

- *Alter Literaturquelle:* Wissenschaftliche Arbeiten, welche vor dem Jahre 2000 veröffentlicht wurden, werden nur bedingt in das systematische Literaturreview aufgenommen. Ausnahmen sind gültige Schlüsselarbeiten als Grundlage von zahlreichen relevanten Forschungen.
- *Sprache:* Es werden ausschließlich Veröffentlichungen in englischer und deutscher Sprache berücksichtigt.

- *Herkunft*: Einschluss von Veröffentlichungen renommierter Universitäten, wissenschaftlicher Zeitschriften und international agierender Unternehmen. Leserbriefe, Blogbeiträge werden von der weiterführenden Analyse ausgeschlossen.

Die Anzahl der in die Volltextanalyse einfließenden wissenschaftlichen Arbeiten konnte durch Vorauswahl anhand der Zusammenfassungen und Fazite reduziert werden, sodass für Thema 1 insgesamt 38 und 20 Veröffentlichungen für Thema 2 ausgewertet werden.

2.3.3. Literaturanalyse und Präsentation

Die in der Literaturlauswertung ausgewählten Veröffentlichungen werden in dieser Phase des systematischen Literaturreviews einer Inhaltsanalyse und Einordnung in das Forschungsgebiet dieser Arbeit unterzogen. Zu diesem Zweck fließen sämtliche Metainformationen und der Volltext jeder einzelnen Publikation in die Betrachtung ein. Die Ergebnisse des Literaturreviews werden anhand definierter Kriterien strukturiert und graphisch dargestellt. Hierfür eignet sich nach [Fet06] die Nutzung eines morphologischen Kastens, welcher die wissenschaftlichen Forschungsberichte in unterschiedliche Charakteristika und Kategorien unterteilt. Die relativen Anzahlen der Publikationen werden in jeder Subkategorie angegeben.

Literaturanalyse und Präsentation Thema 1 Die Systematisierung der 38 Publikationen hinsichtlich des Themas 1 umfasst elf Hauptattribute, siehe Abbildung 2.27.

Aufgrund der rasch fortschreitenden technologischen Entwicklungen wird dem *Jahr der Veröffentlichung*, Punkt 1, eine große Bedeutung beigemessen. 47 % der wissenschaftlichen Publikationen wurden zwischen den Jahren 2010 und 2020 und 32 % ab 2020 veröffentlicht. Zwar werden kontinuierlich moderne Verfahren und leistungsstarke Umsetzungen in den Fachgebieten des maschinellen Lernens und der automatischen optischen Inspektion entwickelt, jedoch existiert eine Latenzzeit bis zur Veröffentlichung wissenschaftlicher Arbeiten bezüglich der Synthese dieser beiden Einzelfachgebiete. Aufgrund dessen werden unter Vorbehalt ebenfalls ältere Forschungsberichte in der Literaturanalyse berücksichtigt.

Die Unterteilung des Literaturkatalogs in Bezug auf den *Fokus*, Punkt 2, der Forschung zeigt, dass mit 42 % die automatische optische Inspektion am häufigsten vertreten ist.

Das dritte Charakteristikum ist das verfolgte *Ziel* der Publikation. Eine reine Konzentration auf das maschinelle Lernen ohne Berücksichtigung von Aspekten der Inspektionstechnologie konnte in 11 % der Fälle ermittelt werden. Die digitale Bildverarbeitung und die optische Inspektion stellen mit 18 % und 42 % einen Großteil des Forschungsstands dar. Nicht jeder untersuchte Wissenschaftsbericht formuliert die Zielsetzung von den Autoren

expliziert. In der reinen *Klassifikation* beschränken sich insgesamt 24 % der Forschung auf die Generierung eines optimierten Verfahrens/Modells zur Klassentrennung anhand unterschiedlicher Anforderungen und Attribute. Die Einbettung eines intelligenten Modells in einen praktischen Anwendungsfall und somit die *Prozessoptimierung* ist mit insgesamt 45 % am häufigsten vertreten. Im Gegensatz dazu nehmen lediglich 11 % eine rein theoretische Betrachtungsweise im Rahmen der *Grundlagenforschung* ein.

Die Gestaltung einer Forschungsarbeit erfolgt vor dem Hintergrund einer neutralen *Perspektive* oder einer kritischen Auseinandersetzung mit aktuellen Forschungsmethoden und -inhalten, siehe Punkt 4.

Charakteristik		Kategorie					
1. Jahr der Veröffentlichung		Bis 2010 (21%)		Bis 2020 (47%)		Ab 2020 (32%)	
2. Fokus		ML (11%)	AOI (42%)	SMD-Prozesse (18%)	Digitale BV (18%)	Sonstiges (11%)	
3. Ziel	Formulierung	Expliziert (74%)			Nicht expliziert (26%)		
	Inhalt	Klassif. (24%)	Grundlagenf. (5%)	Prozessopt. (45%)	Klassif. (5%)	Grundlagenf. (11%)	Prozessopt. (11%)
4. Perspektive		Neutral (45%)			Position (55%)		
5. Daten	Inhalt	Numerisch (24%)		Optisch/Bilder (37%)		Konglomerat (39%)	
	Herkunft	Simuliert (3%)	Prozess (21%)	Simuliert (5%)	Prozess (21%)	Simuliert (8%)	Prozess (24%)
6. Zielgruppe		Allgemein (13%)		Forschung (16%)		Praxis/Produktion (71%)	
7. Resultat		Theor. Wissen (16%)		Algorithmus (11%)	Methode (53%)	Prozessopt. (20%)	
						allg. (13%)	herstellersp. (7%)
8. Validierung	Formulierung	Expliziert (68%)			Nicht expliziert (32%)		
	Umgebung	Theorie (34%)		Praxis (34%)		Theorie (18%)	
						Praxis (13%)	
9. Transferbarkeit		Expliziert (32%)			Nicht expliziert (68%)		
10. Zukünftige Forschung		Expliziert (58%)			Nicht expliziert (42%)		
11. Literatur-analyse	Auswahl	Expliziert (87%)			Nicht expliziert (13%)		
	Umfang	Schlüsselarbeit (26%)		Repräsentativ (61%)		Selektiv (13%)	
						Vollständig (0%)	

Abbildung 2.27.: Morphologischer Kasten Thema 1 nach [Fet06]

Die in den Forschungen verwendeten *Daten* und deren Herkunft, siehe Punkt 5 in Abbildung 2.27, sind für die Interpretation der praktischen Umsetzbarkeit relevant. Die Generierung und Anwendung neuer Methoden stützt sich hierbei entweder auf rein numerische Daten beziehungsweise optische AOI-Daten (Bilder) oder deren Vereinigung, wobei eine weitere Unterteilung in Bezug auf die Datenherkunft (simuliert/aus realem Prozess gewonnen) vorliegt.

Die Adressaten, Punkt 6 *Zielgruppe* in Abbildung 2.27, an die die wissenschaftlichen Arbeiten gerichtet sind, unterscheiden sich hauptsächlich in ihren Kernkompetenzen voneinander.

Punkt 7 fasst die Eigenschaften der *Resultate* zusammen. Erreichte Prozessverbesserungen lassen sich weiterhin in herstellereinspezifische (7 %) oder allgemein anwendbare (13 %)

Optimierungsschritte segmentieren. Als dominierend erweist sich in der Charakterisierung die Vorstellung neuartiger Methoden mit 53 %.

Da die Problemstellung dieser Dissertation die Einbettung von ML-Verfahren in die Qualitätsanalyse von Leistungselektroniken beinhaltet, werden der Validierungsform und -art neuer Modelle, Punkt 8 *Validierung*, eine besonders hohe Bedeutung zugewiesen. Die Nachweisführung der Modellperformance und -genauigkeit teilt sich hierbei in eine theoretische und praktische Durchführung auf.

Die *Transferbarkeit* der Ergebnisse/Methoden auf andere Forschungsbereiche oder abweichende Prozesse (andere System-Hersteller, Übertragung Ergebnisse AXI/SPI auf AOI, Unterschiede der digitalen Bildverarbeitung) wird in Punkt 9 dargestellt.

Punkt 10 unterscheidet die Literatur anhand der Angabe und Beschreibung der *zukünftigen Forschung*.

Die Aussagekraft und wissenschaftliche Gültigkeit von Forschungsberichten steht in engem Verhältnis zu deren Literatúrauswahl und der Qualität der zugrundeliegenden Quellen, siehe Punkt 11 *Literaturanalyse* in Abbildung 2.27.

Die Einzigartigkeit des Forschungsfeldes und der Problemstellung des Themas 1 dieser Arbeit bleibt auch nach Durchführung eines systematischen Literaturreviews bestehen. Nicht jede Publikation enthält zielführende Informationen. Die Ursachen hierfür sind vielfältig, wie beispielsweise nicht auf diese Problemstellung übertragbare AOI-Featurewerte, veraltete ML-Algorithmenfamilien oder stark divergierende und unüberbrückbare Prozessunterschiede.

Vorstellung relevanter Literatur [Gao11] stellt eine alternative Inspektionsmethode, basierend auf der Definition von Subregionen in AOI-Bildern, der Extraktion von Featurewerten und anschließender Generierung von Klassifikationsmodellen, vor. Die Bildakquisition erfolgt unter Zuhilfenahme einer hemisphärischen trichromatischen Beleuchtung und stellt somit den ersten Prozessunterschied zu dieser Forschungsarbeit dar. Die auf dem Bild dargestellten SMD-Chipbauteile werden in [Gao11], ähnlich der digitalen Bildverarbeitung der von Continental genutzten AOI-Systeme, Unterregionen aufgeteilt (Maskierung). Diese Unterregionen dienen der Extraktion von Featurewerten für jeden der drei Farbkanäle. Sowohl die unabhängige Verarbeitung von drei Farbkanälen als auch die mathematischen Ansätze der Featurewertbildung und letztendlich deren numerische Repräsentation physikalischer Eigenschaften unterscheiden sich von dieser Arbeit. Ein optimierter AdaBoost-Algorithmus erlaubt die Auswahl der performantesten Featurewerte für jede Subregion. Anhand der Zuteilung der relevantesten Featurewerte wird für jede Subregion ein Klassifikator trainiert und die Einteilung des Bauteilzustandes in eine Fehlerkategorie durch die Verarbeitung der Ergebnisse der Subregionen in einem finalen

Entscheidungsbaum durchgeführt. Die Klassifikation der Lötstelle in sechs mögliche Lötfehlerkategorien wird in [Gao11] umgesetzt und erreicht eine Korrektklassifikationsrate von bis zu 97,3 %.

Die wissenschaftliche Ausarbeitung einer ML-Anwendung zur Identifizierung von Pseudofehlern in SMT-Fertigungsprozessen wird in [Sch20] untersucht und vorgestellt. Analog zur Vorgehensweise in dieser Forschungsarbeit werden Messwerte zur Generierung von ML-Modellen verwendet. Die Unterschiede bestehen jedoch in der Anzahl und Aussagekraft der verwendeten Featurewerte. [Sch20] berücksichtigt lediglich vier numerische Messwerte und drei Metainformationsdaten (Subpanel, Komponententyp, Komponenten-Identifikationsnummer). Diese Dissertation verarbeitet die gesamten von dem AOI-System generierten Messwerte aus den Kategorien Klassifikation, Geometrie, Bildverarbeitung und allgemeine Analyseinstellungen. [Sch20] vergleicht die Klassifikationsleistung drei unterschiedlicher ML-Methoden (neuronales Netzwerk, Random Forest und kNN-Algorithmus) miteinander, wobei der Random Forest-Klassifizierer mit 94 % die höchste Accuracy erzielt. Die Extrahierung von AOI-Prüfplanparametern zwecks AOI-Prozessoptimierung und Verbesserung der Einteilung in Echtfehler und fehlerfreien Zustand ist nicht Bestandteil von [Sch20].

[Shi19] untersucht die Einflussfaktoren auf die Entscheidungsfindung in der automatischen optischen Inspektion und nutzt die erhobenen Fehler- sowie Bilddaten zur Prozessparametrierung im Rahmen der Qualitätsbeurteilung von TFT-LCD-Erzeugnissen (im Gegensatz zu oberflächenmontierten Leistungselektroniken in dieser Arbeit). Die Metainformationen, Anzahl und Charakteristika (welche physikalische Eigenschaft repräsentiert wird) der verwendeten Featurewerte ist in [Shi19] nicht ersichtlich, sodass lediglich die Methodik dieser wissenschaftlichen Arbeit berücksichtigt werden kann. Ein Naive-Bayes-Algorithmus, eine logistische Methode und eine Clusteranalyse (k-Means) dienen der experimentellen Versuchsdurchführung in [Shi19].

Eine Verbindung von ML-Anwendungen in der SMT-Produktion und der Technologie des Edge-Cloud-Computings wird in [Deu06] hergestellt. Messwerte der Pasteninspektion (SPI) dienen als Eingangsdaten für die Generierung und Validierung von Algorithmen des überwachten Lernens und umfassen Support Vector Machines, Naive-Bayes-Klassifikatoren, die logistische Regression, Random Forests, kNN-Clustering und neuronale Netze. Der Trainingsvorgang findet in einer Cloud-Umgebung statt. Die Bereitstellung des erstellten Modells erfolgt durch ein lokales Edge-System, auf welches die Fertigungssysteme in Echtzeit zugreifen, die Prozessdaten unverzüglich in der intelligenten ML-Anwendung verarbeitet und Klassifikationsergebnisse generiert werden können.

Literaturanalyse und Präsentation Thema 2 Die Strukturierung und thematische Einordnung des Themas 2 erfolgt anhand ähnlicher Charakteristika, jedoch wird statt der Validierung von ML-Modellen eine Einflussuntersuchung des Montagewinkels auf bestimmte Defektarten in Punkt 7 inkludiert. Die graphische Darstellung des morphologischen Kastens in Abbildung 2.28 zeigt, dass 65 % der untersuchten Publikationen zwischen den Jahren 2010 und 2020 veröffentlicht wurden.

Die in dem Literaturkatalog behandelten Kernthemen in Punkt 2 *Fokus* befassen sich zu 50 % mit dem reinen Lötprozess (Reflowlöten, Dampfphasenlöten), gefolgt von einer Untersuchung des gesamten SMT-Prozesses (30 %). Lediglich 10 % der Forschungsberichte thematisieren den Einfluss des Layouts von Leiterplatten.

Die *Zielsetzung*, Punkt 3, konnte in zwei Unterkategorien, Prozessoptimierung und Grundlagenforschung eingeteilt werden.

Die *Perspektive* der Forschungsarbeiten weist in 80 % der Analysen einen neutralen Standpunkt auf, siehe Punkt 4.

Die am häufigsten adressierte *Zielgruppe* stellt in Punkt 5 die praxisbezogene Produktion mit 60 % dar.

Eine ausgewogene Unterteilung ist in Punkt 6 *Resultat* erkennbar. 50 % der analysierten Quellen beinhalten die Erhebung von theoretischen Erkenntnissen und weitere 50 % unterbreiten Vorschläge zur Prozessoptimierung, welche entweder die Parametrierung bestehender Prozesse oder die Implementierung von alternativen Teilprozessen umfassen.

Der Begriff *Montagewinkel*, Punkt 7, wird in 5 % der Literatur erwähnt. Eine wissenschaftliche Ausarbeitung, die behandelt, wie der Entstehungsmechanismus des Tombstoneeffekts oder die Selbstzentrierung von Chip-Komponenten während des Lötvorgangs durch deren Montagewinkel beeinflusst werden, konnte im Rahmen des Literaturreviews nicht ausreichend wissenschaftlich fundiert identifiziert werden.

90 % der in den Quellen erreichten Erkenntnisse werden durch experimentelle Untersuchungen erlangt, die sich aus Simulationen, realen Prozessen oder deren Kombination zusammensetzen, Punkt 8 *Erkenntnisgrundlage*.

Die *Transferbarkeit* und *zukünftige Forschung* in den Punkten 9 und 10 werden in einer Mehrzahl der Forschungsarbeiten expliziert angegeben.

Die Unterteilungen der *Literaturanalyse* der untersuchten Publikationen werden in Punkt 11 wiedergegeben.

Vorstellung relevanter Literatur In [III19] wird der Zusammenhang zwischen thermischen Abschattungsprozessen induziert durch große SMD-Bauteile und dadurch möglicherweise verursachte Tombstone-Defekte von SMD-Chipkomponenten experimen-

Charakteristik		Kategorie					
1. Jahr der Veröffentlichung		Bis 2010 (20%)		Bis 2020 (65%)		Ab 2020 (15%)	
2. Fokus		Bestück. (5%)	PCB (10%)	Löten (50%)	AOI (5%)	Gesamter SMD-Proz. (30%)	
3. Ziel	Formulierung	Expliziert (70%)			Nicht expliziert (30%)		
	Inhalt	Grundlagenf. (45%)		Prozessopt. (25%)	Grundlagenf. (5%)	Prozessopt. (25%)	
4. Perspektive		Neutral (80%)			Position (20%)		
5. Zielgruppe		Allgemein (5%)		Forschung (35%)		Praxis/Produktion (60%)	
6. Resultat		Theoretisches Wissen (50%)			Prozessoptimierung (50%)		
					Parametrierung (25%)	Alt. Teilprozess (25%)	
7. Einfluss Winkel		Expliziert (5%)			Nicht expliziert (95%)		
8. Erkenntnis- grundlage	Versuchsart	Experimentell (90%)				Theoretisch (10%)	
	Methodik						
9. Transferbarkeit		Simulation (10%)		Realer Proz. (65%)		Konglomerat (15%)	
10. Zukünftige Forschung		Expliziert (55%)			Nicht expliziert (45%)		
11. Literatur- analyse		Expliziert (60%)			Nicht expliziert (40%)		
Auswahl	Umfang	Expliziert (85%)			Nicht expliziert (15%)		
		Schlüsselarbeit (15%)	Repräsentativ (70%)		Selektiv (10%)	Nicht angegeben (5%)	

Abbildung 2.28.: Morphologischer Kasten Thema 2 nach [Fet06]

tell untersucht. Zu diesem Zweck wurden Test-Leiterplatten mit abweichenden Layout-Konfigurationen entworfen und hergestellt. Die Layoutplanung umfasst hierfür die Anordnung von Chip-Bauteilen der Gehäusegruppen 0805 und 0603 um massereichere 4036-PET-Folienkondensatoren und XAL-1010-Induktoren in unterschiedlichen Abständen und Orientierungen sowie die Herstellung von Kontroll-Leiterplatten gleicher Anordnung der Chip-Komponenten in Abwesenheit der erwähnten massereichen SMD-Bauteile. Nach dem Kondensationslötvorgang konnte kein Zusammenhang zwischen den Bauteilorientierungen und einem vermehrten Auftreten von Tombstone-Defekten hergestellt werden. [Ill19] Die wissenschaftlichen Ausarbeitungen bezüglich der Ursachenforschung der Aufrichtung von Chip-Bauteilen in SMT-Produktionsprozessen erfolgen in [Str13] unter Berücksichtigung des Oberflächen-Finishings von Leiterplatten. In [Sza13] wird eine Einflussanalyse der Leiterplatten-Schrumpfung während der thermischen Belastung und darauffolgende Ausdehnung im Anschluss an den Lötvorgang untersucht.

2.4. Grundkonzeptionierung der Forschungstheorie

Die im Rahmen des Literaturreviews gewonnenen Erkenntnisse dienen der Eingrenzung und Inhaltsdefinition des Forschungsfeldes dieser Arbeit. Die kritische Analyse von Antworten auf Forschungsfragen, die in anderen Publikationen bereits gestellt und untersucht wurden, erlaubt eine effiziente und zielgerichtete Formulierung eigener wissenschaftli-

cher Hypothesen.

Die in der Aufgabenstellung geforderte Extraktion von Prüfplanparametern der optischen Inspektion stellt das Kernthema dieser Forschung dar. In der Vergangenheit konnten durch die Verwendung von Verfahren aus dem Bereich ML/DL effiziente Methoden zur Prozessverbesserung und -unterstützung im Bereich der Elektronikfertigung erzielt werden. Diese Zusammenhänge und Forschungsergebnisse sind die Grundlage für die Formulierung der Hypothesen H1–H5 (Thema 1 in Kapitel 2.3).

Die Untersuchung der im Rahmen des AOI-Prozesses erhobenen Daten mit Klassenannotation unter Berücksichtigung der Orientierungswinkel der oberflächenmontierten Bauteile und der Fehlercodierung wird in Hypothese H6 (Thema 2 in Kapitel 2.3) thematisiert. Eine Gegenüberstellung der Performance-Metriken binärer und multikategorier Modelle des überwachten Lernens stellt den Abschluss und eine Beantwortung der letzten Forschungsfrage H7 dar.

Hypothese 1 (H1):

Die von AOI-Systemen generierten Messwerte weisen statistisch signifikante Unterschiede hinsichtlich ihres Bauteilzustandes auf.

Hypothese 2 (H2):

Die von AOI-Systemen generierten Messwerte folgen dem zentralen Grenzwertsatz.

Hypothese 3 (H3):

Es existieren von AOI-Systemen generierte Featurewerte, welche einen gesteigerten Einfluss auf das Klassifikationsergebnis (im Vergleich zu anderen Featurewerten) ausüben.

Hypothese 4 (H4):

Die Nutzung von Prüfplänen mit Grenzwerten, welche im Rahmen von intelligenten mathematischen Modellen ermittelt werden, hat eine bessere Klassentrennung mit geringeren Pseudo- und Falschklassifizierungsraten als konventionell angelegte Prüfpläne zur Folge.

Hypothese 5 (H5):

Die Unterstützung des konventionellen regelbasierten Inspektionsverfahrens durch moderne Deep-Learning-Methoden bewirkt eine Reduzierung der Pseudo- und Falschklassifizierungsraten.

Hypothese 6 (H6):

*Die von AOI-Systemen generierten Messwerte weisen statistisch signifikante Unterschiede hinsichtlich ihrer multikategorialen **Fehlerklassenzugehörigkeit** und **Orientierungswinkel** auf der Leiterplatte auf.*

Hypothese 7 (H7):

Die Performance-Metrik von multikategoriellen Modellen, die auf Basis von AOI-Messwerten generiert werden, übersteigen aufgrund der hinzugefügten Informationen die Leistung von binären Modellen, welche den Bauteilzustand in Echtfehler und Gutteil einordnen.

2.5. Vorgehensmodell und Methodik der Forschung

Unter Berücksichtigung des aktuellen Forschungsstandes und der definierten Arbeits-hypothesen konnte die Konzeptionierung sowie Ablaufplanung dieser Arbeit realisiert werden, siehe Abbildung 2.29. Diese beinhaltet eine Prozessanalyse, wodurch Verbesserungspotenzial identifiziert wird. Im Anschluss erfolgt die Formulierung der Aufgabenstellung. Die Durchführung des systematischen Literaturreviews basiert auf den Prozessinformationen und der definierten Problemstellung. Nach erfolgreicher Eingrenzung des Forschungsfeldes folgt die Akquise von Prozessdaten und die Generierung geeigneter intelligenter Modelle aus dem Bereich ML/DL.

Im Rahmen der Datenakquise erfolgt die Entwicklung eines softwaregesteuerten Lösungsansatzes zur Erhebung von AOI-Produktionsdaten mit der Anforderung, keine Prozessstörungen der SMT-Fertigungslinie zu verursachen. Die Datenstruktur und -formate sind hierbei so zu wählen, dass die erhobenen Daten den weiteren Forschungsschritten effizient zugeführt werden. Darüber hinaus ist die kontinuierliche Aufrechterhaltung höchster Datenqualität zu gewährleisten. Dies betrifft vor allem die Korrektheit der Klassenannotationen sowie Homogenität und Konsistenz der gesamten Datenpopulation.

Als zentraler Bestandteil der gesamten Forschungsaufgabe werden fünf SMD-Bauteilformen anhand unterschiedlicher Selektionskriterien ermittelt und im Anschluss deren Produktionsdaten erhoben sowie weiteren Untersuchungen zur Verfügung gestellt. Neben der Montagehäufigkeit werden bei der Auswahl der Testkomponenten/-produkte die Komplexität der Prüfstrategie (Anwendung zusätzlicher 3D-Inspektionsmethoden) sowie die aktuelle und zeitliche Entwicklung der Fehlerhäufigkeiten (fcpmo- und dpmo-Raten) berücksichtigt, siehe Kapitel 3.

Die Bewertung des Verbesserungspotenzials durch optimierte Prüfpläne und den Einsatz moderner ML-Modelle kann mithilfe der Etablierung einer lokalen Simulationsumgebung eines Inline-AOI-Systems ohne den Eingriff in reale Produktionsprozesse realisiert werden.

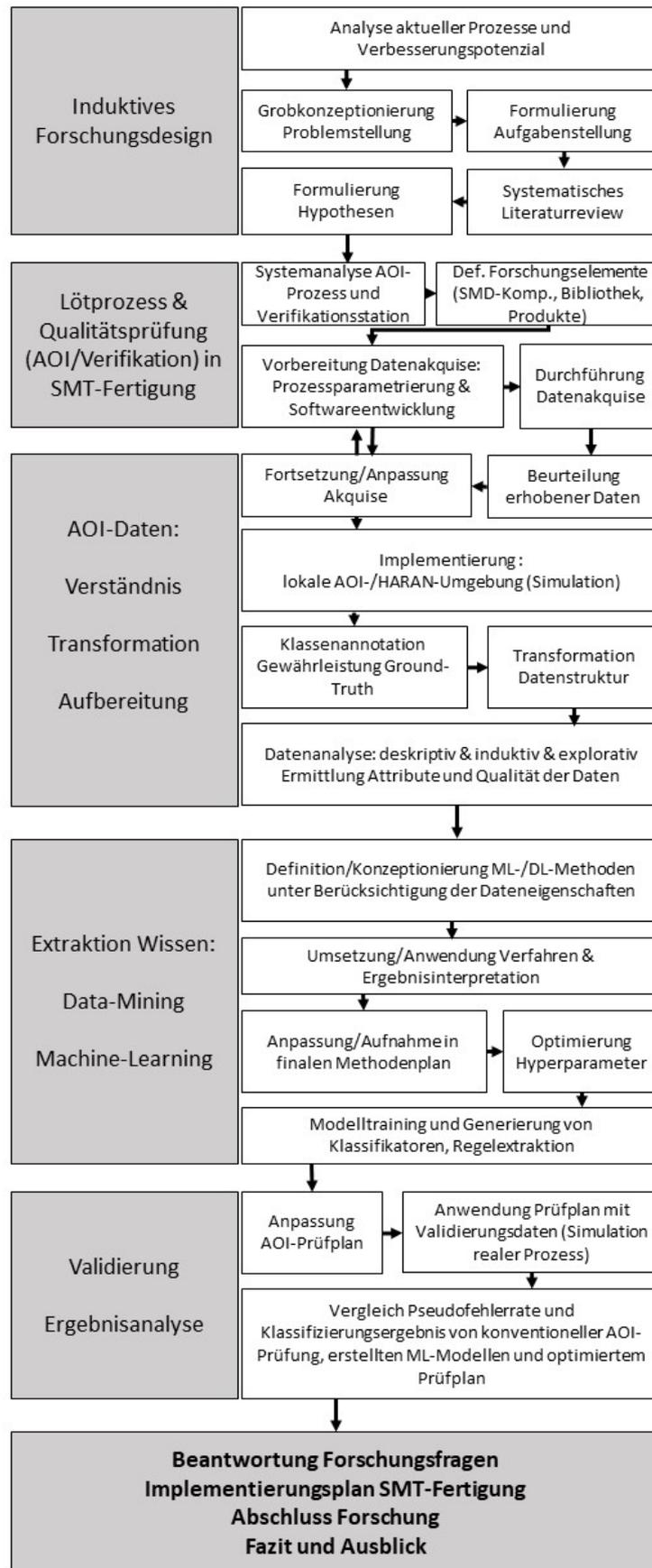


Abbildung 2.29.: Ablaufplanung der Forschungsinhalte

3. Datenakquise

Die Entwicklung von robusten Methoden zwecks Erhebung von SMT-Produktionsdaten, welche die an sie gestellten Qualitätskriterien zu jedem Zeitpunkt der Erhebung erfüllen, wird unter dem Begriff der Datenakquise zusammengefasst. Diese setzt sich aus zwei Kernaufgaben zusammen und beinhaltet einerseits die Ermittlung der zu untersuchenden SMD-Bauteile durch die Vereinigung unterschiedlicher Anforderungskriterien und andererseits die technische Umsetzung der Erfassung von Produktionsdaten. Eine detaillierte Beschreibung der Vorgehensweisen und Ergebnisvorstellung erfolgen in den Unterkapiteln 3.1 und 3.2.

3.1. Identifikation des Untersuchungsgegenstands

Die breite Produktpalette des Standortes Karben der Continental Automotive GmbH verhindert die Aufnahme des gesamten SMD-Bauteil-Portfolios in die Untersuchungen dieser Forschungsarbeit, sodass eine Reduzierung der Forschungsgegenstände erforderlich ist. Aufgrund der Beschaffenheit und Datenstrukturierung des AOI-Prozesses erfolgt zunächst die Ermittlung der globalen Prüfbibliothek und im Anschluss die Festlegung der zugehörigen Gehäuseformen unter Berücksichtigung unterschiedlicher Selektionskriterien, die im Folgenden näher beschrieben werden.

Verteilung des Produktionsvolumens Zu Beginn der Bauteilidentifikation erfolgt die Analyse der geplanten und bereits gefertigten Stückzahlen eines jeden Produktes. Zu diesem Zweck wird die vollständige, zur SMT-Produktion gehörige Produktpalette anhand ihres Anteils an der Gesamtproduktion bewertet. Nur Produkte, die über einen Zeitraum von mindestens 6 Monaten in hoher Stückzahl gefertigt werden, erreichen die nächste Analysestufe.

Fehlerhäufigkeit Die im vorigen Prozessschritt identifizierten Highrunner-Produkte des SMT-Produktionssegmentes werden unter Angabe ihrer Zugehörigkeit zu der übergeordneten, globalen Prüfbibliothek in Bezug auf die Echtfehler- und Pseudofehlerraten der auf dem jeweiligen Produkt oberflächenmontierten Bauteile untersucht. Hierbei fließt der

zeitliche Verlauf des Fehleraufkommens in die Untersuchung ein. Bauteile und Produkte mit großen zeitlichen Intervallen, in denen keine oder geringe Fehlerraten auftreten, und einzelne Peaks mit hohen absoluten Fehlerhäufigkeiten werden von der weiteren Analyse ausgeschlossen. Einzelne Peaks in der Fehleruntersuchung stellen einmalige Ereignisse (beispielsweise Musteranläufe) dar und weisen ein geringes Verbesserungspotential des Prozesses auf.

Prüfstrategie In jeder globalen Bauteilbibliothek sind die Prüfstrategien der einzelnen SMD-Bauteile nach ihren Gehäusebezeichnungen sortiert und hinterlegt. Ausnahmebehandlungen und Alternativprüfungen werden entweder global oder lokal auf dem Inspektionssystem definiert und gespeichert. Die Auswahl des Prüfplans sieht den Ausschluss von Alternativprüfungen vor, um die Datenhomogenität zu gewährleisten und einen hohen Generalisierungsgrad der trainierten Modelle zu erreichen. Moderne 3D-AOI-Analysen sind in die Forschung dieser Arbeit aufzunehmen.

Auswahl SMD-Komponenten und Prüfstrategie Im Anschluss an die Festlegung der SMD-Komponenten, der globalen Prüfbibliothek sowie der darin enthaltenen Analysealgorithmen ist die Eingrenzung der berücksichtigten AOI-Prüfsysteme erforderlich. In diesem Selektionsschritt steht die Gewährleistung einer linien- und maschinenübergreifenden Erzeugung von numerischen Messwerten, welche eine gültige Datenpopulation bilden, im Rahmen der optischen Inspektion im Vordergrund. Diese Anforderung betrifft sowohl den Prozess der Bildakquise als auch die daraufhin angewendeten Bildverarbeitungsmethoden und deren Parametrierung. Zu diesem Zweck werden ausschließlich AOI-Systeme des Unternehmens Viscom AG mit einheitlichem Sensorkopfaufbau beziehungsweise Bildakquisecharakteristika in den Forschungsplan aufgenommen.

Durch eine während des Erhebungsvorganges kontinuierliche Überwachung des Kalibrierungszustandes und die Durchführung von Systemjustierungen wird die Aufrechterhaltung der Datenintegrität gesichert. Lokale und somit ungültige Abschattungs-/Alternativprüfungen für spezielle Prüffenster werden durch Filtereinstellungen in beiden Datenerhebungsmethoden, siehe Kapitel 3.2, von der Messwertintegration in das Versuchsdesign ausgeschlossen. Des Weiteren wird eine identische Softwareversionierung aller ausgewählten Prüfsysteme gefordert. Ausschließlich AOI-Systeme der Produktreihen 6065 und 3088UC finden in der Datenerhebung Verwendung.

Es werden fünf SMD-Gehäuseformen, siehe Abbildung 3.1, in den Forschungsplan aufgenommen. Dabei handelt es sich um die vier Chip-Komponenten mit den Bezeichnungen R0402, R0603, C0402 und C0603, siehe Teilabbildungen 3.1a–d, und den Transistor mit der Gehäusebezeichnung SOT23, Abbildung 3.1e. In Tabelle 3.1 sind Informationen über die Prüfstrategie der Versuchskomponenten enthalten.

Die Anzahl der Bauteile pro Einzelleiterplatte wird mithilfe der relativen Häufigkeit $h_n(A)$ angegeben. Die Gesamtanzahl der verwendeten AOI-Analysealgorithmen variiert zwischen vier bei der Prüfung der Gehäuseform R0603 und sechs bei den SMD-Komponenten C0402 und SOT23. In Tabelle 3.1 wird die Anzahl der Vorlauffenster in Form einer Subanalyse der GENR-Prüffamilie angegeben. Diese weisen keine eigenständige Analyse auf, sondern ermöglichen die Bildaufnahme der ROI in einer definierten Bildaufnahmesituation und die Nutzung dieser optischen Informationen in weiteren Algorithmen (Multi-Image-Analyse). Obwohl der SMD-Chip R0603 keine 3D-Analyse aufweist, ist er Bestandteil der Forschung, da die erhöhten Fehlerraten eine Optimierung fordern. Die optische Inspektion der vier SMD-Chipkomponenten sieht eine orthogonale Lötstellen- und Bauteilprüfung vor, während diese im Falle einer Inspektion des Transistors SOT23 um eine geneigte Analyse erweitert wird.

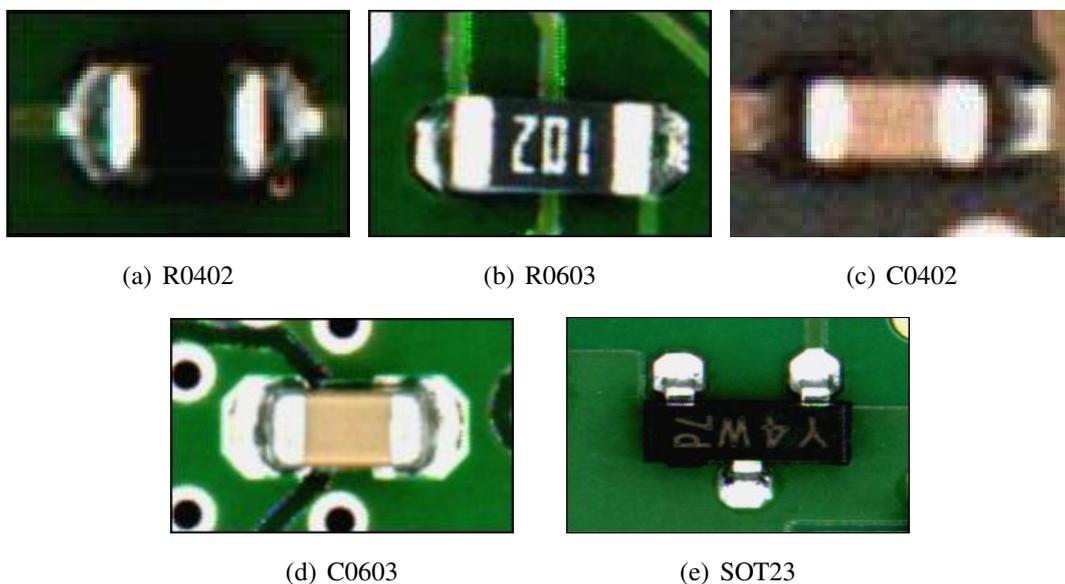


Abbildung 3.1.: Beispielbilder Prüfmuster (a)–(d) SMD-Chips; (e) SMD-Transistor

Tabelle 3.1.: Eigenschaften Prüfmuster

Eigenschaften Prüfmuster	$h_n(A)$	# Alg./GENR	Verlinkung?	3D?	Orth./angl.
C0402	0,135	6/3	ja	ja	ja/nein
C0603	0,111	5/2	ja	ja	ja/nein
R0402	0,176	5/2	ja	ja	ja/nein
R0603	0,097	4/2	ja	nein	ja/nein
SOT23	0,016	6/3	ja	ja	ja/ja

Um Fertigungsinformationen der Continental Automotive GmbH zu schützen, werden die in der Prüfstrategie definierten Analyse-Algorithmen in Tabelle 3.2 anonymisiert angegeben. Diese Bezeichnungen finden im gesamten Verlauf der Forschungsarbeit Verwendung. Bei den Analysen A, B, D und E handelt es sich um orthogonale 2D- beziehungsweise geneigte 2,5D-AOI-Verfahren (Analyse-D). Analyse-C ist ein Vertreter der 3D-Analysemethoden, die mithilfe eines Streifenlichtprojektors im AOI-System und Phasenverschiebungsverfahren eine Höhenkarte der SMD-Komponente erstellen und anschließende Beurteilungen über die Höhe, Koplanarität und Ausrichtung des Bauteils realisieren.

Analyse-A wird zur Anwesenheitsprüfung von Lötmaterial und SMD-Chipbauteilen verwendet. Die umfangreichere Analyse-B, ebenfalls für Chipkomponenten, untersucht die Bauteilanwesenheit/-position und -rotation. Der SMD-Transistor mit der Gehäusebezeichnung SOT23 wird den Analysen-D und -E unterzogen. Die Lötstellenkontrolle und Pinvermessung in Analyse-D wird durch Akquisition von geneigten Bildaufnahmen und anschließenden Bildverarbeitungsprozessen durchgeführt. Analyse-D überprüft die Anwesenheit, Verschiebung, Drehlage, Anwesenheit von Lötmaterial und -brücken sowie die Bauteilpolarität.

Die Featurewerte von AOI-Inspektionsalgorithmen mit der gleichen Bezeichnung weisen identische Zuordnungen zu den physikalischen Lötstellen-/Bauteilmerkmalen auf, so entspricht beispielsweise Featurewert 5 des Algorithmus B für alle vier SMD-Chipkomponenten der gemessenen X-Verschiebung des Bauteils im Prüffenster. Unterschiedliche Algorithmen besitzen abweichende Messwert-Bezeichnungs-Zuweisungen.

Tabelle 3.2.: Analysealgorithmen Prüfmuster

Prüfmuster	AOI-Algorithmen
C0402	Analyse-A Analyse-B Analyse-C (3D)
C0603	Analyse-A Analyse-B Analyse-C (3D)
R0402	Analyse-A Analyse-B Analyse-C (3D)
R0603	Analyse-A Analyse-B
SOT23	Analyse-C (3D) Analyse-D Analyse-E

3.2. Datenflussplan und Beschreibung der Prozessstufen

Übersicht Im Anschluss an die Festlegung der Prüfbibliothek einschließlich der Prüfpläne und der fünf SMD-Testkomponenten wird die technische Umsetzung der Datenakquise in der SMT-Serienfertigung entwickelt. Diesbezüglich steht die Erfüllung der

Anforderungen an die Datenqualität, insbesondere die Gewährleistung der Groundtruth der Klassenannotationen, im Vordergrund. Zu diesem Zweck werden unterschiedliche Softwarelösungen entwickelt und angepasst. Die Kombination der Einzelsoftwareprogramme ermöglicht die Realisierung der Datenakquise dieser Forschungsarbeit, siehe Abbildung 3.2. Hierbei teilt sich die Generierung der Datenbasis in zwei Extrahierungsmöglichkeiten auf. Einerseits werden die Featurewerte simultan zum AOI-Prozess inline erfasst, indem eine Applikation lokal auf dem jeweiligen AOI-System etabliert wird. Diese Methode hat den Vorteil, dass sämtliche Metainformationen und Messwerte mit Fail-Flags jedes Prüffensers der optischen Inspektion pro Mehrfachnutzen in einer separaten JSON-Datei aggregiert werden. Jedoch hat die dauerhafte Aktivierung dieser Software auf zahlreichen AOI-Systemen der SMT-Serienfertigung einen hohen Bedarf an Speicherplatz zur Folge. Die Sammlung einer ausreichenden Anzahl von Echtfehlern würde aufgrund der geringen dpmo-Raten der SMT-Fertigung des Standortes Karben eine permanente Aktivierung des Online-Lösungsansatzes erfordern. Des Weiteren beinhalten die generierten JSON-Ergebnisdateien keine Bildinformationen und Klassifizierungen der Verifikationsstation, sondern lediglich die Fail-Flags des AOI-Prozesses. Bedingt durch diese überwiegenden Nachteile findet die Online-Akquise nur für die Erfassung der Featurewerte fehlerfreier, nicht am AOI-System ausgefallener Prüffenster (i.O.-Klasse) Verwendung. Die zwangsläufige Neuentwicklung eines weiteren Verfahrens für die Erhebung von Echt- und Pseudofehlerdaten resultiert in der Offline-Methode, welche in Abbildung 3.2 durch einen Programmablaufplan dargestellt wird.

