

# Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie

Bachelorarbeit  
von

Roman Ungefuk

An der Fakultät für Informatik  
Institut für Anthropomatik und Robotik (IAR)

Erstgutachter:	Prof. Dr.-Ing. Rüdiger Dillmann
Betreuender Mitarbeiter:	Dipl.-Inform. Darko Katic

Bearbeitungszeit: 15. Oktober 2015 – 14. April 2016



Ich versichere hiermit, die vorliegende Bachelorarbeit selbstständig und ohne fremde Hilfe angefertigt zu haben. Die verwendeten Hilfsmittel und Quellen sind im Literaturverzeichnis vollständig aufgeführt.

Karlsruhe, den 14. April 2016



# Abstract

Die Laparoskopie bietet viele Vorteile für Patienten, ist aber schwierig für Chirurgen. Um die Nachteile in der Laparoskopie zu minimieren, werden computergestützte Assistenzsysteme entwickelt. MEDIASSIST ist ein solches computergestütztes Assistenzsystem und wird zur Zeit am Karlsruher Institut für Technologie am Institut für Anthropomatik und Robotik (IAR) von der Forschungsgruppe "Computerassistierte Chirurgie" entwickelt. Hiermit können den Chirurgen und ihren Assistenten wichtige Zusatzinformationen zur Verfügung gestellt werden. Eine dieser Informationen ist die, die Zeit bis zum nächsten Schritt in der Operation, oder wie lange es noch dauert, bis die Operation beendet wird. Diese Information kann zum Beispiel hilfreich für die Assistenten der Chirurgen sein, um rechtzeitig die für den nächsten Operationsschritt benötigten Instrumente oder den nächsten Patienten für die Operation vorzubereiten. In der Realität spielen die gewonnenen Sekunden manchmal eine große Rolle, um die Behandlung der Patienten zu verbessern. Deshalb ist es wichtig, die Dauer der Schritte in der Laparoskopie und allgemein in der Chirurgie vorherzusagen. Dafür wurden in der Arbeit einige Methoden entwickelt und durch die Leave-One-Out-Kreuzvalidierung evaluiert. Das beste Ergebnis wurde mit der Methode erreicht, die den Random Forest nutzt. Dieser verwendet zum Trainieren die Dauer von den vorherigen Schritten und ihre Reihenfolge in der Operation. Das Ergebnis hat einen Durchschnitt der Gesamtfehlerquote pro Historie von 26.36% und wurde mit 17 Annotationen und mit der Historiegröße 5 erreicht. Die Historiegröße ist die Anzahl der Schritte, die beim Konstruieren des Featurevektors miteinbezogen wird. Die Annotationen sind die Informationen über die Rektumresektion. Die Gesamtfehlerquote ist der prozentuale Anteil des Gesamtfehlers zur Dauer der evaluierten Operation.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Bedeutung der Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie . . . . .	1
1.2	Zielsetzung . . . . .	2
1.3	Gliederung der Arbeit . . . . .	2
<b>2</b>	<b>Bisherige Methoden zur Dauervorhersage</b>	<b>3</b>
<b>3</b>	<b>Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie</b>	<b>5</b>
3.1	Durchschnittsmethode . . . . .	6
3.2	Lineare Regression . . . . .	7
3.3	Random Forest Methoden . . . . .	7
3.3.1	Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	8
3.3.2	Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	9
3.3.3	Korrelation zwischen Schritten und Features mit der Schrittdauer . . . . .	9
<b>4</b>	<b>Implementierung</b>	<b>11</b>
4.1	Durchschnittsmethode . . . . .	11
4.2	Lineare Regression . . . . .	12
4.3	Random Forest Methoden . . . . .	14
4.3.1	Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	16
4.3.2	Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	18
4.3.3	Korrelation zwischen Schritten und Features mit der Schrittdauer . . . . .	19
<b>5</b>	<b>Evaluation</b>	<b>23</b>
5.1	Verwendeter Datensatz . . . . .	23
5.2	Durchschnittsmethode . . . . .	26
5.3	Lineare Regression . . . . .	30
5.4	Random Forest Methoden . . . . .	33
5.4.1	Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	33
5.4.2	Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation . . . . .	36
5.4.3	Korrelation zwischen Schritten und Features mit der Schrittdauer . . . . .	38
5.5	Vergleich aller Methoden . . . . .	40
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>41</b>

<b>Abbildungsverzeichnis</b>	<b>44</b>
<b>Tabellenverzeichnis</b>	<b>45</b>
<b>Literaturverzeichnis</b>	<b>48</b>







# 1. Einleitung

Im folgenden Kapitel wird zuerst die Bedeutung der Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie erläutert. Danach werden die Ziele der vorliegenden Bachelorarbeit definiert und die Aufbau der Arbeit vorgestellt.

## 1.1 Bedeutung der Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie

Die Laparoskopie oder die laparoskopische Chirurgie ist eine Operationsart, die mit Hilfe eines optischen Instruments im Inneren des Bauchs durchgeführt wird [1]. Die Laparoskopie ist ein Zweig der minimal-invasiven Chirurgie (MIC) oder Bildschirmchirurgie. Die minimal-invasive Chirurgie ist die Operationen oder die Untersuchungen im Inneren des Menschen mit möglichst wenig traumatischen Eingriffen [2]. Die Operationen im Bauchraum sind besonders gut hierfür geeignet. Die Operation Rektumresektion ist keine Ausnahme. Sie ist auch gut für den minimal-invasiven Eingriff geeignet. Bei der Laparoskopie werden kleine Schnitte auf dem Bauch gemacht und dadurch wird ein Zugang für das Endoskop und die anderen chirurgischen Instrumente geschaffen. Die Zugänge werden durch sogenannte Trokare gelegt. Um die Bauchdecke von den Organen zu heben, um so die Arbeit für den Chirurgen zu erleichtern, wird der Bauchraum mit Kohlenstoffdioxid gefüllt. Anschließend werden die chirurgischen Instrumente und das Endoskop durch die Trokare eingeführt.

Die Vorteile der minimal-invasiven Chirurgie sind offensichtlich. Das sind die kleine Operationswunde, die zu schneller Rekonvaleszenz nach der Operation führt, bessere kosmetische Ergebnisse und einen erhöhten Komfort für den Patienten. Neben den Vorteilen gibt es auch einige Nachteile. Die Nachteile betreffen vor allem den Chirurgen. Er hat keinen direkten Hand-Auge-Kontakt. Dadurch, dass das Bild auf dem Bildschirm zweidimensional ist, erschwert das den Chirurgen ihre Arbeit. Oft haben die Chirurgen auch nicht das direkte Tastgefühl.