Vorstellung Offline-Methode Trotz des erhöhten Komplexitätsgrades der Offline-Methode erfüllt diese das gesamte Anforderungsprofil mit gleichzeitig höchster Flexibilität. Ein weiterer Vorteil liegt in der ermöglichten Zugriffsnahme auf archivierte Ergebnisdateien, bestehend aus Metainformationen, Bilddaten und Klassenannotationen durch den Operator der Verifikationsstation. Jeder Ausfall am AOI-System bewirkt die Erzeugung von drei Ergebnisdateien, welche im Anschluss dem Datenarchiv hinzugefügt werden. Die VPC-Datei dient der statistischen Überwachung und Erfassung des Fertigungsgeschehens, BB- und RES-Dateien umfassen sämtliche numerische und Bildinformationen der AOI-Prüfung und der anschließenden Verifikation.

Unter Zuhilfenahme der VPC-Software des Unternehmens Viscom (Statistische Prozesskontrolle) werden die eindeutigen Identifikationsnummern von inspizierten Mehrfachnutzen anhand vielfältiger Filtereinstellungsmöglichkeiten ermittelt und eine Prozessanalyse erstellt. Beispiele für die Filtereinstellungsmöglichkeiten sind die Angabe der Produktionszeit oder des Zeitintervalls, Prüfpläne, Gehäuseformen, Fehlercode, Analysealgorithmus mit Submodus oder die Maschinenbezeichnung.

Die in einer Paretoanalyse ermittelten Barcodes ermöglichen die Bereitstellung der zu-

gehörigen BB-/RES-Dateien mithilfe der Software RES-Finder. Um die Korrektheit der Klassenannotation zu gewährleisten wird jede einzelne Ergebnisdatei einer erneuten Verifikation unterzogen, indem eine Simulation der Software HARAN der Verifikationsstation lokal auf einem Computer getrennt von der Serienfertigung implementiert wird. Die in diesem Prozessschritt aktualisierten und kontrollierten Ergebnisdateien werden zunächst in eine erforderliche Datenstruktur umgewandelt (BB-Merger) und anschließend der Viscom-Online-Programmier-Station (VOPS) zur Verfügung gestellt. Dabei handelt es sich um die Simulation eines Inline-AOI-Systems, welches die realitätsgetreue Nachbildung der Prüfsoftware, Analysealgorithmen und Prozessabläufe darstellt. In dem letzten Prozessschritt der Offline-Methode der Datenakquise werden durch die VOPS-Umgebung und den Testprüfplan aus der SMT-Fertigung die AOI-Featurewerte extrahiert.

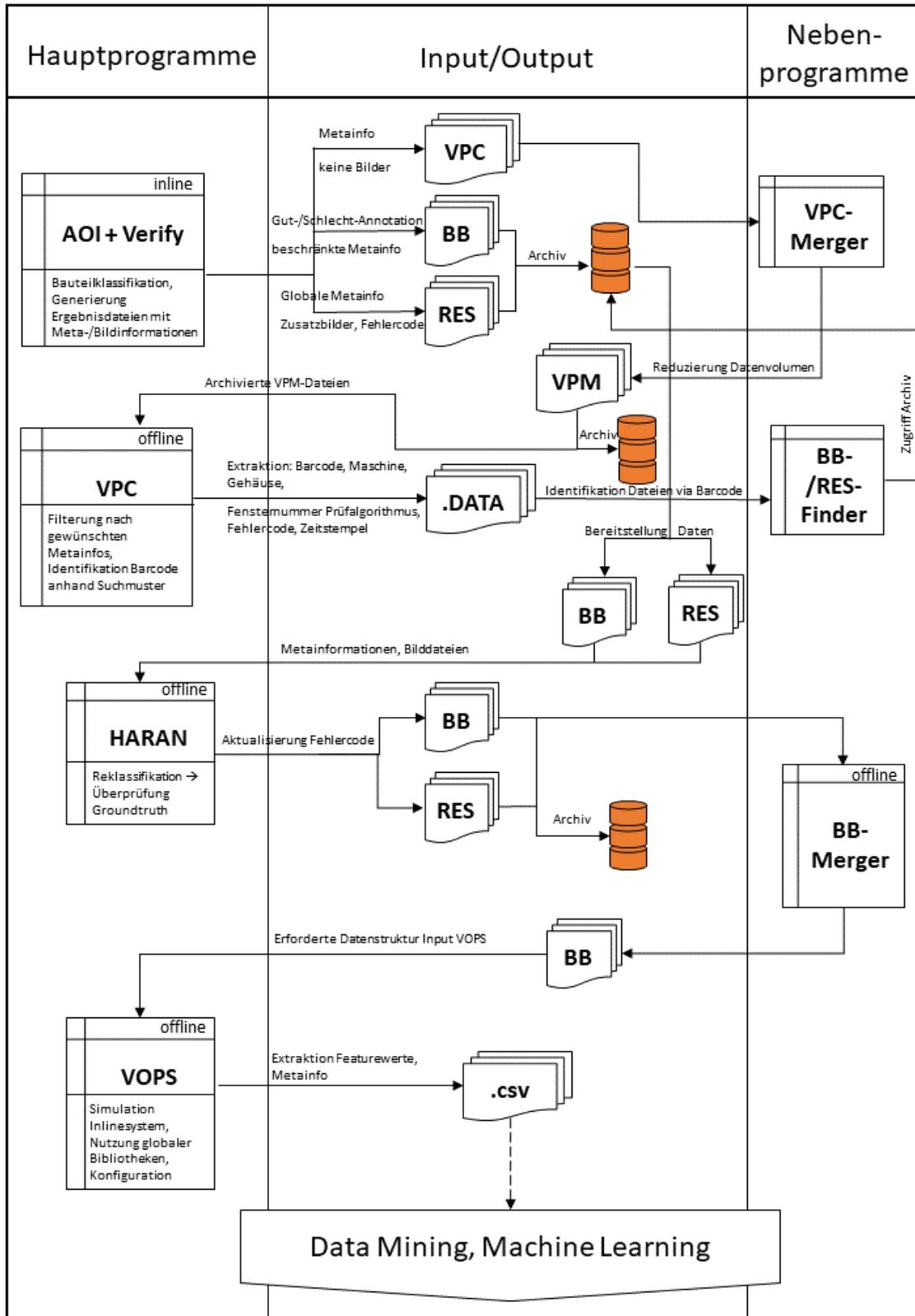


Abbildung 3.2.: Ablaufplanung der Datenakquisition

3.3. Datenstruktur und Inhalte der AOI-Informationen

Nach Ablauf des Erhebungszeitraumes werden die Daten bezüglich der ebenfalls akquirierten Metainformationen auf die Erfüllung der Qualitätsanforderungen überprüft und bereinigt. Infolgedessen werden doppelte, unvollständige oder inkonsistente Datenpunkte ausgeschlossen. Die Korrektheit der Klassenannotationen wird durch eine 100%-Sichtprüfung gewährleistet.

Der Bearbeitung des Hauptforschungsziels der Prüfplanoptimierung und featurebasierten AOI-Klassifizierung, durch die Arbeitshypothesen H1–H5 zusammengefasst, liegen Datensätze mit binären Klassenannotationen zugrunde. In dieser Betrachtung umfasst die i.O.-Klasse die im AOI-Prozess identifizierten Pseudofehler und die Ergebnisse von Prüfungen, welche ohne Auffälligkeit durchgeführt wurden und somit den Anforderungen der regelbasierten Klassifikation entsprechen. Datenpunkte der n.i.O.-Klasse entsprechen realen Löt- oder Bauteilfehlern.

Das Nebenforschungsgebiet (Hypothesen H6 und H7) thematisiert die Identifikation der Existenz eines Einflusses des Orientierungswinkels von SMD-Chipkomponenten auf die Entstehung von Tombstone-Effekten anhand der numerischen Messwerte der AOI-Prüfung. Zu diesem Zweck ist die Selektion von Datenpunkten mit Annotationen des Tombstone-Fehlercodes im Rahmen einer multikategoriellen Klassenunterscheidung erforderlich.

Jeder finale Datensatz D setzt sich jeweils aus der Anzahl der Prüfungen m , 128 unterschiedlichen Merkmalspalten und dem Klassenlabel y zusammen, sodass $D \in \mathbb{R}^{m \times 128}$ mit $y \in \mathbb{R}^{127 \times 1}$ gilt.

Tabelle 3.3 enthält die Zeilenanzahl (Anzahl Datenpunkte) jedes Subdatensatzes hinsichtlich des Themas 1. Aufgrund der Prozesseigenschaften der optischen Inspektion, im Besonderen die Ablaufstruktur der digitalen Bildverarbeitung des regelbasierten Klassifikationsverfahrens, ist es notwendig, die Datenpunkte nach ihrer SMD-Gehäuseform und des Analyse-Submodus getrennt zu erfassen, sodass aus der Datenakquise 14 Datensätze resultieren. In die weiterführenden ML-Anwendungen fließen ausgewogene Datensätze ein, sodass stets die gleiche Anzahl von Datenpunkten in jeder Klasse vorliegt und die Angabe der EF-Anzahlen genügt.

Die relative Häufigkeitsverteilung der auftretenden Löt-/Bauteilfehler hinsichtlich der fünf Testkomponenten und der vorliegenden Fehlercodes wird in Abbildung 3.3 dargestellt. Die Defektkategorien *Versatz* (FC 510), *Bauteil fehlt* (FC 520) und *Tombstone* (FC 560) sind insbesondere für die SMD-Gehäuseformen C0402 und C0603 am zahlreichsten vertreten und werden in den folgenden Kapiteln weiteren statistischen Analysen unterzogen.

Als Konsequenz der feingranularen Aufteilung der Datensätze nach Orientierungswinkel, Fehlercode, Analysesubmodus und SMD-Gehäuseform entstehen Datensätze mit geringen Beobachtungsmengen pro Datenfeld. Demzufolge werden lediglich die Daten der Gehäuseformen C0402 und C0603 für die multikategoriellen Untersuchungen verwendet. Die resultierenden Datensatzgrößen sind in den Tabellen im Anhang A.1 aufgeführt.

Tabelle 3.3.: Datensatz SMD-Testkomponenten – Anzahl Echtfehler – Binäre Klassenannotation i.O./n.i.O.

Bauform	Analyse	Anzahl EF-Datenpunkte
R0402	Analyse-A	360
	Analyse-B	799
	Analyse-C	1214
R0603	Analyse-A	249
	Analyse-B	388
C0402	Analyse-A	688
	Analyse-B	1487
	Analyse-C	1424
C0603	Analyse-A	924
	Analyse-B	2030
	Analyse-C	1897
SOT23	Analyse-C	419
	Analyse-D	1287
	Analyse-E	338

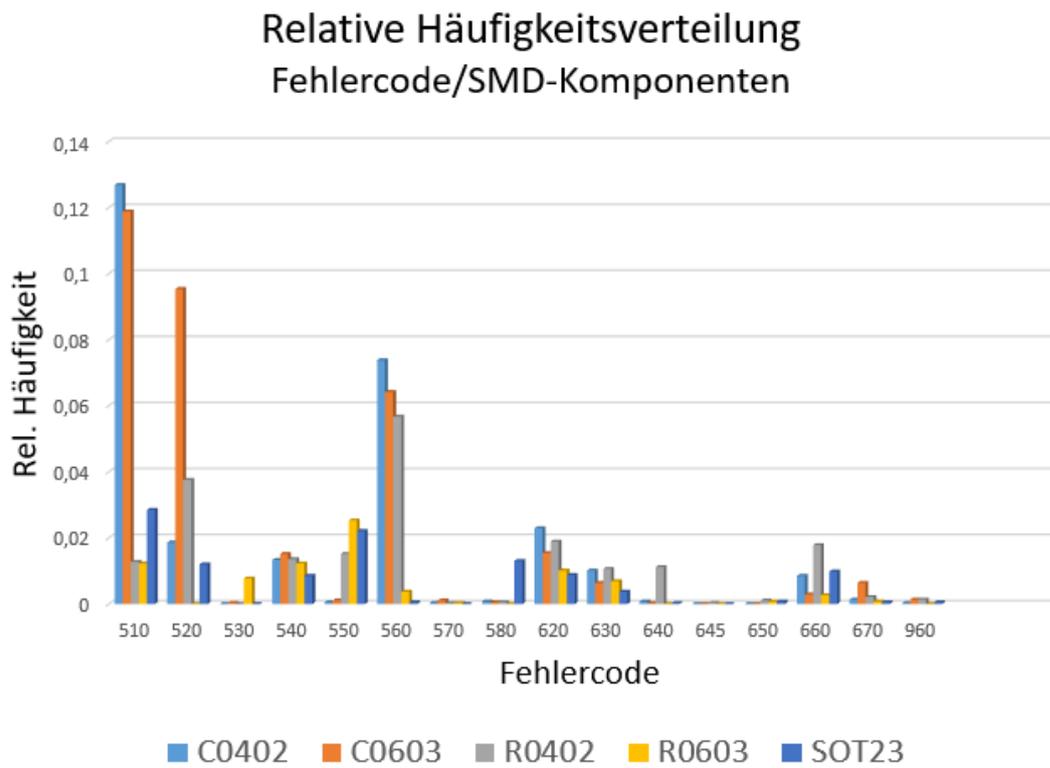


Abbildung 3.3.: Relative Häufigkeitsverteilung – Berücksichtigung sämtlicher Fehlercodes

4. Statistische Datenanalyse

Das folgende Kapitel beinhaltet die Methoden- und Ergebnisvorstellung der statistischen Untersuchung der im vorigen Abschnitt 3.3 erhobenen Datensätze. Zu diesem Zweck werden einzelne Statistikverfahren zu übergeordneten Kategorien gleicher Untersuchungsgegenstände zusammengefasst. Eine weitere Kapitelsegmentierung erfolgt anhand der beiden formulierten Forschungsaufgaben, die einerseits das Thema 1 (binäre Klassifikation und Prüfplanoptimierung) als Hauptforschungsgegenstand und andererseits die Nebenuntersuchung im Rahmen des Themas 2 (datenbasierte Einflussanalyse Orientierungswinkel/Tombstonefehler) beinhalten.

Die Datenexploration beginnt in Kapitel 4.1 zunächst mit der numerischen Ermittlung der Maße der zentralen Tendenzen, unterstützt durch graphische Darstellungsformen. Kernthema des zweiten Unterabschnittes 4.2 ist die Bestimmung der Korrelationskoeffizienten aller AOI-Messwerte untereinander und die Anwendung der Feature-Selektionsverfahren, welche in Kapitel 2.2.1.4 beschrieben werden. Die Resultate der Anpassungstests in Kapitel 4.3 dienen der Methodenauswahl und -durchführung von Dependenzanalysen in Kapitel 4.4. Diese beinhalten die Untersuchung und Identifikation signifikanter Unterschiede und Zusammenhänge innerhalb der Daten. Abschließend werden den relevantesten Featurewerten in Abschnitt 4.5 die korrespondierenden physikalischen Bauteil- und Lötstelleneigenschaften zugeordnet.

4.1. Maße der zentralen Tendenz und Datenvisualisierung

Dieses Kapitel beinhaltet die Vorstellung der Maße der zentralen Tendenz einer binären und multikategoriellen Untersuchungsreihe.

4.1.1. Binäre Hauptuntersuchung

Die Klassenannotationen der Datensätze bezüglich des Themas 1 liegen im direkten Anschluss an die Datenakquise und Reklassifizierung (Sicherstellung Groundtruth) in drei unterschiedlichen Klassen vor. Dabei handelt es sich um die Klassen der Echtfehler (bei

Identifikation eines Löt-/Bauteilfehlers), der Pseudofehler (Ausfall AOI, Bauteilzustand zulässig) und des i.O.-Zustandes (Pass AOI, Bauteilzustand zulässig).

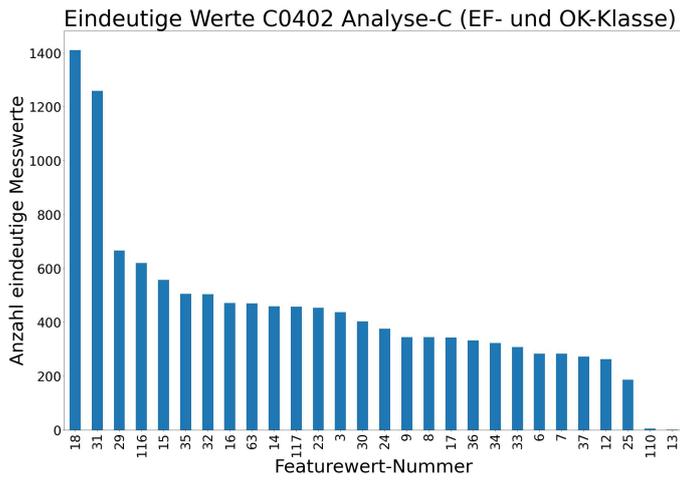
Aufgrund der geforderten Klassifikationsergebnisse nachfolgender ML-Verfahren werden die ternären Informationen durch Vereinigung der beiden fehlerfreien Zustandsklassen (PF und i.O.) zur OK-Klasse in binäre Datensätze transformiert. Gegenstand der statistischen Untersuchungen sind sowohl die ternären Ausgangsdaten als auch die generierten binären Datensätze als Eingangsgröße für ML-Methoden.

In keinem der untersuchten AOI-Analysealgorithmen wird die maximale Anzahl von 128 zu besetzenden Featurewert-Zuordnungen erreicht. Mit steigendem Komplexitätsgrad und Prüftiefe des korrespondierenden Bildverarbeitungs-Verfahrens sowie der Parametrierung des AOI-Prozesses steigt die Anzahl der mit numerischen Messwerten gefüllten Featurewert-Spalten in dem Datensatz. Im Gegensatz dazu wird nicht genutzten Feature-spalten der Wert 0 zugeordnet. Diese Prozesseigenschaft resultiert in verminderten und unterschiedlichen Merkmalsanzahlen innerhalb der Datensätze.

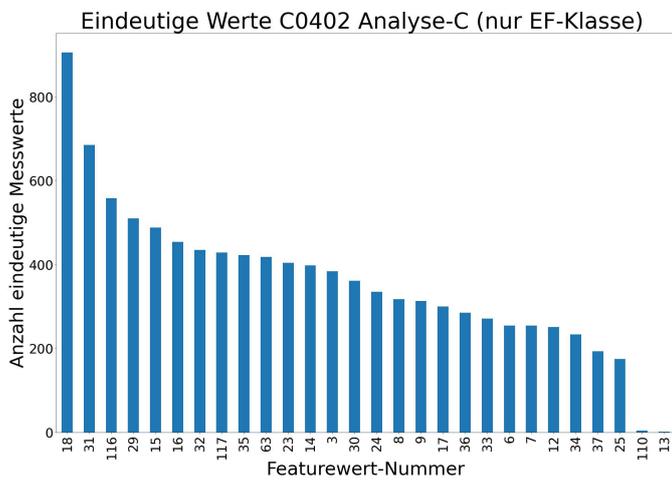
Analyse-A findet Anwendung bei der Lot- und Pastenanwesenheitsprüfung zweipoliger SMD-Komponenten, sodass aufgrund der geringen Prüftiefe und der simplen Prüfmethodik lediglich drei Messwert-Spalten besetzt sind.

Kardinalität Zu Beginn der statistischen Datenexploration ist dem Versuchsplan eine quantitative Untersuchung der erhobenen AOI-Informationen beigelegt. Nach ersten Sichtungen der Datensätze konnten Merkmalsspalten, die eine geringe Anzahl unterschiedlicher Messwerte enthielten, identifiziert werden. Aus diesem Grund wird für jede Kombination anhand des Analyse-Algorithmus, des Bauteils und der Fehlerklassenanzahl (ternär, binär) die Kardinalität der eindeutigen Elemente erhoben.

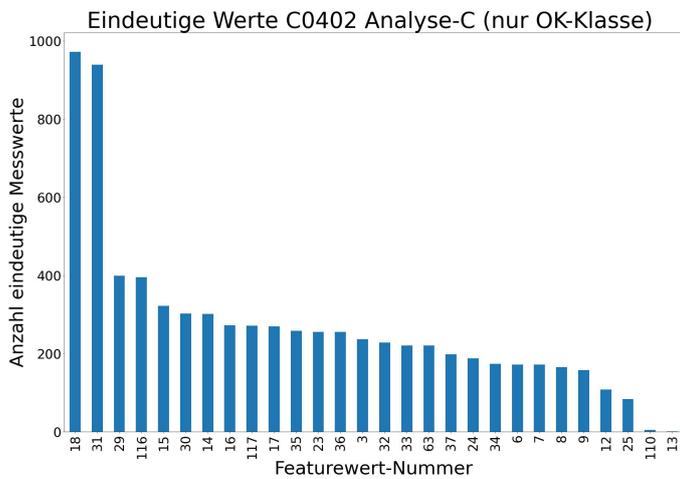
Der binäre Datensatz der Analyse-C und des Bauteils C0402 dient in Abbildung 4.1 als Informationsgrundlage. In jeder Teilgraphik weisen die Featurewerte 110 und 13 die geringsten Kardinalitäten mit einer Anzahl von 5 beziehungsweise 2 eindeutigen Messwertstufen auf. Im Gegensatz dazu erreicht Featurewert 18 in Abbildung 4.1 a eine Kardinalität von 1411 bei 2849 Datenpunkten. Die Standardabweichung beträgt 980,13. Die starke Ausprägung der Messwertstreuung weist auf eine geringe Repräsentativität des arithmetischen Mittelwertes und die Existenz statistisch ungültiger Ausreißer hin. Die nach den Fehlerklassen getrennte Kardinalitätsanalyse in den Subgraphiken 4.1 b und c ermöglicht die Feststellung, dass die in den Gesamtdaten beobachtete breite Messwertverteilung durch die EF-Klasse generiert wird. Die Kardinalitätsergebnisse der OK-Klasse in Abbildung 4.1 c der Featurewerte liegen im Vergleich zur EF-Klasse in Abbildung 4.1 b auf einem niedrigeren Niveau.



(a) Gesamter Datensatz



(b) EF-Klasse



(c) OK-Klasse

Abbildung 4.1.: Diagramme Kardinalität eindeutiger Messwertstufen – Analyse-C – C0402 – zwei Klassen (EF, OK)

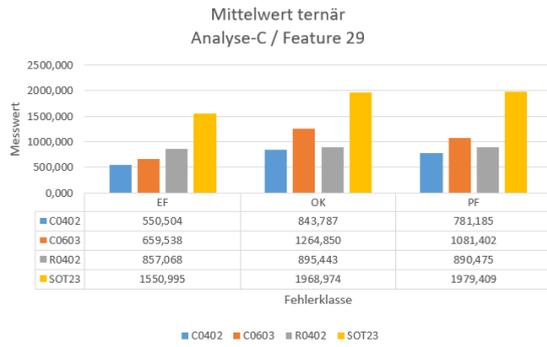
Dispersion und zentrale Tendenz In dieser Untersuchungsreihe erfolgt die Ermittlung von statistischen Kennwerten hinsichtlich der Lage- und Streuungseigenschaften einzelner Featurewerte. Sowohl Abweichungen der zentralen Tendenz als auch Gleichheit der Dispersionskennzahlen werden als Hinweise für bestehende signifikante Unterschiede zwischen den einzelnen Klassen interpretiert und stellen darüber hinaus Voraussetzungen für nachfolgende Hypothesentests dar.

Die Mittelwerte, Mediane, Standardabweichungen und deren Visualisierung mithilfe von Boxplot-Whisker-Diagrammen werden für die fünf SMD-Versuchskomponenten, vierzehn AOI-Subalgorithmen und jeden Featurewert in einer ternären (EF, PF, OK) und binären (EF, OK) Analyse bestimmt. Die Ergebnisse der Analyse-C dienen in den nachfolgenden Abbildungen und Erläuterungen als repräsentatives Beispiel für die gesamte Untersuchung. Abbildung 4.2 beinhaltet Mittelwerte, Standardabweichungen und Mediane von Featurewert Nr. 29 sowie die entsprechenden graphischen Darstellungen bezüglich der binären und ternären AOI-Informationen. Im Vergleich der Mittelwertkennzahlen sind für das Gehäuse R0402 sowohl im ternären als auch binären Datensatz die Klassenunterscheidbarkeiten zwischen der EF-Klasse und der OK-Klasse sowie PF-Klasse auf einem geringen Niveau. Die maximalen Mittelwertdifferenzen in der ternären und binären Untersuchungsreihe betragen für R0402 4,3% beziehungsweise 4%. Die Analyse der restlichen Bauteilformen erreicht höhere Mittelwertdifferenzen, wie 47,9% in der ternären und 44,2% in der binären Betrachtung des SMD-Bauelements C0603.

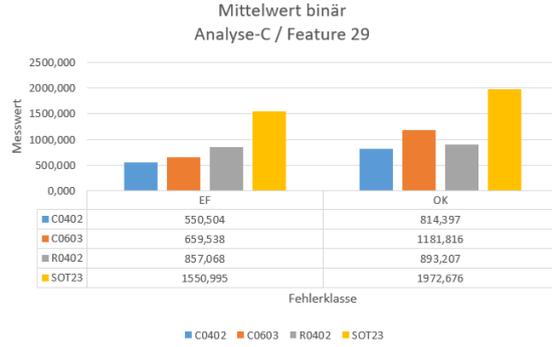
Die Standardabweichung der EF-Klasse weist in jeder Untersuchungsreihe deutliche Unterschiede gegenüber den OK- und PF-Klassen in Abbildung 4.2 c und d auf. Unterschiedliche Medianniveaus sind am stärksten ausgeprägt für das Bauelement C0603 nachweisbar und betragen 51,6% (ternär; EF-Klasse und OK-Klasse) sowie 49,7% (binär; EF-Klasse und OK-Klasse). Analog zu den Ergebnissen der Mittelwertuntersuchung erreicht auch hier das SMD-Gehäuse R0402 die geringsten Differenzen.

Im Gegensatz zu den nachweisbaren Unterschieden zwischen der EF-Klasse und den Datensätzen mit zulässigem Bauteil-/Lötstellenzustand sind derart eindeutige Differenzen im Vergleich der OK- und PF-Klassen nicht erkennbar.

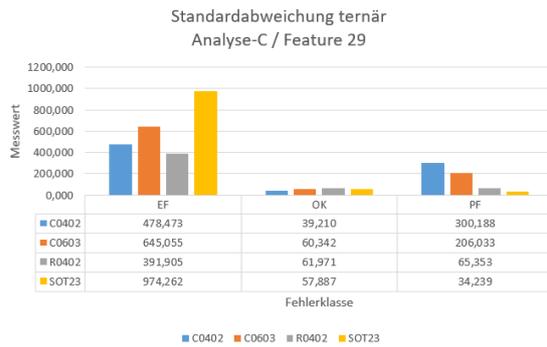
Diese Annahmen werden in den durchgeführten Dependenzversuchen in den folgenden Unterkapiteln überprüft.



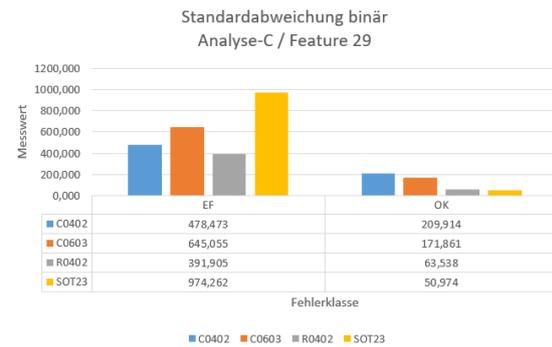
(a) Mittelwert drei Klassen



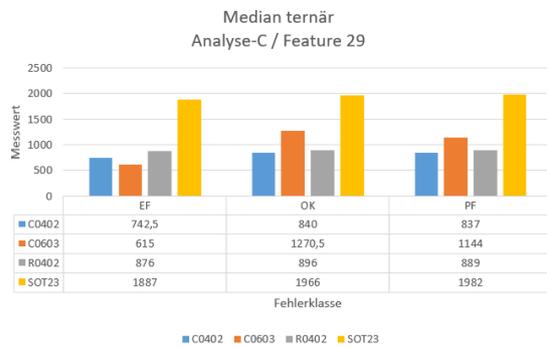
(b) Mittelwert zwei Klassen



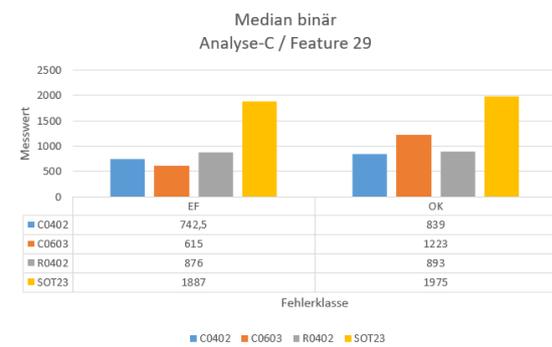
(c) Standardabweichung drei Klassen



(d) Standardabweichung zwei Klassen



(e) Median drei Klassen



(f) Median zwei Klassen

Abbildung 4.2.: Diagramme zentrale Tendenz – Analyse-C – Feature 29

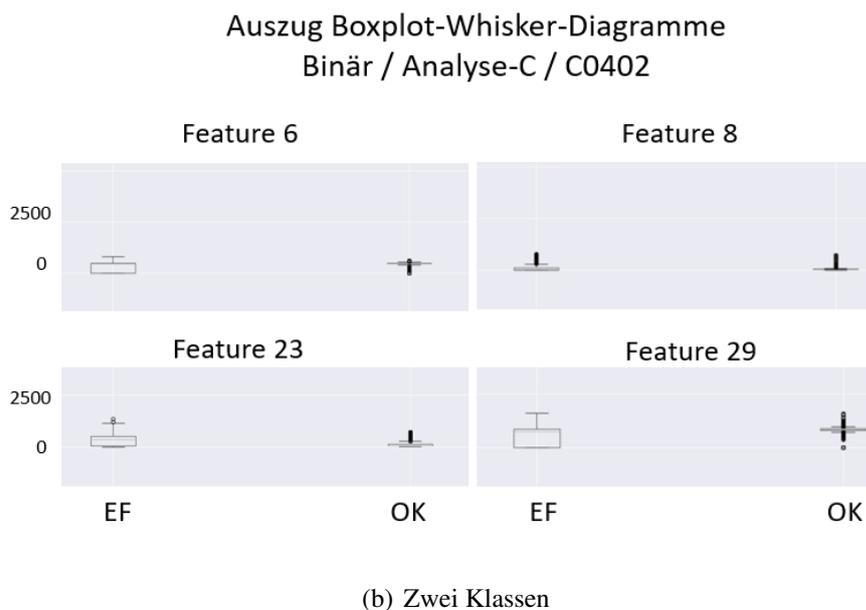
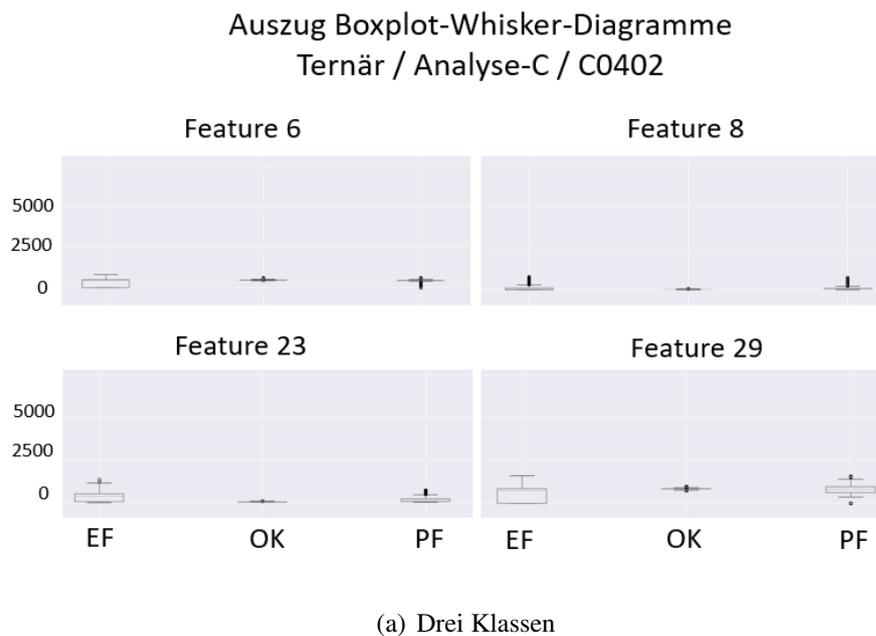
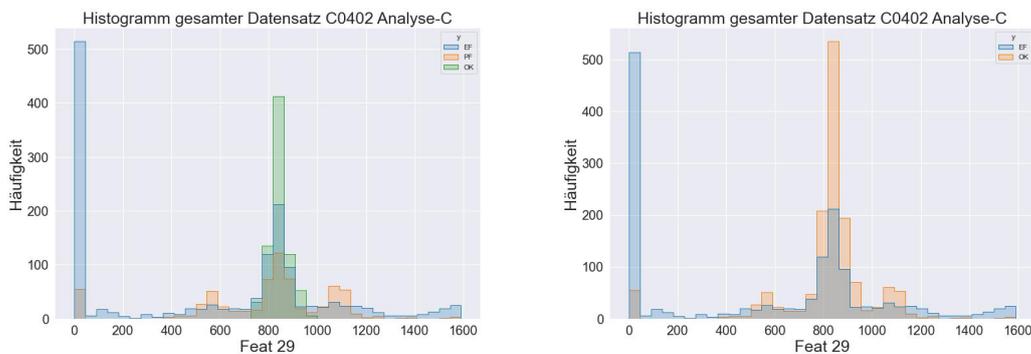


Abbildung 4.3.: Boxplot-Whisker-Diagramme – Analyse-C – C0402 – Feature 6, 8, 23, 29

Eine weitere Auffälligkeit liegt in dem vermehrten Auftreten von Ausreißern, verbunden mit geringen Interquartilsabständen. Dies ist ein Hinweis auf die Zugehörigkeit zu einer anderen Population. Mithilfe von Boxplot-Whisker-Diagrammen, siehe Abbildung 4.3, und Histogrammen, siehe Abbildungen 4.5 für Feature 6 und 4.4 für Feature 29, wird die Erfassung dieser Verteilungseigenschaften ermöglicht. Die Verteilungen getrennt durch die Klassenannotation der ternären und binären Untersuchungsreihen für Feature 29 wei-

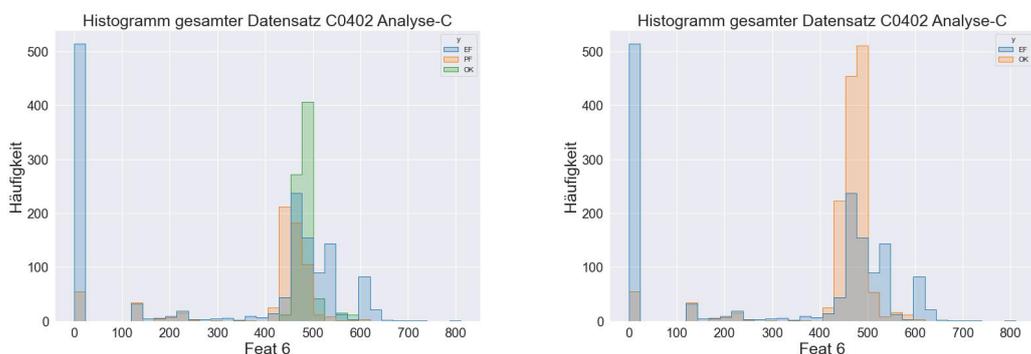
sen in der vorliegenden Skalierung optisch nicht unterscheidbare Mittelwertlagen auf. Die Breite der Verteilung ist, analog zu den Ergebnissen der Kardinalitätsanalyse, in der EF-Klasse am stärksten ausgeprägt. Eine erhöhte Anzahl von Datenpunkten mit dem Messwert 0 sind ebenfalls nachweisbar. Die Form der Histogramme stellen Anzeichen für die Existenz von symmetrischen Messwertverteilungen dar. Überdies wird anhand der op-



(a) Drei Klassen

(b) Zwei Klassen

Abbildung 4.4.: Histogramme – Analyse-C – C0402 – Feature 29



(a) Drei Klassen

(b) Zwei Klassen

Abbildung 4.5.: Histogramme – Analyse-C – C0402 – Feature 6

tischen Informationen durch die Generierung von Histogrammen einzelner Featurewerte und zweidimensionalen Streudiagrammen, siehe Abbildung 4.6, eine Einschätzung der Klassentrennbarkeit von zwei unterschiedlichen Featurewerten getroffen.

Da moderne ML-Algorithmen in der Lage sind, hochdimensionale Daten und darin enthaltene Informationen effizient zu verarbeiten, stellt die Analyse der vorgestellten Diagramme keine abschließende Bewertung dar. In der Messwertanalyse anhand von Histogrammen und KDE-Diagrammen (englisch Kernel Density Estimation), entlang der Hauptdiagonalen in Abbildung 4.6a und b, sind vielfach Überlappungen der Klassen erkennbar. In den Streudiagrammen in Abbildung 4.6, welche einen Messwertvergleich von

zwei Featurewerten untereinander repräsentieren, bilden sich jedoch Cluster von Datenpunkten gleicher Klassenannotation. Diese Cluster sind nicht voneinander getrennt, sondern weisen ebenfalls Überlappungen auf.

Der Anhang A.2.1 dieser Ausarbeitung enthält ergänzende Diagramme und Auswertungen weiterer Analysealgorithmen und SMD-Bauteile.

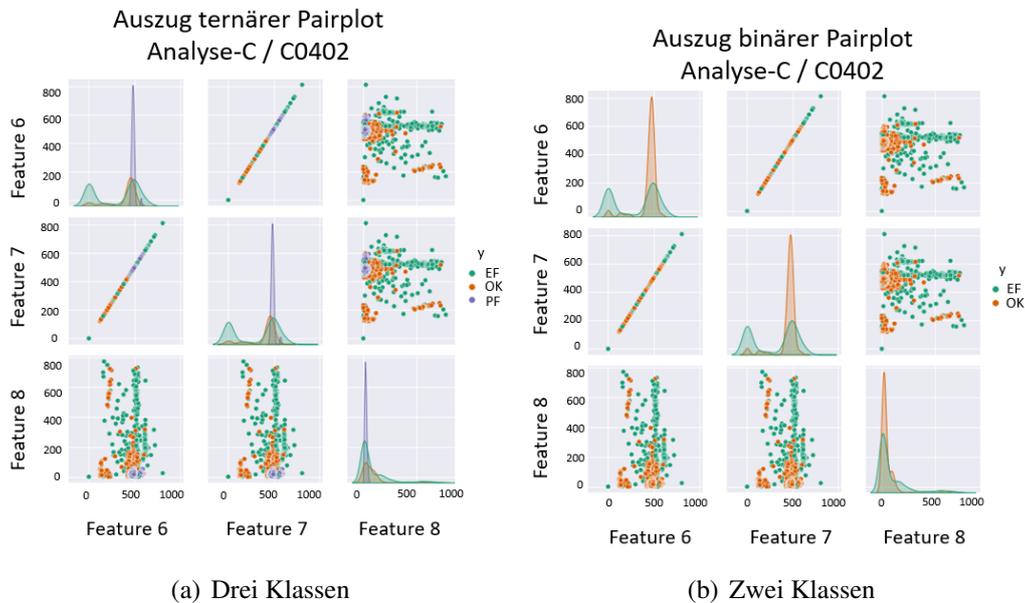


Abbildung 4.6.: Streudiagramme – Analyse-C – C0402 – Feature 6, 7, 8

4.1.2. Multikategorielle Nebenuntersuchung

Quantitative Analyse Die Datensätze der Nebenuntersuchung, Thema 2 (Einfluss Montagewinkel auf Tombstone-Häufigkeit), setzen sich aus EF-Informationen mit FC-Annotation und einer OK-Klasse, welche PF-Daten und AOI-Featurewerte ohne Auffälligkeit während des optischen Inspektionsprozesses enthält, zusammen. Darüber hinaus kann eine weitere Unterteilung der Daten anhand ihrer Montagewinkel erfolgen, sodass hinsichtlich der beiden Kategorisierungsgrößen (FC und Montagewinkel) die in Abbildung 4.7 dargestellte Häufigkeitsverteilung entsteht. Abhängig von dem Bestückungsplan der in dieser Untersuchung berücksichtigten Produkte und dem Datenbereinigungsvorgang im Anschluss an die Datenerhebung resultieren stark voneinander abweichende Anzahlen von Datenpunkten innerhalb der Subklassen. Obgleich Abbildung 4.7 die drei am häufigsten vertretenen EF-Fehlerklassen (Versatz, Bauteil fehlt und Tombstone), die OK-Klasse des Bauteils und den Analysealgorithmus mit der höchsten Anzahl an Daten berücksichtigt, entsteht eine unregelmäßige Häufigkeitsverteilung. In dieser sind alle vier Fehlerklassen nicht in jeder Winkelklasse vertreten. Infolgedessen und unter Berücksichtigung der Voraussetzungen statistischer Testverfahren werden zwei der fünf SMD-

Bauteile, C0402 und C0603, der Datenexploration unterzogen.