Um diese Nachteile zu minimieren, werden computergestützte Assistenzsysteme entwickelt. So ein computergestütztes Assistenzsystem ist MEDIASSIST und wird zur Zeit am Karlsruher Institut für Technologie am Institut für Anthropomatik und Robotik (IAR) von der Forschungsgruppe "Computerassistierte Chirurgie" entwickelt. Durch so ein Assistenzsystem können den Chirurgen und ihren Assistenten wichtige Zusatzinformationen zur Verfügung gestellt werden. Eine dieser Informationen ist die Information über die Zeit bis

zum nächsten Schritt in der Operation, oder wie lange es noch dauert, bis die Operation beendet wird. Diese Information kann zum Beispiel hilfreich für die Assistenten der Chirurgen sein, um rechtzeitig die für den nächsten Operationsschritt benötigten Instrumente oder den nächsten Patienten für die Operation vorzubereiten. In der Realität spielen die gewonnenen Sekunden eine große Rolle, um Operationen effizient durchführen zu können und die Behandlung der Patienten zu verbessern. Deshalb ist es wichtig, die Dauer der Schritten in der Laparoskopie und allgemein in der Chirurgie vorherzusagen.

## 1.2 Zielsetzung

Das Ziel dieser Bachelorarbeit ist die Entwicklung eines Verfahrens zur Dauervorhersage der Schritte in der Laparoskopie. Dafür werden mehrere Methoden entwickelt. Folgende Konzepte wurden untersucht:

- arithmetisches Mittel
- lineare Regression
- maschinelles Lernen mit Hilfe von Random Forest.

Für die Entwicklung der Methoden werden die Daten aus den Annotationen verwendet. Bei der Operation Rektumresektion des menschlichen Phantoms wurden diese Annotationen in einem CSV-Format von den Medizинern aus dem Universitätsklinikum Heidelberg aufbereitet und zur Verfügung gestellt. Am Ende dieser Arbeit sollten die entwickelten Methoden evaluiert werden und durch das Analysieren der Ergebnisse für eine oder mehrere Methoden für die Dauervorhersage der Schritten in der Laparoskopie entschieden werden.

## 1.3 Gliederung der Arbeit

Im Folgenden werden die sechs Kapitel dieser Bachelorarbeit vorgestellt:

- Einleitung
- Bisherige Methoden zur Zeit- und Dauervorhersage
- Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie
- Implementierung
- Evaluation
- Zusammenfassung und Ausblick .

Im Kapitel Einleitung 1 wird auf die Bedeutung dieser Bachelorarbeit eingegangen, Ziele gesetzt und die Gliederung dieser Bachelorthesis erläutert. Im nächsten Kapitel 2 werden einige wissenschaftliche Arbeiten zusammengefasst und erläutert. Das Kapitel 3 beschäftigt sich mit den Methoden zur Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie. In diesem Teil der Arbeit wird auf die Methoden eingegangen, die in dieser Thesis entwickelt oder benutzt wurden. In den Kapiteln 4 und 5 wird erläutert, wie diese Methoden implementiert und evaluiert wurden und die Ergebnisse der Evaluation werden präsentiert und verglichen. Im letzten Kapitel 6 wird die Bachelorarbeit zusammengefasst und ein Ausblick gegeben.

## 2. Bisherige Methoden zur Dauervorhersage

Um die bisherige Methoden zur Zeit- und Dauervorhersage zu erläutern, werden in diesem Kapitel einige wissenschaftliche Artikel zusammengefasst. Für die Dauervorhersage und Zeitvorhersage gibt es bereits verschiedene wissenschaftliche Artikel, Methoden und Ansätze.

René Schult et. al machen Vorhersagen über Operationsdauer, um die Effizienz der Planung in der Chirurgie zu verbessern [3]. Um diesen Mangel zu vermeiden, wurde eine Methode für die Dauerabschätzung der Operation auf Basis von Anästhesie-Protokollen entwickelt. In vielen Krankenhäusern werden Patienten nach der OP vom Operationsraum in einen anderen Raum gebracht und dort gelassen bis die Wirkung der Betäubung nachlässt. Die Autoren beschäftigen sich mit dem Szenario, dass kein weiterer Raum nach der Operation zur Verfügung steht. Infolgedessen bleibt der Operationsraum länger mit dem Patient belegt, der noch unter Narkose steht. Die Ärzte müssen warten, bis der Patient von der Narkose aufwacht. Das Ziel ist es, die Zeit, während der der Patient unter der Narkose ist, zu minimieren. Der Ansatz besteht darin, dass man die Zielvariable intelligent diskretisiert und anschließend mit verschiedenen Algorithmen so klassifiziert, dass der Mensch es interpretieren kann.

Das System besteht aus folgenden Punkten:

- Persönliche Daten und Medikamente aus den Anästhesie-Protokollen werden als Input verwendet.
- Die Zielvariable wird folgendermaßen spezifiziert: Dauer der Operation = Belegung des Operationsraums pro Operation.
- Zur Diskretisierung der Daten werden drei Methoden verwendet:
  - TUBE (englisch Tree based unsupervised bin estimator) [4]
  - EWIB (englisch Equal Width Interval Binning) [5]
  - K-Means [6]
- Zur Klassifizierung verwendet man folgende Algorithmen:
  - Das Arithmetische Mittel [7] wird als Referenzwert verwendet.
  - Der Naive-Bayes-Algorithm [8].

- Die Entscheidungsbaumklassifizierung mit ID3-Algorithmus [9].
- Die Entscheidungsbaumklassifizierung mit J.48 [10].
- Das Random Forest Klassifizierungsverfahren [11].
- Bei der Evaluation wurden die Diskretisierungs- und Klassifizierungsmethoden kombiniert analysiert. Zur Evaluation wurde eine Methode implementiert. Diese Methode spiegelt die Belegung des Operationsraums wieder.
- Die Implementierung des Werkzeuges zur automatisierten Vorhersage der Operationsdauer. Für die Prognose der OP-Dauer wurde aus der wissenschaftlichen Arbeit von René Schult et. al die beste Kombination aus der Diskretizations- und Klassifizierungsmethoden gewählt. Drei Diskretisierungs- und vier Klassifizierungsmethoden wurden evaluiert. Da das beste Ergebnis bei der Kombination ID3-Algorithmus für die Klassifizierung und mit dem K-Means-Algorithmus für die Diskretisierung erreicht wurde, entschied man sich auch bei der Implementierung des Werkzeuges für diese Kombination.