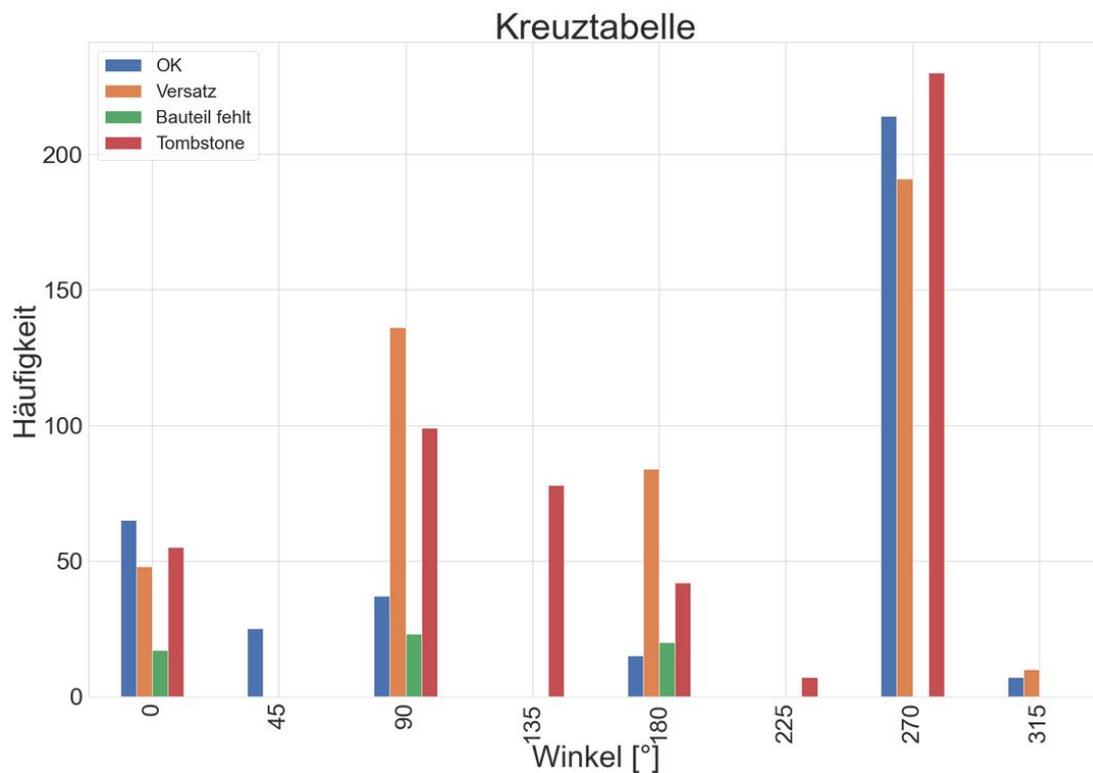
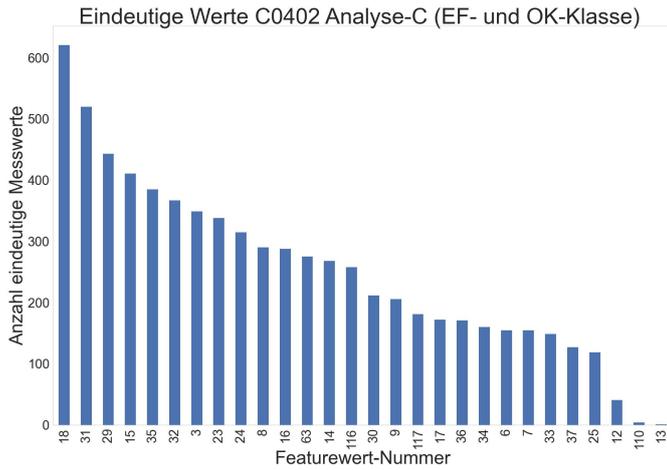
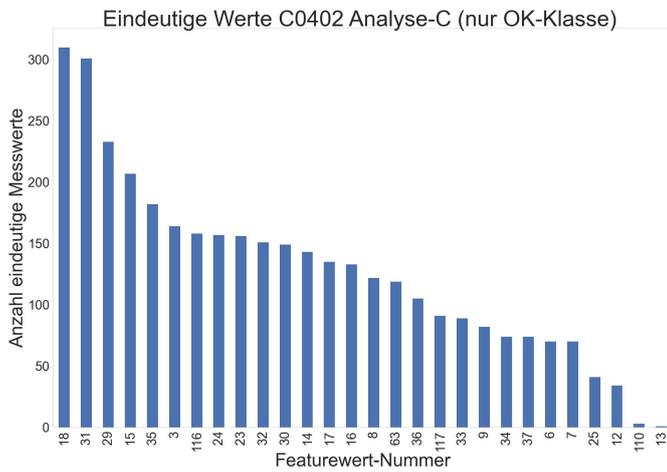


Abbildung 4.7.: Absolute Häufigkeitsverteilung nach Winkel und FC – Analyse-C – C0402 – vier Klassen

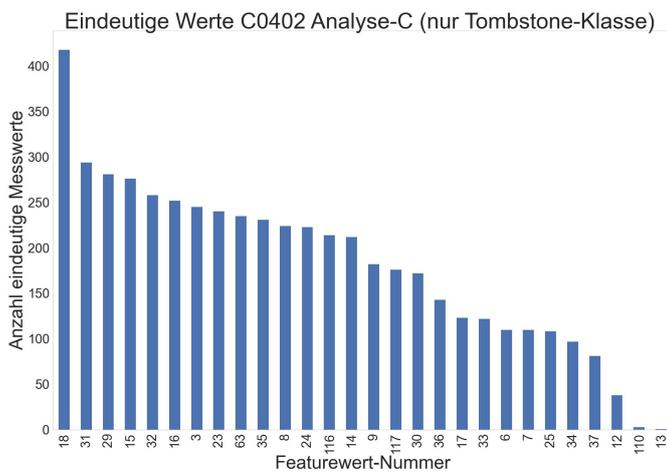
Die Ergebnisse der Kardinalitätsuntersuchungen, siehe Abbildung 4.8, weisen eine vergleichbare Rangfolge der Featurewerte im Vergleich zur Untersuchung des Themas 1 auf. Analog zu jenem Resultat weisen die Featurewerte 18 und 31 die höchste Anzahl eindeutiger Messwerte innerhalb des gesamten und nach OK-/Pseudofehler-Klasse unterteilten Datensatzes auf. Des Weiteren erreichen die Featurewerte mit den Nummerierungen 110 und 13 die letzten Rangplatzierungen. Dieser Zusammenhang wird ebenfalls in den übrigen Untersuchungen der verbliebenen Analyse-Algorithmen und des Bauteils C0603 nachgewiesen.



(a) Gesamter Datensatz



(b) OK-Klasse



(c) Tombstone-Klasse

Abbildung 4.8.: Diagramme Kardinalität eindeutiger Messwertstufen – Analyse-C – C0402 – zwei Klassen (OK, Tombstone)

Dispersion und zentrale Tendenz Aufgrund der gesteigerten Anzahl an Untersuchungsreihen bedingt durch die beiden Untersuchungsmerkmale (FC und Montagewinkel) werden auch in dieser Ergebnisvorstellung Auszüge aller generierten Resultate vorgestellt.

Die Boxplot-Whisker-Diagramme, welche auf Grundlage der binären Variable des Fehlercodes (OK-/Tombstoneklasse; FC 0 und 560) erzeugt werden, sind in Abbildung 4.9 dargestellt. Feature 6, 8 und 29 weisen durch optische Beurteilung keine relevanten Unterschiede der Lage- und Dispersionsmetriken auf. Die Mediane des Featurewertes 23 weichen stärker voneinander ab. Die Höhe der Boxen sowie die Lage des Medians innerhalb dieser Boxen und somit die Variabilität beziehungsweise Symmetrie der AOI-Messwerte unterscheiden sich voneinander. Analog zum Thema 1 treten in der Mehrzahl aller Untersuchungen Ausreißer auf.

Auszug Boxplot-Whisker-Diagramme
Analyse-C / C0402

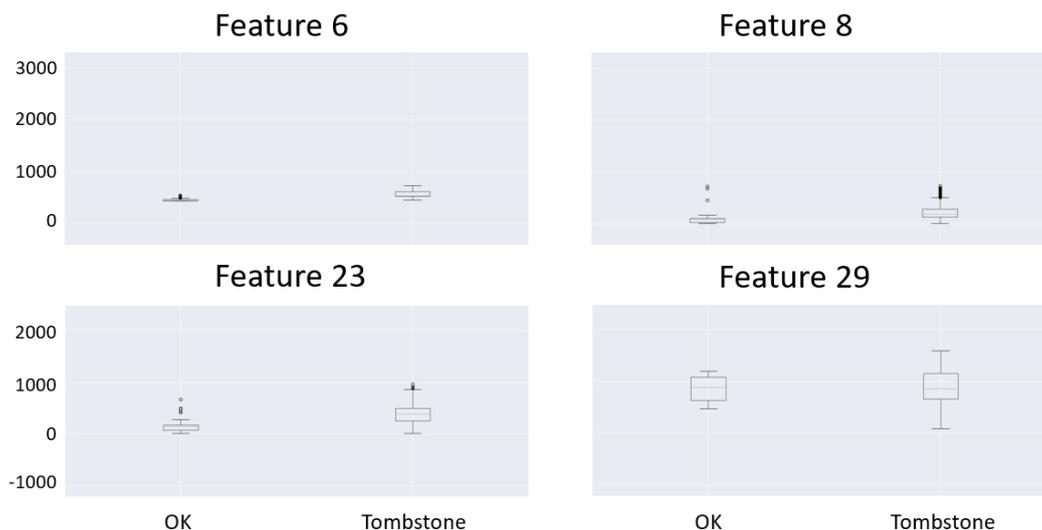


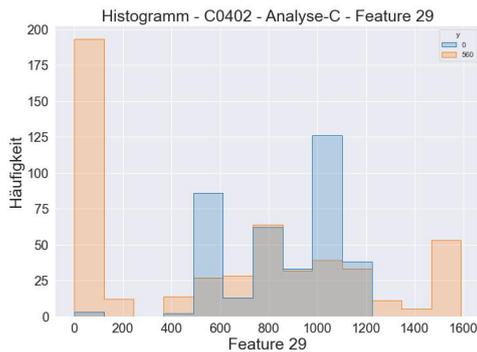
Abbildung 4.9.: Boxplot-Whisker-Diagramme – Analyse-C – C0402 –
Feature 6, 8, 23, 29 – zwei Klassen (OK, Tombstone)

Eine visuelle Analyse der Gruppenunterscheidbarkeit wird durch die Histogramme in den Abbildungen 4.10 und 4.11 ermöglicht. Beide Graphiken unterscheiden sich in ihrer zugrundeliegenden Datenfiltrierung voneinander.

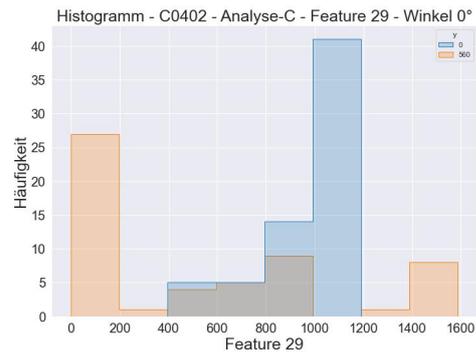
Die unterschiedlich eingefärbten Gruppen werden in Abbildung 4.10 anhand des Montagewinkels selektiert. In Subdiagramm 4.10a sind die am stärksten besetzten Winkelklassen enthalten, während in den Diagrammen 4.10b und c jeweils nur ein Winkel in den korrespondierenden Daten vorliegt. Jede Untergraphik weist breite Überlappungsbereiche auf, welche die Durchführung einer Klassentrennung aufgrund der optischen Informatio-

nen erschwert beziehungsweise verhindert.

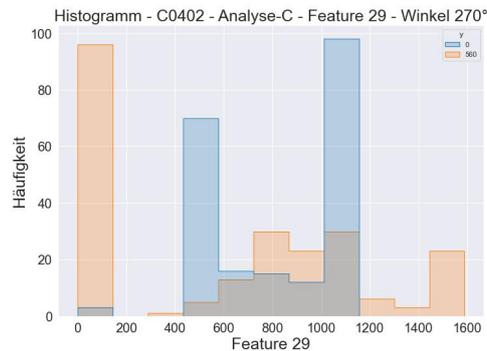
Eine Verbesserung der Trennbarkeit wird durch die Datenfilterung hinsichtlich der FC-Annotation in Abbildung 4.11 nicht erreicht. Dies betrifft sowohl die Vereinigung der FC-Klassen OK und Tombstone in Subgraphik 4.11a als auch eine FC-getrennte Histogrammerstellung, siehe Subdiagramme b und c. Die Abweichungen der Höheninformationen resultieren in den unterschiedlichen Mengen von Datenpunkten innerhalb des Datensatzes und tragen keine Informationen bezüglich der Gruppenunterscheidbarkeit.



(a) Winkel $0^\circ, 45^\circ, 90^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$

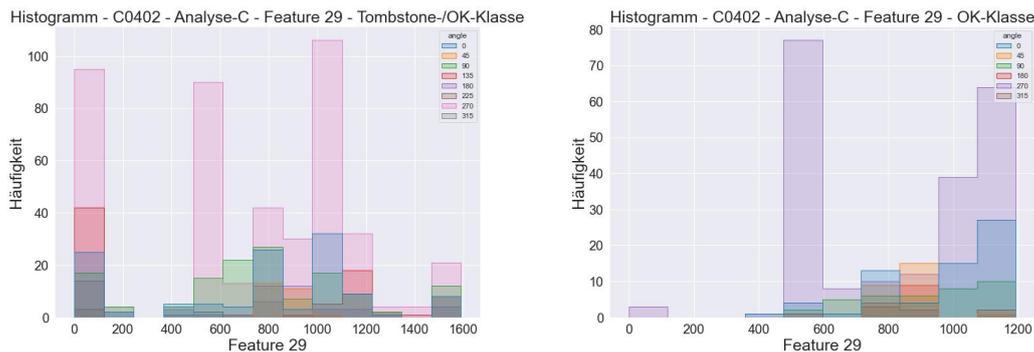


(b) Winkel 0°



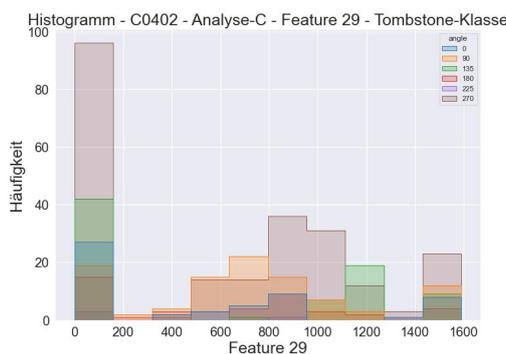
(c) Winkel 270°

Abbildung 4.10.: Histogramme nach Winkel – Analyse-C – C0402 – Feature 29



(a) Alle Winkel

(b) OK-Klasse (FC 0)

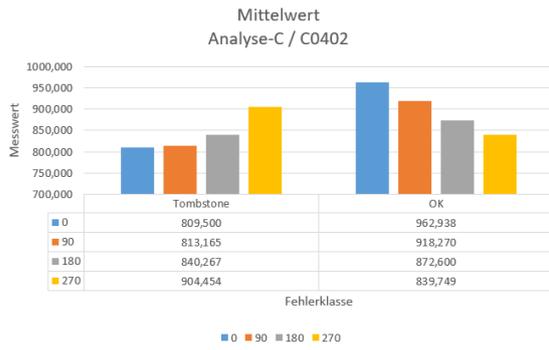


(c) Tombstone-Klasse (FC 560)

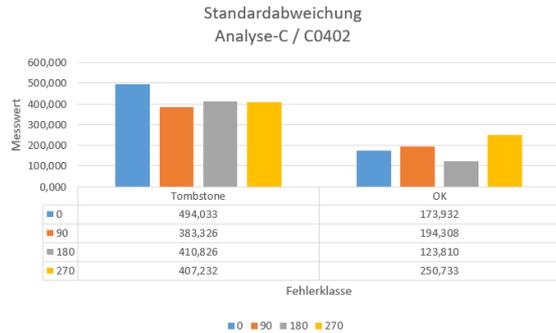
Abbildung 4.11.: Histogramme nach Fehlercode – Analyse-C – C0402 – Feature 29

Die graphischen und numerischen Darstellungen der Mittelwerte, Standardabweichungen und Mediane getrennt nach FC und Montagewinkel werden in Graphik 4.12 dargestellt. In der Analyse der Mittelwerte sind sowohl Unterschiede zwischen den Fehlerklassen (OK/Tombstone) als auch innerhalb einer Fehlerklasse unter Berücksichtigung der Winkelinformationen nachweisbar. Die Standardabweichung liegt innerhalb einer Fehlerklasse für jeden betrachteten Montagewinkel auf vergleichbaren Niveaus.

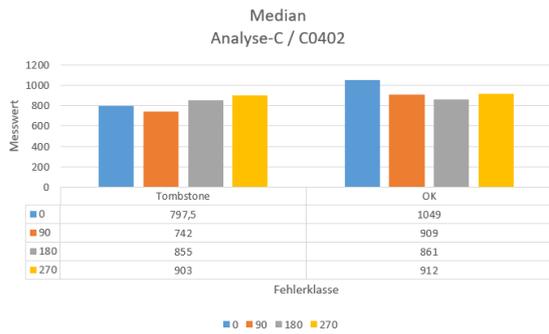
Die Darstellung der Datenpunkte in Form eines Streudiagramms, siehe Abbildung 4.13, verbietet auch in dieser Darstellungsform die Identifikation bestehender Strukturen und Cluster. Es kommt in sämtlichen Streudiagrammen aller möglichen Featurewertpaarungen zur Ausbildung einer zusammenhängenden Datenpunkt-Wolke mit einer hohen Packungsdichte. Innerhalb dieser Gebilde ist die Verteilung der Datenpunkte gleicher Klassenannotation unregelmäßig. Übergeordnete Muster und eindeutig voneinander unterscheidbare Abgrenzungen der Datenpunkte gleicher Klassenannotation sind nicht nachweisbar.



(a) Mittelwert

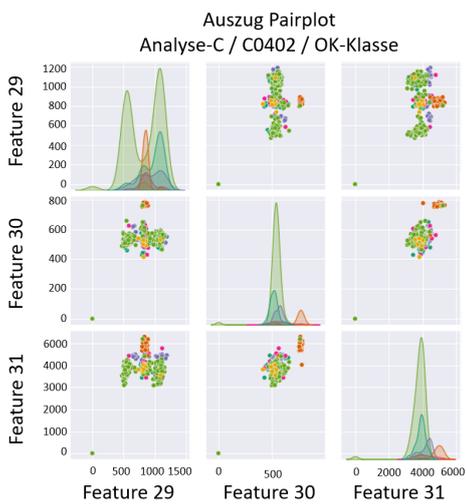


(b) Standardabweichung

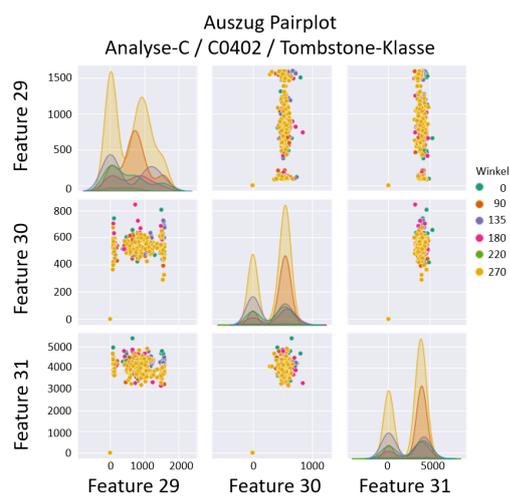


(c) Median

Abbildung 4.12.: Diagramme zentrale Tendenz – Analyse-C – C0402 – Feature 29 – zwei Klassen (PF, Tombstone)



(a) OK-Klasse (FC 0)



(b) Tombstone-Klasse (FC 560)

Abbildung 4.13.: Streudiagramme nach Winkel und FC – Analyse-C – C0402 – Feature 29, 30, 31

4.2. Maße des Zusammenhangs und Feature-Ranking

In diesem Abschnitt umfasst die Datenexploration zwei Untersuchungsreihen. Zunächst rückt die Ermittlung der Ausprägung des statistischen Zusammenhangs der Featurewerte untereinander in den Vordergrund. Dieser wird durch den Korrelationskoeffizienten jeder Featurewert-Paarung gemessen. Des Weiteren erfolgt die Ermittlung relevanter Merkmale, welche durch ihren hohen Informationsgehalt einen gesteigerten Einfluss auf das Klassifikationsergebnis ausüben, mithilfe der in Abschnitt 2.2.1.4 vorgestellten Feature-Selektionsverfahren.

4.2.1. Binäre Hauptuntersuchung

Featurebasierte Inter-Korrelationsstudie und distanzbasierte Clusteranalyse Der Korrelationskoeffizient ist ein standardisiertes Maß für die Stärke des Zusammenhangs zwischen Merkmalen des Datensatzes. Unter Berücksichtigung des vorliegenden Skalenniveaus und der in den vorigen Kapiteln nachgewiesenen Dateneigenschaften erfolgt die Wahl des zu bestimmenden Korrelationskoeffizienten. Bedingt durch den Nachweis von Ausreißerwerten sowie die vorliegenden Rangbindungen innerhalb der Datenstruktur empfiehlt sich die robuste Berechnung der Korrelation nach Kendalls τ . [Ken38]

Des Weiteren erfolgt die Visualisierung der Korrelationsbeziehungen zwischen den Featurewerten in Form eines Dendrogramms, welches sich einer distanzbasierten Clusteranalyse des UPGMA-Algorithmus (englisch Unweighted Pair Group Method with Arithmetic Mean) nach [Sci22] bedient.

Das in Abbildung 4.14 wiedergegebene Heatmap-Diagramm enthält eine numerische und farbliche Visualisierung der Korrelationskoeffizienten aller Featurewert-Paarungen des gesamten binären Datensatzes bezüglich der AOI-Analyse-C und dem SMD-Bauteil C0402. Diese farbliche Darstellung in Form einer unteren Dreiecksmatrix erlaubt eine intuitive, effiziente Interpretation der Zusammenhänge. Die binären und ternären Datensätze sowie die zugehörigen Subdatensätze, die jeweils nur eine Fehlerklasse enthalten, weisen einheitliche Farbmuster und Messwertverhältnisse auf. Dieser Zusammenhang deutet auf die Existenz einer homogenen Korrelationsverteilung hin. Bei einem nachgewiesenen perfekten positiven Zusammenhang (Korrelationskoeffizient = 1) liegen entweder identische oder stark redundante Messwerte innerhalb einer Messwert-Paarung vor. Diese irrelevanten Informationen werden im Rahmen der Datenbereinigung beseitigt. Es wird die Generierung von Datensätzen, bestehend aus einer Untermenge relevanter Merkmale, mit geringen Korrelationskoeffizienten (nahe Null) angestrebt. Diese Korrelationsei-

genschaften werden darüber hinaus durch UPGMA-distanzbasierte Dendrogramme, in Abbildung 4.15 für Analyse-C und Gehäuseform C0402, nachgewiesen. Eine starke Korrelation, wie beispielsweise zwischen Feature 6 und 7 ($\rho = 1$) oder Feature 3 und 8 ($\rho = 0,9$), ist durch eine geringe Clusterdistanz gekennzeichnet, wobei schwache statistische Zusammenhänge (Feature 24 und 16 mit $\rho = 0,0067$) maximale Werte in der Distanzanalyse zur Folge haben.

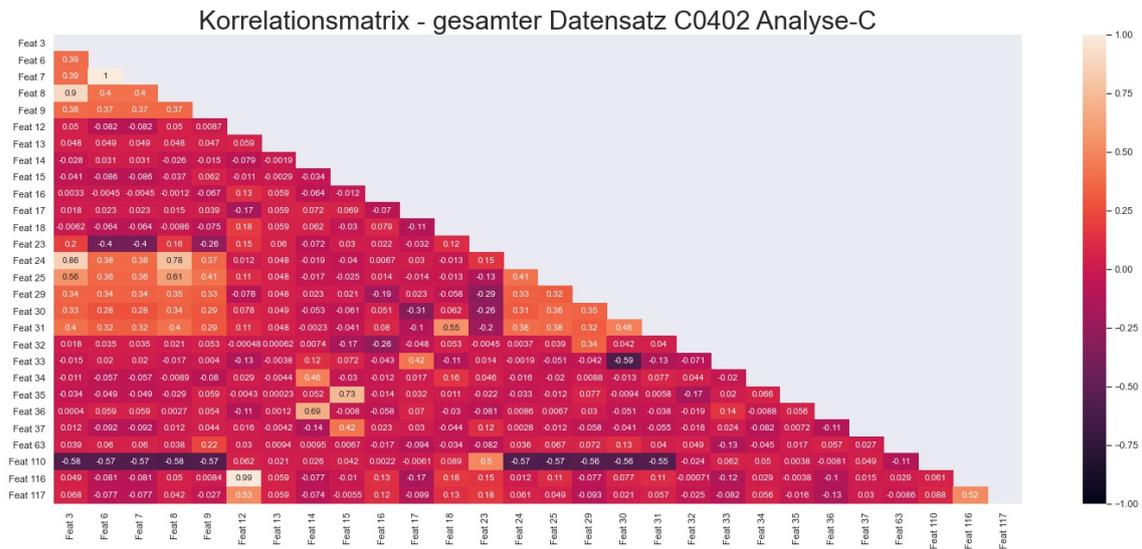


Abbildung 4.14.: Heatmap-Diagramm – binär – Analyse-C – C0402

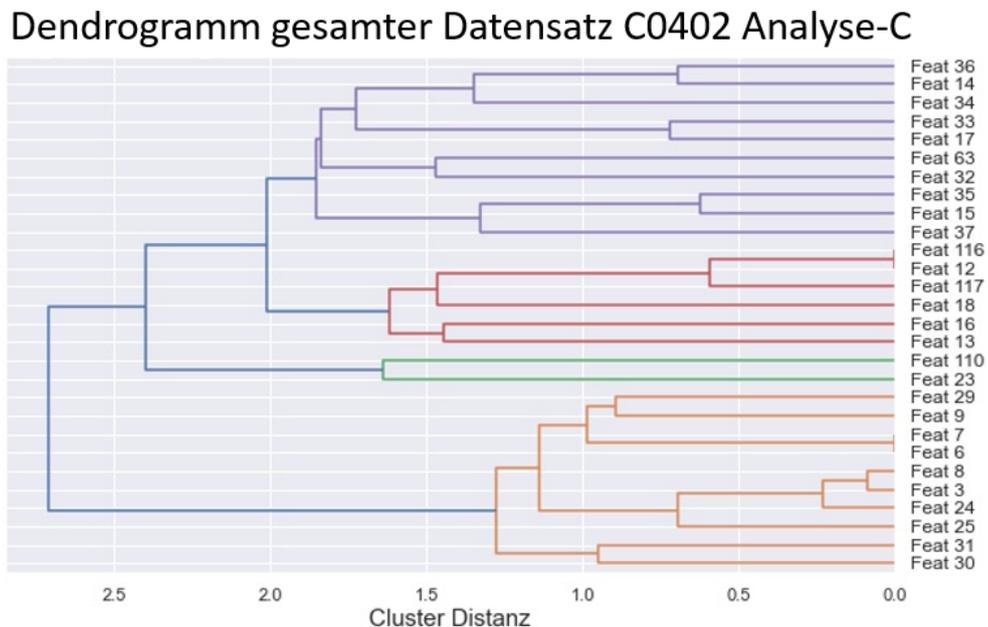


Abbildung 4.15.: Dendrogramm – binär – Analyse-C – C0402

Feature-Ranking Die Ergebnisse der in Kapitel 2.2.1.4 vorgestellten Methoden des Feature-Ranking für den binären Datensatz der Analyse-C und Prüfmuster C0402 werden in Tabelle 4.1 vorgestellt. Der Boruta-Algorithmus wird mit zwei unterschiedlichen Backbones (Random Forest und Extra-Tree) realisiert. Ebenfalls der Entscheidungsbaum-Methodik entstammend findet ein XG-Boosting-Verfahren Anwendung. Allen drei Modellen liegt eine standardisierte Wichtigkeitskennzahl zugrunde, sodass ein methodenübergreifender Ergebnisvergleich trotz unterschiedlicher mathematischer Modellierungen durchführbar ist.

In diesem Beispiel belegen die Featurewerte der Nummerierungen 6, 9, 23, 29 der Entscheidungsbaum-gestützten Algorithmen obere Platzierungen in dem Ranking. Die Ergebnisse des SOAP-Verfahrens widersprechen dieser Verteilung. Ähnliche Zusammenhänge treten ebenfalls in den Untersuchungsreihen anderer AOI-Analysealgorithmen und SMD-Bauteile in Erscheinung.

Tabelle 4.1.: Feature-Ranking – binär – Analyse-C – C0402

Boruta ET	Imp.	Boruta RF	Imp.	SOAP	Imp.	XGBoost	Imp.
Feat 23	1	Feat 9	1	Feat 18	1	Feat 23	1
Feat 6	0,952	Feat 23	0,854	Feat 31	0,941	Feat 9	0,944
Feat 7	0,773	Feat 7	0,531	Feat 29	0,405	Feat 6	0,255
Feat 29	0,771	Feat 6	0,521	Feat 116	0,341	Feat 8	0,163
Feat 30	0,683	Feat 8	0,46	Feat 35	0,336	Feat 12	0,128
Feat 9	0,681	Feat 30	0,427	Feat 63	0,334	Feat 35	0,124
Feat 110	0,67	Feat 3	0,408	Feat 32	0,334	Feat 24	0,119
Feat 31	0,496	Feat 29	0,4	Feat 15	0,324	Feat 3	0,113
Feat 3	0,417	Feat 63	0,392	Feat 117	0,298	Feat 17	0,109
Feat 24	0,415	Feat 24	0,281	Feat 3	0,255	Feat 30	0,109
Feat 8	0,372	Feat 31	0,275	Feat 9	0,238	Feat 117	0,101
Feat 63	0,298	Feat 110	0,224	Feat 24	0,233	Feat 34	0,093
Feat 32	0,255	Feat 14	0,191	Feat 34	0,231	Feat 16	0,085
Feat 15	0,167	Feat 32	0,165	Feat 23	0,231	Feat 33	0,075
Feat 35	0,166	Feat 15	0,127	Feat 16	0,223	Feat 15	0,074
Feat 12	0,156	Feat 35	0,113	Feat 12	0,22	Feat 14	0,074
Feat 116	0,153	Feat 33	0,11	Feat 8	0,19	Feat 116	0,072
Feat 14	0,139	Feat 116	0,107	Feat 14	0,185	Feat 18	0,069
Feat 16	0,131	Feat 25	0,105	Feat 25	0,133	Feat 110	0,064
Feat 18	0,123	Feat 37	0,105	Feat 7	0,124	Feat 31	0,062
Feat 33	0,121	Feat 36	0,101	Feat 17	0,124	Feat 29	0,057
Feat 117	0,119	Feat 12	0,091	Feat 37	0,122	Feat 25	0,054

Feat 17	0,113	Feat 117	0,091	Feat 6	0,122	Feat 63	0,052
Feat 37	0,104	Feat 17	0,089	Feat 36	0,121	Feat 36	0,051
Feat 36	0,094	Feat 18	0,082	Feat 33	0,119	Feat 32	0,05
Feat 34	0,093	Feat 16	0,08	Feat 30	0,114	Feat 37	0,044
Feat 25	0,093	Feat 34	0,045	Feat 110	0,003	Feat 13	0
Feat 13	0	Feat 13	0	Feat 13	0	Feat 7	0

4.2.2. Multikategorielle Nebenuntersuchung

Featurebasierte Inter-Korrelationsstudie und distanzbasierte Clusteranalyse Die Korrelationsstudie des Themas 2, welche neben den AOI-Featurewerten die Winkelinformationen der Prüffenster sowie multinominale EF-Klassenannotationen enthält, wird analog zur vorigen Methodik in Abschnitt 4.2.1 durch ein Heatmap-Diagramm und das numerische Resultat des Verfahrens nach Kendalls τ realisiert.

Die Ergebnisvorstellung erfolgt in Abbildung 4.16 bezüglich der OK-Klasse und in Abbildung 4.17 hinsichtlich der Tombstone-Klasse. Der Anhang A.2.5 enthält Diagramm A.14, welches nach der Winkelkategorie 270° und der OK-Klasse erstellt wird. Sowohl in diesem beispielhaften Auszug als auch in der Gesamtheit erhobener Korrelations-Maßzahlen kann in dieser Versuchsreihe, im Gegensatz zur Bearbeitung des Themas 1, keine einheitliche, datensatzübergreifende Korrelationssituation nachgewiesen werden. Eine mögliche Ursache für geringe Abweichungen ist die Existenz von statistisch zulässigen, stochastischen Featurewert-Streuungen. Eine weitere Ursache liegt möglicherweise in den unterschiedlichen Anzahlen von Datenpunkten in den Subdatensätzen. So wird zwischen den Featurewerten 8 und 23 in Abbildung 4.16 eine Korrelation von $\rho = 0,83$ und in Abbildung 4.17 von $\rho = 0,068$ ermittelt. Solche gegensätzlichen Zusammenhangsmetriken treten ebenfalls in der Bewertung anderer Analyse-Algorithmen sowie Winkel- und Fehlercode-Kombinationen auf.

Eine Bestätigung der Existenz dieser inhomogenen Korrelationswechselbeziehungen wird durch die Beurteilung der Dendrogramme in den Abbildungen 4.18 und 4.19 erbracht. Die Distanz zwischen den Clustern und deren Subcluster der Dendrogramme unterscheiden sich sowohl in ihrer Ausprägung als auch hinsichtlich der darin enthaltenen Featurewerte und deren Zusammenfassung zu übergeordneten Clustern voneinander.

Die im oberen Textabschnitt erwähnten Featurewerte 8 und 23 sind in Abbildung 4.18 in Clustern mit geringem Abstand zueinander enthalten. Demgegenüber wird die Distanzermittlung in Abbildung 4.19 erst durch die Bildung und den anschließenden Vergleich von übergeordneten Clustern, deren jeweilige Elemente die Merkmale 8 und 23 sind, ermöglicht. In dieser Betrachtung liegt ein maximaler Abstandswert auf der Abszisse vor.

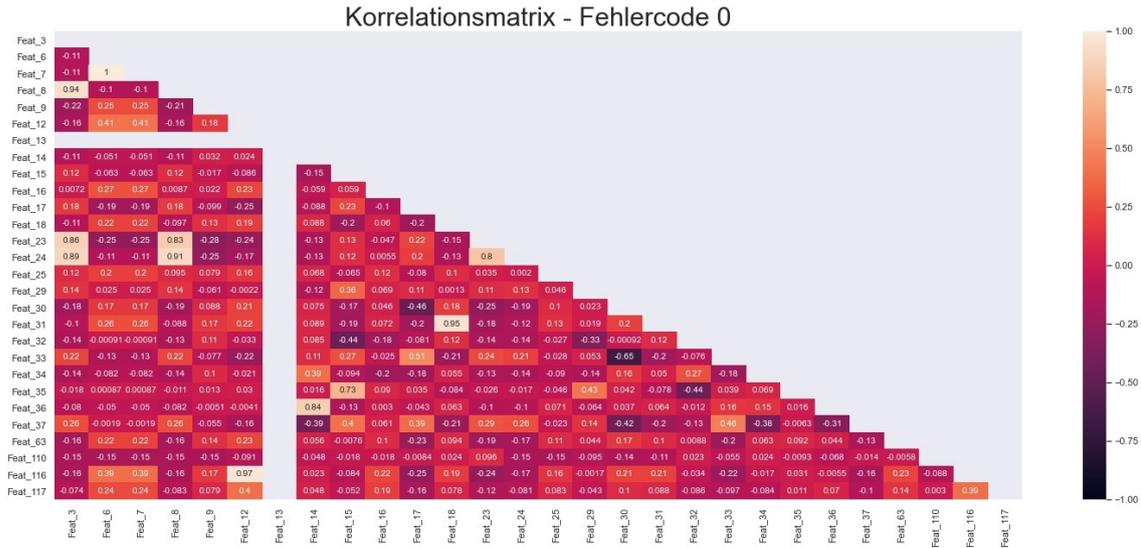


Abbildung 4.16.: Heatmap-Diagramm nach FC – Analyse-C – C0402 – OK-Klasse (FC 0)

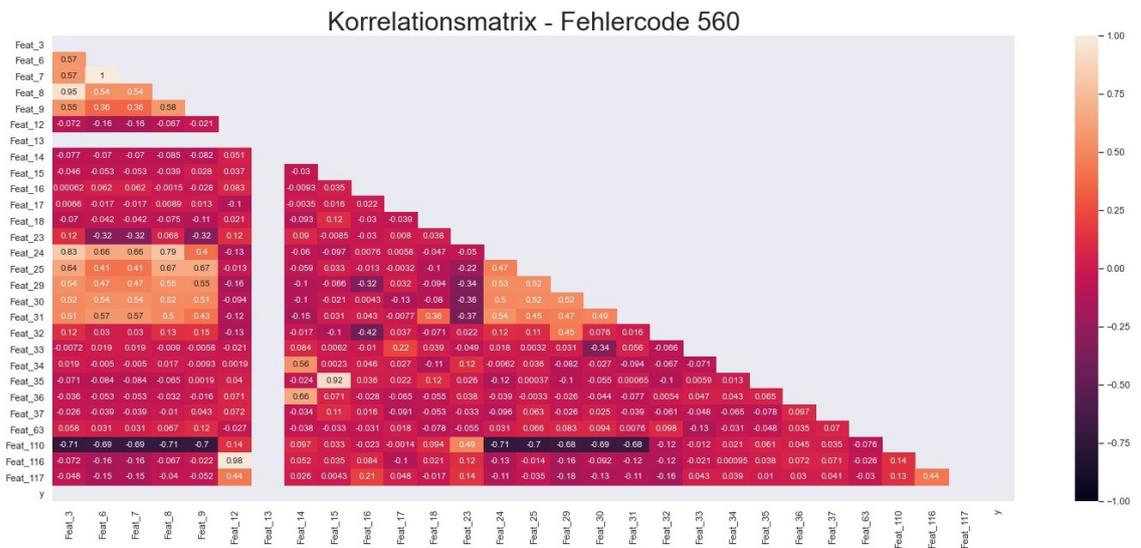


Abbildung 4.17.: Heatmap-Diagramm nach FC – Analyse-C – C0402 – Tombstone-Klasse (FC 560)

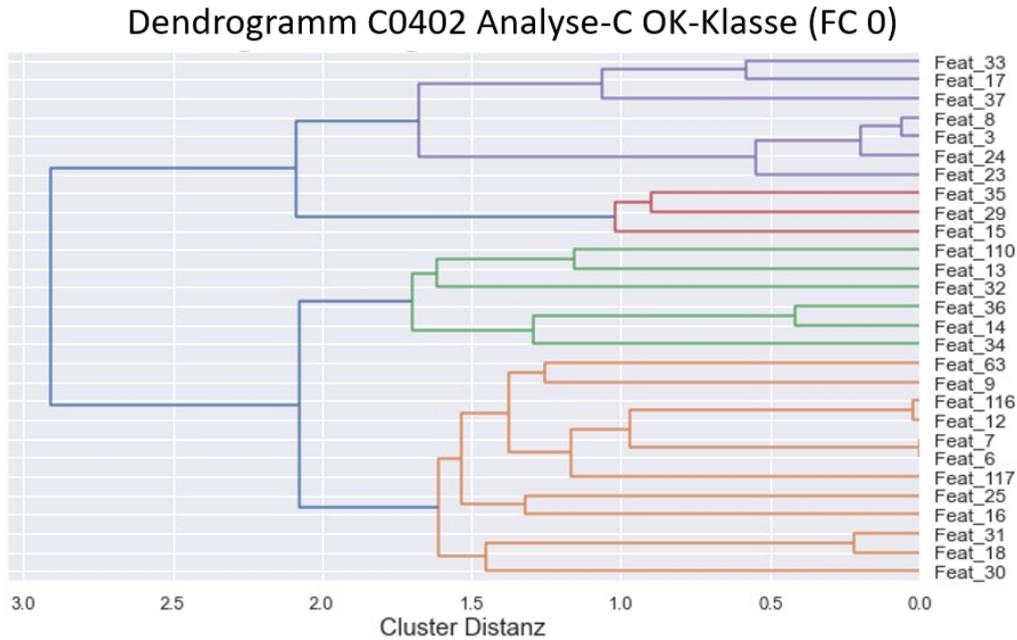


Abbildung 4.18.: Dendrogramm nach FC – Analyse-C – C0402 – OK-Klasse (FC 0)

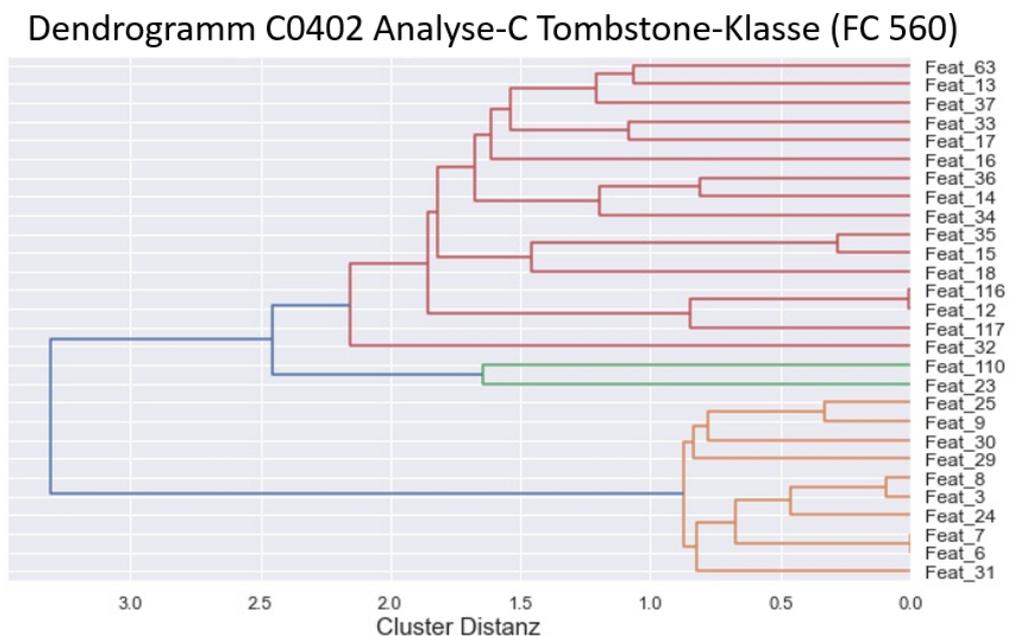


Abbildung 4.19.: Dendrogramm nach FC – Analyse-C – C0402 –
Tombstone-Klasse (FC 560)

Feature-Ranking Die Resultate der Feature-Ranking-Methoden für den Datensatz, bestehend aus der OK- und Tombstone-Klasse des Bauteils C0402 und der Analyse-C, sind in Tabelle 4.2 aufgetragen. Die Merkmale mit den Nummerierungen 3, 6, 7, 23

und 29 besetzen in dieser Analyse hinsichtlich der beiden Ausführungen des Boruta-Verfahrens die oberen Wichtigkeitsstufen. Die Ergebnisse des Feature-Ranking von Thema 1 und 2 sind ähnlich und weisen eine vergleichbare Rangplatzierung der Feature-Werte auf. Verletzungen der Ergebnisähnlichkeit von Thema 1 und 2 treten vereinzelt in wenigen Subdatensätzen auf.