Ein anderer Ansatz wurde von Kayis et al. in [12] präsentiert. Der Artikel beschäftigt sich in der erste Linie mit der Verbesserung der Operationsdauervorhersage. Die Prognose der Operationsdauer basiert auf der weitverbreiteten Methode, die im Folgenden beschrieben wird. Die Prognose der Dauer für eine bestimmte Operationsart wird basierend auf den Erfahrungsdaten von den letzten 5 Operationen gemacht, die ein bestimmter Chirurg im letzten Jahr durchgeführt hat. Wenn keine Erfahrungsdaten über den Chirurg für diese Operationsart vorhanden sind, wird die Dauer mithilfe der Daten aus der Praxis eines anderen Chirurgen abgeschätzt [13]. Wenn die Daten der letzten 5 Operationen vorliegen, wird die vorhergesagte Dauer der Operation durch das arithmetische Mittel [7] von der Dauer dieser letzten 5 Operationen dargestellt. Die Fälle, bei denen die Daten aus der Praxis für die letzten 5 Operationen fehlten, wurden in der Arbeit nicht behandelt. Die Evaluation der Methode wurde mit zwei Methoden zur systematischen Messabweichung [14] und mittlerer absoluter Abweichung [15] in Minuten gemacht.

Die systematische Messabweichung wird als mittlere Differenz zwischen Ist- und Sollwert berechnet. Die mittlere absolute Abweichung berechnet man als die mittlere absolute Differenz zwischen Ist- und dem Sollwert. Dieses Maß beschreibt die Verteilung der Fehler und macht die Abschätzung transparenter. Für die Evaluation dieser Methoden wurde die Kreuzvalidierung [16] benutzt.

Für diese Arbeit wurden die Operationsdaten im detaillierten elektronischen Format in einem Zeitrahmen von zwei Jahren gesammelt. Die Daten wurden durch die Berichte von Medizinern bestätigt und umfassen insgesamt 10305 Operationen aus den 7 Operationsräumen im Lucile Packard Children's Hospital.

### 3. Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie

Am Institut für Anthropomatik und Robotik (IAR) am Karlsruher Institut für Technologie (KIT) wird von der Forschungsgruppe "Computerassistierte Chirurgie" ein einheitliches Assistenzsystem für verschiedene chirurgische Eingriffe entwickelt, wie die Cholezystektomie (Gallenblasenentfernung), Dentalimplantologie und Rektumresektion. Diese Arbeit beschäftigt sich mit der Rektumresektion. Man möchte, dass das Assistenzsystem den Chirurgen und ihren Assistenten die übrige Zeit bis zum nächsten Schritt der Operation oder bis zum Ende der Operation vorhersagen kann. Dies kann für Assistenten hilfreich sein, um zum Beispiel die für den nächsten Schritt relevanten chirurgische Instrumente vorzubereiten. Im Endeffekt können die Sekunden, die man dadurch in der Operation gewinnt, in einem kritischen Moment des chirurgischen Eingriffes dem Patienten das Leben retten. Die Vorhersage der Dauer der Operation kann auch beispielsweise zur Bewertung der Auszubildenden herangezogen werden, um zu überprüfen, ob sie schnell genug sind oder noch weiter ihre Professionalität steigern müssen, um ein ausreichendes Maß an Schnelligkeit und Qualität zu erreichen.

Um zu prognostizieren, wie viel Zeit man noch bis zum nächsten Schritt hat und/oder wie lange die Schritte noch dauern, werden Angaben zu den Schritten aus den Annotationen benötigt. In diesem Kapitel geht es um die Daten von der Operation auf dem menschlichen Phantom, die von Chirurgen aus Heidelberg gemacht wurden. Für die Operation Rektumresektion auf dem Phantom wurden von Chirurgen 14 Schritte konzipiert. Die Schritte sind

- Check for metastasis and adhesions
- Mobilization sigmoid
- Mobilization descending colon
- Mobilization splenic flexure
- Inspection colon
- Lancing retroperitoneum
- Delineating of vessels
- Division of artery

- Division of vein
- Opening of lesser pelvic peritoneum
- Dissection rectum
- Transect rectum
- Salvage rectum
- Visual inspection lesser pelvis.

Abbildung 3.1 gibt einen Überblick über eine der möglichen Vorgehensweisen zur Vorhersage von Schritten bei der Operation Rektumresektion. In dieser Arbeit verwendet man zur Vorhersage das Verfahren des maschinellen Lernens. Diese Methode basiert auf der künstlichen Gewinnung der Wissensbasis aus den Erfahrungsdaten [17]. Mehr über maschinelles Lernen in der Medizininformatik kann man dem Buch Machine Learning in Healthcare Informatics [18] entnehmen. Zuerst werden die Annotationen geeignet aufbereitet, die in der Form mehrerer CSV-Dateien vorliegen. Das Programm, das die Vorhersage der Schrittdauer macht, wird einfachheitshalber Prophet genannt. Dann wird der Prophet mit den aufbereiteten Daten trainiert. Nachdem der Prophet trainiert wurde, kann er als Eingabe den Featurevektor bekommen. Mit Hilfe des Featurevektors entscheidet der Prophet, wie lange der Schritt dauert. Je nachdem was für eine Methode man zur Prognose nimmt, verändert sich auch das Trainieren des Propheten.

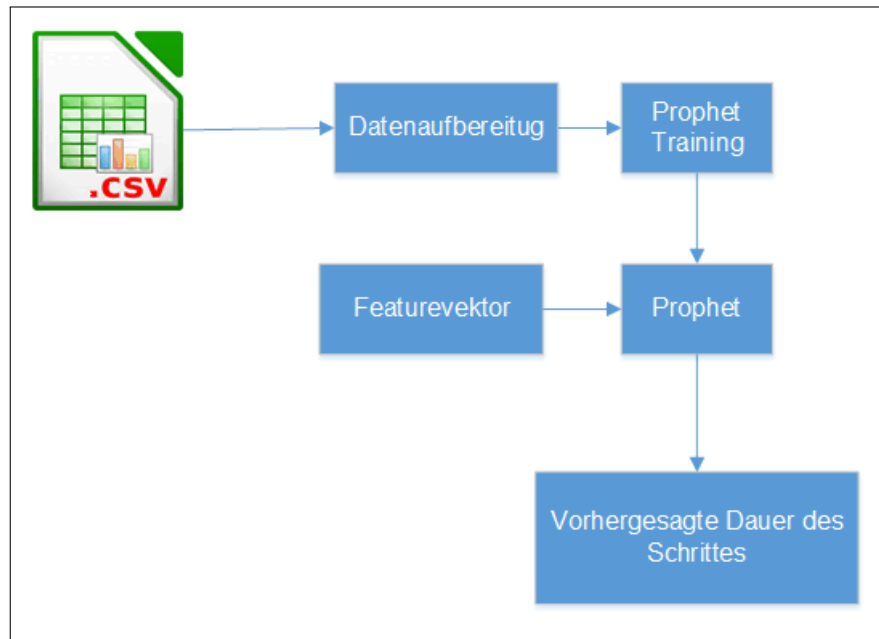


Abbildung 3.1: Ablauf der Dauervorhersage

In dieser Bachelorarbeit wurden mehrere Methoden erstellt, die im Folgenden näher beschrieben werden.

### 3.1 Durchschnittsmethode

Der Durchschnitt in dem Verfahren zur Prädiktion wird als arithmetisches Mittel verstanden. Bei  $n$  Schritten wird der Wert mit der Formel

$$\frac{1}{n} \sum_{i=1}^n X_i \quad (3.1)$$



berechnet. Wobei die Variable  $X_i$  die Werte der Schrittdauer annimmt. Mit Hilfe dieses Wertes kann man grobe Aussagen über die Grundgesamtheit machen oder Entscheidungen treffen, wie zum Beispiel wieviel Geld man im Monat durchschnittlich in einem bestimmten Haushalt ausgeben darf. Die Vorteile des Verfahrens sind die intuitive Verständlichkeit und der geringe Implementierungsaufwand. Das arithmetische Mittel wird deshalb als eine Methode in dieser Arbeit verwendet. Der große Nachteil der Methode ist, dass die Prognose der Schrittdauer zu ungenau sein kann. Die Methode ist gut für die Vorhersage der Schritte geeignet, die kleinere Abweichungen vom arithmetischen Mittel aufweisen. Bei den größeren Abweichungen der Schrittdauer liefert die Methode entsprechend größere Fehler. Näheres zur Ungenauigkeit wird im Kapitel Evaluation 5 gesagt. Zur Berechnung des arithmetischen Mittels der Schrittdauer werden die Daten benötigt. Dazu haben die Chirurgen des Uniklinikums der Stadt Heidelberg mit Hilfe von Annotationssoftware an einem menschlichen Phantom die Annotationen aufbereitet. Im Programm werden aus den Annotationen die benötigten Daten extrahiert und daraus zuerst die Dauer berechnet und dann das arithmetische Mittel für jeden Schritt der OP ermittelt. Das Maß dient dann als Basis zur Prognose. Im Kapitel Implementierung 4.1 wird die Implementierung und Vorgehensweise bei den Berechnungen genauer beschrieben. Die Evaluation dieser Methode wird im Kapitel Evaluation 5.2 erläutert.

## 3.2 Lineare Regression

Die lineare Regression [19] ist eine Konzeption, die bei der Regressionsanalyse benutzt wird. Zur Analyse wird die sogenannte Regressionsgerade verwendet. Bei einer solcher Analyse mit Hilfe einer Regressionsgerade wird versucht, eine Variable durch eine oder mehreren Variablen auszudrücken. Es wird eine Funktion ermittelt, die die Abhängigkeit der Variablen beschreibt. Die Voraussetzung für eine solche Analyse ist, dass die Variablen einen linearen Zusammenhang haben. Zum Beispiel wird die Variable  $Y = \text{systolischer Blutdruck}$  durch die Variablen  $X_1 = \text{Alter}$ ,  $X_2 = \text{Geschlecht}$ ,  $X_3 = \text{Rauchen}$  mit Hilfe der linearen Regression untersucht. Um einen linearen Zusammenhang von Variablen festzustellen, wird in dieser Arbeit der Korrelationskoeffizient [20] verwendet. Die Variablen haben einen linearen Zusammenhang, wenn der Korrelationskoeffizient ungleich null ist.

Die Idee ist, wenn die Schritte einen linearen Zusammenhang haben, die lineare Regression für die Prognose der Dauer zu nutzen. Im Unterkapitel Verwendeter Datensatz 5.1 des Kapitels Evaluation 5 werden die Schritte auf den linearen Zusammenhang untersucht. Es wurde festgestellt, dass alle Schritte einen kleinen oder großen Zusammenhang aufweisen. Das wurde auf Basis des Bravais-Pearson-Korrelationskoeffizienten [20] herausgefunden. Um bessere Ergebnisse in der Prognose zu erzielen, wurden bei der Implementierung die Schritte mit den besten Korrelationskoeffizienten herausgesucht. Aus den Daten für die zwei mit bestem Korrelationskoeffizient zusammenhängenden Schritten wurde eine Regressionsfunktion berechnet, die zur Prognose verwendet wurde. Problematisch ist es mit dem 1. Schritt, da er keinen Vorgänger hat. Das Problem wurde mit Hilfe der Durchschnittsmethode im Kapitel 3.1 gelöst. Die vorhergesagte Dauer des 1. Schrittes ist sein Durchschnitt. Die Kapitel 4.2 und Evaluation 5.3 beschreiben die genauere Vorgehensweise bei der Berechnung und der Evaluation und repräsentieren somit die Ergebnisse dieser Evaluation.

## 3.3 Random Forest Methoden

Ein Random Forest [11] ist ein Fachausdruck aus dem Bereich des maschinellen Lernens [17]. Es handelt sich dabei um ein Klassifizierungsverfahren [21]. Mehrere Entscheidungsbäume [22] sind ein Teil dieses Verfahrens. Die Entscheidungsbäume in diesem Verfahren

sind unkorreliert. Die Entscheidungsbäume werden durch einen Algorithmus randomisiert trainiert. Daher stammt der Name für diese Methode. Jeder Baum in diesem Wald darf eine Entscheidung treffen. Die entgültige Entscheidung gehört zur Klasse mit den meisten Stimmen. Mit dieser Methode werden sowohl Klassifizierungsprobleme als auch Regressionsprobleme gelöst [11]. In dieser Bachelorarbeit wird der Random Forest für das Lösen des Regressionsproblems eingesetzt, da die Dauer der Schritte in der OP nicht als endliche Menge von Objekten dargestellt werden kann.

### 3.3.1 Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

Zum Trainieren der Random Forests werden Informationen über die Schritte verwendet, wie die Korrelation zwischen den Schritten, ihre Dauer und Position in der Operation. Dazu wird ein Featurevektor aus diesen Informationen erstellt, der dem Random Forest übergeben wird. Ein Element dieses Vektors kann beispielsweise wie folgt aussehen: 2;67000;7;255000. Die Werte 2 und 7 geben dabei die Position der Schritte in der OP an und dienen als eindeutige IDs der Schritte. Die ID mit der Nummer 2 entspricht dem zweiten Schritt der Operation, "Mobilization sigmoid". Die ID mit der Nummer 7 entspricht dem siebten Schritt "Delineating of vessels". Mithilfe der Annotationen wurden diese Schritte zuvor als diejenigen identifiziert, die den größten absoluten Korrelationskoeffizienten und den besten linearen Zusammenhang aufweisen. Die Einträge 67000 und 255000 repräsentieren jeweils die Dauer des zweiten und siebten Schrittes in Millisekunden. Mit dem Featurevektor kann der Random Forest so trainiert werden, dass dieser die Dauer des siebten Schrittes vorhersagt. Die IDs der einzelnen Schritte wurden wie folgt festgelegt:

- ID 1 - Check for metastasis and adhesions
- ID 2 - Mobilization sigmoid
- ID 3 - Mobilization descending colon
- ID 4 - Mobilization splenic flexure
- ID 5 - Inspection colon
- ID 6 - Lancing retroperitoneum
- ID 7 - Delineating of vessels
- ID 8 - Division of artery
- ID 9 - Division of vein
- ID 10 - Opening of lesser pelvic peritoneum
- ID 11 - Dissection rectum
- ID 12 - Transect rectum
- ID 13 - Salvage rectum
- ID 14 - Visual inspection lesser pelvis.