Des Weiteren wird die Variable des *Montagewinkels* ebenfalls in die Betrachtung integriert, sodass anhand jenes Rangplatzes eine Beurteilung des Einflusses des Orientierungswinkels der SMD-Chipkomponente auf den resultierenden Bauteilzustand ermöglicht wird. Die geringen Wichtigkeitsmetriken der entscheidungsbaum-basierten Methoden, insbesondere des als effizient eingestuften Boruta-Algorithmus, sind ein Indiz für eine geringe Einflussnahme. Eine Zusammenfassung der Wichtigkeitsmetriken hinsichtlich der Winkelklassen sämtlicher multikategorierter Untersuchungsreihen ist in Tabelle A.8 dem Anhang A.2.6 beigelegt.

Tabelle 4.2.: Feature-Ranking multikategorierll mit FC 0 und 560 – Analyse-C – C0402

Boruta ET	Imp.	Boruta RF	Imp.	SOAP	Imp.	XGBoost	Imp.
Feat 6	1	Feat 23	1	Winkel	1	Feat 23	1
Feat 7	0,91	Feat 6	0,925	Feat 110	0,974	Feat 6	0,327
Feat 23	0,515	Feat 7	0,92	Feat 13	0,972	Feat 17	0,159
Feat 24	0,298	Feat 3	0,445	Feat 12	0,301	Feat 9	0,158
Feat 3	0,287	Feat 8	0,433	Feat 17	0,094	Feat 37	0,096
Feat 8	0,277	Feat 29	0,387	Feat 6	0,06	Feat 12	0,093
Feat 31	0,25	Feat 30	0,28	Feat 7	0,06	Feat 24	0,056
Feat 29	0,244	Feat 24	0,267	Feat 34	0,056	Winkel	0,055
Feat 30	0,237	Feat 16	0,18	Feat 25	0,056	Feat 35	0,053
Feat 110	0,221	Feat 9	0,18	Feat 36	0,039	Feat 32	0,052
Feat 32	0,147	Feat 31	0,176	Feat 37	0,038	Feat 15	0,049
Feat 16	0,136	Feat 110	0,14	Feat 117	0,038	Feat 30	0,047
Feat 117	0,105	Feat 117	0,139	Feat 30	0,035	Feat 63	0,036
Feat 35	0,104	Feat 32	0,123	Feat 33	0,03	Feat 34	0,029
Feat 9	0,101	Feat 17	0,116	Feat 116	0,026	Feat 16	0,029
Feat 17	0,094	Feat 25	0,085	Feat 24	0,021	Feat 3	0,028
Feat 12	0,092	Feat 35	0,08	Feat 14	0,019	Feat 25	0,025
Feat 15	0,079	Feat 12	0,069	Feat 9	0,019	Feat 14	0,024
Feat 25	0,075	Feat 116	0,065	Feat 16	0,019	Feat 33	0,023
Feat 116	0,074	Feat 14	0,061	Feat 8	0,018	Feat 116	0,023
Winkel	0,067	Feat 36	0,061	Feat 63	0,015	Feat 18	0,021
Feat 14	0,051	Feat 33	0,054	Feat 3	0,014	Feat 31	0,019

Feat 63	0,045	Feat 15	0,049	Feat 32	0,013	Feat 36	0,019
Feat 36	0,041	Feat 63	0,045	Feat 23	0,011	Feat 8	0,013
Feat 37	0,04	Feat 37	0,039	Feat 35	0,011	Feat 117	0,011
Feat 34	0,037	Winkel	0,023	Feat 15	0,01	Feat 29	0,002
Feat 33	0,031	Feat 34	0,023	Feat 29	0,006	Feat 13	0
Feat 18	0,03	Feat 18	0,019	Feat 31	0,002	Feat 110	0
Feat 13	0	Feat 13	0	Feat 18	0	Feat 7	0

4.3. Anpassungstests

Eine Vielzahl von Hypothesentests, Dependenzanalysen sowie ML-Anwendungen setzen die Existenz von normalverteilten Stichproben beziehungsweise Daten voraus. Zu diesem Zweck wird durch einen Anpassungstest nach Kolmogorov-Smirnow die Übereinstimmung einer theoretischen mit einer empirischen, durch den Datensatz generierten Verteilungsfunktion ermittelt. [Leo17]

Die Erstellung von Q-Q- sowie P-P-Diagrammen (englisch Quantile-Quantile; Probability-Probability) dient der visuellen Unterstützung der Beurteilung des Grades der Deckungsgleichheit beider Testfunktionen. Analog zu den vorangegangenen Ergebnisvorstellungen dient auch in diesem Kapitel der Datensatz der Analyse-C und des Bauteils C0402 als Grundlage der in den folgenden Abbildungen und tabellarischen Auflistungen dargestellten Resultate.

4.3.1. Binäre Hauptuntersuchung

Die Überprüfung der in dem Datensatz zugrunde liegenden theoretischen Verteilung mithilfe des Kolmogorov-Smirnov-Tests (Abkürzung KS-Test) und der Erstellung von Q-Q- und P-P-Diagrammen erfolgt für jeden Subdatensatz (binär/ternär) und jede Zustandsklasse (EF/PF/OK) getrennt. Zunächst wird jeder Datensatz, welcher nach den Filterkriterien *Bauteil*, *AOI-Analyse* und *Featurewert* erstellt wird, anhand der vorliegenden Fehlerklasse unterteilt. Die daraus resultierenden Datensätze werden einem KS-Test unterzogen. Aufgrund dessen kann eine Aussage über die Angehörigkeit zu Populationen unterschiedlicher Verteilungsfunktionen für die einzelnen Klassen getroffen werden.

Die durch den KS-Test ermittelten Statistik-Kennwerte, welche einerseits dem p-Wert des Hypothesentests und andererseits dem Distanzmaß als Differenz der kumulativen Verteilungsfunktionen der Stichprobe und der theoretischen Verteilung entsprechen, werden in Tabelle 4.3 aufgeführt. Diese stellt einen Auszug der untersuchten theoretischen Wahr-

scheinlichkeitsverteilungen dar.

Unter Zuhilfenahme des gewählten Signifikanzniveaus von 5 %, sodass $\alpha = 0,05$ gilt, und des kritischen Distanzwertes d_α wird über die Annahme oder Ablehnung der Nullhypothese entschieden. Diese geht davon aus, dass eine zur theoretischen Verteilung deckungsgleiche empirische Stichprobe vorliegt und diese aus der gleichen Grundgesamtheit stammt. [Leo17] Für d_α gilt nach [Mas51]:

$$d_\alpha = \frac{\sqrt{-0,5 \ln\left(\frac{\alpha}{2}\right)}}{\sqrt{N}} = \frac{1,36}{\sqrt{N}} = 0,036 \quad (4.1)$$

Mit:

α : Signifikanzniveau Hypothesentest, $\alpha = 0,05$

N : Anzahl Elemente in Stichprobe, $N = 1424$

In jeder Untersuchungsreihe der Originaldaten werden konstante Verletzungen der maximal zulässigen Distanzmaße und der Irrtumswahrscheinlichkeiten (p-Wert) festgestellt und die Nullhypothese abgelehnt. Dieses konstante Ergebnis lässt die Schlussfolgerung zu, dass die Mehrheit der gesamten AOI-Messwertinformationen keiner der untersuchten Verteilungen, im Besonderen der Normalverteilung, folgen. Für nachfolgende Dependenzanalysen sind nicht-parametrische Testverfahren anzuwenden. Diese Erkenntnisse werden durch Hinzunahme und Interpretation der Q-Q- und P-P-Diagramme, siehe Abbildung 4.20, erhärtet. Zwar sind in dem Anpassungsdiagramm 4.21 zahlreiche Verteilungsfunktionen in der Lage, eine optisch positiv bewertete Annäherung an die empirischen Messdaten zu erreichen, jedoch verbieten die erhobenen Differenzkennwerte eine Annahme der Nullhypothese.

Im Rahmen einer Wiederholung der Anpassungstests auf Basis bereinigter Datensätze, aus denen die statistischen Ausreißer und 0-Zahlenwerte entfernt werden, wird ein erhöhter Anteil von normalverteilten Featurewerten nachgewiesen, siehe Tabellen A.3 und A.4 im Anhang.

Zusätzlich erbringt eine geringe Anzahl von Untersuchungsreihen den Beweis, dass die Stichproben einer Population anderer theoretischer Wahrscheinlichkeitsverteilungen entstammen, siehe Anhang A.2.3 in Abbildung A.12 für die OK-Klasse des Features 9 aus der ternären Analyse-C und Bauteil C0402, mit p-Werten des KS-Tests von $p_{foldNorm} = 0,164$ und $p_{Gompertz} = 0,219$ für die gefaltete Normalverteilung und die Gompertz-Verteilung.

Tabelle 4.3.: KS-Test – Analyse-C – C0402 – OK-Klasse – Feature 6

Verteilung	p-Wert	Distanz
Cauchy	2,84E-15	0,109
Burr Typ XII	8,85E-63	0,223
Pearson Typ III	1,56E-99	0,280
Hypergeom. Gauss	6,63E-103	0,285
Beta	3,51E-104	0,287
Weibull_max	2,13E-118	0,305
Gen. Logistisch	6,95E-157	0,350
Log. Normal	4,38E-163	0,357
Normal	1,10E-163	0,358
Gefaltete Normal	1,10E-163	0,358
Exp. Normal	1,07E-163	0,358
Chi	7,71E-165	0,359
Gen. Normal	1,65E-172	0,367
Inv. Gamma	1,03E-183	0,378
Gamma	1,60E-185	0,380
Gen. halb Logistisch	1,43E-186	0,381
Alpha	5,33E-188	0,382
Inv. Weibull	3,69E-192	0,387
Inv. Beta	1,11E-201	0,396
Halb Cauchy	4,03E-227	0,419
Inv. Normal	8,68E-242	0,432
Gilbrat	2,25E-244	0,434
Gen. Inv. Normal	1,92E-286	0,468

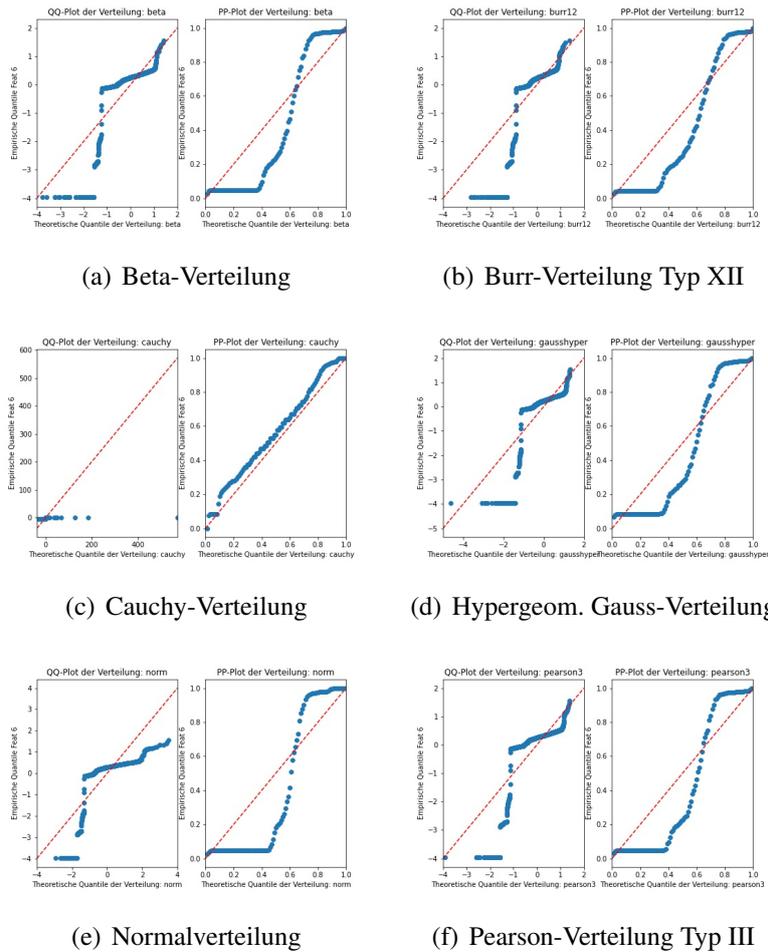


Abbildung 4.20.: Q-Q-Diagramme (links); P-P-Diagramme (rechts) – Analyse-C – C0402 – Feature 6 – OK-Klasse



Abbildung 4.21.: Anpassungsdiagramm – Histogramm – binär – Analyse-C – C0402 – Feature 6

4.3.2. Multikategorielle Nebenuntersuchung

Konform zur Verteilungsprüfung der Hauptuntersuchung (Thema 1) werden ebenfalls in dieser vollfaktoriellen Untersuchungsreihe die Subdatensätze aller möglichen Einstellkombinationen, anhand der beiden Einstellgrößen *Fehlercode* und *Winkel*, auf eine gemeinsame Ursprungspopulation mit theoretischen Wahrscheinlichkeitsverteilungen untersucht. Ein Repräsentant für die Einzeluntersuchungen, in denen die Resultate des Anpassungstests die Normalverteilung als Grundgesamtheit bestätigen, wird im Rahmen der Analyse des bereinigten Datensatzes hinsichtlich Bauteil C0402, Algorithmus-C, OK-Klasse (FC 0) und dem Winkel 270° vorgestellt. Dem Anhang A.2.4 ist eine tabellarische Auflistung jenes KS-Anpassungstests in A.5 und der entsprechenden Q-Q-/P-P-Diagramme in A.13 beigelegt.

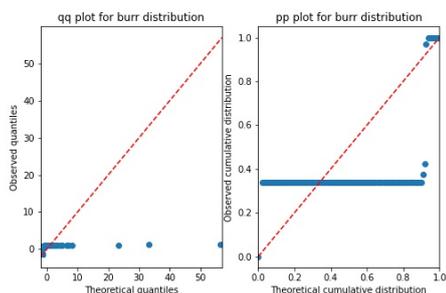
Die Entfernung statistisch irrelevanter Datenpunkte resultiert, analog zu Abschnitt 4.3.1, in einer Zunahme der p-Werte und Abnahme der Distanzmetriken der KS-Anpassungstests, worauf eine Bestätigung der Nullhypothese in mehreren Anwendungsfällen folgt. Sowohl die Unterteilung der Datensätze anhand ihrer Fehlercode- und Winkelinformationen, als auch die Datenbereinigung bewirken einen Anstieg der Anteile normalverteilter Featurewerte innerhalb der Untersuchungsreihen im Vergleich zu binären und ternären Datensätzen, siehe hierzu Tabellen A.6 und A.7 im Anhang.

Jedoch existieren auch vereinzelte Subdatensätze, welche keiner theoretischen Verteilung folgen. Das Resultat des KS-Tests der Originaldaten für das Bauteil C0402, Analyse-C, Winkel 270° und die Tombstone-Klasse ist in Tabelle 4.4 dargestellt. Mit einem Signifikanzniveau von $\alpha = 0,05$ lässt sich die empirische Stichprobenverteilung des Feature 6 für Bauteil C0402 und Analyse-C durch keine der ausgewählten, theoretischen Verteilungen erklären. Die Nullhypothese wird in jedem Untersuchungsfall abgelehnt. Korrespondierende Q-Q- sowie P-P-Diagramme sind Abbildung 4.22 zu entnehmen.

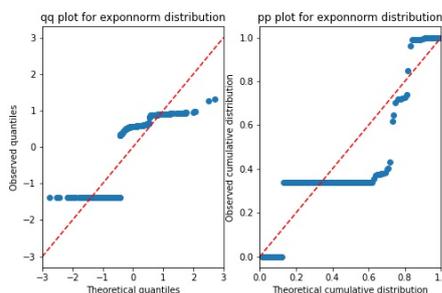
Tabelle 4.4.: KS-Test – multikategoriell – Analyse-C – C0402 – Tombstone-Klasse – Winkel 270° – Feature 6 – vor Datenbereinigung

Verteilung	p-Wert	Distanz
Gen. Inv. Gauss	1,34E-06	0,144
Inv. Weibull	1,42E-07	0,155
Burr	1,22E-07	0,155
Exp. Normal	7,45E-08	0,157
Fisk	2,31E-08	0,163
Mielke	1,93E-08	0,164
Inv. Gamma	6,28E-09	0,168
Log. Normal	6,15E-09	0,169

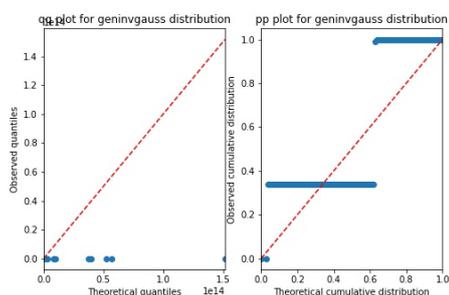
Rez. Inv. Normal	6,15E-09	0,169
Inv. Beta	6,02E-09	0,169
Alpha	5,06E-09	0,169
Gamma	4,17E-09	0,170
Pearson Typ III	4,17E-09	0,170
Exp. Weibull	1,13E-09	0,176
Beta	4,83E-10	0,179
Chi	2,29E-10	0,182
Burr Typ XII	1,71E-10	0,183
Weibull_min	1,52E-10	0,184
Gompertz	1,81E-13	0,208
Gen. halb Logistisch	6,54E-14	0,212
Gen. Norm	3,64E-14	0,214
Cauchy	1,85E-14	0,216
Gefaltete Normal	1,54E-14	0,217



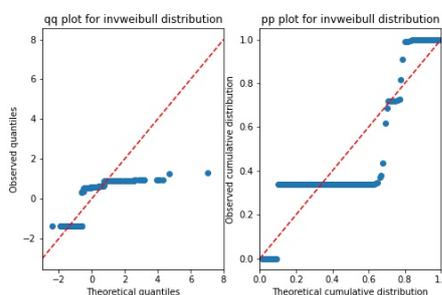
(a) Burr-Verteilung



(b) Exponentielle Normalverteilung



(c) Generalisierte inverse Gauss-Verteilung



(d) Inverse Weibull-Verteilung

Abbildung 4.22.: Q-Q-Diagramme (links) & P-P-Diagramme (rechts) – Analyse-C – C0402 – Tombstone-Klasse – Winkel 270° – Feature 6 – vor Datenbereinigung

4.4. Dependenzanalyse

Das Kapitel 4.4 bildet den Abschluss der statistischen Analyse. Die Intention liegt in der Ermittlung von Populationsunterschieden der AOI-Messwerte hinsichtlich ihrer Fehlerklasse. Im vorangegangenen Abschnitt konnte nicht in allen Untersuchungsreihen eine Normalverteilung innerhalb der Datensätze nachgewiesen werden, sodass sämtliche Methodiken den nicht-parametrischen Testverfahren entstammen.

4.4.1. Binäre Hauptuntersuchung

Da in den gesamten Informationen statistische Ausreißer existent sind, werden die binären Subdatensätze zunächst einem Median-Test zwecks Ermittlung der vorliegenden Gruppenunterscheidbarkeit bezüglich des Gesamtmedians der Stichprobe unterzogen. Die Ergebnisse werden für jede Merkmalspalte separat erhoben, siehe Tabelle 4.5. Die Ablehnung der Nullhypothese, welche von einer gemeinsamen Ursprungspopulation ausgeht, wird anhand der Unterschreitung des Signifikanzniveaus $\alpha = 0,05$ beziehungsweise der Überschreitung des kritischen χ^2_{krit} -Wertes für einen Freiheitsgrad $d_f = 1$, wodurch $\chi^2_{krit} = 3,48146$ entspricht, definiert. Die Featurewerte mit der Nummerierung 17 und 37 weisen daher keine signifikanten Unterschiede anhand ihrer Klassenannotation vor.

Tabelle 4.5.: Median-Test – Analyse-C – C0402 – zwei Klassen (EF, OK)

Feature-Nummer	χ^2 -Wert	p-Wert	Unterschied signifikant?
17	0,028	0,8660	nein
37	2,188	0,1391	nein
18	6,397	0,0114	ja
15	6,484	0,0109	ja
16	10,025	0,0015	ja
32	12,609	0,0004	ja
3	12,806	0,0003	ja
8	12,808	0,0003	ja
9	15,332	9,02E-05	ja
24	19,388	1,07E-05	ja
31	31,801	1,71E-08	ja
116	32,651	1,10E-08	ja
12	32,657	1,10E-08	ja
35	43,812	3,61E-11	ja
117	48,314	3,63E-12	ja
34	54,505	1,55E-13	ja

6	59,323	1,34E-14	ja
7	59,323	1,34E-14	ja
33	62,820	2,26E-15	ja
25	71,550	2,70E-17	ja
63	108,943	1,67E-25	ja

Die Auflistung A.9 in Anhang A.2.7 enthält die Resultate des U-Tests nach Mann-Whitney, welcher in [Leo17] als teststärker, jedoch anfälliger für Testungenauigkeiten bei auftretenden Ausreißerwerten, eingestuft wird. Anhand jener Dependenzanalyse wird die Einteilung des Featurewertes 17 ebenfalls als nicht signifikant eingestuft. Dementgegen wird diese Einordnung um die Featurewerte 9, 18, 33, 35 und 63 erweitert. Ähnliche Ergebnisse konnten in sämtlichen binären Subdatensätzen erhoben werden. Eine geringe Anzahl von, in den meisten Anwendungsfällen als irrelevant für das Klassifikationsergebnis anhand der Feature-Ranking-Verfahren eingestuften, Featurewerten erfüllt die Voraussetzungen zur Annahme der Nullhypothese, wohingegen die Mehrzahl der Merkmale eine deutliche Signifikanz besitzt.

Im weiteren Verlauf der Zusammenhags-Untersuchung werden die ternären Datensätze (EF, PF, OK) einem H-Test nach Kruskal & Wallis, dessen Ergebnisse im Anhang A.2.7 in Tabelle A.10 dargestellt werden, unterzogen. Im Anschluss wird zur Minderung der α -Fehler-Inflation und Identifikation der Subklassen-Paarung mit vorliegenden, signifikanten Testmetrik-Unterschieden ein Post-hoc-Test nach Dunn-Bonferroni, Tabelle A.11, durchgeführt.

Als abschließende Analyse wird die binäre logistische Regression hinsichtlich der zweiklassigen AOI-Informationen und eine multinominale logistische Regression hinsichtlich der ternären Datensätze angewendet. Siehe Abbildung A.15 und A.16 im Anhang A.2.7. In den Ergebniszusammenfassungen werden die p-Werte der einzelnen unabhängigen Variablen entnommen. Der durch die unabhängigen Variablen erklärte Anteil an der Gesamtvarianz des logistischen Modells Pseudo- R^2 und die restlichen Gütekriterien der Modell-Performance weisen eine geringe Annäherung an die Strukturen und Zusammenhänge der Ausgangsdaten auf. Aus diesem Grund wird den Ergebnissen der logistischen Regression gegenüber den Resultaten der featurebasierten Tests (Median, Kruskal-Wallis und Mann-Whitney) eine geringere Aussagekraft zugeordnet.

4.4.2. Multikategorielle Nebenuntersuchung

Der Versuchsplan der Dependenzanalysen des Themas 2 sieht ebenfalls die Anwendung eines Mediantests und U-Tests nach Mann-Whitney auf die Datensätze mit den Fehler-

klassen OK und Tombstone vor. Die untenstehende Tabelle 4.6 und die Auflistung A.12 in Anhang A.2.8 stellen die Resultate dieser Methoden für die Fehlerklassen Tombstone und OK, das Bauteil C0402 und die Analyse-C dar. Die Featurewerte 9, 18, 24 und 34 weisen in beiden Prüfungen Testmetriken auf, sodass die Existenz eines signifikanten Unterschieds statistisch nicht belegt werden kann. Dem Featurewert 24 wird in Tabelle 4.2 in dem Feature-Ranking-Verfahren nach Boruta ET und Boruta RF ein Wichtigkeitskennwert von 0,298 beziehungsweise 0,267 zugewiesen. Die Rangplatzierungen der anderen Featurewerte (9, 16, 18, 34, 37) sind ebenfalls niedrig.

Tabelle 4.6.: Median-Test – Analyse-C – C0402 – zwei Klassen (OK, Tombstone)

Feature-Nummer	χ^2 -Wert	p-Wert	Unterschied signifikant?
34	0,048	0,827	nein
24	0,375	0,540	nein
32	0,469	0,493	nein
18	1,127	0,288	nein
16	2,040	0,153	nein
9	2,335	0,127	nein
37	2,372	0,124	nein
25	4,098	0,043	ja
31	11,040	0,001	ja
63	14,861	0,0001	ja
8	21,921	2,84E-06	ja
3	21,962	2,78E-06	ja
30	30,277	3,75E-08	ja
35	38,220	6,32E-10	ja
29	42,177	8,34E-11	ja
15	46,852	7,66E-12	ja
33	64,971	7,60E-16	ja
6	83,706	5,74E-20	ja
7	83,706	5,74E-20	ja
116	92,020	8,58E-22	ja
12	92,899	5,50E-22	ja
36	97,692	4,89E-23	ja
117	98,200	3,78E-23	ja
14	104,964	1,24E-24	ja
17	114,620	9,53E-27	ja
110	127,241	1,65E-29	ja
23	385,498	7,91E-86	ja

Weiteren Dependenzanalysen liegen Datensätze zugrunde, in denen der Montagewinkel auf der Leiterplatte als abhängige Antwortvariable definiert wird. Da sich in diesem Fall die Datenstruktur aus mehreren unabhängigen Stichproben zusammensetzt, findet der nicht-parametrische Kruskal-Wallis-Test Anwendung. Dieser erzielt je nach untersuchtem Fehlercode, siehe Tabelle A.13 für FC 0 und Tabelle A.15 für FC 560 in Anhang A.2.8, Unterschiede der nicht signifikanten Featurewerte. Mithilfe der Korrekturen nach Dunn-Bonferroni, siehe Tabellen A.14 und A.16 Anhang A.2.8, werden die Gruppen mit signifikanten Abweichungen ermittelt. Darüber hinaus werden durch die Median- und Mann-Whitney-Tests (Untersuchung nach FC) sowie Kruskal-Wallis-Verfahren (Untersuchung nach Winkel) die Featurewerte mit hohen Rangplatzierungen in den Wichtigkeitskennzahlen als signifikant eingestuft.

Den Abschluss der statistischen Datenexploration des Themas 2 bildet die Durchführung der logistischen Regressions-Methode. Diese prüft die Existenz und Ausprägung eines Zusammenhangs zwischen mehreren unabhängigen Variablen und einer kategorialen, binären Antwortvariable (OK-/Tombstone-Klasse). Die Interpretation der Modellkennwerte, siehe Abbildung A.17 im Anhang A.2.8, steht im Widerspruch zu den Ergebnissen der Kruskal-Wallis-Analysen. Die logistische Regression weist den Winkelgruppierungen als Dummy-Variable kodiert in der Spalte $P > |z|$ p-Werte zu, welche die Existenz signifikanter Unterschiede ausschließen. Aufgrund der Teststärke und der Methodik dieser Verfahren wird den Dependenzanalysen nach Kruskal-Wallis, Mann-Whitney und dem Median-Test eine höhere Relevanz der erhobenen Resultate beigemessen, da in diesen Anwendungen eine Zuordnung der AOI-Messwerte in Gruppen nach dem jeweiligen Filterkriterium (Winkel, FC) erfolgt. Die logistische Regression bewertet die Winkelspalten als eigenständige Variablen.

4.5. Zuordnung relevanter Featurewerte zu Lötstellencharakteristika

Die Zuweisung der in den vorangegangenen Kapiteln untersuchten Featurewerte zu ihren Untersuchungsgegenständen, welche einerseits die AOI-Parametrierungen und andererseits Lötstellen-/Bauteileigenschaften repräsentieren, kann Tabelle 4.7 entnommen werden. Aus Gründen des Technologieschutzes der Continental Automotive GmbH werden die Einheiten der Messwerte nicht angegeben und lediglich ein stark reduzierter Auszug des gesamten Featurewert-Spektrums vorgestellt.

Die in Kapitel 4.1.1 als Träger eines geringen Informationsumfangs identifizierten Fea-

turewerte mit der Nummerierung 13 und 110 werden auch in diesem Kapitel durch das repräsentierte AOI-Prüfungsmerkmal als irrelevant eingestuft. Hierbei handelt es sich um die Anzahl von Eingangsbildern, die Analyse-C zur Verfügung gestellt werden, und die Fehlerkodierung im Falle eines auftretenden, unerwarteten Fehlers während der optischen Inspektion.

Die minimale mittlere Bauteilhöhe (Feature 6), die vertikale Bauteilneigung (Feature 8) und der Schwerpunkt des Bauteils in der x-Koordinate (Feature 29) weisen in allen Untersuchungsreihen der fünf Test-Gehäuseformen hohe Platzierungen des Feature-Rankings auf.

Dem Anhang A.2.9 sind analog zu obiger Betrachtung die tabellarischen Zuweisungsschlüssel der Analysen-A, B, D und E beigelegt.

Tabelle 4.7.: Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-C

Feature-Nummer	AOI-Messobjekt
6	Minimale mittlere Bauteilhöhe
7	Maximale mittlere Bauteilhöhe
8	Vertikale Bauteilneigung
9	Horizontale Bauteilneigung
13	Anzahl Eingangsbilder Analyse
18	Z-Schwerpunkt Referenzpunkt
23	Tombstonehöhe
29	Schwerpunkt des Bauteils in x-Koordinate

4.6. Zwischenfazit statistische Datenanalyse

Die Zusammenstellung und Inhalte des Methodenplans für die statistische Analyse der Strukturen, Unterschiede und Zusammenhänge in den erhobenen, numerischen AOI-Messwerten erlauben eine ganzheitliche Betrachtung mit einem hohen Abdeckungsgrad an ermittelten Kennzahlen. Da in diesem Kapitel zwei Hauptthematiken behandelt werden, teilt sich die Formulierung des Zwischenfazits in zwei Unterabschnitte auf. Zunächst erfolgt die Ergebniszusammenfassung der fehlercode-neutralen binären Datensätze im Rahmen der Hauptuntersuchung als Vorstufe zu ML-Anwendungen. Im Anschluss werden die Resultate der multikategorialen Einflussanalyse des SMD-Bauteil-Montagewinkels auf die Ausbildung von Tombstonedefekten vorgestellt.

Hauptuntersuchung Thema 1 Die Resultate der statistischen Untersuchungen, welche quantitative und qualitative Aussagen über die Eigenschaften der zugrundeliegen-

den AOI-Informationen ermöglichen, weisen eine Reihe von Gemeinsamkeiten auf. Zum einen stellt sich die Ermittlung der Kardinalität eindeutiger Messwertstufen des gesamten Datensatzes und daraufhin der nach Fehlerklassen gefilterten Subdatensätze als zuverlässige Methodik zur ersten Analyse der Datenstrukturen heraus. Diese Resultate werden durch Hinzunahme prüf- und löttechnischer Zusammenhänge bestätigt.

Des Weiteren erweisen sich die implementierten Varianten des Boruta-Algorithmus und die XGBoost-Methode in dieser Aufgabenstellung als zuverlässige und robuste Feature-Ranking-Verfahren, deren Ergebnisse den nachfolgenden Dependenz- und Anpassungstests zugrunde gelegt werden. Die Ergebnisse des SOAP-Algorithmus weisen Widersprüche zu den Dependenzanalysen auf. Der KS-Anpassungstest und die graphische Interpretation von Q-Q- sowie P-P-Diagrammen weisen in der Mehrzahl der Untersuchungen verteilungsfreie Datensätze auf. Konträr dazu wird in einer erhöhten Anzahl von Untersuchungsreihen auf Basis bereinigter Messwerte die Nullhypothese, welche eine gemeinsame Population postuliert, angenommen. Oftmals handelt es sich bei den nachgewiesenen theoretischen Wahrscheinlichkeitsverteilungen um Variationen/Familien der Gauss-Verteilung, wie beispielsweise die gefaltete sowie die generalisierte Gauss-Verteilung, oder um asymmetrische Parametrierungen der Beta-, Gompertz- und Weibull-Verteilung.

Bedingt durch die Nichterfüllung der Normalverteilungs-Voraussetzung finden im Rahmen der Dependenzanalyse zwecks Überprüfung der Klassenunterscheidbarkeit nicht-parametrische Verfahren Anwendung. In zahlreichen Versuchsreihen werden für die Merkmale mit hohen Ranking-Platzierungen signifikante Klassenunterschiede nachgewiesen. Die Durchführung von Post-hoc-Verfahren der ternären Datensätze bestätigt die Ergebnisse und ermöglicht eine Identifikation der signifikanten Fehlerklassen-Paarungen. Featurerwerte, welche das Signifikanzniveau übersteigen und dadurch eine Annahme der Nullhypothese erreichen, weisen oftmals geringe Wichtigkeitswerte auf. Des Weiteren ist es in zahlreichen Anwendungsfällen möglich, diese durch quantitativ beziehungsweise prüftechnisch basierte Kausalformulierungen aus der Betrachtung auszuschließen.

Nebenuntersuchung Thema 2 Die Unterteilung der AOI-Informationen in Subdatensätze hinsichtlich des Fehlercodes und Montagewinkels bewirkt die Generierung von Messwerttabellen mit stark reduzierten Anzahlen von Datenpunkten, sodass die Durchführung von statistischen Analysen und ML-Anwendungen für einige Bauteile und AOI-Algorithmen nicht durchführbar ist. Des Weiteren wird die Ergebniszuverlässigkeit durch kleine Stichproben beeinträchtigt. Aus diesem Grund werden die Versuchsplanung und -durchführung zur Bearbeitung des zweiten Themas angepasst. Ausgehend von fünf Testbauteilen werden zwei SMD-Chipkomponenten, C0402 und C0603, sowie die in den Subdatensätzen vorliegenden Winkel- beziehungsweise FC-Paarungen, welche die Mindestanzahl an Datenpunkten für die jeweiligen Verfahren vorweisen, in den Versuchsplan

integriert.

Bedingt durch die identischen zugrunde liegenden Prüfpläne und deren Parametrierung, weisen die Ergebnisse der Kardinalitätsuntersuchung der Themen 1 und 2 starke Übereinstimmungen der Feature-Rangfolgen und Ausprägungen der Quantität eindeutiger Messwerte auf.

In der Histogrammdarstellung der Featurewerte entstehen in der Mehrzahl der Anwendungsfälle, sowohl durch die Filterung nach dem Montagewinkel als auch dem Fehlercode, Überlappungsbereiche. Die visuelle Trennung unterschiedlicher Klassen mithilfe von Streudiagrammen wird durch die Ausbildung diffuser Datenwolken erschwert.

Unregelmäßigkeiten werden im Vergleich der Resultate der Subdatensätze in der Korrelations- und Dendrogrammanalyse nachgewiesen. Diese treten vereinzelt in der Feature-Ranking-Untersuchung auf, jedoch besteht eine Analogie der Ergebnisse zu den Rangplatzierungen des Themas 1. Eine positive Beeinflussung aufgrund der granularen Selektion von Subdatensätzen wird in der Durchführung von Anpassungstests, im Besonderen der Normalverteilungs-Prüfung, erreicht. Durch die Filterung der Datensätze nach Fehlercode und Montagewinkel werden zugrundeliegende Verteilungen nachgewiesen. Die Anzahl der Datensätze, welche einer theoretischen Wahrscheinlichkeitsverteilung folgen, wird durch die Durchführung von Datenbereinigungs-Verfahren erhöht. Die Anwendung von nicht-parametrischen Dependenzanalysen erfolgt aufgrund des Nachweises von nicht normalverteilten Featurewerten in den Datensätzen. Sämtliche Verfahren waren in der Lage, signifikante Unterschiede in den AOI-Informationen (nach FC und Winkel gefiltert) nachzuweisen. Die Winkelinformationen, siehe Tabelle A.8, erreichen in den Feature-Ranking-Methoden niedrige Rangplätze und somit geringe Wichtigkeitskennzahlen, sodass ein Widerspruch zu den Resultaten der Dependenzanalysen, welche die Existenz von signifikant unterscheidbaren Winkelklassen nachweisen, besteht.

5. Einsatz des maschinellen Lernens und neuronalen Netzarchitekturen in der Lötstelleninspektion

In diesem Kapitel erfolgt die Planung, Umsetzung sowie Ergebnisvorstellung von Methoden aus den Bereichen des maschinellen und tiefen Lernens. Diese Untersuchungsinhalte entsprechen der vierten und fünften Ebene (Wissensextraktion und Validierung/Ergebnisanalyse) des Forschungsplanes in Abbildung 2.29, Kapitel 2.5.

In der ersten Gliederungsstufe wird dieses Kapitel in zwei Abschnitte unterteilt. Die Bestimmung und fertigungstechnische Beurteilung von Prüfplanparametern der AOI-Lötstelleninspektion mithilfe von Verfahren des erklärbaren Lernens wird in Unterkapitel 5.1 vorgestellt. Abschnitt 5.2 beinhaltet die Ergebnisvorstellung der Realisierung von intelligenten Klassifikationsverfahren des maschinellen sowie tiefen Lernens binärer und multikategorieller Datensätze.

5.1. Anwendung von Verfahren des erklärbaren Lernens

Die in diesem Abschnitt vorgestellten Verfahren entstammen der Algorithmenfamilie des erklärbaren Lernens (XAI) und werden in diesem Forschungsprojekt für die Generierung von Regeln in den Methodenplan aufgenommen. Diese Regeln liegen in Form einer Wenn-Dann-Implikation vor und ermöglichen eine merkmalsbasierte Trennung anhand der Klassenannotationen der AOI-Informationen. Im Anschluss erfolgt eine Anpassung der Prüfpläne auf Grundlage der erhobenen Feature-Grenzwerte. Zur Regelextraktion finden sechs Verfahren Anwendung, von denen vier auf einer Architektur von Entscheidungsbäumen basieren und zwei Verfahren rein heuristische Methoden verwenden. Die Vorstellung der verwendeten Algorithmen und des Methodenplans erfolgt in Abschnitt 5.1.1.

Für jeden der 14 originalen Prüfpläne (der fünf SMD-Gehäuseformen und den AOI-Analysealgorithmen) werden neue Prüfpläne, welche die aus den XAI-Methoden generierten Prüfplanparameter aus Abschnitt 5.1.2 enthalten, erstellt. Die sich daraus ergebenden neuen Prüfpläne werden einer Evaluierung ihrer Klassifizierungsgüte in SMT-Fertigungsumgebung unterzogen, siehe Abschnitt 5.1.3.

5.1.1. Aufstellung Methodenplan erklärbares Lernen

Detaillierte Vorstellungen der implementierten Algorithmen sind Kapitel 2.2.1.3 des informationstechnischen Bezugsrahmens zu entnehmen. In den nachfolgenden Tabellen sind die Hauptunterschiede und Funktionsweise der sechs XAI-Algorithmen und die ihnen zugrunde liegenden Quellen der wissenschaftlichen Publikationen und technischen Umsetzungen eingetragen.

Der RRules-Algorithmus und der Ripper-Algorithmus sind in Tabelle 5.1 enthalten und stellen zwei heuristische Regellerner in dem Methodenplan dieses Kapitels dar. Die von diesen Verfahren generierten Regelsätze ermöglichen in [Mos19] die Formulierung von vereinfachten, nicht-hierarchischen Regeln, demnach werden zwei Regellerner dieser Verfahrenskategorie in den Methodenplan aufgenommen.

Tabelle 5.1.: Methodenplan: Entscheidungsbaum-unabhängige XAI-Algorithmen

Name	Kurzbeschreibung	Quelle
Ripper	Generierung Regelsatz durch Iterationschleife. Foil-Informationsgewinn mit anschließender Kürzung der Einzelregeln (Maximierung Fehlermetrik). Beschreibungslänge Abbruchkriterium.	Wissenschaftliche Ausarbeitung in [Coh95]. Python-Implementierung nach [Mos19].
RRule	Iterative, heuristische Definition von Attribut-Wert-Paaren und Ableitung von geltenden Bedingungen im Datensatz. Verzicht auf Basismodell.	Wissenschaftliche Ausarbeitung und Implementierung nach [Pal21].