Für fast jeden Schritt wird zuerst der Schritt herausgesucht, mit dem der beste lineare Zusammenhang besteht. Anschließend wird der Featurevektor konstruiert und damit der Random Forest für den entsprechenden Schritt trainiert. Die Ausnahme bildet der erste Schritt, da dieser keine vorherigen Schritte in der OP hat. Eine Zeile aus den Trainingsdaten für den ersten Schritt könnte folgendermaßen aussehen: 0;0;1;73272. Die führt jedoch

nicht zum Ziel, da der Random Forest für die Vorhersage nicht genügend Informationen enthalten wird. Deshalb fiel die Entscheidung bei der Vorhersage für Schritt 1 auf eine andere Methode. Es handelt sich um die in Kapitel 3.1 erläuterte Durchschnittsmethode. Im Teil Implementierung 4.3.1 und Evaluation 5.4.1 dieser Thesis wird erklärt, wie diese Methode implementiert und evaluiert wurde. Die Ergebnisse der Evaluation findet man in den vorgenannten Kapiteln.

### 3.3.2 Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

In diesem Unterkapitel handelt es sich wieder wie im vorherigen Unterkapitel um eine Methode mit dem Random Forest. In diesem Verfahren verwendet der Random Forest wieder die Informationen von den vorherigen Schritten. Der Unterschied zur vorherige Methode besteht darin, dass der Featurevektor zum Trainieren und zur Vorhersage anders konstruiert wird. Der Fachausdruck Historiegröße bedeutet im Zusammenhang mit dem Random Forest in dieser Bachelorarbeit die Information über die vorherige Schritte. Historiegröße 3 zum Beispiel bedeutet, dass der Featurevektor aus 3 IDs von den Schritten 3.3.1 und ihren Dauern besteht. Um den Random Forest zu trainieren und die Information für den siebten Schritt zu liefern, kann eine Zeile aus den Trainingsdaten der Historiegröße 3 folgendermaßen 5;33000;6;86000;7;255000 aussehen. Die Zahlen sind die IDs von den Schritten 3.3.1, die im vorherigen Kapitel erläutert wurden und der Rest ist die Dauer der Schritte in Millisekunden. In diesem Verfahren wurde der Random Forest mit den Historien von 1 bis 14 trainiert und evaluiert. Wenn es in der Historie noch keine Information über die vorherigen Schritten vorhanden ist, werden die Elemente des Featurevektors mit Nullen gefüllt. Folgendermaßen 0;0;0;0;1;55000;2;340000 kann zum Beispiel eine Zeile aus den Trainingsdaten bei der Historie 4 konstruiert werden. Obwohl die Nullwerte keine Information enthalten, sind sie zum Trainieren notwendig, weil das Programm ohne diese Werte einen Fehler auswerfen würde. Der Featurevektor hat immer eine gleiche Länge. Die Erläuterungen zur Implementierung und zu den Ergebnissen der Evaluierung entnimmt man dem Kapitel Implementierung 4.3.2 und Evaluation 5.4.2.

### 3.3.3 Korrelation zwischen Schritten und Features mit der Schrittdauer

Das Verfahren, in dem der Random Forest nur Informationen verwendet, wie die Korrelation zwischen den Schritten und ihre Dauer, ist dem Verfahren aus dem Unterkapitel Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation 3.3.1 sehr ähnlich. Der Featurevektor enthält nur numerische Werte die Dauer der Schritte. Die Vereinbarung welcher Schrittvorhersager trainiert wird, wird im Programmablauf festgehalten, damit die Informationen der Schritte nicht vermischt werden. Eine Zeile aus den Trainingsdaten 67000;77000 enthält zum Beispiel keine Informationen, wie welche Schritte die Dauer 67000 und 77000 Millisekunden haben. Die Informationen über den linearen Zusammenhang der Schritte fehlen auch in diesem Featurevektor. In dieser Methode werden diese fehlenden Angaben durch das Programm selbst geregelt. Die Informationen, wie so ein Programm implementiert wurde und die Evaluationsergebnisse sind detaillierter im Kapitel Implementierung 4.3.3 und Evaluation 5.4.3 erklärt.



## 4. Implementierung

In diesem Kapitel wird beschrieben, wie die Methoden aus dem Kapitel Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie 3 sowohl für das Assistenzsystem als auch die Evaluation programmiert wurde. Für das Live-System wird die beste Methode zur Prognose der Schrittdauer benötigt. Um die Methoden auf ihre Güte zu untersuchen und um die beste Methode festzustellen, wurde auch die Evaluation implementiert. In den nächsten Unterkapiteln wird auf die Programmierung jeder Methode eingegangen und sie wird detaillierter erklärt.

### 4.1 Durchschnittsmethode

In Abbildung 4.1 ist die Implementierung des Propheten, der die Durchschnittsmethode für die Prognose verwendet, durch ein Klassendiagramm beschrieben. Die Klasse Surgery und StepTimePredictor sind die Programmeinstiegspunkte, um die Dauer des bestimmten Schrittes zu prognostizieren. In dieser Arbeit wird zuerst für jede Annotation eine neue Instanz der Klasse Surgery gebildet. Dann wird eine Sammlung der Klasse ArrayList erzeugt. Zu der Sammlung der Klasse ArrayList werden Objekte der Klasse Surgery hinzugefügt, die vorher gebildet wurden. Die Vorhersage der Dauer erfolgt durch die Methode predict der Klasse StepTimePredictor. Die Instanz der Klasse (StepTimePredictor) bekommt bei der Initialisierung die Sammlung von Surgery-Objekten übergeben. Die Methode (predict) kann Werte vom Typ String annehmen. Sie liefert die sinnvolle vorhergesagte Dauer nur für die 14 Schritte ( Siehe 3 ) zurück. Sonst wird der Wert 0 geliefert. Die ganze Initialisierung kann in der Main-Methode von einer beliebigen Klasse vorgenommen werden. Dies könnte auch als Webservice realisiert werden. Im laufenden Assistenzsystem kann die Methode so erweitert werden, dass das arithmetische Mittel der Schrittdauer mit neuen, aus dem System gewonnenen Daten berechnet wird.