Eine Vielzahl der Verfahren des erklärbaren Lernens weist Architekturen auf, welche Entscheidungsbäume oder deren Variationen als mathematisches Grundkonstrukt verwenden. Die eingesetzten Vertreter dieser Verfahrenskategorie werden in Tabelle 5.2 betitelt und charakterisiert.

Tabelle 5.2.: Methodenplan: Entscheidungsbaum-basierende XAI-Verfahren

Bezeichnung	Kurzbeschreibung	Quelle
Anchor	Post-hoc-Verfahren. Generiert Beschreibungen (Anchors) in Form von Wenn-Dann-Regeln (welche Features sind nicht änderbar für gleiche Prognose). Coverage gibt an, auf wie viele benachbarte Instanzen Regel zutrifft. Basiert in dieser Arbeit auf einem XGBoost-Modell.	Wissenschaftliche Ausarbeitung in [Kau13]. Implementierung in Python nach [Gue18].
DT	Aufteilung Datensatz in mehrschichtiger Baumstruktur. Maximierung Informationsgewinn bei Aufteilung. Intuitive Regelextraktion durch Pfadanalyse des Entscheidungsbaumes.	Implementierung in Python nach [B ⁺ 11].
Rulefit	Lineares Modell aus Originalmerkmalen und Entscheidungsbaum-Regeln. Regularisierungsverfahren reduzieren Modellkomplexität. Basiert in dieser Arbeit auf Gradient-Boosting-Verfahren.	Wissenschaftliche Ausarbeitung nach [Pop08]. Implementierung in Python nach [Mol16].
SkopeRules	Erweiterung RuleFit-Algorithmus. Frühe Entfernung von Regeln mit niedriger Performance, Doppelungen. Nutzt statt Lasso-Regularisierung den F1-Score. Basiert in dieser Arbeit auf Ensemblemethode aus Entscheidungsbäumen.	Wissenschaftliche Ausarbeitung nach [Pop08]. Implementierung in Python und Anpassung nach [Doc16].

5.1.2. Regelextraktion Etablierung AOI-Prüfmethodik

Training und Regelextraktion Die Hyperparameteroptimierung der Modelle wird im Rahmen einer Rastersuche (mit Kreuzvalidierung) durchgeführt. Die definierte Hyperparameterkombination dient der Basismodell-Generierung. Im Anschluss können durch die umgesetzten XAI-Verfahren für jeden Featurewert im Datensatz numerische Grenzwerte ermittelt werden. Diese Grenzwerte sollen eine optimale Klassentrennung des Datenraumes ermöglichen. Tabelle 5.3 enthält einen Ergebnisauszug des SkopeRules-Algorithmus hinsichtlich SMD-Komponente C0402 und AOI-Analyse-C. Als Metrik und Kontrollkennwert der Klassifikationsgüte der einzelnen Mengenpartitionen finden in diesem Auszug die Precision und der Recall Anwendung. Weitere Bewertungsmaße der generierten Regeln und Grenzwerte sind der Abdeckungsgrad, siehe die Algorithmus-Vorstellung

des Anchor-Verfahrens in Kapitel 2.2.1.3 auf Seite 40, der F1-Score und die Accuracy.

Tabelle 5.3.: Auszug Ergebnisse Skope-Algorithmus – C0402 – Analyse-C – nach Precision sortiert

Featurewert	Vorschrift	Grenzwert	Precision [%]	Recall [%]
Feat 7	<=	419.5	91,5	43,8
Feat 6	<=	419.5	91,5	43,8
Feat 31	<=	2884.0	91,2	37,4
Feat 110	>	3.5	91,2	36,9
Feat 8	<=	0.5	90,2	37,4
Feat 29	<=	501.5	90,0	43,6
Feat 3	<=	0.5	90,0	37,4
Feat 117	>	661.0	88,1	10,5
Feat 24	<=	0.5	87,7	37,6
Feat 30	<=	476.0	87,5	47,9

Die Entscheidungsregeln der Decision Trees können anhand der gerichteten, binären Baumstruktur ausgehend von dem Wurzelknoten abgeleitet werden. Die Komplexität der zur Regelextraktion herangezogenen Entscheidungsbäume wird zwecks Gewährleistung der Eindeutigkeit von Feature-Grenzwert-Zuordnungen durch Hyperparametrierung reduziert. Zu diesem Zweck erfolgt eine Anpassung des Abbruchkriteriums zur Bildung neuer Knoten/Ebenen in der Baumarchitektur mithilfe der maximalen Baumtiefe. Abbildung 5.1 stellt einen Auszug des Entscheidungsbaumes für Bauteil C0402 und Analyse-C dar. Das Attribut, welches eine Maximierung der Metrik des Informationsgewinnes erreicht und anhand dessen eine weitere Wenn-Bedingung der Gesamtregel gebildet wird, ist in diesem Beispiel Featurewert 7.

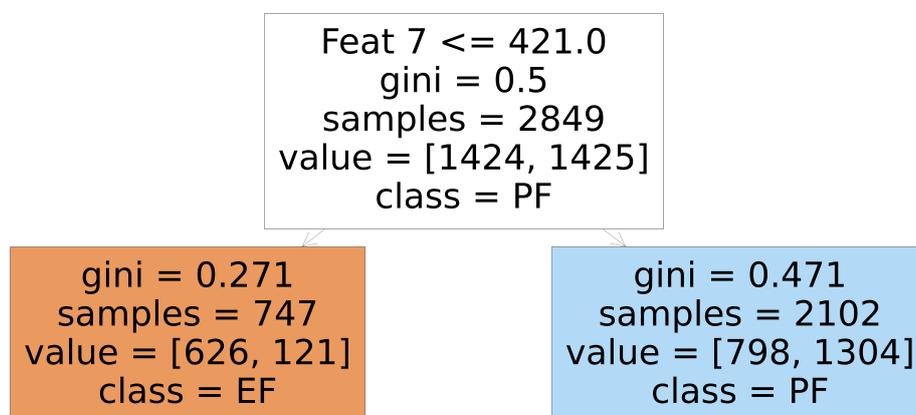


Abbildung 5.1.: Auszug Visualisierung Decision Tree – Aufteilung Knoten anhand Featurewert 7 – C0402 Analyse-C

Sowohl die Bewertungsmetriken als auch die qualitativen sowie quantitativen Charakteristika der von dem RRules-Algorithmus erstellten Wenn-Dann-Vorschriften verbieten dessen Verwendung in weiteren Untersuchungsreihen und der Prüfplanoptimierung. Die Bedingungsteile der Einzelregeln des gesamten Regelsatzes bestehen nicht, wie gefordert, aus numerischen Größenvergleichen in Form von Ungleichungen der Featurewerte (beispielsweise *Wenn Featurewert 7 größer als Wert x*), sondern aus einer Gleichung (*Wenn Featurewert 7 = x*). Der Regelsatz eines Modells setzt sich aus einer Vielzahl von einzelnen Regeln zusammen, wie beispielsweise 1176 Regeln für Bauteil C0402 und Analyse-C mit einer Test-Accuracy von 69,4 %. Diese lokalen und zu stark an die Trainingsdaten angepassten Regeln verhindern eine Generalisierung des Modells. Aus diesen Gründen wird der RRules-Algorithmus aus weiteren Versuchsplänen entfernt.

Erstellung neuer Prüfpläne Dieses Unterkapitel beinhaltet die Synthese aus der Anwendung von XAI-Verfahren sowie der Extraktion von Regelsätzen und deren Integration in einen produktionsbezogenen Inspektionsprozess von Lötstellen. Die XAI-Verfahren besitzen keinerlei Informationen und Kenntnisse bezüglich des AOI-Prozesses und dessen Prüfmethodik. Als Schnittstelle zwischen den rein informationstechnischen Regeldefinitionen und der Neuparametrierung bestehender AOI-Prüfalgorithmen dient die ingenieurwissenschaftliche Prozesskenntnis. Diese erlaubt eine Übersetzung der Regelsätze in Prüfplanparameter.

Die Aufgabenstellung dieser Forschungsarbeit fordert die Entwicklung einer anwendungsbezogenen Prozessverbesserung ohne eine Veränderung der AOI-Bedienersoftware und darin implementierte Funktionen vorzunehmen. Die Änderungen sollen sich dabei auf die dem Anwender möglichen Parametrierungen und Erstellung von Prüfmethodiken und -plänen beschränken. Aufgrund vielfältiger Eigenschaften der Bedienersoftware, der digitalen Bildverarbeitung sowie der regelbasierten Klassifikation des AOI-Prozesses entstehen somit Limitierungen und Widersprüche zu den erhobenen Regelsätzen.

Die symmetrische SMD-Chipbauform (Kondensatoren/Widerstände) und die relative Lage der Anschlusskappen zueinander werden in der AOI-Prüfmethodik berücksichtigt. Die Verfahren der digitalen Bildverarbeitung und die Parametrierung der Klassifikationskennwerte nutzen die vorliegende Bauteilsymmetrie. Daraus resultieren unterschiedliche Herausforderungen der Implementierung der XAI-Regelsätze.

In der Positionskontrolle der zweipoligen SMD-Gehäuseformen sind die Klassifikationsgrenzwerte mehrerer Featurewerte miteinander verknüpft und identisch. Unterschiedliche Kennwerte der XAI-Verfahren müssen aus diesem Grund auf Plausibilität überprüft werden. Da nur einer dieser möglichen Kennwerte im Prüfplan etabliert werden kann, ist entweder einer dieser Pläne auszuwählen oder mehrere unabhängige Prüfpläne zu erstellen und deren Klassifikationsgüten miteinander zu vergleichen.

Unabhängig von der Bauteilsymmetrie existieren weitere Wechselbeziehungen unterschiedlicher Prüfplanparameter. Diese treten daher auch in der Prüfmethodik des Transistors SOT23 auf. Beispielsweise sind die minimal und maximal zulässigen Bauteilhöhen in der Höhenkontrolle auf die Angabe der mittleren Bauteilhöhe und einen Toleranzbereich zurückzuführen.

Des Weiteren stimmen die Genauigkeiten, beziehungsweise Datentypen, der von den XAI-Verfahren extrahierten Grenzwerte und der von der Prüfsoftware in den Prüfplänen zu verarbeitenden Zahlen nicht überein. In Tabelle 5.3 sind Gleitkommazahlen enthalten. Die AOI-Analyse-Algorithmen können jedoch ausschließlich Ganzzahlen verarbeiten. Die Definition der Parametervektoren ist in den AOI-Konfigurationsdateien festgelegt. Dort weist eine Vielzahl von Parametern unterschiedliche numerische Schrittweiten auf, in denen eine Änderung des betreffenden Parameters möglich ist. Zusätzlich zu den Schrittweiten sind Minimal- und Maximalwert des Parameters in der entsprechenden Konfiguration hinterlegt. Des Weiteren ist aus jedem Parameter, dem von den XAI-Algorithmen ein negativer Kennwert zugeordnet wird, eine prozesstechnisch logische Prüfvorschrift abzuleiten.

5.1.3. Klassifizierungsgüte der generierten Prüfpläne in Fertigungsumgebung

Die in Kapitel 5.1.2 generierten Prüfpläne werden hinsichtlich ihrer Klassifizierungsgüte auf einen Validierungsdatensatz angewendet und den Ergebnissen der Originalprüfpläne gegenübergestellt. Zu diesem Zweck werden neue, klassenannotierte Ergebnisdateien (Bildbasen) aus der SMT-Serienfertigung, deren Groundtruth durch Reklassifizierung gewährleistet ist, verwendet. Die Implementierung der angepassten Prüfpläne erfolgt nicht in realen Produktionslinien und AOI-Systemen, sondern unter Zuhilfenahme eines Offline-Programmierplatzes als gleichwertige Simulation. Hierbei handelt es sich um eine digitale Abbildung der AOI-Bedienersoftware und korrespondierenden Maschinenvarianten, samt der darin integrierten Analyse-Algorithmen und regelbasierten Lötstellenklassifikation. Vorteil dieses Vorgehens ist die Erhebung von produktionstechnischen Resultaten ohne aktiv in die Serienfertigung einzugreifen.

Im Anschluss an die Durchführung eines Prüfplans können die AOI-Verifikationsergebnisse erhoben und analysiert werden. Da die Klassenannotation der Bildbasen eine Einteilung des Bauteil- und Lötstellenzustandes in n.i.O. (EF-Klasse) und i.O. (OK-Klasse) beinhaltet, ist es möglich, eine Konfusionsmatrix als Grundlage für die Leistungsmessung anhand unterschiedlicher Metriken zu bilden. Die binäre 2×2 -Konfusionsmatrix repräsentiert die AOI-Klassifikation unter Berücksichtigung der tatsächlichen Klasseneinteilung. Tabelle 5.4 enthält die ermittelte Konfusionsmatrix des Originalprüfplans hinsichtlich

Bauteil C0402 und Analyse-C.

Tabelle 5.4.: Konfusionsmatrix – Originalprüfplan – C0402 – Analyse-C

Konfusionsmatrix		Vorhersage	
		OK	EF
Groundtruth	OK	40	502
	EF	10	495

In der folgenden Ergebnisvorstellung finden die Metriken Precision, Recall, F1-Score und Accuracy Anwendung. Sowohl die auf Entscheidungsbäumen basierenden Regelsätze als auch die durch den heuristischen Ripper-Algorithmus generierten Prüfpläne können in Untersuchungsreihen Verbesserungen der Klassifikationsgüte erreichen.

Ergebnisse Analyse-A Analyse-A weist eine niedrige Komplexität und daraus resultierende geringe Anzahl Featurewerte auf, sodass sowohl die Regelextraktion als auch deren Überführung in Prüfplanparameter und die anschließende Prüfplananpassung erleichtert werden.

Eine Zusammenfassung der Bewertungsmetriken ist in Abbildung 5.2 dargestellt und umfasst vier SMD-Chipbauteile sowie fünf XAI-Verfahren und den Originalprüfplan. Dem Recall der EF-Klasse wird eine größere Bedeutung als der Precision zugeordnet, da die Sensitivität, mit der vorhandene EF im Datensatz vorhergesagt werden, in dieser Auslegung der Aufgabenstellung relevant ist.

Der Recall der Bauteile C0402, C0603 und R0402 wird durch die Umsetzung der XAI-Regelsätze im Vergleich zur originalen AOI-Prüfung verbessert, wobei der RuleFit-Algorithmus das effektivste Verfahren ist. Im Gegensatz dazu vermag der Originalprüfplan bezüglich Bauteil R0603 einen Recall von 100 % zu erreichen. Dieses Resultat wird von keinem der XAI-Verfahren erreicht.

In der Analyse der Precision-Metrik erreicht der Ripper-Algorithmus für die Bauteile C0402, C0603 und R0402 die stärkste Zunahme der Klassifizierungsleistung. Die Resultate der Komponente R0603 widersprechen dieser Aussage, da der Ripper-Algorithmus eine Verschlechterung der Metrik verursacht und die Skope-Methode den stärksten Leistungsanstieg des Bauteils R0603 aufweist.

Der Ripper-Algorithmus ermöglicht laut der Accuracy-Bewertung hinsichtlich der Bauteile C0402, C0603 und R0402 mit großen Differenzen zu den restlichen XAI-Verfahren die Realisierung der größten Anteile korrekter Vorhersagen des Modells. Es wird darüber hinaus in allen Anwendungen der XAI-Regelsätze (ausgenommen RuleFit für C0402 und R0402) eine Zunahme der betreffenden Metrik nachgewiesen. Analog zu der Precision-

Evaluierung des Bauteils R0603 versagt auch hier der Ripper-Algorithmus. Die Anwendung der Decision Tree-, Skope-, RuleFit- und Anchor-Verfahren bewirkt Verbesserungen der Accuracy-Metrik des Bauteils R0603 zwischen 43 % (Decision Tree) und 47 % (Skope).

Das F1-Bewertungsmaß als harmonisches Mittel der Precision und des Recalls stellt aufgrund der Aufgabenstellung, Dateneigenschaften sowie Rahmenbedingungen und Anforderungen an den Klassifikator die final zu beurteilende Metrik dar. In sämtlichen Untersuchungsreihen der Chip-Bauteile ist eine Verbesserung des F1-Scores nachzuweisen, wobei für Bauteil C0402 der Decision Tree, Bauteil C0603 RuleFit, Bauteil R0402 Skope und R0603 Anchor das größte Leistungswachstum erreichen. Die ausgewählten XAI-Methoden sind für Analyse-A und die Versuchskomponenten als robuste, leistungsstarke Verfahren zur Verbesserung der Prüfpläne einzustufen.



Abbildung 5.2.: Ergebnisse Prüfplannerstellung – Analyse-A

Ergebnisse Analyse-B Die Prüfmethodik und Eigenschaften der AOI-Analyse-B resultieren in einem komplexen Übersetzungsprozess der XAI-Regelsätze in anwendbare

AOI-Prüfpläne. Es existieren zusammenhängende, sich gegenseitig beeinflussende Verbände von Featurewerten und korrespondierenden Grenzwerten. Dies betrifft sowohl das Matching-Verfahren der digitalen Bildverarbeitung zur Lötstellen- sowie Bauteilerkennung als auch die Definition der maximal zulässigen Bauteilverdrehung. Des Weiteren erschwert die Existenz einer Vielzahl von Matchingfenstern, Beleuchtungen sowie möglichen Analyse-Iterationsstufen und die hohe Klassifikationsleistung des Originalprüfplans die Umsetzung der XAI-Vorschriften. Die Ergebniszusammenfassung der Prüfplan-Anpassung von Analyse-B ist in Abbildung 5.3 dargestellt.

In der graphischen Beurteilung der Precision-Metrik ist eine geringe Leistung der XAI-Verfahren abzuleiten. Lediglich der Ripper-Algorithmus bewirkt für Bauteil C0603 ein Wachstum des Bewertungsmaßes um 6,9 %. In den restlichen Untersuchungsreihen wird im Rahmen der XAI-basierten Prüfpläne im Vergleich zum Ausgangsprüfplan eine Reduzierung des Anteils der korrekten Vorhersagen nachgewiesen und erreicht für Bauteil R0603 und den Skope-Algorithmus eine maximale Abnahme um 81 %.

Im Gegensatz dazu stehen die Ergebnisse der EF-Sensitivität in Abbildung 5.3b. Eine perfekte EF-Identifikation liegt für die Bauteile C0402 (Skope, Decision Tree), C0603 (RuleFit) und R0402 (Ripper) vor. Mit einer EF-Recallrate von 99,42 % hinsichtlich Bauteil C0402 und der RuleFit-/Anchor-Verfahren wird auch in dieser Versuchsreihe eine Leistungserhöhung nachgewiesen. Die angepassten AOI-Prüfverfahren sind somit in der Lage, einen hohen Anteil von Echtfehlern in dem Datensatz zu identifizieren.

Unter Zuhilfenahme der Konfusionsmatrizen können die negativen Änderungen der Accuracy-Maßzahlen begründet werden. Zwar weisen die Modelle hohe EF-Recallraten auf, jedoch ist die Anzahl der Falschpositiv-Instanzen (Vorhersage EF, Groundtruth PF) ebenfalls auf einem hohen Niveau. Aufgrund dieses vermehrten Aufkommens von Pseudofehlern durch das AOI-System wird der Accuracy-Wert des gesamten Klassifikators negativ beeinflusst. In jeder Versuchsreihe, ausgenommen C0603 und Ripper-Algorithmus mit einer Verbesserung von 5,8 %, werden Leistungsabfälle von bis zu 72,7 % (R0603 Anchor) ermittelt.

Ebenso wird die Berechnung des F1-Scores durch die positiv eingestuften Recallraten und negativen Precision-Kennzahlen beeinflusst. Die Lötstelleninspektion des Bauteils C0603 erreicht durch den Ripper-Algorithmus eine Verbesserung um 5,7 %. Die Ergebnisauswertung des Bauteils R0402 und des Skope-Verfahrens weist eine F1-Score-Steigerung um 11,5 % im Vergleich zum originalen Prüfplan auf.

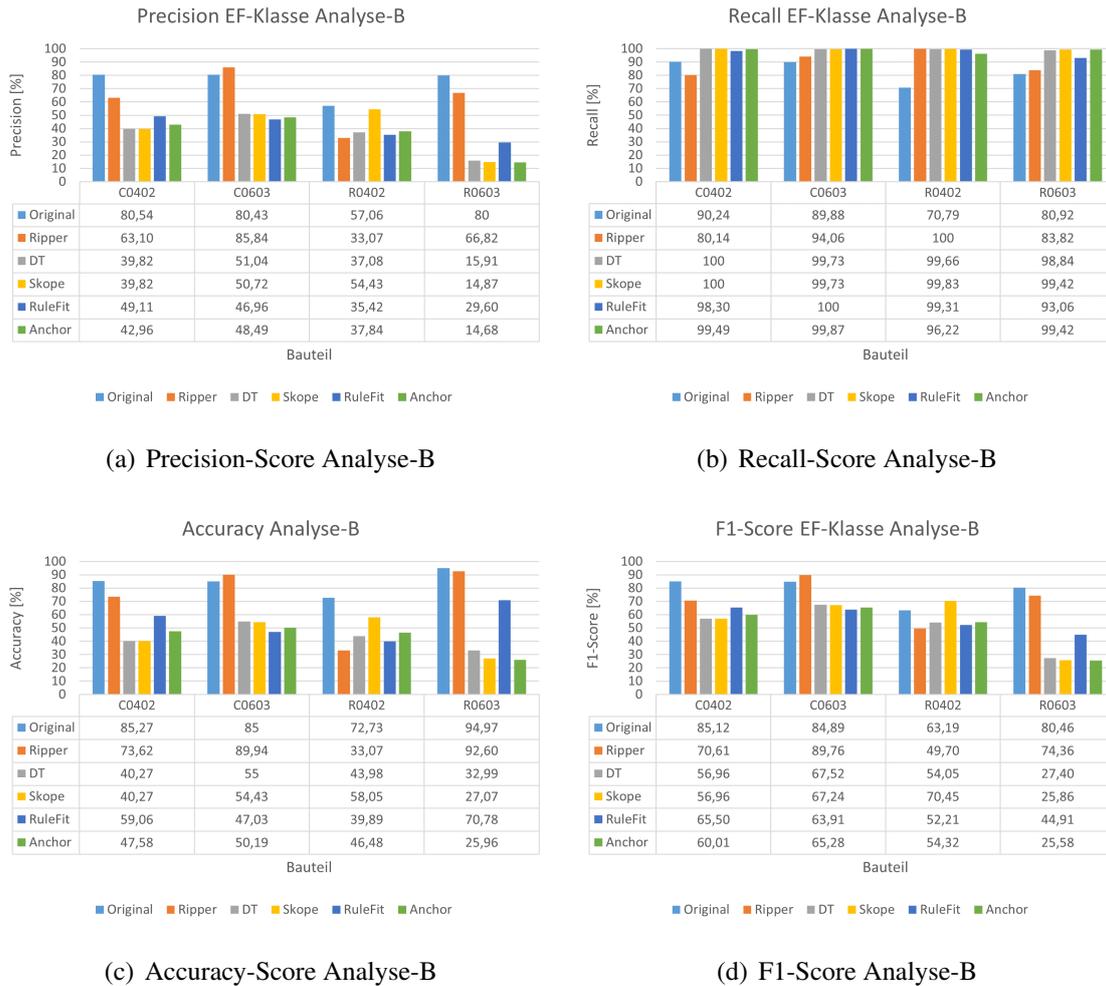


Abbildung 5.3.: Ergebnisse Prüfplanerstellung – Analyse-B

Ergebnisse Analyse-C Die dreidimensionale Analyse-C wird im Rahmen einer Anwesenheits- und Positionskontrolle sowie der Bestimmung der Koplanarität von SMD-Bauteilen verwendet und weist den höchsten Komplexitätsgrad der in dieser Forschungsarbeit verwendeten AOI-Algorithmen auf.

Die Grenzwerte der Höhenkontrolle dieser mehrschrittigen AOI-Analyse werden durch Kombination des zulässigen Toleranzintervalls der Bauteilhöhe mit der mittleren Bauteilhöhe bestimmt. Infolgedessen sind die Ausprägungen der minimalen und maximalen Bauteilhöhe miteinander verbunden, wodurch eine Limitierung des Flexibilitätsgrades der Anpassungsmöglichkeiten des Prüfplans entsteht.

Analog zur vorigen Ergebnisevaluierung treten auch in Analyse-C EF-Recallraten von 100 % auf. Die auf Entscheidungsbäumen basierenden XAI-Verfahren erreichen in fast jedem Anwendungsfall perfekte EF-Sensitivitäten. Der heuristische Ripper-Algorithmus (SOT23 ausgeschlossen) erreicht nicht diese Kennzahlen. Die Veränderungen der Prüfpläne durch die XAI-Regelsätze bringen vermehrt konservative AOI-Prozesse hervor. Echt-

fehler werden perfekt identifiziert, jedoch zum Nachteil der Pseudofehlerrate. Jede Instanz der OK-Klasse wird von dem AOI-System als fehlerhaft klassifiziert. Die entsprechenden Konfusionsmatrizen bestätigen dies, siehe Tabelle 5.5. Aufgrund dieser Ergebnisse sind die Bewertungsmetriken in einer Vielzahl von Untersuchungsreihen identisch. Die Sensitivität der OK-Klasse des Originalprüfplans, hergeleitet aus der Konfusionsmatrix in Tabelle 5.4, ist mit 7,97 % auf einem niedrigen Niveau. Bedingt durch die hohen Recallraten fallen die reziproken Precision-Niveaus deutlich geringer aus. Lediglich das Bauteil R0402 erreicht durch das Decision Tree-Verfahren eine Zunahme von 8,3 % und sämtliche XAI-Verfahren des Bauteils SOT23 geringe Erhöhungen zwischen 0,5 % (Ripper, DT, Skope, RuleFit) und 1,4 % (Anchor) des Precisionwertes.

Die Accuracy- und F1-Bewertung weist ebenfalls für sämtliche XAI-Verfahren des Bauteils SOT23 eine geringe Leistungserhöhung auf. Die Prüfgenauigkeiten des Bauteils R0402 für Ripper und Decision Tree und die F1-Rate nur für Decision Tree erlauben eine bessere Klassifikationsgüte im Vergleich zum Originalprüfplan. Alle generierten Prüfpläne der SMD-Komponenten C0402 und C0603 sind mit einer Verschlechterung der Prüfmetriken verbunden. Das heuristische Ripper-Verfahren und die Entscheidungsbaum-basierenden Methoden sind in dieser AOI-Analyse als gleichwertig einzustufen.

Tabelle 5.5.: Konfusionsmatrix – RuleFit-Algorithmus – C0402 – Analyse-C

Konfusionsmatrix		Vorhersage	
		OK	EF
Groundtruth	OK	0	542
	EF	0	505

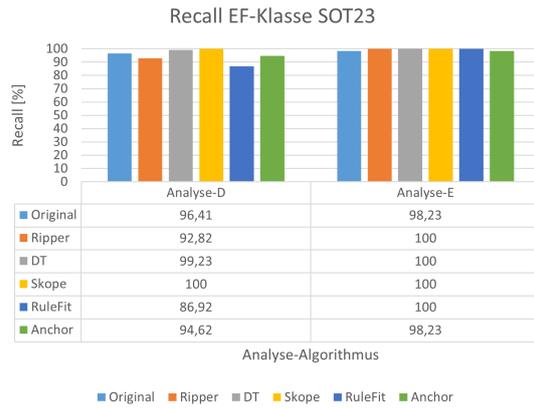
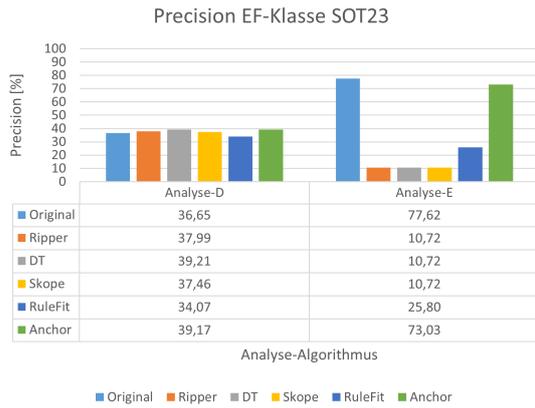


Abbildung 5.4.: Ergebnisse Prüfplanerstellung – Analyse-C

Ergebnisse Analysen-D und -E Die Integration der XAI-Regelsätze in die Prüfpläne der Analysen-D und -E wird durch die Prüfstruktur erleichtert und eine flexible Anpassung der Grenzwertparameter ermöglicht.

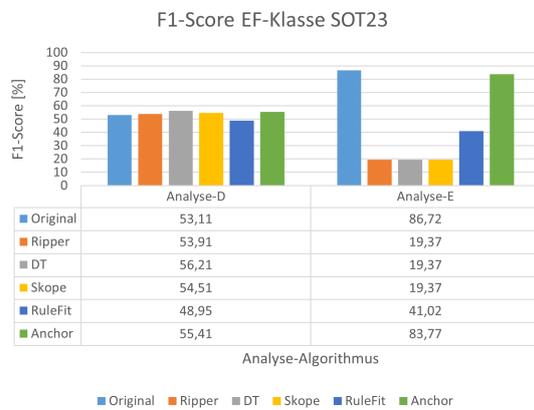
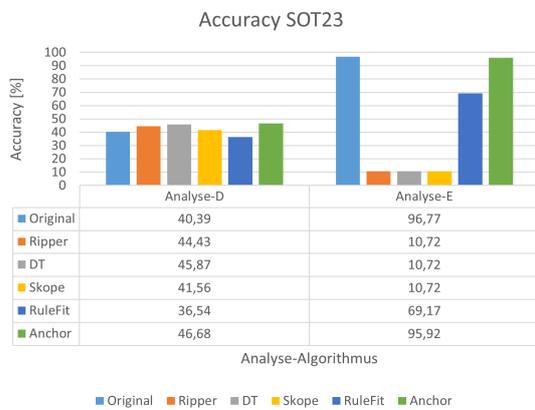
Die EF-Recallraten beider AOI-Analysen konnten durch XAI-Verfahren verbessert werden. Der Skope-Algorithmus erreicht in beiden Untersuchungsreihen eine Echtfehleridentifikation von 100 %. Analyse-E bringt in der Evaluierung der Precision ausschließlich Verschlechterungen hervor. Im Gegensatz dazu werden bei Analyse-D Verbesserungen von bis zu 6,9 % nachgewiesen. Lediglich der RuleFit-Algorithmus ruft eine Abnahme der Bewertungszahl Precision hervor.

Der F1-Score und die Accuracy der XAI-basierten Prüfpläne von Analyse-E sind in jeder Untersuchung geringer als die des Originalprüfplans. Analyse-D kann (außer RuleFit) in beiden Metriken Verbesserungen gegenüber dem Originalprüfplan induzieren.



(a) Precision-Score – SOT23 – Analysen D und E

(b) Recall-Score – SOT23 – Analysen D und E



(c) Accuracy-Score – SOT23 – Analysen D und E

(d) F1-Score – SOT23 – Analysen D und E

Abbildung 5.5.: Ergebnisse Prüfplanerstellung – SOT23 – Analysen D und E

5.2. Anwendung von Verfahren des überwachten Lernens

Dieses Kapitel umfasst in Abschnitt 5.2.1 die Vorstellung der verwendeten Klassifikationsverfahren aus den Bereichen des maschinellen und tiefen Lernens sowie die Ergebnisvorstellung in Abschnitt 5.2.2. Die Modellierung und Generierung von Klassifikatoren werden einerseits anhand eines binären Datensatzes, mit den Klassenannotationen EF-/OK-Klasse, und andererseits mithilfe multikategoriemer Datensätze realisiert. In der multikategoriemer Anwendung wird die EF-Klasse spezifiziert und durch die Angabe der Fehlercodes werden die Löt-/Bauteilfehler in die Informationsbasis aufgenommen.

5.2.1. Aufstellung Methodenplan überwachtetes Lernen

Im Verlauf der Untersuchung des Forschungsstandes zwecks Modellierung von Klassifikatoren numerischer, klassenannotierter Informationen werden zahlreiche Algorithmen auf die Testdatensätze angewendet. Im Anschluss an die Entfernung ineffektiver und nicht performanter Modelle entsteht der in Tabelle 5.6 dargestellte Methodenplan.

Es finden sieben Verfahren des Supervised-Learning Anwendung. Die ersten beiden Verfahren, das Perzeptron und TabNet, entstammen dem Bereich des tiefen Lernens und setzen sich aus Variationen neuronaler Netzwerkarchitekturen zusammen. Vertreter von Methoden des maschinellen Lernens sind Random Forest, Support Vector Machine und XGBoost. Verfahren, welche auf genetischen Evolutionsalgorithmen basieren, sind der MOEFS-Algorithmus (englisch Multi Objective Evolutionary Fuzzy System) und der GANN-Algorithmus des Framework PyGad.

Die Bestimmung der Hyperparameter eines jeden Algorithmus wird durch Rastersuche realisiert und eine Steigerung der Bewertungsmetriken erreicht.

Tabelle 5.6.: Methodenplan AOI-Klassifikatoren – binär und multikategoriemer

Bezeichnung	Kurzbeschreibung	Quelle
Perzeptron	Mehrlagiges neuronales Netz mit Hiddenlayern. Feedforward und fully connected. Dropoutschichten mit 50 % Rate und Batchnormalisierung.	Eigene Netzarchitektur. Implementierung in Keras.
Tabnet	Von Google entwickeltes neuronales Netzwerk. Auf tabellarische Daten spezialisiert. Encoder-Decoder-Architektur mit Feature-Transformations-Blöcken.	Wissenschaftliche Ausarbeitung nach [Pfi19]. Implementierung in Python.

RF	Ensemble-Modell aus Einzelklassifikatoren (Decision Tree).	Python-Implementierung nach [B ⁺ 11].
SVM	Ermittlung Hyperebene und Maximierung Margin mithilfe von Stützvektoren.	Implementierung in Python nach [B ⁺ 11].
XGBoost	Gradient-Boosting-Methode auf Basis von Entscheidungsbäumen.	Implementierung in Python.
MOEFS	Evolutionsalgorithmus mit multikriterieller Optimierung. Klassifikation und Generierung von Fuzzy-Regelsatz.	Wissenschaftliche Ausarbeitung und Implementierung in Python nach [Duc20] und [Duc16].
GANN	Neuronale Netzwerkarchitektur. Trainingsprozess auf Basis eines genetischen Algorithmus.	Implementierung in Python nach [Gad21].

5.2.2. Validierung binärer und multikategoriereller Klassifikatoren

Die im vorigen Kapitel vorgestellten ML-/DL-Verfahren werden den sowohl auf die binären als auch auf die multikategoriiellen Datensätze angewendet. In der binären Versuchsreihe, in welcher die Daten mit den beiden Klassen Echtfehler und i.O. annotiert sind, finden sämtliche fünf SMD-Versuchsbauteile Verwendung. Aufgrund der feineren Granularität im Falle einer multikategoriiellen Fehlerklasseninformation und der daraus entstehenden Abnahme von Datenpunkten pro Fehlerkategorie werden die Bauteile C0402 und C0603 den Versuchen unterzogen. Neben der Ergebnisvorstellung erfolgt darüber hinaus ein direkter Vergleich der Bewertungsmetriken der binären und multikategoriiellen Modellierungen. Analog zur XAI-Ergebnisanalyse werden die Performance-Maßzahlen der Accuracy, des F1-Scores, der Precision und des Recall in den binären Versuchsreihen und der F1-Score in den multikategoriiellen Versuchsreihen präsentiert.

Ergebnisse Analyse-A Die Hyperparameteroptimierung und Modellierung der binären Klassifikatoren hinsichtlich Analyse-A erreicht in fünf Fällen Recallraten von 100 %, siehe Abbildung 5.6b. Der GANN-Algorithmus weist für alle vier SMD-Chipbauteile die höchsten Recallwerte auf. Die Metriken des Random Forest und der SVM werden am schlechtesten für die Bauteile R0402 beziehungsweise R0603 (mit 70,83 % und 70 %) bewertet. Der Ausprägung F1-Score als harmonisches Mittel des Recalls und der Precision wird die größte Bedeutung beigemessen. Die Maximalwerte des F1-Scores, siehe Abbildung 5.6d, sind für die Bauteile C0402 durch die Methoden nach RF, nach Bauteil C0603

und MOEFS, nach R0402 und TabNet sowie nach R0603 und XGBoost nachweisbar. Die Gesamtleistung aller Versuchsreihen übersteigt in keiner Anwendung einen F1-Score oder Accuracy von 90 %, weshalb eine Implementierung in der SMT-Fertigung ohne weitere Maßnahmen mit erhöhten Falschklassifizierungs- und Pseudofehlerraten verbunden wäre.

Der Vergleich der F1-Kennzahlen binärer und multikategorieller Modelle in Abbildung 5.8 dokumentiert einen Leistungsabfall der multikategoriellen Klassifikatoren in jeder Versuchsreihe. Dieser erstreckt sich von einem minimalem F1-Score 10,98 % (C0402 GANN-Verfahren) bis zum Maximalwert 51,35 % (C0402 Perzeptron).

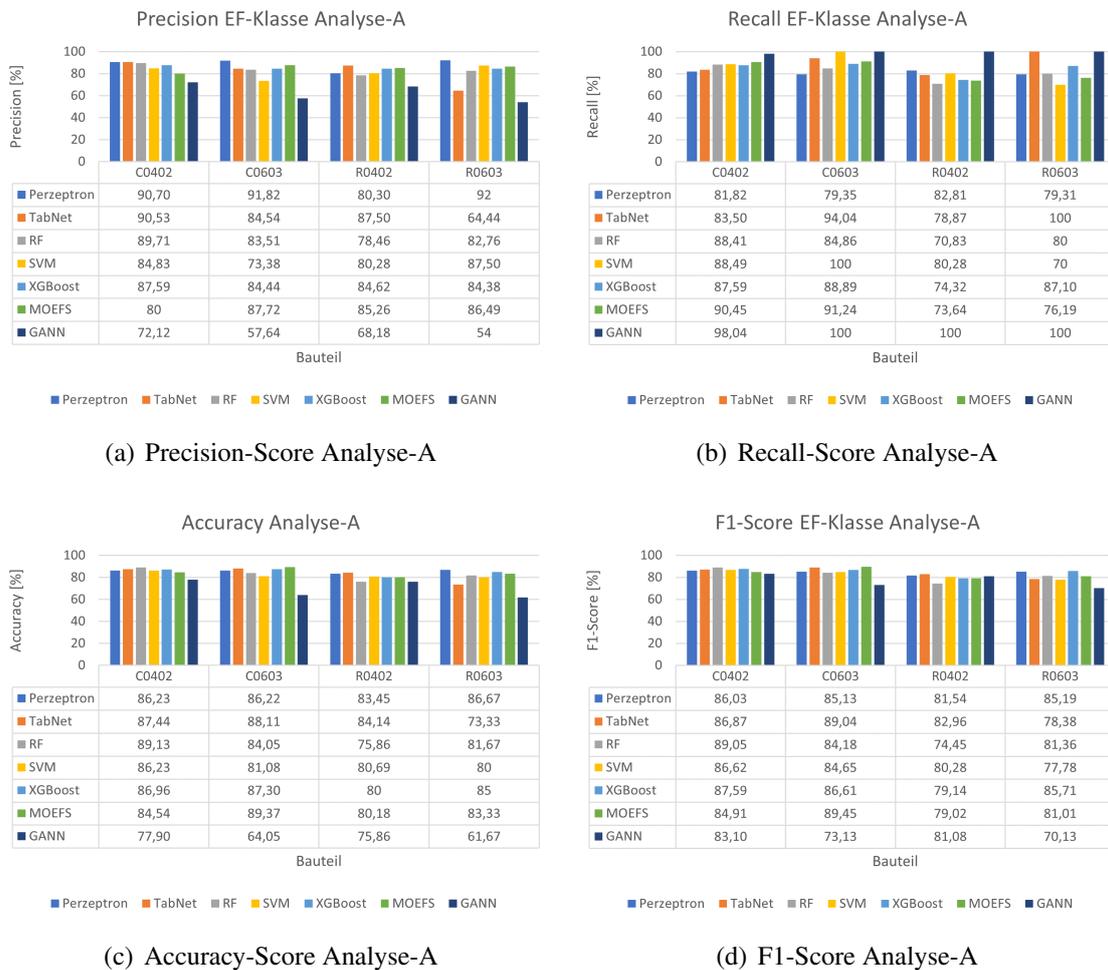


Abbildung 5.6.: Ergebnisse Klassifikatoren – binär – Analyse-A

Ergebnisse Analyse-B Im Gegensatz zu den Resultaten bezüglich Analyse-A liegen die Performance-Metriken der Analyse-B auf einem höheren Niveau, siehe Abbildung 5.8. Sowohl die Recall- als auch die Precisionraten betragen, bis auf wenige Ausnahmefälle (Recall: R0603 SVM und MOEFS sowie Precision: R0603 MOEFS, GANN und Perzeptron), mehr als 90 %. Dies lässt auf hohe Anteile von Richtig-Positiv-Raten mit

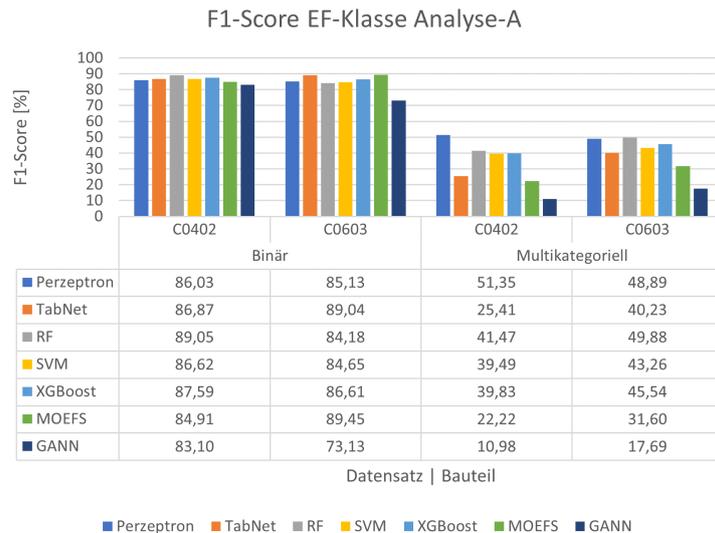


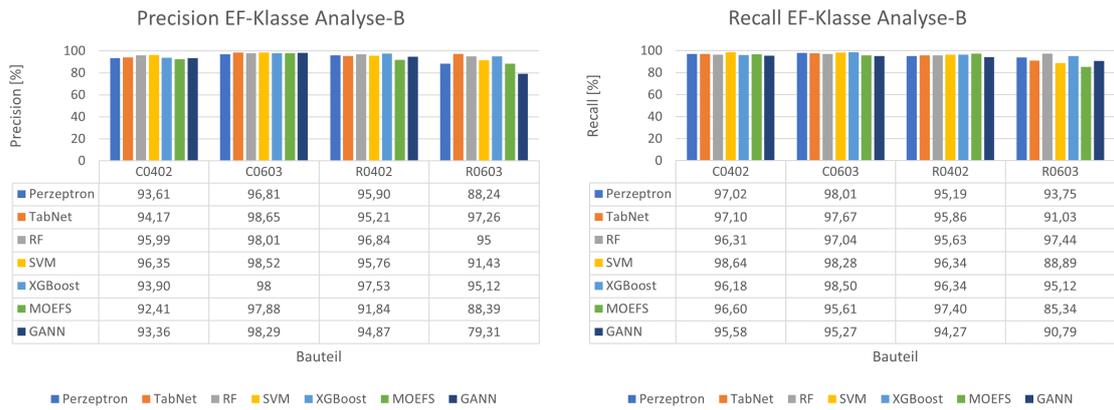
Abbildung 5.7.: Vergleich F1-Score – binär und multikategoriiell – C0402 und C0603 – Analyse-A

gleichzeitiger hoher Sensitivität schließen und spiegelt sich in der Accuracy und dem F1-Score wider. Die zuverlässigste Vorhersage erreichen die Bauteile C0402 und C0603 mit der Methode SVM (97,48 % und 98,4 %), R0402 mit XGBoost (96,93 %) sowie R0603 mit RF (96,2 %). Eine Implementierung in der SMT-Fertigung ist hinsichtlich Analyse-B und den Versuchsbauteilen als zuverlässig einzustufen.

In Abbildung 5.8 ist, analog zu Analyse-A, ein Leistungsabfall der multikategoriiellen Klassifikatoren im Vergleich zu den binären Modellen zu verzeichnen, jedoch fällt dieser geringer aus. Der multikategorielle Ansatz erreicht einen F1-Score bis zu 88,68 % (C0402 SVM). Das maschinelle Lernen, mit den Verfahren SVM, RF und XGBoost, ist in der Lage, Prognosen über 80 % des F1-Scores aufzustellen. Der GANN-Algorithmus weist ebenfalls die schlechtesten Klassifikationsgüten von 22,97 % und 34,86 % bei den SMD-Chips C0402 und C0603 auf. Des Weiteren wird auf die negativen Bewertungskennzahlen der Perzeptron-Modelle und der MOEFS-Methode hingewiesen.

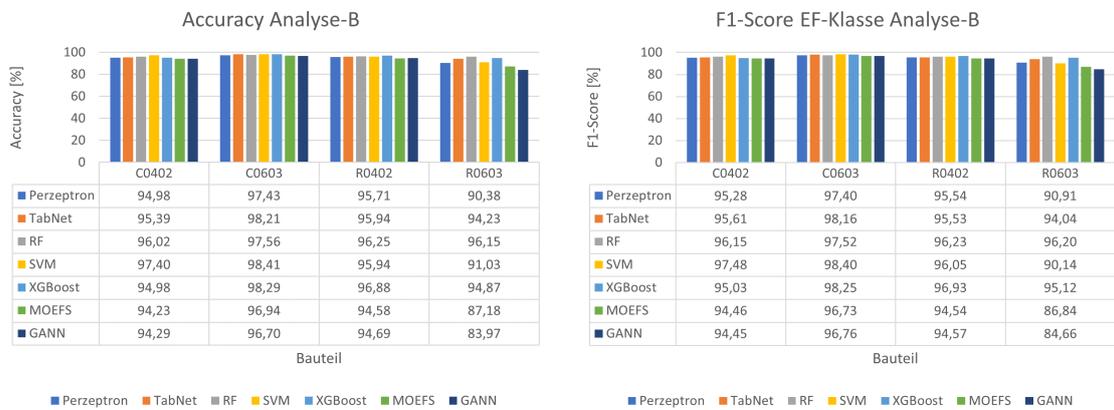
Ergebnisse Analyse-C Die Ergebnisvorstellung der binären Klassifikatoren in Abbildung 5.10 dokumentiert sowohl graphisch als auch tabellarisch homogene und hohe Kennzahlen der vier Bewertungsmaßzahlen. Das Bauteil SOT23 erreicht in der Evaluierung der Accuracy-Metrik und des F1-Scores in der Mehrzahl der Untersuchungen Maximalwerte. Für die neuronalen Netzwerkarchitekturen Perzeptron und TabNet werden Accuracyraten von 98,21 % sowie 99,4 % und F1-Werte von 98,16 % beziehungsweise 99,43 % ermittelt. Im Vergleich dazu sind anhand der Daten des Bauteils C0402 die geringsten Klassifikationsgüten dieses AOI-Analyse-Algorithmus nachweisbar.

Die Differenzen des F1-Scores der binären und multikategoriiellen Modelle, siehe Abbil-



(a) Precision-Score Analyse-B

(b) Recall-Score Analyse-B



(c) Accuracy-Score Analyse-B

(d) F1-Score Analyse-B

Abbildung 5.8.: Ergebnisse Klassifikatoren – binär – Analyse-B

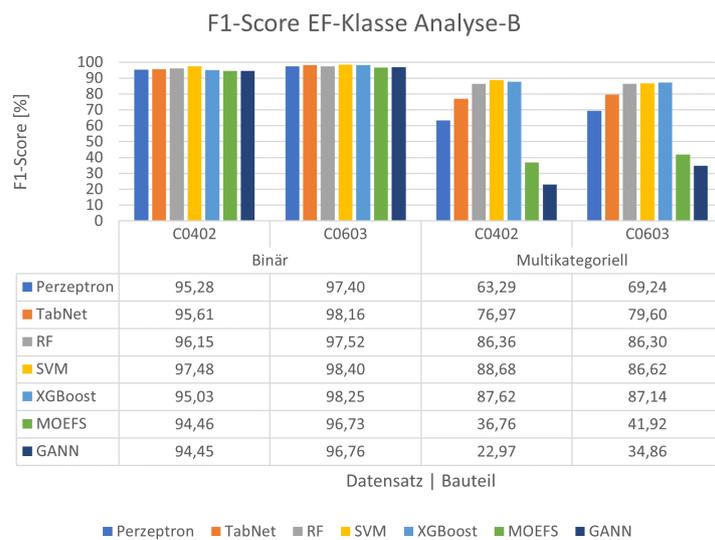
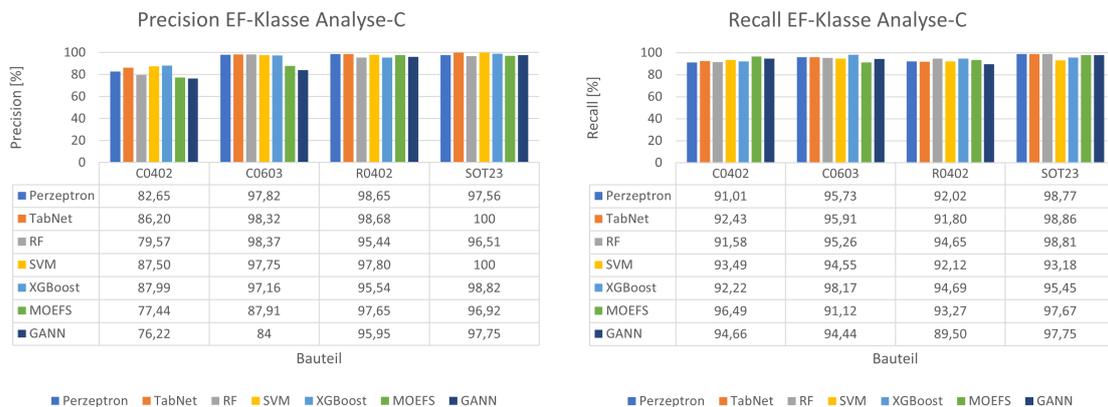


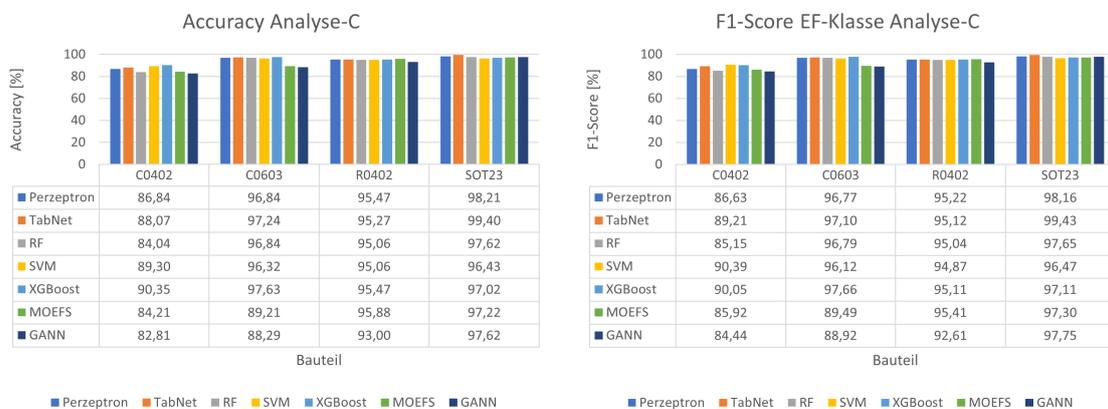
Abbildung 5.9.: Vergleich F1-Score binär und multikategoriiell – Analyse-B
 Vergleich F1-Score binär und multikategoriiell – C0402 und C0603 – Analyse-B

derung 5.11, sind die geringsten aller durchgeführten Gegenüberstellungen (siehe Analyse-A und Analyse-B). Erstmals wird durch multikategorielle Modelle eine höhere Klassifikationsleistung als in der binären Modellierung erreicht, siehe Bauteil C0402 und der RF-Algorithmus sowie die XGBoost-Methode und Steigerungen von 8,9 % sowie 3,4 %. Der GANN-Algorithmus verzeichnet hingegen Einbußen des multikategoriellen Klassifikationsvermögens im Vergleich zu dem binären F1-Score zwischen 23,7 % (C0402) und 33,1 % (C0603). Abschließend sind die drei Vertreter des maschinellen Lernens (RF, SVM und XGBoost) sowie die TabNet-Architektur als effizient zu beurteilen. Leistungsschwach sind die genetischen Evolutions-Verfahren.



(a) Precision-Score Analyse-C

(b) Recall-Score Analyse-C



(c) Accuracy-Score Analyse-C

(d) F1-Score Analyse-C

Abbildung 5.10.: Ergebnisse Klassifikatoren – binär – Analyse-C

Ergebnisse Analysen-D und -E In der Ergebnisvorstellung der Bewertungsmetriken von Analyse-C und -D des Transistors SOT23 in Abbildung 5.12 sind die Recallraten von Analyse-E auf einem niedrigeren Niveau. Demgegenüber ist der Anteil korrekter EF-Vorhersagen (Precision) von Analyse-E erhöht. In den Metriken der Accuracy und des F1-Scores, welche das Klassifikationsmodell gänzlich beschreiben, sind die Ergebnisse

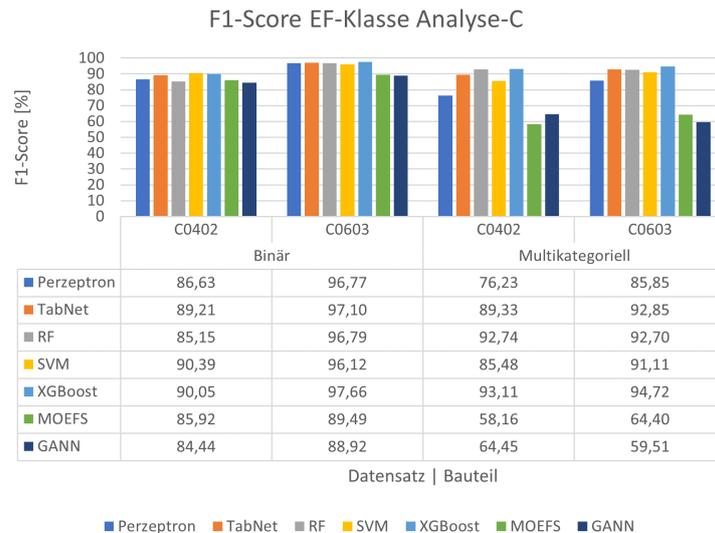


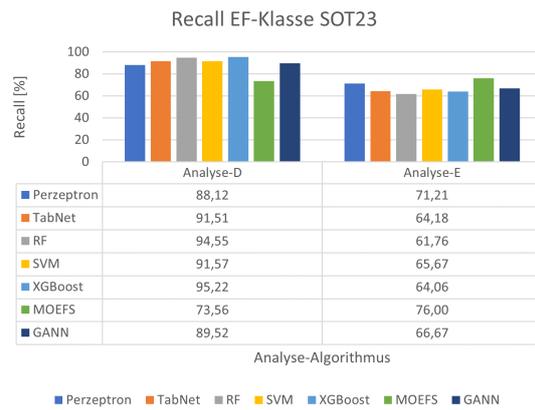
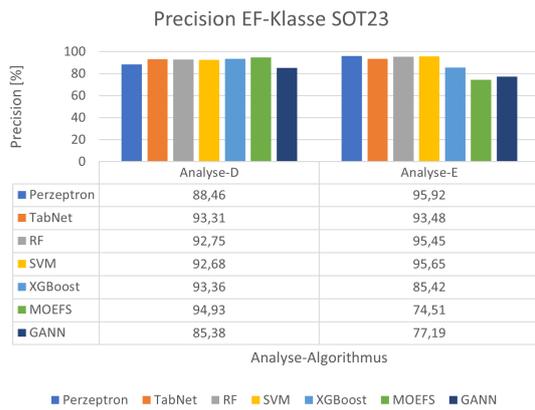
Abbildung 5.11.: Vergleich F1-Score binär und multikategoriiell – C0402 und C0603 – Analyse-C

von Analyse-E auf einem geringen Niveau, welches die Empfehlung einer Implementierung ohne weitere Verbesserungsmaßnahmen der Modelle verhindert. Der XGBoost-Algorithmus bezüglich Analyse-D (Accuracy 94,37 %) und das Perzeptron-Modell hinsichtlich Analyse-E (Accuracy 84,56 %) stellen die leistungsstärksten Klassifikatoren dar.

5.2.3. Gewährleistung der Modellgültigkeit

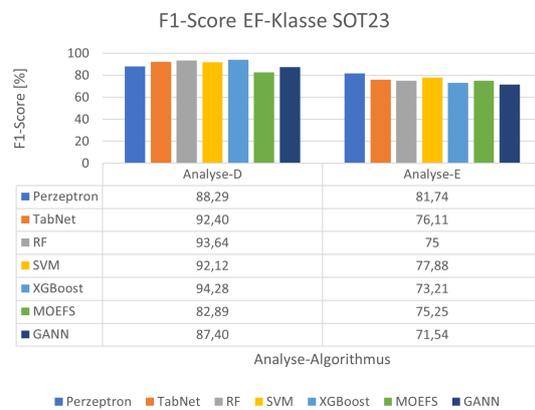
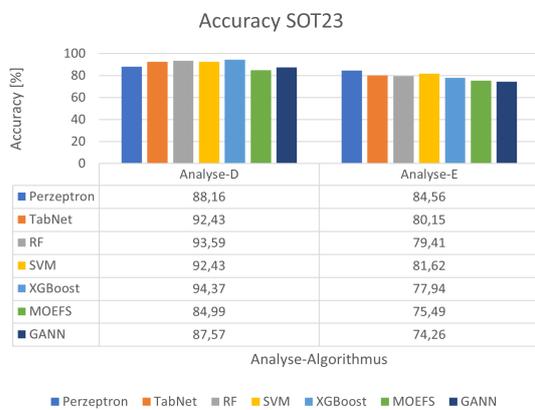
Eine Implementierung von Klassifikatoren in der SMT-Fertigung erfordert die Entwicklung eines Konzepts zur Sicherstellung der kontinuierlichen Aktualität der Modelle. In diesem Kapitel wird eine Methodik vorgestellt, die temporalen Veränderungen der Daten entgegenwirkt. Voraussetzung für diese Methodik ist, dass keine Verletzung der Datenkonsistenz durch Änderungen der Bildaufnahmeparameter oder relevanten Prüfplanparameter vorliegt.

Durch Nachtrainieren der Einzelmodelle und anschließende Zusammenfassung zu einer Blending-Ensemblemethode wird ein Klassifikationssystem generiert, welches die Bewertungsmetriken des leistungsstärksten Einzelmodells in neun von vierzehn Analyse-Algorithmen übertrifft. Abbildung 5.13 enthält eine graphische sowie tabellarische Gegenüberstellung der F1-Kennzahlen des performantesten Einzelklassifizierers und der Ensemblemethode nach Blending auf der Basis von nachtrainierten Lernalgorithmen. Die prozentuale Änderung des F1-Scores ist dem Anhang A.3 in Tabelle A.20 beigelegt.



(a) Precision-Score – SOT23 – Analysen D und E

(b) Recall-Score – SOT23 – Analysen D und E



(c) Accuracy-Score – SOT23 – Analysen D und E

(d) F1-Score – SOT23 – Analysen D und E

Abbildung 5.12.: Ergebnisse Klassifikatoren binär – SOT23 – Analysen D und E

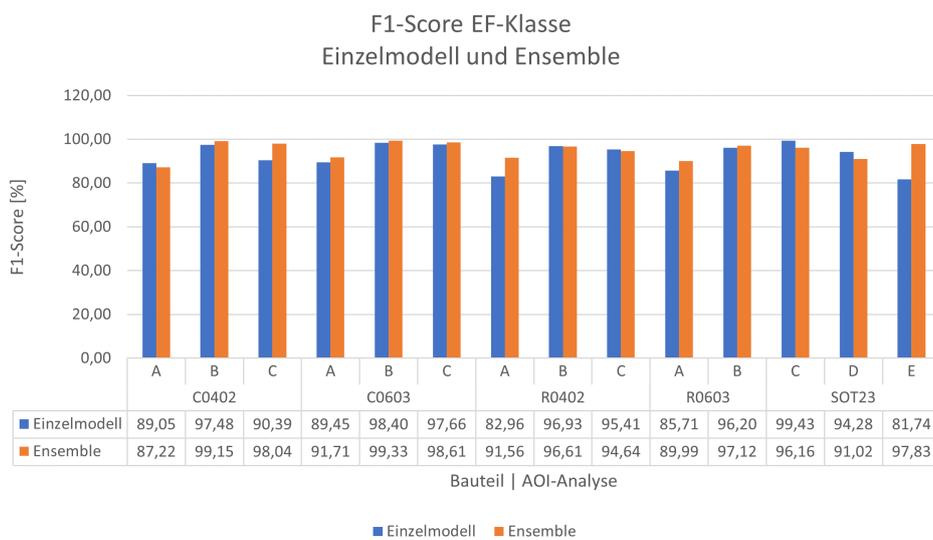


Abbildung 5.13.: Vergleich F1-Score Einzelmodell und Ensemble-Methode – alle Bauteile – alle Analyse-Algorithmen

5.2.4. Zwischenfazit maschinelles und tiefes Lernen

Die Anwendung von Verfahren des maschinellen und tiefen Lernens teilt sich in dieser Forschungsarbeit in zwei Fragestellungen und daraus abgeleitet Untersuchungsschwerpunkte auf. Zum einen erfolgt die Extraktion von Regelsätzen mithilfe von XAI-Verfahren und Übersetzung dieser in AOI-Prüfplanparameter und im Zuge dessen ein direkter Eingriff in den Inspektionsprozess. Die zweite Forschungsthematik befasst sich mit der Generierung von Klassifikatoren, welche nicht in den Prüfplan eingreifen, sondern dem Bediener der Verifikationsstation eine Hilfestellung leisten sollen.

Prüfplananpassung optische Lötstelleninspektion Im Rahmen der theoretischen Literaturforschung werden sechs regelbasierte Verfahren in den Methodenplan aufgenommen, wovon vier Algorithmen auf Varianten von Entscheidungsbäumen basieren und zwei Methoden eine entscheidungsbaum-freie Modellierung verfolgen. Die Ergebnisse des RRules-Algorithmus stellen sich hinsichtlich der Anforderungen der Strukturen und Qualität der Resultate als ineffektiv und nicht zielführend heraus, sodass diese Methode aus der Analyse entfernt wird.

Im Anschluss an die Anpassung der AOI-Prüfpläne werden diese auf neue Validierungsdaten angewendet und die Ergebnisse mit den originalen Prüfplänen verglichen. Eine Mehrzahl der neu konfigurierten Prüfpläne bewirkt die Etablierung von konservativen Prüfstrategien. Echtfehler werden mit einer hohen Zuverlässigkeit erkannt und somit die Echtfehlererkennungsrate positiv beeinflusst, jedoch hat dies eine steigende Pseudofehlerrate zur Folge. Zulässige Lötstellen führen in dem AOI-System zu einem Ausfall und verursachen eine höhere Auslastung der manuellen Verifikation. Die Beurteilung der gesamten Klassifikation durch die Metriken Accuracy und F1-Score führt dennoch in einer Mehrheit der vierzehn Prüfpläne zur Verbesserung der Prognosegüte.

Klassifikatoren des überwachten Lernens Die Verfahren des maschinellen Lernens (Random Forest, Support Vector Machine und XGBoost) erweisen sich in der Auswertung des F1-Scores als die leistungsstärksten Klassifikatoren und erreichen in neun der vierzehn Untersuchungsreihen Höchstwerte. Der genetische, auf einem neuronalen Netzwerk basierende GANN-Algorithmus weist in keinem Anwendungsfall eine maximale Klassifikationsgüte auf. Die Aktualität der Modelle und kontinuierliche Anpassung an Produktionsdaten wird durch Nachtrainings-Prozesse sichergestellt. Eine weitere Leistungssteigerung erzielt die Zusammenfassung der Einzelmodelle zu einer Ensemble-Methode. Die F1-Scores der multikategoriellen Modelle der AOI-Analysen A und B sind in sämtlichen Verfahren geringer als in den binären Modellierungen, während Anwendungsfälle der Analyse-C eine Leistungssteigerung durch die multikategoriellen Daten

erwirken, siehe Abbildung 5.11.

6. Diskussion und kritische Würdigung

Die Erkenntnisse, welche aufgrund der statistischen Beurteilung von klassenannotierten Lötstelleninformationen und Generierung von Klassifikatoren sowie angepassten AOI-Prüfplänen in den Kapiteln 4 und 5 gründen, legitimieren die Beantwortung der Forschungshypothesen, die in Kapitel 2.4 definiert werden. Darüber hinaus werden in Abschnitt 6.2 allgemeine, induktive Lehrsätze aus den Untersuchungsergebnissen und deren Interpretation abgeleitet.

6.1. Ergebnisübertragung auf die Forschungshypothesen

Die aufgestellte Hypothese **H1** fordert die Existenz von signifikanten Unterschieden der Messwertinformationen bezüglich ihres Bauteilzustandes. Im Rahmen der statistischen Datenanalyse werden binäre (EF/OK), ternäre (EF/PF/OK) und multikategorielle (nach FC) Dependenzanalysen durchgeführt. Der Nachweis von signifikanten Unterschieden wird in jeder Untersuchungsreihe erbracht. Dies umfasst eine Mehrzahl von Featurewerten. Dementgegen existiert ein geringer Anteil von Featurewerten in den AOI-Analysen, die anhand ihrer Ergebnisse der Dependenzanalysen keine gemeinsame Ursprungspopulation aufweisen. Jedoch erreichten diese Featurewerte oftmals geringe Wichtigkeitskennzahlen in dem Feature-Ranking und können durch die Hinzunahme von Prozesswissen aus der Betrachtung ausgeschlossen werden.

Die Forschungshypothese **H2** beinhaltet die Annahme, dass die AOI-Messwerte dem zentralen Grenzwertsatz folgen. Die Ergebnisse der statistischen Anpassungstests zeigen, dass die originalen AOI-Messwerte den Nachweis der Normalverteilung erschweren. In einer geringen Anzahl von Featurewerten der binären und ternären Untersuchungen wird die Normalverteilung festgestellt. Durch den Ausschluss von Nullwerten und eine feinere Granularität der Daten, indem der Montagewinkel und Fehlercode in die Betrachtung einfließen, wird der Anteil von normalverteilten Featurewerten in den Anpassungstests deutlich erhöht. Eine Gegenüberstellung der Ergebnisse der KS-Tests für binäre und mul-

tikategorielle Datensätze ist den Tabellen A.3, A.4, A.6 und A.7 im Anhang zu entnehmen.

Mithilfe der Implementierung von vier Verfahren zur Messung der Wichtigkeitskennzahlen konnten Ranglisten generiert werden, die vergleichbare Rangplatzierungen der Featurewerte der binären und fehlercode-basierten Datensätze generieren. Hypothese **H3** wird bestätigt.

Die Beantwortung der Forschungsfrage **H4** bedingt eine getrennte Analyse der PF- und Falschklassifizierungsrate und eine darauf aufbauende Bewertung der gesamten Klassifizierungsgüte. Die Übersetzung der XAI-Regelsätze in Prüfplanparameter und anschließende Prüfplananpassung führt in der Mehrzahl der Untersuchungsreihen zur Bildung von konservativen Inspektionsprozessen. Diese verursachen eine Maximierung der Erkennungsrate von Echtfehlern auf Kosten der Pseudofehlerrate, welche ebenfalls erheblich ansteigt. In der Gesamtbeurteilung der Prognosequalität sind Verbesserungen in Relation zum originalen Prüfplan nachweisbar. Forschungsfrage **H4** wird in mehreren Anwendungsfällen bestätigt.

Die Gültigkeit der Forschungshypothese **H5** wird bewiesen. Durch Ensemblemethoden bestehend aus Einzelklassifikatoren des tiefen und maschinellen Lernens wird in sämtlichen AOI-Prüfungen eine Steigerung der Prognosegüte erreicht. Die Wachstumsfaktoren bezüglich der F1-Scores der originalen Prüfpläne und ML-/DL-Ensembleverfahren sind in Abbildung 6.1 graphisch sowie tabellarisch dargestellt. Die Ergebnisse der statistischen

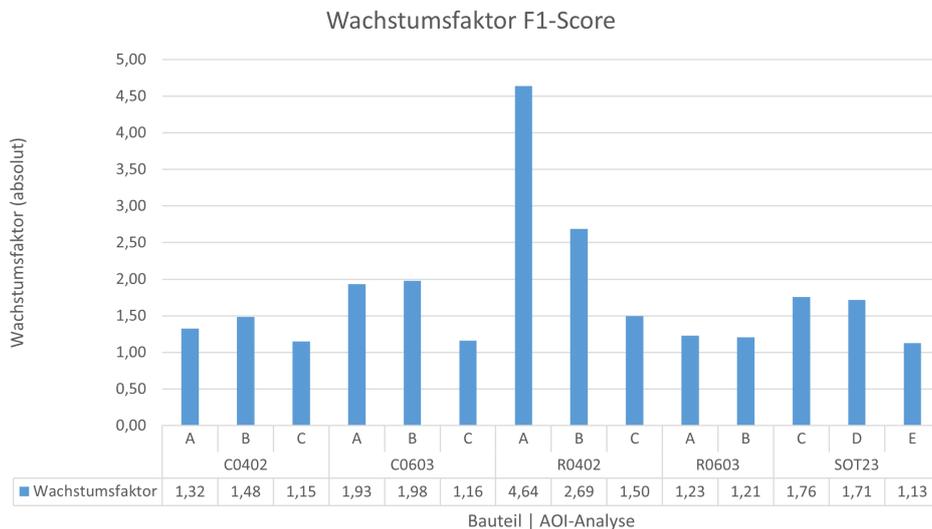


Abbildung 6.1.: Wachstumsfaktor Original-PP und Ensemble-Methode – C0402 und C0603

Datenanalyse weisen die Existenz signifikanter Unterschiede der AOI-Informationen hinsichtlich der Fehlercodierung und Bauteil-Orientierungswinkel auf der Leiterplatte nach. Ein geringer Anteil von Featurewerten, die bei Analyse-B maximal 12,9 % und Analyse-

C 11,5 % betragen, erreicht nicht das kritische Signifikanzniveau. Diesen Featurewerten wird durch die Wichtigkeitsmessung eine niedrige Rangplatzierung zugewiesen. Hypothese **H6** kann somit bestätigt werden, da signifikante Unterschiede in den Daten dokumentierbar sind.

Die in Hypothese **H7** geforderte Leistungszunahme von ML- und DL-Modellen durch multikategorielle Klassenannotation im Vergleich zu binärer Fehlercodierung wird lediglich in drei Untersuchungsreihen der Analyse-C nachgewiesen, siehe Abbildung 5.11. Analyse-A und -B verzeichnen keine Metriksteigerung durch multikategorielle Modellierungen. Jedoch ist auf die reduzierte Datenmenge pro Fehlercode zu verweisen, die ebenfalls einen Einfluss auf die Prognosegüte und Bewertungs-Maßzahlen ausüben kann.

6.2. Ableitung ingenieurwissenschaftlicher Lehrsätze

Zusätzlich zur Beantwortung der Forschungsfragen können aus den gewonnenen Erkenntnissen allgemeine Lehrsätze formuliert werden.

Die Aufgabenstellung dieser Forschungsarbeit beinhaltet die Erhebung und Verarbeitung von numerischen Messwerten des automatischen optischen Inspektionsprozesses. Aufgrund dieser Forderung erfolgt die Konzeption der Datenakquisition, welche folgende Wechselbeziehung zu produktionstechnischen Rahmenbedingungen aufweist.

Die Gewährleistung der Datenkonsistenz erfordert eine statische Produktionssituation mit unveränderlichen Prüfplänen sowie Prozessparametern und steht somit im Widerspruch zur erforderlichen dynamischen Anpassung der Fertigungsprozesse.

Durch die Integration von AOI-Ergebnisdateien, die Bild- und Ergebnisinformationen des Inspektionsprozesses beinhalten, in die Datenerhebung wird die Unabhängigkeit von Prüfplanparametern ermöglicht. Bei einer unveränderlichen Bildaufnahmesituation und Prüfstruktur (Verlinkungen einzelner Analyse-Algorithmen, aktive Analysen) können neue Featurewerte mithilfe der Ergebnisdateien und anschließend neue ML-/DL-Modelle generiert werden.

Die Konzeption geeigneter Dateiformate erhöht den Flexibilitätsgrad, sodass Änderungen der Prüfplanparametrierung durch Anpassung der Featurewerte und Neumodellierung der ML-/DL-Verfahren ausgeglichen werden können.

Als weitere Konsequenz der Aufgabenstellung entsteht der folgende Lehrsatz. Dieser betrifft die erforderliche Struktur der numerischen AOI-Informationen.

Die Granularität der Daten wird durch die featurebasierte Datenerhebung beeinflusst, wodurch eine Unterscheidung hinsichtlich des Bauteils und des Analyse-Algorithmus erforderlich ist und ein erhöhter Aufwand der Datenanalyse, Modellierung und Implementierungsplanung entsteht.

Da die Featurewerte eine numerische Abbildung des komplexen AOI-Prozesses, bestehend aus verschiedenartigen Verfahren der Bildverarbeitung und anschließender regelbasierter Klassifikation, darstellen, wird die Ausprägung der Messwerte durch jede einzelne Analyse beeinflusst.

Die Anzahl und Ausprägung der Featurewerte spiegeln die zugrundeliegende AOI-Analyse, welche mehrschrittige Subprüfungen und verbundene Prüfplanparameter sowie Prüfmuster umfasst, wider und sind in weiteren Verarbeitungsschritten zu beachten.

Die Relevanz produktionstechnischer Informationen und Eigenschaften der Lötstelleninspektion in dieser Forschungsarbeit wird durch den folgenden Lehrsatz wiedergegeben. Die Beurteilung der statistischen Datenanalyse und die Anwendung der aus XAI-Methoden generierten Regelsätze in der AOI-Prüfmethodik erfordern umfassende Kenntnisse der Löttechnologie und der Inspektionsverfahren.

Erst die Synthese von Prozesswissen und Informationstechnik ermöglicht, gültige und plausible Aussagen sowie Resultate zu erheben.

7. Zusammenfassung

Die Aufgabenstellung beinhaltet zunächst die Entwicklung einer Methode zur Erhebung von AOI-Daten aus der SMT-Produktion. Aufgrund der geringen dpmo-Kennzahlen der SMT-Fertigung der Continental Automotive GmbH in Karben und den geltenden Anforderungen an die Datenqualität sowie fertigungsbezogene Rahmenbedingungen werden zwei Methoden konzipiert. Die Offline-Erhebung nutzt archivierte AOI-Ergebnisdateien zwecks Aufbau einer Echtfehler-Datenbank. Inspektionsdateien von Pseudofehlern und i.O.-Prüfungen werden direkt an den AOI-Systemen durch eine Softwarelösung im Rahmen eines online Datenerhebungskonzepts akquiriert. Die Gewährleistung der Datenkonsistenz wird durch informationstechnische und prozessbasierte Kontrollmechanismen sichergestellt. Die Klassenannotation sämtlicher Daten, die in weiteren Forschungsschritten verarbeitet werden, erfahren ein manuelles Relabelling oder eine Bestätigung der bestehenden Klassenannotation.

Die verwendeten AOI-Ergebnisdateien umfassen Bildaufnahmen, die Codierung des Bauteilzustands der AOI-Prüfung und der Verifikationsstation sowie Metainformationen bezüglich des Inspektionsprozesses. Aus diesem Grunde wird die Generierung von binären und multikategoriellen Datensätzen ermöglicht. Die Fehlercodes des Bauteilzustandes können direkt oder durch Zusammenfassung aller Echtfehler zu einer übergeordneten EF-Klasse in dem Datensatz verwendet werden.

Ein weiterer Vorteil der AOI-Ergebnisdateien ist die erhöhte Flexibilität, mit der auf Änderungen der Inspektionsmethodik im dynamischen Fertigungsverlauf reagiert werden kann. Im Falle einer Verletzung der Datenkonsistenz durch Anpassungen der Prüfpläne im Produktionsbetrieb ist es möglich, neue Featurewerte aus den archivierten AOI-Ergebnisdateien hinsichtlich der aktuellen Prüfpläne zu generieren und neue ML-/DL-Modelle zu erstellen. Im Falle von Veränderungen des Inspektionsprozesses sind die archivierten AOI-Dateien jedoch nicht mehr nutzbar. Die Datenkonsistenz ist somit durch nachträgliche Erhebung nicht gewährleistet. Dies betrifft vor allem Änderungen der Bildaufnahmesituation, wie Farb-/Beleuchtungsmodi und Bilddynamik oder der Prüfmethodik. Zusammenfassend ist die Verwendung von Messwerten des AOI-Prozesses zur Realisierung von ML-/DL-Modellen auf die Erhebung einer ausreichenden Datenmenge in einem statischen Fertigungsbetrieb angewiesen. Dies steht oftmals im Widerspruch zur dynamischen Gestaltung der SMT-Produktion.

Die statistische Datenanalyse umfasst unterschiedliche Untersuchungen der binären, ternären und multikategorialen AOI-Datensätze. Der Anteil von Featurewerten, welche einer Population mit theoretischer Verteilung folgen, wird durch Datenbereinigungsmethoden erhöht. Der Einfluss von Einzelfeatures auf das Klassifikationsergebnis wird durch Methoden des Feature-Ranking ermittelt. Die relevantesten Merkmale dieser Importance-Messungen werden durch Hinzunahme von Prozesskenntnissen und -erfahrungen bestätigt. Die Resultate der Dependenzanalysen belegen, dass signifikante Unterschiede in den Daten bezüglich EF-/OK-Klasse und darüber hinaus der Fehlercodierung sowie Orientierungswinkel oberflächenmontierter Bauteile existieren und nachweisbar sind.

Im Rahmen der Untersuchung des Forschungsstandes und Aufstellung der verwendeten ML-/DL-Methoden unter Berücksichtigung der erhobenen Datenmenge erweist sich die binäre Klasseneinteilung als wirksam. Die von XAI-Verfahren ermittelten Regelsätze werden in Prüfplanparameter übersetzt. Zu diesem Zweck ist ein umfassendes Prozesswissen bezüglich Löttechnik und AOI-Analysen erforderlich. Mithilfe der generierten Prüfplanparameter erfolgt die Anpassung bestehender Prüfpläne, die im Anschluss auf ihre Klassifikationsgüte validiert werden. Die fünf verwendeten XAI-Methoden, von denen vier auf Entscheidungsbäumen basieren und ein Verfahren eine heuristische Regelextraktion verfolgt, erreichen Verbesserungen der Klassifikationseigenschaften. Die angepassten Prüfpläne sind oftmals konservativ, wodurch sämtliche Echtfehler in dem Datensatz identifiziert werden, jedoch auf Kosten der Pseudeofehlerrate. Dies schlägt sich in den Metriken bezüglich der Gesamtklassifikation nieder, wodurch in Analyse-B lediglich für Bauteil C0603 eine Verbesserung des F1-Scores im Vergleich zum Originalprüfplan ermittelt wird. Sämtliche SMD-Komponenten der AOI-Analyse-A, Analyse-C (Bauteil R0402 und SOT23) und Analyse-D (SOT23) erreichen eine Zunahme der F1-Scores.

Die regelbasierte Klassifikation der untersuchten AOI-Prozesse limitiert die Flexibilität, mit der Strukturen und Wechselbeziehungen in den Featurewerten in die Entscheidungsfindung integriert werden können. Die Methoden des maschinellen sowie tiefen Lernens sind in der Lage, sämtliche Informationen innerhalb der Datensätze zu nutzen. In den Methodenplan werden sieben verschiedenartige Verfahren aufgenommen. Die Bewertungsmetriken der Einzelmodelle sowie eines Ensemble-Algorithmus, bestehend aus einer Kombination mehrerer Einzelmodelle, erzielen höhere Prognosegüten als der Originalprüfplan. Die Verfahren des maschinellen Lernens (XGBoost, Support Vector Machine und Random Forest) und die neuronale Netzwerkarchitektur des TabNet erweisen sich in den Untersuchungsreihen als leistungsstark.

8. Ausblick

Die Prognosegüte der multikategoriellen ML-/DL-Methoden wird in dieser Forschungsarbeit als gering eingestuft. Des Weiteren reduziert sich die Datenmenge bei Unterteilung der Datensätze anhand ihrer Bauteil-Orientierungswinkel und der Fehlercodierung, so dass eine Modellierung mit modernen Methoden des maschinellen und tiefen Lernens nicht durchführbar ist. Eine mögliche Aufgabenstellung zukünftiger Forschungsarbeiten, welche auf den Ergebnissen dieser Promotionsschrift aufbauen, beinhaltet die Erhebung weiterer Produktionsdaten und den Aufbau einer größeren Datenbasis.

Darüber hinaus ist zu untersuchen, ob Fertigungsdaten vorgelagerter Prozesse der SMT-Fertigungslinie akquiriert und den AOI-Datensätzen zugeordnet werden können. Aufgrund dieser prozessübergreifenden, umfangreicheren Datensätze entstehen neue Forschungsfragen. Diese bestehen darin, Strukturen und Wechselbeziehungen in den Daten zu ermitteln und daraus präventive Maßnahmen zur Entstehung von Echtfehlern abzuleiten. Zusätzlich ist zu untersuchen, ob größere Datensätze einen Anstieg der Klassifikationsgüte von ML-/DL-Modellen hervorrufen.

Die in dieser Arbeit identifizierten signifikanten Unterschiede der Datensätze bezüglich der Bauteil-Orientierungswinkel und Fehlercodierungen, insbesondere der Einfluss des Orientierungswinkels auf den Tombstonedefekt, können durch Hinzunahme von Daten vorgelagerter Produktionsprozesse einer Untersuchung der Einflussfaktoren sowie Ursachen unterzogen werden. Durch die Erhöhung des AOI-Datenumfangs kann neben den Informationen des Orientierungswinkels und des Fehlercodes die Position des Bauteils auf der Leiterplatte in den Datensatz integriert werden. Dies kann entweder durch Einteilung des Mehrfachnutzens oder der Leiterplatte in Segmente oder durch genaue Angabe der X-Y-Position und Ausdehnung des Bauteils erfolgen.