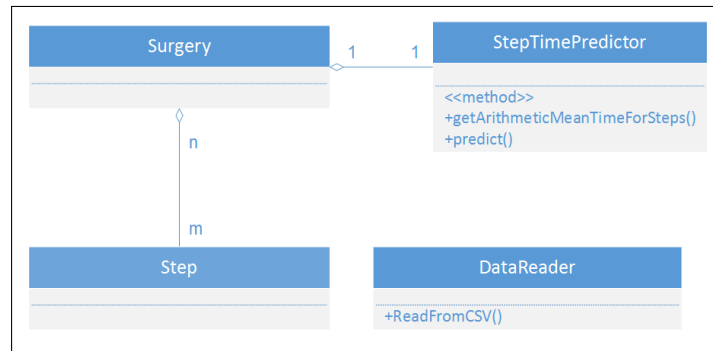


Abbildung 4.1: Klassendiagramm des Propheten auf Basis der Durchschnittsmethode

Abbildung 4.2 enthält das Klassendiagramm, das die Architektur für die Leave-One-Out-Kreuzvalidierung [23] auf Basis der Durchschnittsmethode beschreibt. Der Unterschied zum Klassendiagramm in Abbildung 4.1 liegt daran, dass dieses Klassendiagramm um 2 Klassen erweitert wurde, und zwar um die Klassen **StepsEvaluation** und **Main**. In der main-Methode der Klasse **Main** werden alle benötigten Initialisierungen vorgenommen. Die Klasse **StepsEvaluation** enthält die Methode `evaluate`. Durch sie wird die Evaluierung einer der je nachdem 13 oder 17 Operationen vorgenommen. Bei der Evaluierung werden die absoluten Abstände zwischen der evaluierten Schrittdauer und der vorhergesagten durch den Prophet Dauer des Schrittes gebildet. Die Abstände sind die sogenannten Fehlerwerte, die man für die Bildung des arithmetischen Mittels der Leave-One-Out-Kreuzvalidierung [23] benötigt. Dieser Wert dient der Beurteilung der Methode, die zur Prädiktion der Dauer verwendet wird.

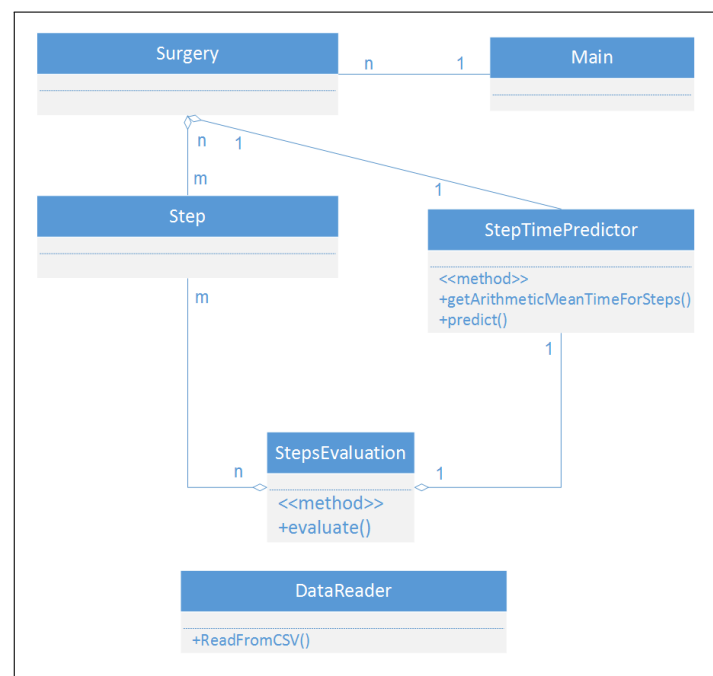


Abbildung 4.2: Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis der Durchschnittsmethode

## 4.2 Lineare Regression

Für die Implementierung dieser Methode, um die Regressionsfunktion zu berechnen, wurden die Klassen **RealMatrix**, **PearsonsCorrelation** und **SimpleRegression** von Apache Com-

mons Math Framework benutzt. Abbildung 4.3 beschreibt durch das UML-Klassendiagramm die Implementierung des Propheten. Die wichtigsten Klassen der Implementierung sind Step und CorrelationCoefficient. Das sind die Programmeinstiegspunkte. Die Klassen DataReader und Helper sind die Klassen, die Hilfsmethoden zur Verfügung stellen. Die CSV-Dateien mit Annotationen werden mit Hilfe der Klasse DataReader durch die Methode ReadFromCSV() gemacht. Viele Berechnungen, die zur Berechnung der Korrelationskoeffizienten und den Regressionsfunktionen der Schritte nötig waren, erfolgten durch die Methoden der Klasse Helper. Der Interface Comparable dient dazu, für jeden Schritt die berechneten Korrelationskoeffizienten zu sortieren. Die Ausnahme ist nur Schritt 1 (siehe 3.2). Nur für die Nachfolger wurden mit dem ersten Schritt die Korrelationskoeffizienten berechnet. In dieser Arbeit wurde die einfache lineare Regression [19] verwendet. Das heißt, um eine Vorhersage der Dauer des Schrittes zu treffen, muss man nur für jeden Schritt die Korrelationskoeffizienten mit seinen Vorgängern berechnen. Für die Prognose nimmt man dann den Schritt mit dem höchsten Korrelationskoeffizient. In der Arbeit wurde die Klasse CorrelationCoefficient so implementiert, dass das Objekt dieser Klasse genau zwei Schritten zugeordnet wird. Das Objekt der Klasse Step hat die Liste mit den Objekten der Klasse CorrelationCoefficient und die Möglichkeit die besten CorrelationCoefficienten auszugeben. Jedes Objekt der Klasse Step hat dann eine Regressionsfunktion. Die Verwendung der Regressionsgerade wird durch die Methode getDuration() geregelt. Um den Propheten im Assistenzsystem zu benutzen, müssen die Namen und die Dauer der vorherigen Schritten gespeichert werden und die Schrittpaare aus den Annotationen gewonnen werden. Im Live-System können die Schrittpaare, Korrelationskoeffizienten und die Regressionsfunktion durch neue Werte aktualisiert und neu berechnet werden.