Angesichts des Verbesserungspotenzials der XAI-Regelsätze sind diese in der zukünftigen Prüfplangestaltung zu berücksichtigen und zu evaluieren. Dies erfordert eine kontinuierliche Überwachung der Datenqualität/-konsistenz und Aktualisierung der XAI-Modelle.

Eine weitere zukünftige Aufgabenstellung ist die Analyse, ob Prognosen der ML-/DL-Modelle der manuellen Klassifikation an der Verifikationsstation als unterstützende Maßnahme eine Verbesserung der Rate des menschlichen Schlupfs bewirken.

Literaturverzeichnis

- [Abb19] Murdoch, James W. & Singh, Chandan & Kumbier, Karl & Yu, Bin & Abbasi-Asl, Reza. *Definitions, methods, and applications in interpretable machine learning*. PNAS October 29, 2019, Volume 116, Issue 44, S. 22071–22080, 2019. DOI: <https://doi.org/10.1073/pnas.1900654116>.
- [Agu02] Ruiz, Roberto & Riquelme, José C. & Aguilar-Ruiz, Jesús. *Projection-based measure for efficient feature selection*. Journal of Intelligent & Fuzzy Systems, Volume 12, Issue 3–4, S. 175–183, 2002. URL: <https://ip.ios.semcs.net/articles/journal-of-intelligent-and-fuzzy-systems/ifs00178>.
- [Alm21] Almit GmbH. *Terminologie des Lötens: Benetzungswinkel (contact angle)*, 2021. [Online: Stand 08.12.2021] URL: <https://www.almit.de/terminologie-des-Lotens/benetzungswinkel/>.
- [Alp19] Alpaydin, Ethem. *Maschinelles Lernen*. Walter de Gruyter GmbH & Co. KG, 2. erweiterte Auflage, 2019. DOI: <https://doi.org/10.1515/9783110617894>.
- [Alp20] Alpha Laytron GmbH. *Testkonzepte*, 2020. [Online: Stand 14.12.2021] URL: <https://www.alpha-laytron.de/testkonzepte/>.
- [Ask95] Pham, D. T. & Askoy, M. S. *RRULES: A Simple Rule Extraction System*. Expert Systems With Applications, Volume 8, Issue 1, S. 59–65, 1995. DOI: [https://doi.org/10.1016/S0957-4174\(99\)80008-6](https://doi.org/10.1016/S0957-4174(99)80008-6).
- [B⁺11] Buitinck et al. *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, Volume 12, S. 2825–2830, 2011. URL: <https://scikit-learn.org/stable/index.html>.
- [Bel05] Bell, Hans. *Reflowlöten: Grundlagen, Verfahren, Temperaturprofile und Lötfehler*. Eugen G. Leuze Verlag, 2005. 1. Auflage.
- [Bis06] Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Information Science and Statistics, Kapitel 7 – Sparse Kernel Machines, Springer Science+Business Media, LLC, 2006.
- [Bog15] Jović, A. & Brkić, K. & Bogunović, N. *A review of feature selection methods with applications*. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), S. 1200–

- 1205, 2015. DOI: <https://doi.org/10.1109/MIPRO.2015.7160458>.
- [Bru09] Bruderreck, Lutz, TechnoLab GmbH. *Chemisch Zinn – Wer spielt mit dem Schmuddelkind? Teil 1*, 2009. [Online; Stand 13.12.2021] URL: http://www.apl-ssc.com/fileadmin/user_upload/News-Archiv/Plus_chem._Sn_TechnoLab.pdf.
- [Coh95] Cohen, William W. *Fast Effective Rule Induction*. Proceedings of the Twelfth International Conference on Machine Learning, Tahoe City, California, 9.–12. Jul. 1995, S. 115–123, 1995. DOI: <https://doi.org/10.1016/B978-1-55860-377-6.50023-2>.
- [Cot17] Biran, Or & Cotton, Courtenay V. *Explanation and Justification in Machine Learning : A Survey*. 2017.
- [Dau11] Strater, Swetlana & Dautfest, Jakob. *TU Darmstadt: Projekt Stochastik – Black Jack und Aktien – Modul Modellierung*, 2011. [Online: Stand 24.02.2022] URL: https://wwwid.mathematik.tu-darmstadt.de/amustud/amu_stud_website/blackjack/Projekt/Download/Modellierung.pdf.
- [Deu06] Schmitt, Jacqueline & Böning, Jochen & Borggräfe, Thorbjörn & Beitinger, Gunter & Deuse, Jochen. *Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing*. Advanced Engineering Informatics, Volume 45, Artikel 101101, 2006. DOI: <https://doi.org/10.1016/j.aei.2020.101101>.
- [Dig21] Digital Today. *Bei künstlicher Intelligenz ist mehr Transparenz gefragt*, 2021. [Online: Stand 02.03.2022] URL: <https://digital-today.de/bei-kuenstlicher-intelligenz-ist-mehr-transparenz-gefragt/>.
- [DIN03a] DIN EN 8580:2003-09: Fertigungsverfahren – Begriffe, Einteilung; Ausgabedatum 2003-09, November 2003.
- [DIN03b] DIN EN ISO 8593-7:2003-09: Fertigungsverfahren Fügen – Teil 7: Fügen durch Löten; Einordnung, Unterteilung, Begriffe; Ausgabedatum 2003-09, November 2003.
- [DIN06] DIN 8514:2006-05: Lötbarkeit; Ausgabedatum 2006-05, May 2006.
- [DIN07] DIN ISO 857-2:2007-03: Schweißen und verwandte Prozesse – Begriffe – Teil 2: Weichlöten, Hartlöten und verwandte Begriffe; Ausgabedatum 2007-03, November 2007.
- [Doc16] SkopeRules Documentation. *SkopeRules Documentation*, 2016. [Online: Stand 18.04.2022] URL: <https://skope-rules.readthedocs.io/en/latest/index.html>.

- [Duc16] Antonelli, Michaela & Marcelloni, Francesco & Ducange, Petro. *Multi-Objective Evolutionary Design of Fuzzy Rule-Based Systems*. HANDBOOK ON COMPUTATIONAL INTELLIGENCE: Volume 2: Evolutionary Computation, Hybrid Systems, and Applications, S. 635–670, 2016. DOI: https://doi.org/10.1142/9789814675017_0018.
- [Duc20] Gallo, Gionatan & Ferrari, Vincenzo & Marcelloni, Francesco & Ducange, Petro. *SK-MOEFS: A Library in Python for Designing Accurate and Explainable Fuzzy Models*. Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020. Communications in Computer and Information Science, Volume 1239, S. 68–81, 2020. DOI: https://doi.org/10.1007/978-3-030-50153-2_6.
- [Dí05] Ruiz, Roberto & Aguilar-Ruiz, Jesús & Riquelme, José C. & Díaz-Díaz, Norberto. *Analysis of Feature Rankings for Classification*. Conference: Advances in Intelligent Data Analysis VI, 6th International Symposium on Intelligent Data Analysis, IDA 2005, Madrid, Spain, 08.–10. Sept. 2005, 2005. DOI: https://doi.org/10.1007/11552253_33.
- [Fay17] Bunker, Rory P. & Fayez, Fadi. *A Machine Learning Framework for Sport Result Prediction*. Applied Computing and Informatics, Volume15, Issue 1, S. 27–33, 2017. DOI: <https://doi.org/10.1016/j.aci.2017.09.005>.
- [Fel13] Feldmann, Klaus. *Handbuch Fügen, Handhaben und Montieren*. Carl Hanser Verlag, 2013. 2. Auflage.
- [Fet06] Fettke, Peter. *State-of-the-Art des State-of-the-Art*. Wirtschaftsinformatik, Volume 48, Issue 4, August 2006, Artikel 257, S. 257–266, Springer, 2006. DOI: <https://doi.org/10.1007/s11576-006-0057-3>.
- [Fri06] Fritzsche, Ralf. *Erstellung von Parameter-Diagrammen in der Automobilindustrie*. ATZ – Automobiltechnische Zeitschrift Volume 108, Issue 6, Juni 2006, S. 492–497, 2006.
- [Gad21] Ahmed F. Gad. *PyGAD: An Intuitive Genetic Algorithm Python Library*, 2021. [Online: Stand 10.01.2022] URL: <https://pypi.org/project/pygad/>.
- [Gao11] Hongwei, Xie & Xianmin, Zhang & Yongcong, Kuang & Gaofei, Ouyang. *Solder Joint Inspection Method for Chip Component Using Improved AdaBoost and Decision Tree*. IEEE Transactions on Components, Packaging and Manufacturing Technology, Volume 1, Issue 12, Dezember 2011, S. 2018–2027, 2011. DOI: <https://doi.org/10.1109/TCPMT.2011.2168531>.
- [Gue18] Ribeiro, Marco T. & Singh, Sameer & Guestrin, Carlos. *Anchors: High-Precision Model-Agnostic Explanations*. Proceedings of the AAAI Con-

- ference on Artificial Intelligence, Volume 32, Issue 1, 2018. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/11491>.
- [GÖ21] GÖPEL electronic GmbH. *Lot- und Sinterpaste inspizieren mit dem SPI Line · 3D*, 2021. [Online: Stand 09.12.2021] URL: <https://www.goepel.com/inspektionsloesungen/spi/lotpasteninspektion/sinterpasteninspektion.html/>.
- [Her17] Hering, Ekbert. *Elektronik für Ingenieure und Naturwissenschaftler*. Springer Vieweg, Berlin, Heidelberg, 2017. 7. Auflage.
- [Hes12] Hesselmann Service GmbH. *Richtlinie 2012/19/EU – WEEE2 des Europäischen Parlaments und des Rates über Elektro- und Elektronik-Altgeräte vom 4. Juli 2012*, 2012. [Online: Stand 13.12.2021] URL: <https://www.elektrogesetz.de/gesetzestexte/weee/>.
- [Ill19] Geczy, Attila & Szalmasi, Daniel & Straubinger, Daniel & Illes, Balazs. *Investigating shadowing and possible tombstoning caused by large SMD components during vapour phase reflow soldering*. IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME) 2019, S. 343–346, 2019. DOI: <https://doi.org/10.1109/SIITME47687.2019.8990744>.
- [Ind17] IPC Association Connecting Electronics Industries. *IPC A 610 DE – Abnahmekriterien für elektronische Baugruppen*. 2017. Ausgabe G.
- [Kau13] Kalyanakrishnan, Shivaram & Kaufmann, Emilie. *Information complexity in bandit subset selection*. JMLR Workshop and Conference Proceedings (Conference on Learning Theory), Issue 30, S. 228–251, 2013. URL: https://www.researchgate.net/publication/287279572_Information_complexity_in_bandit_subset_selection.
- [Ken38] Kendall, Maurice G. *A New Measure of Rank Correlation*. Biometrika, Volume 30, Issue 1/2, Jun. 1938, S. 81–93, 1938. DOI: <https://doi.org/10.2307/2332226>.
- [KI-22] KI-News. *Lasso Regression: Die Shrinkage-Methode einfach erklärt!*, 2022. [Online: Stand 28.02.2022] URL: <https://www.kinews.de/ki-methoden/lasso-regression-einfach-gemacht/>.
- [Kot07] Kotsiantis, Sotiris. *Supervised Machine Learning: A Review of Classification Techniques*. Informatica (Ljubljana) 31, S. 249–268, 2007.
- [Kru05] Kruppa, Werner, STANNOL GmbH & Co. KG. *RoHS, WEEE, ElektroG – Gesetzliche Bestimmungen und Regeln*, 2005. [Online: Stand 13.12.2021].
- [Leo17] Leonhart, Rainer. *Lehrbuch Statistik – Einstieg und Vertiefung*. Hogrefe AG, 2017. 4. überarb. u. erw. Aufl. 2017.

- [Lig13] Morik, Jakob & Ligges, Uwe. *TU-Dortmund – Wissensentdeckung in Datenbanken: 8. Entscheidungsbäume*, 2013. [Online: Stand 25.02.2022] URL: https://www.statistik.tu-dortmund.de/fileadmin/user_upload/Lehrstuehle/Datenanalyse/Wissensentdeckung/Wissensentdeckung-Li-6_2x2.pdf.
- [Mai05] Rokach, Lior & Maimon, Odid. *Decision Trees*. The Data Mining and Knowledge Discovery Handbook, S. 165–192, 2005. DOI: 10.1007/0-387-25465-X_9.
- [Mai18] Maisenhölder, Bernd, Iftest AG. *Höhere Qualität dank Automatischer Optischer Inspektion (AOI) in 3D*, 2018. [Online: Stand 21.02.2022] URL: https://www.all-electronics.de/wp-content/uploads/2018/10/Hoehere-Qualitaet-dank-Automatischer-Optischer-Inspektion-AOI-in-3D_V1.pdf.
- [Mar19] Martin, Florian, Viscom AG. *Qualität aus allen Perspektiven*, 2019. [Online Stand 21.02.2022] URL: https://www.viscom.com/fileadmin/content/products/Optical_Inspection_3D_AOI/S3088_ultra_gold/Viscom_Article_SP_Qualitaet_aus_allen_Perspektiven_EPP_1911_de.pdf.
- [Mas51] Massey Jr., Frank J. *The Kolmogorov-Smirnov Test for Goodness of Fit*. Journal of the American Statistical Association, Volume 46, Issue 253, S. 68–78, 1951. DOI: <https://doi.org/10.2307/2280095>.
- [Mir18] Raschka, Sebastian & Mirjalili, Vahid. *Machine Learning mit Python und Skikit-Learn und Tensorflow*. mitp Verlag GmbH & Co. KG, Frechen, 2018. 2. Auflage.
- [Mit97] Mitchell, Thomas M. *Machine Learning*. McGraw-Hill, Inc., 1997. 1. Auflage.
- [Mol16] Christoph Molnar. *rulefit. Python implementation of the rulefit algorithm*, 2016. [Online: Stand 07.05.2022] URL: <https://github.com/christophM/rulefit>.
- [Mol22] Molnar, Christoph. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2022. URL: <https://christophm.github.io/interpretable-ml-book/>.
- [Mos19] Moscovitz, Ilan. *How to Perform Explainable Machine Learning Classification – Without Any Trees*, 2019. [Online: Stand 20.04.2022] URL: <https://medium.com/@ilan.moscovitz>.
- [Nas17] Nasteski, Vladimir. *An overview of the supervised machine learning methods*. Horizons 4, S. 51–62, 2017. DOI: 10.20544/HORIZONS.B.04.1.17.P05.
- [Pal21] Palliser-Sans, Rafael. *RRULES: An improvement of the RULES rule-based classifier*. 2021. DOI: <https://doi.org/10.48550/arXiv.2106.07296>.

- [PCB21] PCBMay. *AOI-Test*, 2021. [Online: Stand 13.12.2021] URL: <https://www.pcbmay.com/de/Aoi-Test/>.
- [Pfi19] Arik, Sercan Ö. & Pfister, Tomas. *TabNet: Attentive Interpretable Tabular Learning*. Proceedings of the AAAI Conference on Artificial Intelligence, Volume 35, Issue 8, S. 6679–6687, 2019. DOI: <https://doi.org/10.48550/arXiv.1908.07442>.
- [Pol14] Royer, Sullivan & Thil, Stéphane & Talbert, Thierry & Polit, Monique. *Black-box modeling of buildings thermal behavior using system identification*. IFAC Proceedings Volumes, Volume 47, Issue 3, S. 10850–10855, 2014. DOI: <https://doi.org/10.3182/20140824-6-ZA-1003.01519>.
- [Pop08] Friedman, Jerome H. & Popescu, Bogdam E. *Predictive learning via rule ensembles*. Annals of Applied Statistics 2008, Volume 2, No. 3, S. 916–954, 2008. DOI: <https://doi.org/10.1214/07-AOAS148>.
- [PRI20] PRISMA. *PRISMA Flow Diagram*, 2020. [Online: Stand 15.03.2022] URL: <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram>.
- [Rah14] Rahn, Armin. *Handbuch der Prozess- und Lötfehler*. Eugen G. Leuze Verlag, 2014. 1. Auflage.
- [Reh22] Rehm Thermal Systems GmbH. *Konvektionslöten – Vision TripleX*, 2022. [Online: Stand 17.03.2022] URL: <https://www.rehm-group.com/prozesse/konvektionsloeten/vision-triplex.html>.
- [Rud10] Kursa, Miron B. & Jankowski, Aleksander & Rudnicki, Witold. *Boruta – A System for Feature Selection*. Fundamenta Informaticae, Volume 101, Issue 4, S. 271–285, 2010. DOI: <https://doi.org/10.3233/FI-2010-288>, URL: <https://content.iospress.com/articles/fundamenta-informaticae/fi101-4-02>.
- [Sch04] Smola, Alex J. & Schölkopf, Bernhard. *A Tutorial on Support Vector Regression*. Statistics and Computing, Volume 14, Issue 3, August 2004, S. 199–222, 2004. DOI: 10.1023/B:STCO.0000035301.49549.88.
- [Sch05] Schmidhuber, Jürgen. *Grundlagen algorithmischer Informationstheorie – Vorlesung Maschinelles Lernen II (SS 2005)*. Fakultät für Informatik, TUM, 2005. URL: <https://people.idsia.ch/~juergen/alginfobasics.pdf>.
- [Sch20] Thielen, Nils & Werner, Dominik & Schmidt, Konstantin. *A Machine Learning Based Approach to Detect False Calls in SMT Manufacturing*. 43rd International Spring Seminar on Electronics Technology (ISSE) 2020, S. 1–6, 2020. DOI: <https://doi.org/10.1109/ISSE49702.2020.9121044>.
- [Sci22] SciPy Dokumentation Version 1.8.0. *scipy.cluster.hierarchy.linkage*, 2022. [Online: Stand 06.04.2022] URL: <https://docs.scipy.org/doc/scipy/reference/>

generated/scipy.cluster.hierarchy.linkage.html.

- [Shi19] Chen, Yun & Wang, Zijin & Sheng, Shuo & Shi, Guangyi. *Research on Accuracy Improvement Technology of Defect Detection Based on Machine Learning*. IEEE The 2nd International Conference On Mirco/Nano Sensors for AI, Healthcare, and Robotics (NSENS) 2019, S. 19–20, 2019. DOI: <https://doi.org/10.1109/NSENS49395.2019.9293991>.
- [STA21] STANNOL GmbH & Co. KG in Reichelt Elektronik Magazin. *Warum eigentlich beim Handlöten von den „guten“ bleihaltigen Loten auf bleifreie Lote wechseln?*, 2021. [Online: Stand 13.12.2021] URL: <https://www.reichelt.de/magazin/ratgeber/beim-handloeten-auf-bleifreie-lote-wechseln/>.
- [Str13] Dusek, Karel & Brejcha, Michal & Kemler, Andras & Beshajova-Pelikanov, Ivana & Straka, V. *Influence of type of reflow technology and type of surface finish on tomb stone effect*. Proceedings of the 36th International Spring Seminar on Electronics Technology 2013, S. 132–135, 2013. DOI: <https://doi.org/10.1109/ISSE.2013.6648229>.
- [Sza13] Geczy, Attila & Tersztyanszky, Laszlo & Illes, Balazs & Kemler, Andras & Szabo, Andras. *Reducing lead-free soldering failures caused by Printed Circuit Board shrinkage*. IEEE 19th International Symposium for Design and Technology in Electronic Packaging (SIITME) 2013, S. 65–68, 2013. DOI: <https://doi.org/10.1109/SIITME.2013.6743645>.
- [Tec19] Technische Hochschule Aschaffenburg. *Praktikum Aufbau- und Verbindungstechnik – Konvektions-Reflowlöten*, 2019. [Online: Stand 10.12.2021] URL: https://moodle.th-ab.de/pluginfile.php/2571/mod_resource/content/6/Konvektionsloeten%202019.pdf.
- [Tec21] TechnoLab GmbH. *Entnetzung*, 2021. [Online: Stand 08.12.2021] URL: <https://www.technolab.de/de/loetlexikon/smd-hmd/entnetzung.php/>.
- [Van09] Van Der Maaten, Laurens J. P. & Postma, Eric O. & Van Den Herik, H. Jaap. *Dimensionality Reduction: A Comparative Review*. Journal of Machine Learning Research – JMLR. 10, S. 66–71, 2009.
- [VDE19] VDE Prüf- und Zertifizierungsinstitut GmbH. *RoHS-Richtlinie: Weitere Stoffe und Geräte ab 22. Juli 2019 verboten*, 2019. [Online: Stand 13.12.2021] URL: <https://www.vde.com/tic-de/news/2019-1/rohs-richtlinie-erweitert>.
- [Vie21] Vierling Production GmbH. *SMD-Bestückung bis Bauform 01005 – Surface-Mount Technology (SMT) bei VIERLING*, 2021. [Online: Stand 10.12.2021] URL: <https://www.vierling.de/de/ems-dienstleistungen/leiterplatten-bestueckung/smd-bestueckung>.

- [Vis18] Viscom AG. Beispiellinie, 2018. [Online; Stand 15. Juli 2018] URL: <https://www.viscom.com/de/eu-rop/home/>.
- [Vis20] Viscom AG. *Online-Hilfe 2020*. Viscom AG, 2020. Version 14.1.
- [Vis22a] Viscom AG. Systembroschüre S3088 ultra gold, 2022. [Online: Stand 21.02.2022] URL: https://www.viscom.com/fileadmin/content/downloads_brochures/de/Viscom_Brochure_AOI_S3088_ultra_with_gold_de.pdf.
- [Vis22b] Viscom AG. *Die Prozesskontrolle. Für Ihre Qualität*, 2022. [Online: Stand 24.02.2022] URL: <https://www.viscom.com/de/produkte/quality-uplink/>.
- [Vis22c] Viscom AG. *Exzellente Prüfergebnisse aus allen Perspektiven*, 2022. [Online: Stand 21.02.2022] URL: <https://www.viscom.com/de/produkte/optische-inspektion-3d-aoi/>.
- [Vis22d] Viscom AG. *vVerify Software: Zeit einsparen und Prüfergebnisse hoch zuverlässig klassifizieren*, 2022. [Online: Stand 26.06.2022] URL: <https://www.viscom.com/de/produkte/software-loesungen/vverify/>.
- [Wil15] Wild, Paul. *Mit oder ohne Vakuum? Flexible Zwei-in-Eins-Lösung für das Reflowlöten*, 2015. [Online: Stand 10.02.2022] URL: <https://www.all-electronics.de/elektronik-fertigung/flexible-zwei-in-eins-loesung-fuer-das-reflowloeten.html>.
- [Wit18] Wittmann, Katharina. *Evaluierung von Feature-Ranking-Algorithmen zur Identifikation fehlerrelevanter Anlagenparameter beim Spritzgießen*. Masterarbeit mit Sperrvermerk bei Continental Automotive GmbH, Regensburg und Ostbayrische Technische Hochschule Regensburg, Fakultät für Informatik und Mathematik, 2018.
- [Wö12] Wölflick, Peter. *Automatische Optische Inspektion (AOI) in der Elektronikfertigung*. VDE Verlag, 2012. [Konferenz: Elektronische Baugruppen und Leiterplatten – EBL 2012 – Hochentwickelte Baugruppen in Europa – 6. DVS/GMM-Tagung 14.02.2012 – 15.02.2012 in Stuttgart, Deutschland].

A. Anhang

A.1. Multikategorielle Datenakquise

Tabelle A.1.: Aufbau Datensatz C0402 – Berücksichtigung Fehlercode 510, 520, 560 und Pseudofehler

Analyse	Winkel [°]	Fehlercode	Anzahl Datenpunkte
A	0	0	40
		510	7
		520	5
		560	16
	90	0	67
		510	7
		520	19
		560	41
	135	560	88
	180	0	35
		510	2
		520	16
560		25	
225	560	2	
270	0	74	
	510	5	
	560	81	
B	0	0	108
		510	68
		520	16
		560	57
	90	0	251
		510	171
		520	25
		560	99

	135	560	75
		0	97
	180	510	123
		520	22
		560	45
	225	560	7
		0	186
	270	510	163
		560	216
	315	0	37
		510	23
C		0	65
	0	510	48
		520	17
		560	55
	45	0	25
		0	37
	90	510	136
		520	23
		560	99
	135	560	78
		0	15
	180	510	84
		520	20
		560	42
225	560	7	
	0	214	
270	510	191	
	560	230	
315	0	7	
	510	10	

Tabelle A.2.: Datensatz C0603 – Berücksichtigung Fehlercode 510, 520, 560 und Pseudofehler

Analyse	Winkel [°]	Fehlercode	Anzahl Datenpunkte
		0	151
A	0	510	24

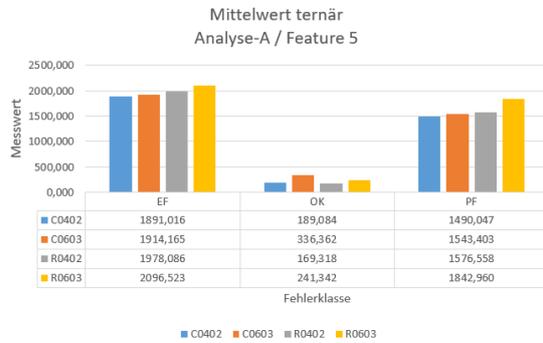
		520	26
		560	14
	45	510	8
		520	33
		560	19
	90	0	31
		510	17
		520	15
		560	14
	135	520	1
		560	1
	180	0	8
		510	8
		520	15
		560	8
	225	0	86
		510	12
		520	50
		560	4
	270	0	50
		510	29
		520	25
		560	16
	315	0	34
		510	18
		560	13
B	0	0	94
		510	111
		520	103
		560	83
	45	510	22
		520	41
		560	56
	90	0	106
		510	71
		520	168
		560	35
	135	510	36
		520	18

		560	5
	180	0	70
		510	69
		520	42
		560	20
	225	0	27
		510	24
		520	33
		560	12
	270	0	252
		510	145
		520	58
		560	37
	315	0	10
		510	50
		560	21
C	0	0	341
		510	111
		520	103
		560	83
	45	510	22
		520	41
		560	54
	90	0	137
		510	71
		520	168
		560	35
	135	0	194
		510	36
		520	18
		560	5
	180	0	71
		510	69
		520	42
		560	20
	225	0	27
		510	24
520		33	
560		12	

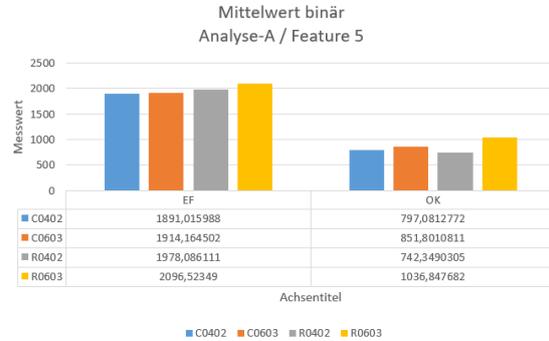
		0	252
	270	510	145
		520	58
		560	37
	315	0	10
		510	50
		560	21

A.2. Statistik

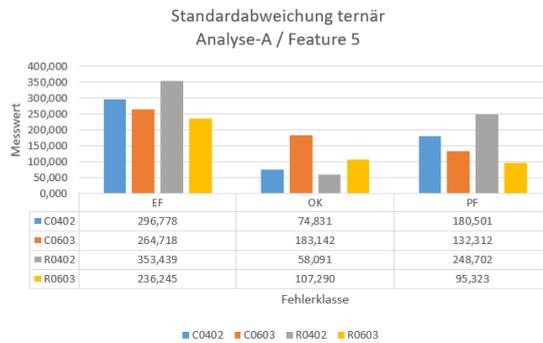
A.2.1. Zentrale Tendenz und Dispersion – Thema 1



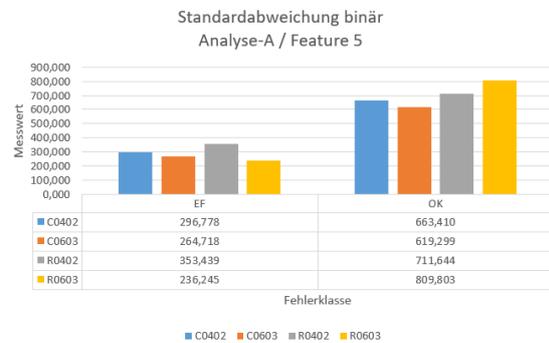
(a) Mittelwert drei Klassen



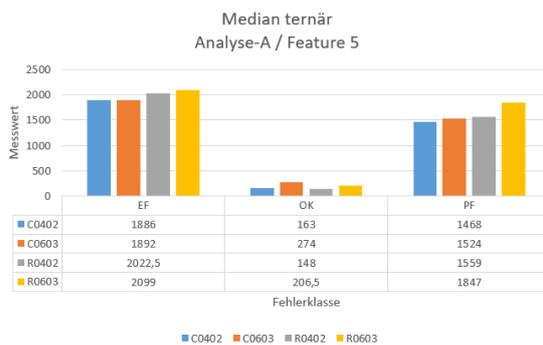
(b) Mittelwert zwei Klassen



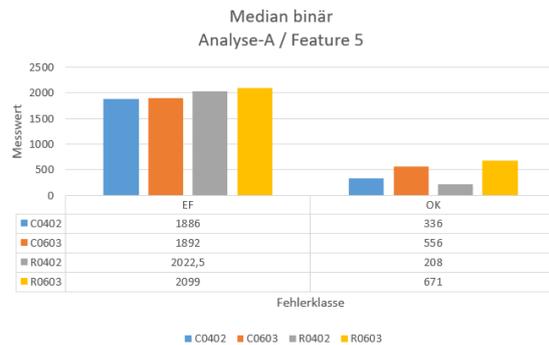
(c) Standardabweichung drei Klassen



(d) Standardabweichung zwei Klassen

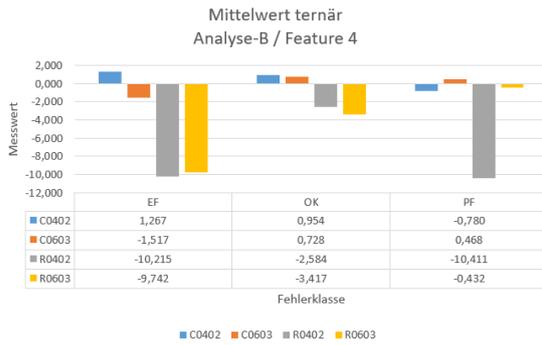


(e) Median drei Klassen

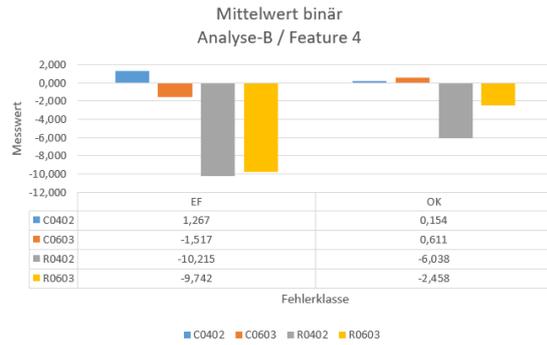


(f) Median zwei Klassen

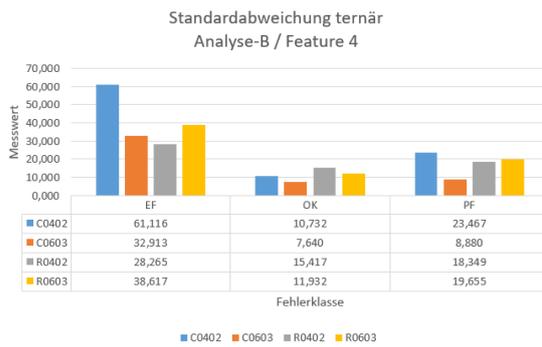
Abbildung A.1.: Diagramme zentrale Tendenz – Analyse-A – Feature 5



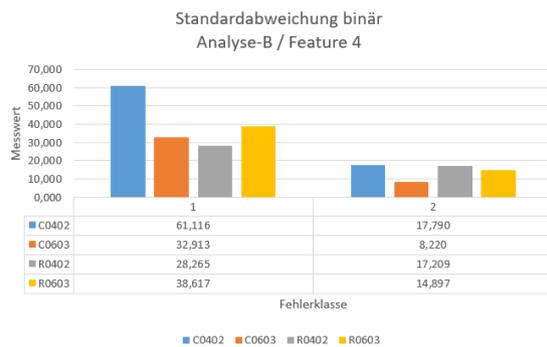
(a) Mittelwert drei Klassen



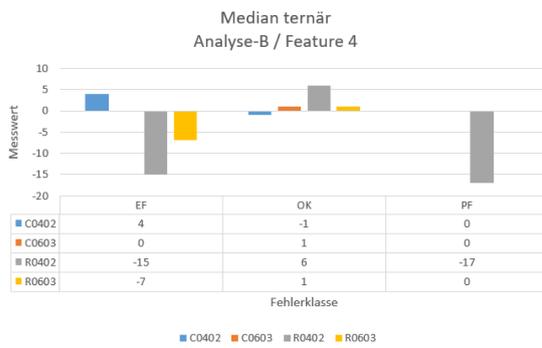
(b) Mittelwert zwei Klassen



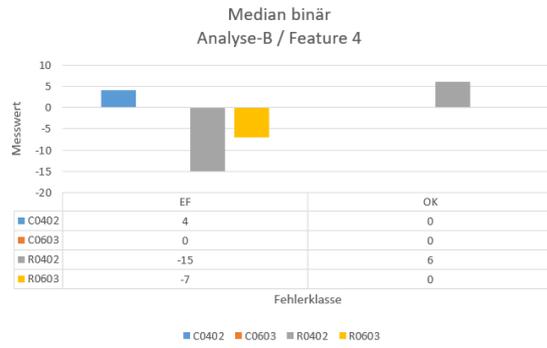
(c) Standardabweichung drei Klassen



(d) Standardabweichung zwei Klassen

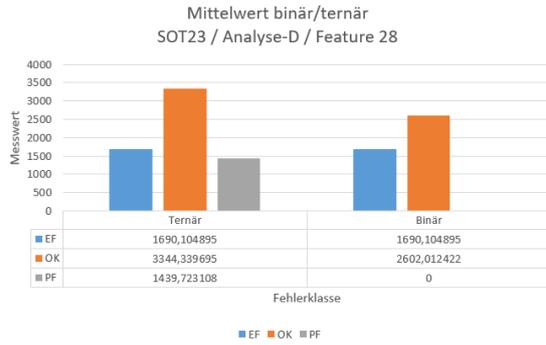


(e) Median drei Klassen

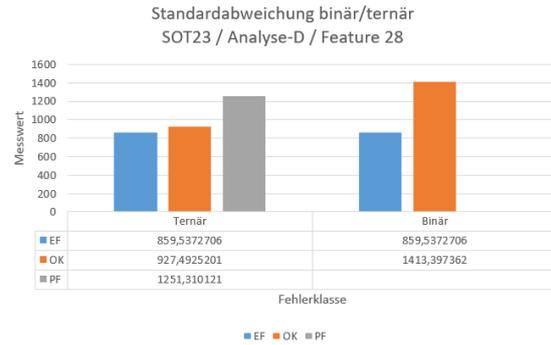


(f) Median zwei Klassen

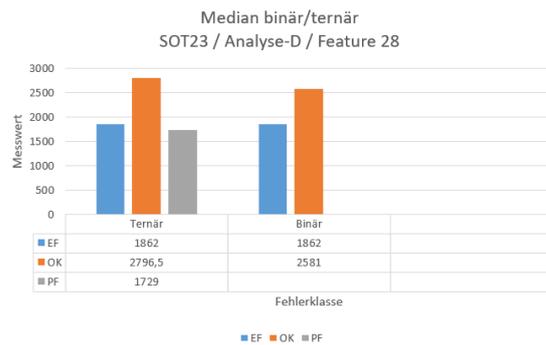
Abbildung A.2.: Diagramme zentrale Tendenz – Analyse-B – Feature 4



(a) Mittelwert zwei und drei Klassen

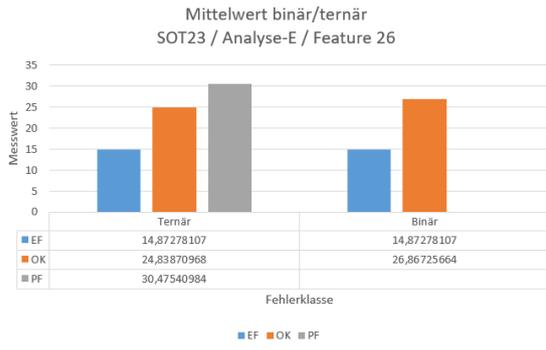


(b) Standardabweichung zwei und drei Klassen

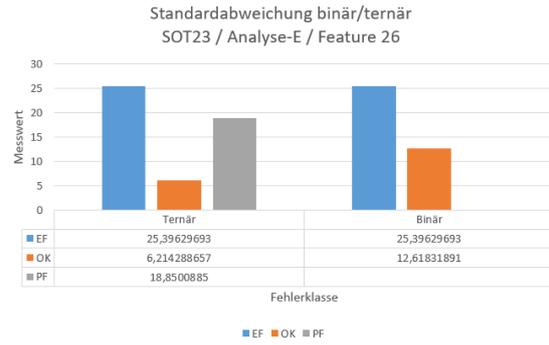


(c) Median zwei und drei Klassen

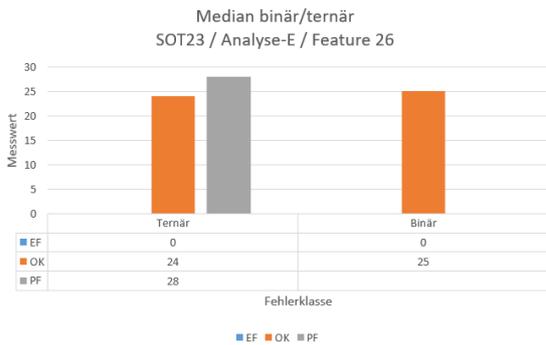
Abbildung A.3.: Diagramme zentrale Tendenz – Analyse-D – Feature 28



(a) Mittelwert zwei und drei Klassen

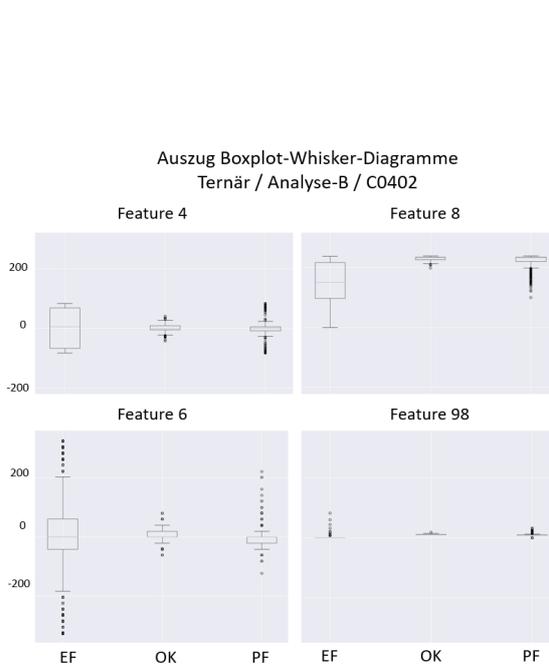


(b) Standardabweichung zwei und drei Klassen

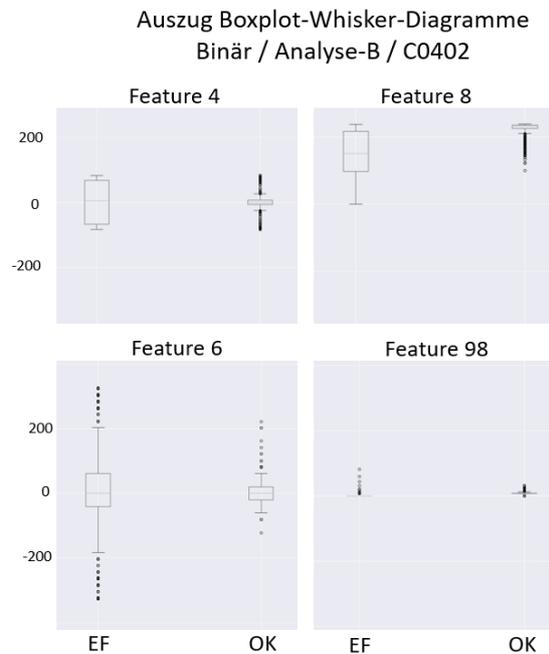


(c) Median zwei und drei Klassen

Abbildung A.4.: Diagramme zentrale Tendenz – Analyse-E – Feature 26



(a) Drei Klassen



(b) Zwei Klassen

Abbildung A.5.: Boxplot-Whisker-Diagramme – Analyse-B – C0402 – Feature 4, 6, 8, 98