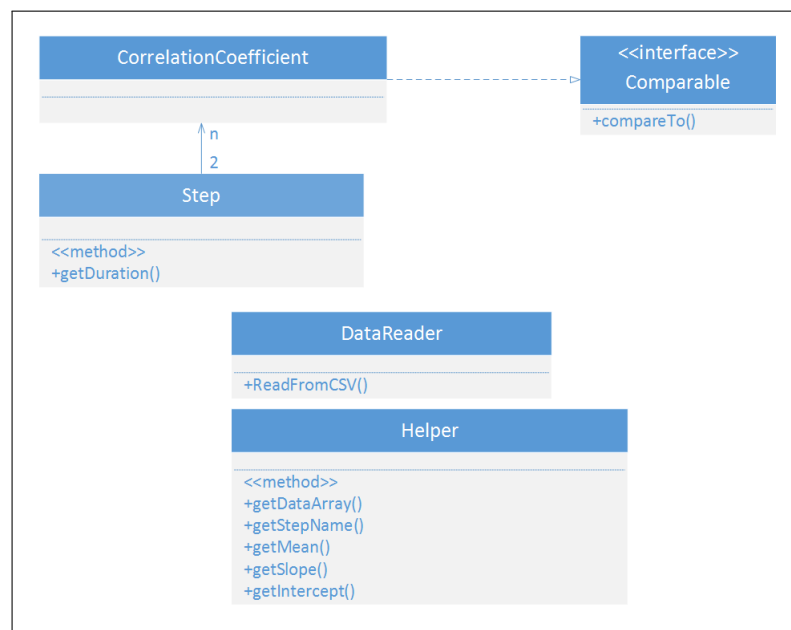


Abbildung 4.3: Klassendiagramm des Propheten auf Basis der linearen Regressionsmethode

In der Abbildung 4.4 wird die Implementierung der Leave-One-Out-Kreuzvalidierung [23] durch ein UML-Klassendiagramm dargestellt. Für die Evaluation wurde auch die Leave-One-Out-Kreuzvalidierung [23] wie bei der Implementierung der Durchschnittsmethode 4.1 verwendet. Der Unterschied zum Diagramm 4.3 besteht darin, dass das Programm um die zwei Klassen SurgeryEvaluation und Main erweitert wurde. Die Main Klasse ist der Einstiegspunkt für die Implementierung der Evaluation. In dieser Klasse, wie schon die Name verrät, befindet sich die Main-Methode. In der Main-Methode werden alle Initialisierungen ausgeführt. Die Klasse SurgeryEvaluation beinhaltet folgende Methoden:

- `getAllFault()`
- `getSurgeryDuration()`
- `getAllFaultRate()`
- `toString()`.

Das Objekt dieser Klasse wird für jede Annotation erzeugt, die schon im Teil zum Durchschnittsmethode 3.1 erwähnt wurde. Die Methode `getAllFault()` gibt den Wert für den Gesamtfehler der evaluierten Operation zurück. Die Dauer der evaluierten OP liefert die Methode `getSurgeryDuration()`. Die Methode `getAllFaultRate()` hat als Ausgabe den Wert der Gesamtfehlerquote. Die Ausgabe aller berechneten Werte der evaluierten Operation als String macht die Methode `toString()`. Der Wert Gesamtfehler ist die Summe alle absoluten Differenzen zwischen der vorhergesagten Dauer des Schrittes und seiner tatsächlichen Dauer. Die Gesamtfehlerquote ist der prozentuale Anteil der Gesamtfehler von der Dauer der evaluierten Operation. Die Ergebnisse der Evaluation werden noch im Unterkapitel Lineare Regression 5.3 im Kapitel Evaluation erläutert.

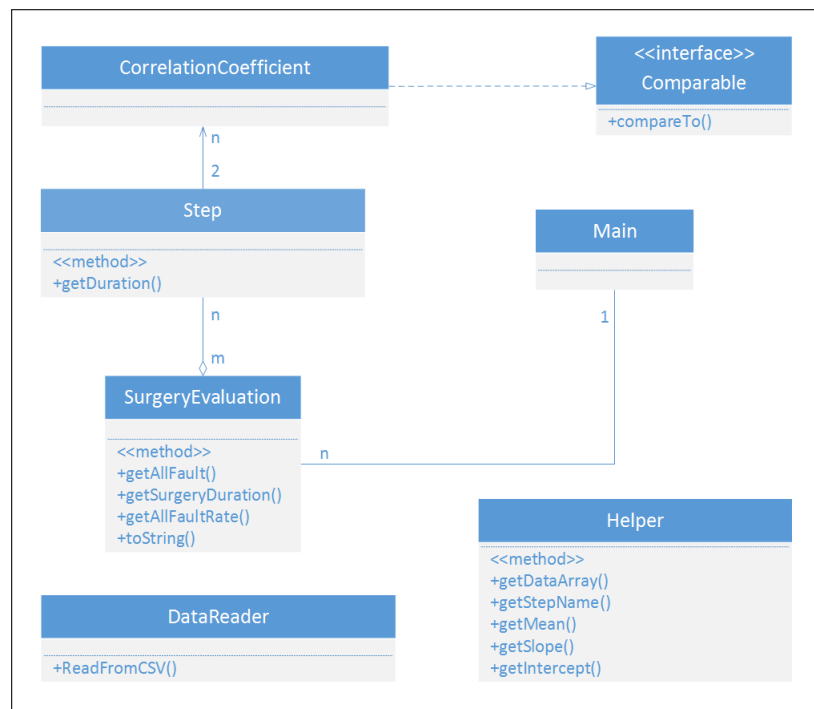


Abbildung 4.4: Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis der linearen Regressionsmethode

### 4.3 Random Forest Methoden

In diesem Unterkapitel werden verschiedene Implementierungen, die einen Random Forest mit verschiedenen Featurevektoren zum Trainieren verwenden, erläutert. Für die Implementierung dieser Methode, die eine Regressionsfunktion berechnet, wurden die Klassen `RealMatrix`, `PearsonsCorrelation` und `SimpleRegression` von Apache Commons Math Framework benutzt. Für Random Forest wurde die Programmbibliothek OpenCV verwendet.

Richtig eingestellte Parameter für den Random Forest sind wichtig, um eine gute Vorhersage für die Dauer des Schrittes zu erzielen. Es sind folgende Parameter:

- `max_depth`



Der Parameter gibt die maximale Tiefe des Baumes an. Wenn die Tiefe zu niedrig gesetzt wird, hat der Baum nicht genug Knoten, um die korrekte Antwort zu geben. Auf einen zu hohen Wert zu setzen, ist auch keine gute Idee. Deshalb werden in dieser Bachelorarbeit für den Parameter verschiedene Werte gesetzt, um ein besseres Ergebnis zu erzielen.

- `min_sample_count`

Der Parameter teilt dem Algorithmus die Mindestanzahl von Samples mit, die er enthalten muss bevor er den Baum zerlegt. Die OpenCV Dokumentation empfiehlt 1% der Gesamtdaten.

- `regression_accuracy`

Ist dieser Parameter gesetzt, behandelt der Algorithmus das Problem als Regressionsproblem. Der Parameter wurde mit dem kleinem Float Wert 0.00000001f initialisiert. Man konnte auch einen anderen kleinen Wert nehmen. Die Hauptsache ist, dass der Parameter nicht auf den Wert 0 gesetzt wird. Sonst rechnet der Algorithmus mit einem Klassifizierungsproblem.