Auszug Boxplot-Whisker-Diagramme
Binär & Ternär / Analyse-A / C0603

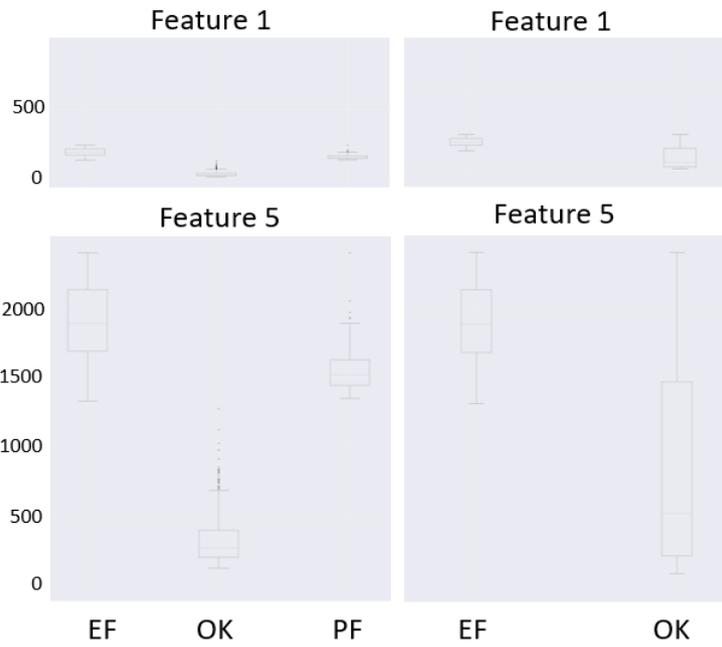
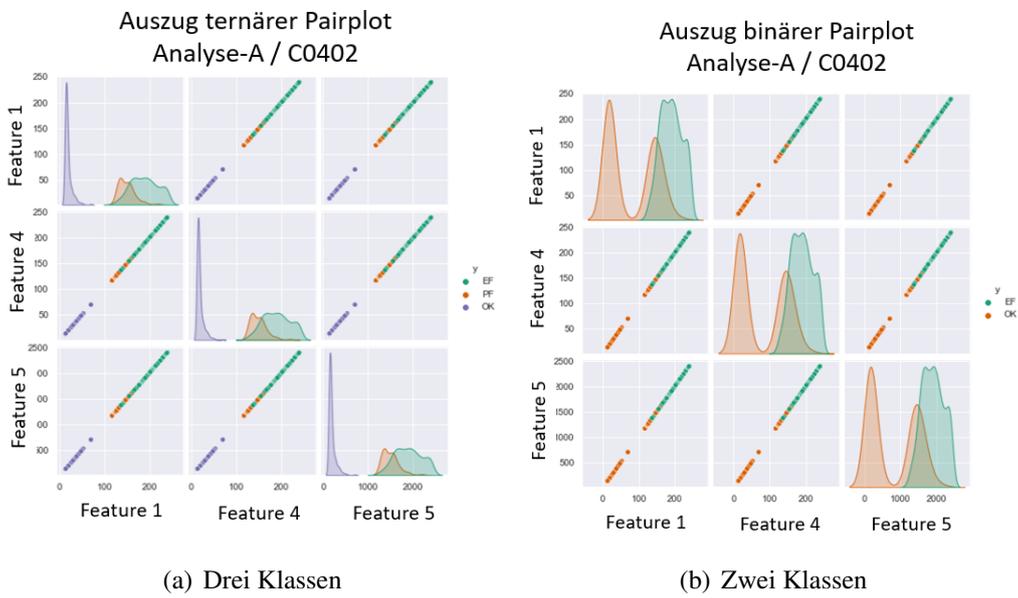


Abbildung A.6.: Boxplot-Whisker-Diagramme – Analyse-A – C0603 – Feature 1, 5



(a) Drei Klassen

(b) Zwei Klassen

Abbildung A.7.: Streudiagramme – Analyse-A – C0402 – Feature 1, 4, 5

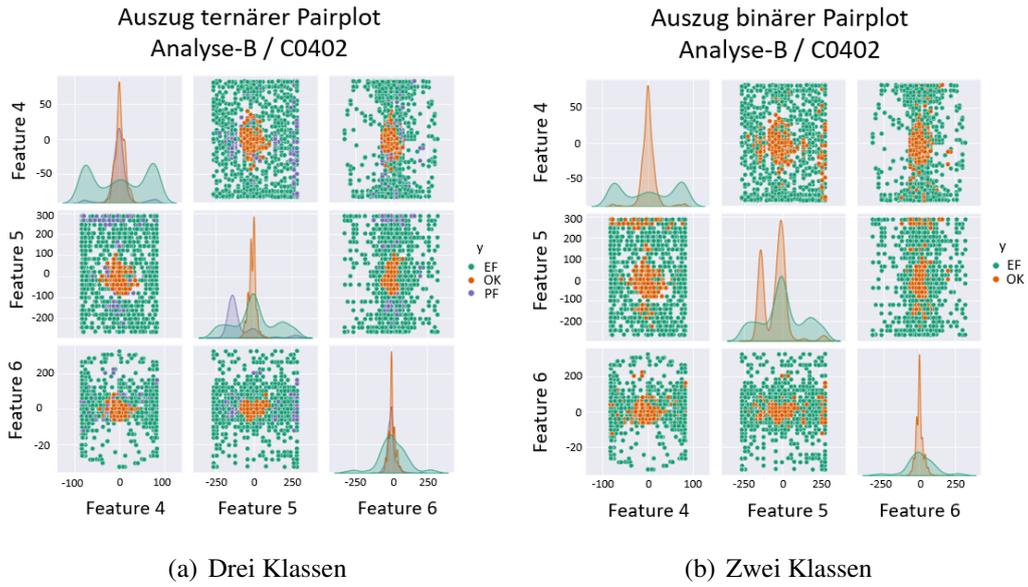


Abbildung A.8.: Streudiagramme – Analyse-B – C0402 – Feature 4, 5, 6

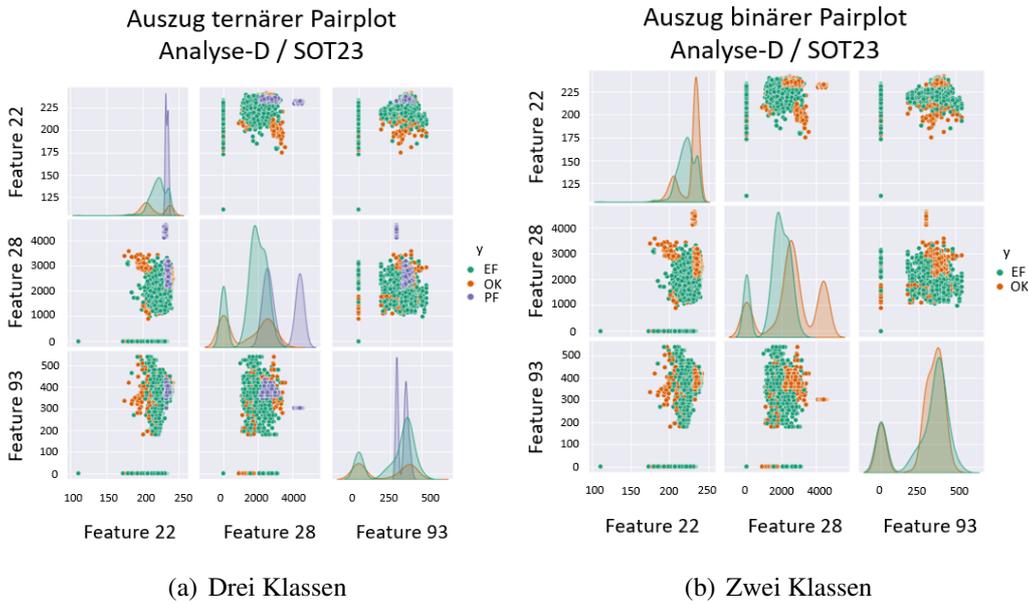


Abbildung A.9.: Streudiagramme – Analyse-D – SOT23 – Feature 22, 28, 93

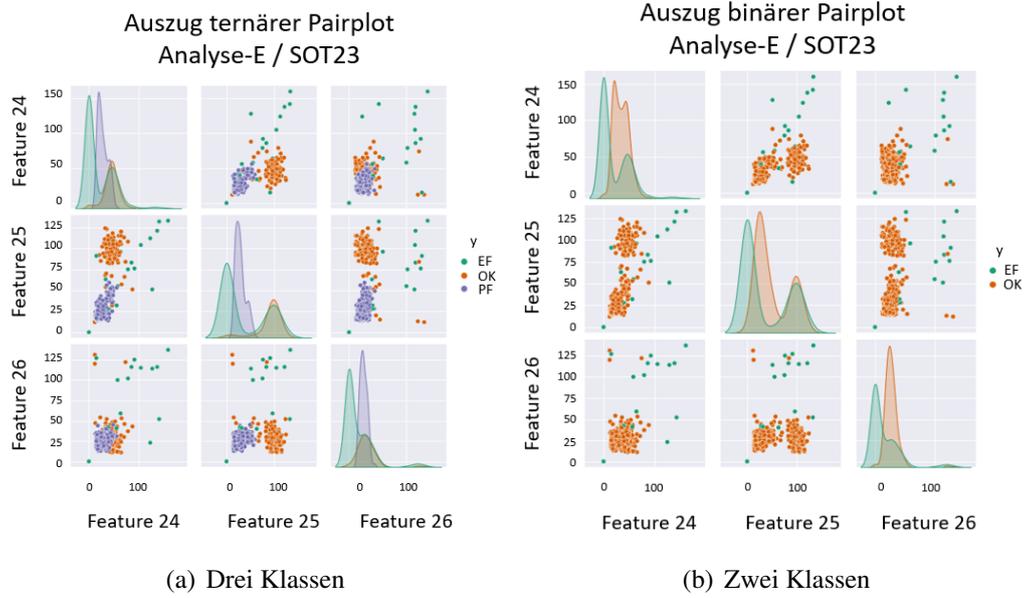
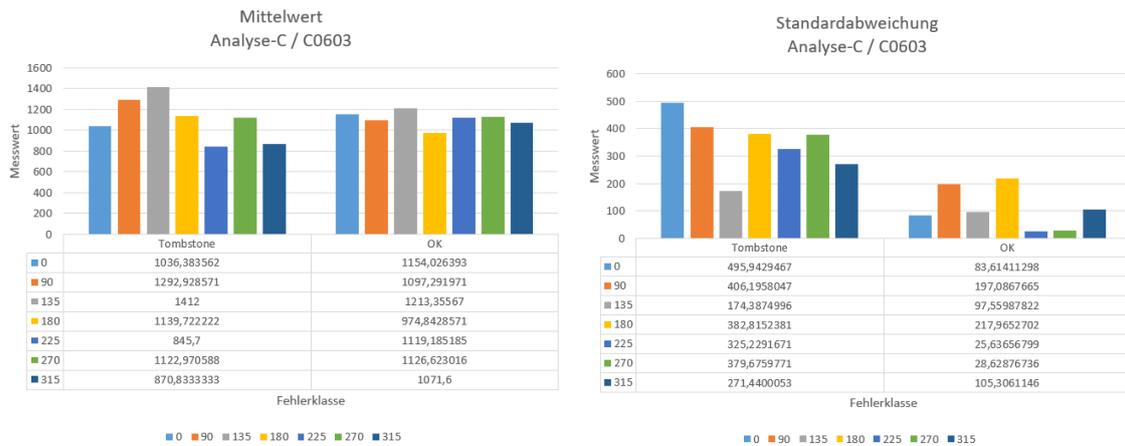


Abbildung A.10.: Streudiagramme – Analyse-E – SOT23 – Feature 24, 25, 26

A.2.2. Zentrale Tendenz und Dispersion – Thema 2



(a) Mittelwert

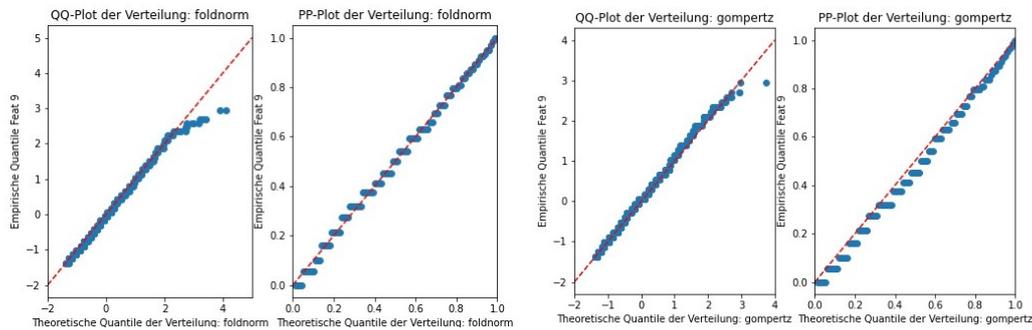
(b) Standardabweichung



(c) Median

Abbildung A.11.: Diagramme zentrale Tendenz – Analyse-C – C0603 – Feature 29 – zwei Klassen (PF, Tombstone)

A.2.3. Anpassungstest – Thema 1



(a) gefaltete Normalverteilung

(b) Gompertz Verteilung

Abbildung A.12.: Q-Q-Diagramme (links) & P-P-Diagramme (rechts) nach Datenbereinigung – ternäre Analyse-C – C0402 – Feature 9 – OK-Klasse

Tabelle A.3.: KS-Test: Anteil [%] normalverteilte Featurewerte – Binär – EF-Klasse

Bauteil	AOI-Analyse	Original	Bereinigt
C0402	Analyse-A	33,3	33,3
	Analyse-B	0	0
	Analyse-C	0	12,1
C0603	Analyse-A	0	0
	Analyse-B	0	4
	Analyse-C	0	7,1
R0402	Analyse-A	0	0
	Analyse-B	0	3
	Analyse-C	0	14,3
R0603	Analyse-A	0	0
	Analyse-B	11,5	23,1
SOT23	Analyse-C	0	10,7
	Analyse-D	0	0
	Analyse-E	0	16

Tabelle A.4.: KS-Test: Anteil [%] normalverteilte Featurewerte – Binär – OK-Klasse

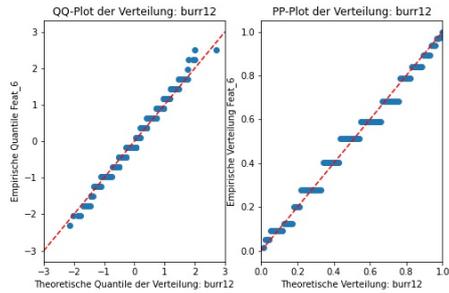
Bauteil	AOI-Analyse	Original	Bereinigt
C0402	Analyse-A	0	28,6
	Analyse-B	0	0
	Analyse-C	0	0

C0603	Analyse-A	0	0
	Analyse-B	0	0
	Analyse-C	0	7,1
R0402	Analyse-A	0	0
	Analyse-B	0	0
	Analyse-C	0	21,4
R0603	Analyse-A	0	0
	Analyse-B	0	0
SOT23	Analyse-C	0	32,1
	Analyse-D	0	0
	Analyse-E	4	0

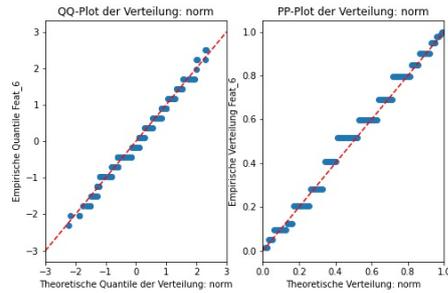
A.2.4. Anpassungstest – Thema 2

Tabelle A.5.: KS-Test – Analyse-C – C0402 – OK-Klasse – Winkel 270 ° – Feature 6 –
nach Datenbereinigung

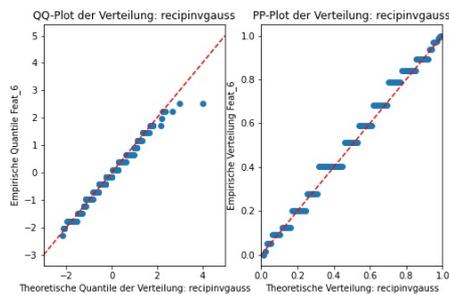
Verteilung	p-Wert	Distanz
Weibull_min	0,324	0,0676
Reziproke inv. Normal	0,318	0,0679
Burr Typ XII	0,317	0,0680
Weibull_max	0,314	0,0682
Pearson Typ III	0,312	0,0683
Hypergeom. Gauss	0,307	0,0686
Gen. inv. Gauss	0,304	0,0688
Inv. Gamma	0,297	0,0692
Log. Normal	0,297	0,0692
Gamma	0,296	0,0692
F	0,295	0,0693
Exp. Weibull	0,291	0,0696
Alpha	0,290	0,0696
Beta	0,277	0,0705
Gen. Gauss	0,210	0,0753
Fisk	0,201	0,0761
Exp. Normal	0,184	0,0776
Normal	0,157	0,0801
Gompertz	0,129	0,0832
Gen. Logistisch	0,089	0,0887



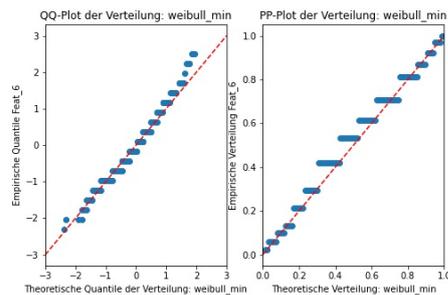
(a) Burr Typ XII Verteilung



(b) Normalverteilung



(c) Reziproke inverse Gaussverteilung



(d) Minimale Weibull-Verteilung

Abbildung A.13.: QQ-Diagramme (links); PP-Diagramme (rechts) – Analyse-C – C0402 – OK-Klasse – Feature 6 – nach Datenbereinigung

Tabelle A.6.: KS-Test: Anteil [%] normalverteilte Featurewerte – Multikategoriell – Tombstone-Klasse – Winkel 90°

Bauteil	AOI-Analyse	Original	Bereinigt
C0402	Analyse-A	100	100
	Analyse-B	9,1	30,3
	Analyse-C	42,9	71,4
C0603	Analyse-A	0	0
	Analyse-B	65,4	69,2
	Analyse-C	46,4	82,1

Tabelle A.7.: KS-Test: Anteil [%] normalverteilte Featurewerte – Multikategoriell – OK-Klasse – Winkel 90°

Bauteil	AOI-Analyse	Original	Bereinigt
C0402	Analyse-A	100	100
	Analyse-B	0	6,1
	Analyse-C	75	78,6

C0603	Analyse-A	100	100
	Analyse-B	7,7	34,6
	Analyse-C	25	61,5

A.2.5. Korrelationsanalyse – Thema 2

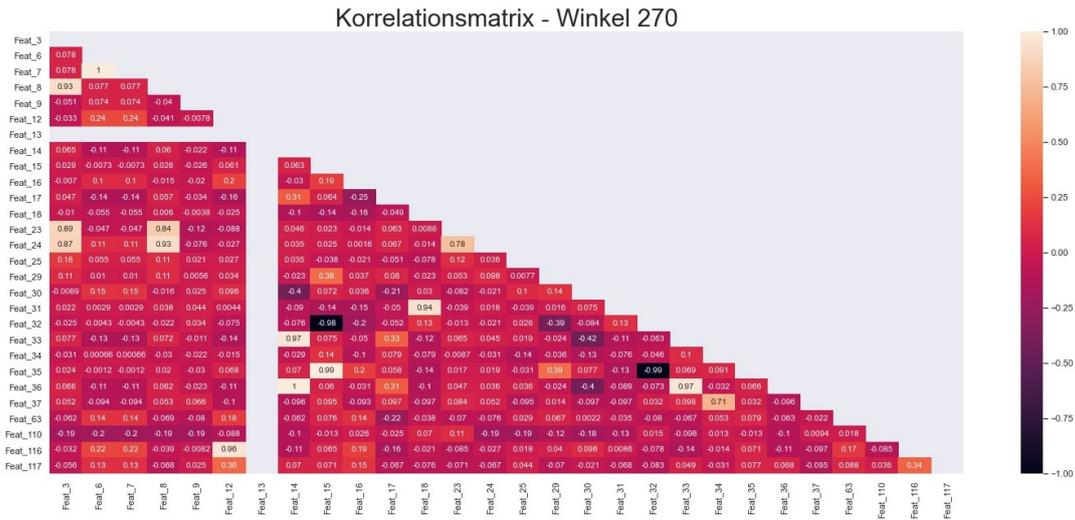


Abbildung A.14.: Heatmap-Diagramm – multikategoriell – Analyse-C – C0402 – OK-Klasse (FC 0) – Winkel 270 °

A.2.6. Feature-Ranking – Thema 2

Tabelle A.8.: Importance-Kennzahlen Winkelinformationen – multikategoriell mit FC 0 und 560

Algorithmus	Bauteil/Analyse			
	C0402		C0603	
	B	C	B	C
Boruta ET	0,138	0,067	0,154	0,133
Boruta RF	0,084	0,023	0,113	0,062
SOAP	0,166	1	0,042	0,031
XGBoost	0,024	0,055	0,005	0,019

A.2.7. Abhängigkeitsanalyse – Thema 1

Tabelle A.9.: U-Test von Mann-Whitney – Analyse-C – C0402 – zwei Klassen (EF, OK)

Feature-Nummer	U-Wert	p-Wert	Unterschied signifikant?
35	1009375,5	0,403	nein
18	998316,5	0,229	nein
17	991803,5	0,150	nein
63	989585	0,126	nein
9	988063,5	0,112	nein
33	985286,5	0,090	nein
13	1011037,5	0,013	ja
34	968092	0,005	ja
16	922841,5	1,45E-05	ja
32	917249	4,11E-06	ja
3	915418	2,84E-06	ja
36	917891	2,45E-06	ja
8	914289,5	2,22E-06	ja
24	897184	3,83E-08	ja
15	887405,5	2,91E-09	ja
14	874427	7,00E-11	ja
12	835232	1,47E-16	ja
116	834842	1,32E-16	ja
6	832782	4,56E-17	ja
7	832782	4,56E-17	ja
117	824943	2,82E-18	ja
37	829946,5	2,94E-24	ja
25	776440	4,97E-28	ja
31	717664	2,54E-42	ja
30	695512	1,54E-48	ja
29	693321,5	3,54E-49	ja
110	685860,5	6,51E-103	ja
23	453531,5	2,68E-145	ja

Tabelle A.10.: Kruskal-Wallis-Test – Analyse-C – C0402 – drei Klassen (EF, OK, PF)

Feature-Nummer	H-Wert	p-Wert	Unterschied signifikant?
17	4,147	0,1257	nein
33	4,580	0,1013	nein
13	6,006	0,0496	ja
34	9,661	0,0080	ja
32	10,177	0,0062	ja
16	10,196	0,0061	ja
18	10,277	0,0059	ja
63	12,906	0,0016	ja
14	29,519	3,88E-07	ja
35	39,826	2,24E-09	ja
36	45,936	1,05E-10	ja
116	53,279	2,69E-12	ja
12	53,606	2,28E-12	ja
9	58,489	1,99E-13	ja
25	72,891	1,48E-16	ja
117	74,628	6,23E-17	ja
15	92,620	7,72E-21	ja
8	109,262	1,87E-24	ja
3	112,617	3,51E-25	ja
31	119,215	1,29E-26	ja
24	130,829	3,89E-29	ja
30	132,846	1,42E-29	ja
29	134,651	5,76E-30	ja
37	143,545	6,75E-32	ja
6	355,617	6,01E-78	ja
7	355,617	6,01E-78	ja
110	381,628	1,35E-83	ja
23	819,514	1,10E-178	ja

Tabelle A.11.: Dunn-Bonferroni-Test – binär – Analyse-C
 Signifikante Unterschiede Dunn-Bonferroni-Test – obere Dreiecksmatrix
 Feature-Nummerierungen – untere Dreiecksmatrix Anzahl Hits – C0402 – Analyse-C –
 drei Klassen (EF, OK, PF)

Fehlerklasse	EF	OK	PF
EF		6, 7, 9, 12, 14, 15, 16, 23, 25, 29, 30, 31, 32, 35, 36, 37, 110, 116, 117	3, 6, 7, 8, 9, 12, 14, 23, 24, 25, 29, 30, 31, 34, 35, 36, 37, 63, 110, 116, 117
OK	19		3, 6, 7, 8, 9, 14, 15, 18, 23, 24, 34, 35, 36, 37, 63, 110, 117
PF	21	17	

Results: Logit

```

=====
Model:                Logit                Pseudo R-squared: 0.364
Dependent Variable:  y                AIC:                2562.6081
Date:                2022-04-01 12:00    BIC:                2711.4674
No. Observations:   2848                Log-Likelihood:    -1256.3
Df Model:           24                LL-Null:           -1974.1
Df Residuals:       2823                LLR p-value:       1.2415e-288
Converged:          1.0000                Scale:             1.0000
No. Iterations:     8.0000
=====

```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Feat 7	0.0090	0.0015	6.0717	0.0000	0.0061	0.0119
Feat 8	0.0190	0.0023	8.3827	0.0000	0.0146	0.0235
Feat 9	0.0101	0.0008	12.3055	0.0000	0.0085	0.0118
Feat 13	-3.4854	0.5612	-6.2110	0.0000	-4.5852	-2.3855
Feat 14	-0.0218	0.1390	-0.1568	0.8754	-0.2943	0.2507
Feat 15	-0.0225	0.1412	-0.1596	0.8732	-0.2992	0.2541
Feat 16	-0.0011	0.0005	-2.0731	0.0382	-0.0021	-0.0001
Feat 17	0.0046	0.0009	5.0901	0.0000	0.0029	0.0064
Feat 18	-0.0001	0.0001	-0.8154	0.4148	-0.0002	0.0001
Feat 23	0.0057	0.0024	2.3201	0.0203	0.0009	0.0104
Feat 24	-0.0089	0.0028	-3.1679	0.0015	-0.0144	-0.0034
Feat 25	-0.0357	0.0041	-8.6247	0.0000	-0.0438	-0.0276
Feat 29	-0.0010	0.0004	-2.7328	0.0063	-0.0018	-0.0003
Feat 30	0.0013	0.0013	1.0402	0.2983	-0.0012	0.0038
Feat 31	-0.0001	0.0001	-1.1549	0.2481	-0.0004	0.0001
Feat 32	0.0003	0.0005	0.4895	0.6245	-0.0008	0.0013
Feat 33	-0.0027	0.0014	-2.0042	0.0451	-0.0054	-0.0001
Feat 34	0.0191	0.1390	0.1373	0.8908	-0.2534	0.2916
Feat 35	0.0217	0.1412	0.1536	0.8780	-0.2550	0.2983
Feat 36	0.0192	0.1390	0.1381	0.8902	-0.2533	0.2917
Feat 37	0.0316	0.1412	0.2235	0.8231	-0.2452	0.3084
Feat 63	-0.0027	0.0006	-4.7172	0.0000	-0.0039	-0.0016
Feat 110	1.1650	0.2383	4.8882	0.0000	0.6979	1.6321
Feat 116	0.0002	0.0001	1.5942	0.1109	-0.0000	0.0004
Feat 117	0.0012	0.0003	4.5430	0.0000	0.0007	0.0018

=====

Abbildung A.15.: Binäre logistische Regression – Analyse-C – C0402 – zwei Klassen
 (EF, OK)

Results: MNLogit

```

=====
Model:                MNLogit                Pseudo R-squared: 0.640
Dependent Variable:  y                        AIC:                3458.1072
Date:                2022-04-01 12:10       BIC:                3712.5007
No. Observations:   4272                    Log-Likelihood:    -1689.1
Df Model:           38                        LL-Null:           -4693.3
Df Residuals:       4232                    LLR p-value:       0.0000
Converged:          1.0000                    Scale:             1.0000
No. Iterations:     17.0000
=====

```

y = 0	Coef.	Std.Err.	t	P> t	[0.025	0.975]
Feat 7	0.0402	0.0106	3.7786	0.0002	0.0193	0.0610
Feat 8	-0.1653	0.0835	-1.9803	0.0477	-0.3289	-0.0017
Feat 9	-0.1948	0.0176	-11.0545	0.0000	-0.2293	-0.1603
Feat 14	-0.0449	0.0063	-7.1895	0.0000	-0.0572	-0.0327
Feat 15	0.0226	0.0070	3.2204	0.0013	0.0089	0.0364
Feat 16	0.0234	0.0043	5.4147	0.0000	0.0149	0.0318
Feat 17	0.0360	0.0067	5.4088	0.0000	0.0230	0.0491
Feat 18	0.0079	0.0011	6.9135	0.0000	0.0057	0.0102
Feat 23	-0.3282	0.0337	-9.7350	0.0000	-0.3943	-0.2622
Feat 24	0.2453	0.0487	5.0342	0.0000	0.1498	0.3407
Feat 25	0.4125	0.0873	4.7250	0.0000	0.2414	0.5836
Feat 29	0.0096	0.0036	2.7151	0.0066	0.0027	0.0166
Feat 30	-0.0860	0.0112	-7.6583	0.0000	-0.1080	-0.0640
Feat 31	-0.0065	0.0012	-5.3626	0.0000	-0.0089	-0.0041
Feat 32	0.0110	0.0047	2.3649	0.0180	0.0019	0.0202
Feat 33	-0.1088	0.0136	-7.9757	0.0000	-0.1356	-0.0821
Feat 35	-0.0551	0.0092	-5.9676	0.0000	-0.0733	-0.0370
Feat 63	0.0145	0.0070	2.0705	0.0384	0.0008	0.0282
Feat 116	-0.0128	0.0032	-3.9455	0.0001	-0.0192	-0.0065
Feat 117	0.0024	0.0042	0.5601	0.5754	-0.0059	0.0106

y = 1	Coef.	Std.Err.	t	P> t	[0.025	0.975]
Feat 7	-0.0050	0.0009	-5.4206	0.0000	-0.0069	-0.0032
Feat 8	-0.0087	0.0015	-5.6320	0.0000	-0.0118	-0.0057
Feat 9	-0.0032	0.0005	-7.0094	0.0000	-0.0041	-0.0023
Feat 14	0.0024	0.0005	4.9508	0.0000	0.0015	0.0034
Feat 15	-0.0061	0.0012	-5.2602	0.0000	-0.0084	-0.0039
Feat 16	0.0000	0.0004	0.0714	0.9430	-0.0008	0.0009
Feat 17	0.0002	0.0007	0.3452	0.7299	-0.0011	0.0015
Feat 18	-0.0001	0.0001	-1.3381	0.1809	-0.0003	0.0001
Feat 23	-0.0018	0.0007	-2.4962	0.0126	-0.0033	-0.0004
Feat 24	0.0041	0.0011	3.6956	0.0002	0.0019	0.0063
Feat 25	0.0131	0.0025	5.1442	0.0000	0.0081	0.0181
Feat 29	0.0012	0.0003	4.2072	0.0000	0.0006	0.0017
Feat 30	0.0017	0.0011	1.5302	0.1260	-0.0005	0.0038
Feat 31	0.0004	0.0001	4.0095	0.0001	0.0002	0.0006
Feat 32	-0.0008	0.0004	-2.1032	0.0355	-0.0016	-0.0001
Feat 33	0.0015	0.0013	1.1439	0.2527	-0.0010	0.0039
Feat 35	0.0068	0.0012	5.6088	0.0000	0.0044	0.0092
Feat 63	0.0018	0.0003	5.5321	0.0000	0.0012	0.0024
Feat 116	0.0000	0.0001	0.1420	0.8871	-0.0001	0.0001
Feat 117	-0.0010	0.0002	-5.4033	0.0000	-0.0014	-0.0006

=====

Abbildung A.16.: Multinomiale logistische Regression – Analyse-C – C0402 – drei Klassen (EF, PF, OK)

A.2.8. Dependenzanalyse – Thema 2

Tabelle A.12.: U-Test von Mann-Whitney – Analyse-C – C0402 – zwei Klassen (OK, Tombstone)

Feature-Nummer	U-Wert	p-Wert	Unterschied signifikant?
34	61274	0,445	nein
63	61100,5	0,433	nein
9	60339	0,325	nein
37	60309	0,261	nein
24	58532	0,130	nein
18	57458,5	0,064	nein
15	54240,5	0,003	ja
16	53826,5	0,002	ja
33	53093	0,001	ja
32	52226	0,0002	ja
35	52054	0,0002	ja
25	51682	0,0001	ja
3	51542	9,270E-05	ja
8	51201	5,542E-05	ja
31	48763,5	8,99E-07	ja
14	45385	7,605E-10	ja
6	42440,5	4,775E-13	ja
7	42440,5	4,775E-13	ja
29	42422,5	4,570E-13	ja
30	41958,5	1,274E-13	ja
36	42000,5	5,674E-14	ja
17	33541,5	9,015E-26	ja
12	33337	3,451E-26	ja
116	33209,5	2,444E-26	ja
110	41187,5	4,528E-30	ja
117	31010,5	2,849E-30	ja
23	12158	2,875E-76	ja

Tabelle A.13.: Kruskal-Wallis-Test – Analyse-C – C0402 – nach Winkel 0 °, 90 °, 180 °, 270 ° – OK-Klasse

Feature-Nummer	H-Wert	p-Wert	Unterschied signifikant?
35	0,47	0,976	nein
110	2,68	0,613	nein
29	13,49	0,009	ja
32	16,62	0,002	ja
63	28,37	1,05068E-05	ja
25	33,05	1,16721E-06	ja
117	40,55	3,32434E-08	ja
8	47,06	1,47895E-09	ja
3	49,12	5,50597E-10	ja
9	51,52	1,73724E-10	ja
24	52,79	9,44465E-11	ja
15	62,22	9,91327E-13	ja
23	75,21	1,79947E-15	ja
34	80,09	1,66356E-16	ja
36	80,45	1,40025E-16	ja
18	81,72	7,52326E-17	ja
31	86,93	5,90708E-18	ja
30	99,97	9,97603E-21	ja
33	100,94	6,19196E-21	ja
116	101,92	3,84614E-21	ja
12	102,64	2,69106E-21	ja
16	103,87	1,47734E-21	ja
14	105,58	6,36179E-22	ja
17	115,42	5,08518E-24	ja
6	192,28	1,71632E-40	ja
7	192,28	1,71632E-40	ja
37	194,47	5,78887E-41	ja

Tabelle A.14.: Signifikante Unterschiede Dunn-Bonferroni – obere Dreiecksmatrix
 Feature-Nummerierungen – untere Dreiecksmatrix Anzahl Hits – C0402
 – Analyse-C – OK-Klasse – Winkel 0 °, 90 °, 180 °, 270 °

Winkel	0	90	180	270
0		3, 8, 9, 12, 16, 17, 18, 23, 24, 30, 31, 33, 34, 37, 116	3, 8, 9, 16, 17, 23, 24, 30, 34, 37	3, 6, 7, 8, 12, 16, 17, 24, 25, 29, 30, 32, 33, 34, 36, 37, 116, 117
90	15		0	6, 7, 9, 12, 16, 17, 18, 23, 25, 30, 31, 63, 116, 117
180	10	0		3, 6, 7, 8, 9, 12, 17, 18, 23, 24, 116, 117
270	18	14	12	

Tabelle A.15.: Kruskal-Wallis-Test – Analyse-C – C0402 – Winkel 0 °, 90 °, 135 °, 180 °, 270 ° – Tombstone-Klasse

Feature-Nummer	H-Wert	p-Wert	Unterschied signifikant?
36	3,47	0,48	nein
29	3,59	0,46	nein
33	4,10	0,39	nein
116	5,30	0,26	nein
12	5,48	0,24	nein
17	5,87	0,21	nein
37	6,84	0,14	nein
3	8,24	0,08	nein
8	8,32	0,08	nein
25	9,51	0,05	ja
24	10,21	0,04	ja
63	13,51	0,009	ja
9	14,38	0,006	ja
31	15,44	0,004	ja
30	18,21	0,0011	ja
117	18,81	0,0009	ja
110	21,56	0,0002	ja
6	23,74	8,996E-05	ja
7	23,74	8,996E-05	ja
23	28,19	1,14268E-05	ja
14	28,38	1,043E-05	ja
32	38,64	8,270E-08	ja
35	42,70	1,196E-08	ja
15	44,50	5,05E-09	ja
34	56,36	1,689E-11	ja
18	78,44	3,731E-16	ja
16	80,36	1,463E-16	ja

Tabelle A.16.: Signifikante Unterschiede Dunn-Bonferroni-Test – obere Dreiecksmatrix
 Feature-Nummerierungen – untere Dreiecksmatrix Anzahl Hits – C0402 –
 Analyse-C – Tombstone-Klasse – Winkel 0 °, 90 °, 135 °, 180 °, 270 °

Winkel	0	90	135	180	270
0		6, 7, 18, 23, 34, 110	16, 32, 117	18	14, 15, 18, 32, 34, 35, 63
90	6		25, 6, 7, 9, 14, 15, 16, 18, 23, 25, 30, 32, 34, 35, 110, 117	0	23, 31, 32
135	3	16		16, 18, 30, 34, 117	6, 7, 14, 15, 16, 18, 34, 35, 117
180	1	0	5		
270	7	3	9	0	

Results: Logit

```

=====
Model:                Logit                Pseudo R-squared:  0.954
Dependent Variable:   y                    AIC:               105.7270
Date:                2022-03-31 17:06       BIC:               246.0029
No. Observations:    682                    Log-Likelihood:    -21.864
Df Model:            30                      LL-Null:           -472.73
Df Residuals:        651                    LLR p-value:       2.6462e-170
Converged:           0.0000                  Scale:             1.0000
No. Iterations:      35.0000

-----

```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Feat_6	0.4239	0.1394	3.0404	0.0024	0.1506	0.6972
Feat_8	0.0409	0.0314	1.3032	0.1925	-0.0206	0.1025
Feat_9	0.0412	0.0173	2.3825	0.0172	0.0073	0.0752
Feat_12	2.0948	1.7502	1.1969	0.2314	-1.3356	5.5252
Feat_13	-24.8416	21907.8999	-0.0011	0.9991	-42963.5363	42913.8532
Feat_14	3.0748	1.7409	1.7663	0.0773	-0.3372	6.4869
Feat_15	5.4033	2.4407	2.2138	0.0268	0.6196	10.1870
Feat_16	0.0107	0.0114	0.9401	0.3472	-0.0116	0.0330
Feat_17	-0.1421	0.0529	-2.6886	0.0072	-0.2457	-0.0385
Feat_18	-0.0071	0.0035	-2.0529	0.0401	-0.0140	-0.0003
Feat_23	0.2130	0.0727	2.9299	0.0034	0.0705	0.3555
Feat_24	-0.2203	0.0755	-2.9203	0.0035	-0.3682	-0.0725
Feat_25	-0.2775	0.0982	-2.8248	0.0047	-0.4700	-0.0850
Feat_29	0.0030	0.0095	0.3165	0.7516	-0.0157	0.0217
Feat_30	-0.1709	0.0677	-2.5238	0.0116	-0.3036	-0.0382
Feat_31	-0.0048	0.0042	-1.1613	0.2455	-0.0130	0.0033
Feat_32	-0.0012	0.0117	-0.1033	0.9177	-0.0242	0.0218
Feat_33	-0.0897	0.0473	-1.8951	0.0581	-0.1825	0.0031
Feat_34	-3.0953	1.7466	-1.7721	0.0764	-6.5186	0.3281
Feat_35	-5.3936	2.4364	-2.2137	0.0268	-10.1689	-0.6183
Feat_36	-3.0318	1.7348	-1.7476	0.0805	-6.4320	0.3684
Feat_37	-5.4210	2.4418	-2.2201	0.0264	-10.2069	-0.6352
Feat_63	-0.0328	0.0126	-2.6029	0.0092	-0.0576	-0.0081
Feat_110	0.6662	2.3912	0.2786	0.7805	-4.0205	5.3530
Feat_116	-0.1938	0.1748	-1.1088	0.2675	-0.5365	0.1488
Feat_117	0.0197	0.0174	1.1363	0.2558	-0.0143	0.0537
angle_0	40.6554	43815.7952	0.0009	0.9993	-85836.7252	85918.0360
angle_90	44.2096	43815.7953	0.0010	0.9992	-85833.1710	85921.5903
angle_135	61.4524	43826.5735	0.0014	0.9989	-85837.0532	85959.9580
angle_180	84.9086	218877.3658	0.0004	0.9997	-428906.8453	429076.6625
angle_270	41.9817	43815.7952	0.0010	0.9992	-85835.3989	85919.3623

Abbildung A.17.: Binäre logistische Regression – Analyse-C – C0402 – zwei Klassen (OK, Tombstone) – Winkel 0 °, 90 °, 135 °, 180 °, 270 °

A.2.9. Zuordnung Featurewert – AOI-Messgröße

Tabelle A.17.: Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-A

Feature-Nummer	AOI-Messgröße
1	Oberer Grauwert
4	Unterer Grauwert
5	Oberer Grauwert

Tabelle A.18.: Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-B

Feature-Nummer	AOI-Messobjekt
1	Gesamtmatchwert
3	Machtwert rechts
4	Verdrehung Bauteil
6	Verschiebung Bauteil x-Koordinate
8	min. seitlicher Matchwert
23	Klassifikation Fenster A
24	Klassifikation Fenster B
25	Klassifikation Fenster C
63	Anzahl Eingangsbilder Analyse

Tabelle A.19.: Zuordnung Featurewert zu Messgröße AOI-Prüfung Analyse-D & Analyse-E (nur SOT23)

Analyse	Feature-Nummer	AOI-Messobjekt
Analyse-D	1	Pinversatz
	4	Abweichung Pinbreite
	18	Ergebnis Lückenanalyse
	22	Ergebnis Pinfindung
	94	Pinlänge
Analyse-E	12	Bauteilfindung
	25	Ergebnis Findung Pin B
	41	Verbiegung Pin A
	42	Verbiegung Pin A
	60	Verschiebewert der Vor-Analyse (Link-Funktion)

A.3. Maschinelles und tiefes Lernen

Tabelle A.20.: Steigerung F1-Score Einzelmodell und Ensemble-Methode – binär

Bauteil	AOI-Analyse	Steigerung F1-Score [%]
C0402	A	-2,05
	B	1,72
	C	8,46
C0603	A	2,53
	B	0,94
	C	0,97
R0402	A	10,36
	B	-0,33
	C	-0,81
R0603	A	4,99
	B	0,96
SOT23	C	-3,29
	D	-3,46
	E	19,68