- `use_surrogates`

Den Parameter verwendet man bei fehlenden Daten. Bei den Simulationen, bei denen der Parameter aktiviert ist, werden bessere Ergebnisse erzielt. In dieser Arbeit wird der Parameter auf den Wert false gesetzt.

- `calc_var_importance`

Mit dem Parameter werden die Gewichte der Exemplare im Baum gesetzt. Da in dieser Arbeit alle Exemplare gleich wichtig sind, wird der Parameter mit dem Wert false initialisiert.

- `nactive_vars`

Der Standardwert ist  $\sqrt{\text{Anzahl der Merkmale}}$  und wird durch die Initialisierung mit 0 erzielt.

- `term_crit`

Der Parameter ermöglicht, das Einlernen des Random Forest abubrechen. Durch diesen Parameter werden vor allem folgende Parameter eingestellt:

- Anzahl der Bäume

Der Parameter gibt die maximale Anzahl der Bäume im Random Forest an. Je mehr Bäume der Random Forest hat, desto besser sind die Ergebnisse. Die Vorhersage wird aber langsamer beim Erhöhen der Anzahl der Bäume. In dieser Arbeit wird der Parameter auf 100 Bäume eingestellt.

- Forest accuracy

Der Parameter gibt eine ausreichende Genauigkeit des OOB-Fehlers (Out of bag error). In der Arbeit wurde der Parameter auf den Wert 0.01f gesetzt.

Die Werte für die Parameter, die mit den unterschiedlichen Werten initialisiert werden, werden in den nächsten Kapitel genannt. Bei den Methoden, die im Random Forest die IDs der Schritte 3.3.1, also die kategorischen Werte, verwenden, muss man die Variable

varType vom Typ Mat auf 1 setzen. Wenn das nicht gemacht wird, wird der Random Forest alle Werte im Featurevektor als numerische Werte interpretieren und falsche Ergebnisse für die Vorhersage der Dauer der Schritte liefern. In den nächsten Unterkapitel wird auf die Implementierungen der Methode mit Random Forest eingegangen.

### 4.3.1 Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

In diesem Unterkapitel werden die UML-Klassendiagramme aus den Abbildungen 4.5 und 4.6 beschrieben. Das Klassendiagramm des Prophets auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes und ihre Position in der OP verwendet. Abbildung 4.5 beschreibt die Implementierung des Prophets, der schon im Kapitel 3.3.1 erläutert wurde. Der Kern dieses Propheten besteht aus den zwei Klassen Step und RandomForest. Die Klassen Helper und DataReader stellen nur die Hilfsmethoden zur Verfügung. Das Objekt der Klasse RandomForest enthält 14 Objekte der Klasse Step, da es 14 Schritte in der OP Rektumresektion gibt. Jeder Step hat genau einen Prophet(RandomForest-Objekt). Durch das Objekt der Klasse RandomForest wird mit Hilfe der Methode train() der Random Forest trainiert. Der Random Forest gibt nach dem Trainieren die Möglichkeit, die Vorhersage für jeden Schritt durch die Methode predict() zu geben. Die Methode getSecondStepNameWithBestCorrCoeff() aus der Klasse Step dient dazu, dem Propheten den besten linearen Zusammenhang der Schritte für das Trainieren mitzuteilen, um bessere Ergebnisse für die Vorhersage zu erzielen.

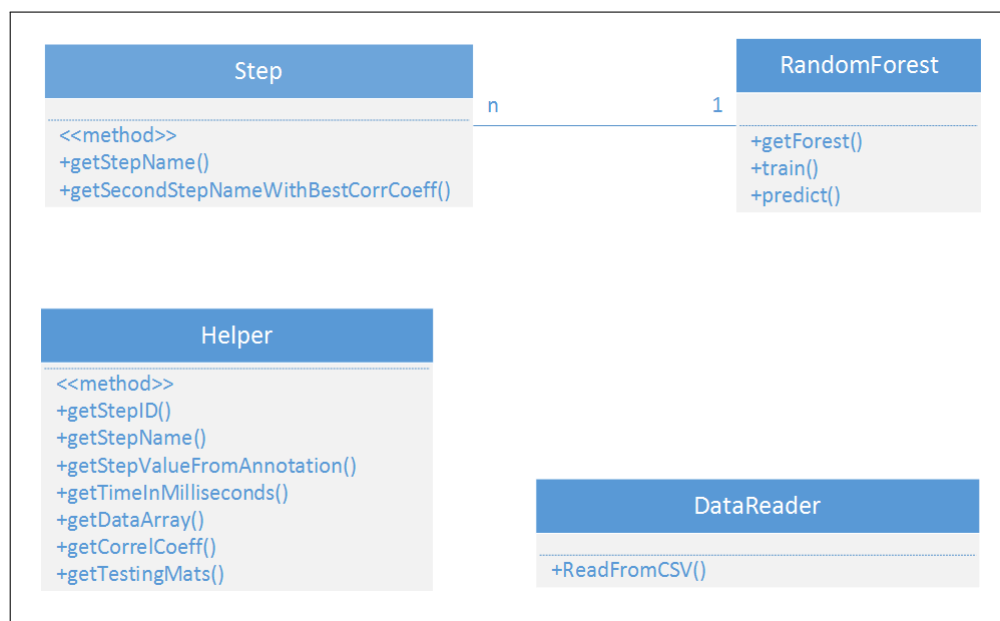


Abbildung 4.5: Klassendiagramm des Prophets auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes und ihre Position in der OP verwendet.

Die Implementierung der Leave-One-Out-Kreuzvalidierung wird im Klassendiagramm der Abbildung 4.6 beleuchtet. Der Unterschied zum Prophet ist, dass das Programm um den zwei Klassen EvaluateSurgery und Main erweitert wird. In der Klasse Main werden die Initialisierungen für die evaluierten Operationen durch die Bildung der Objekte der Klasse EvaluateSurgery vorgenommen. Jedes Objekt der Klasse EvaluateSurgery enthält eine Instanz der Klasse RandomForest, die wiederum die Objekte der Klasse Step enthält. Die Klasse EvaluateSurgery beinhaltet folgende Methoden:

- getSurgeryName()
- getSurgeryDuration()
- getAllFault()
- getAllFaultRate()
- toString().

Ein Objekt dieser Klasse wird für jede Annotation erzeugt, die schon im Teil Kapitel 3.1 erwähnt wurde. Die Methode `getSurgeryName()` liefert den Name der evaluierten Operation(Annotation). Die Dauer der evaluierten OP liefert die Methode `getSurgeryDuration()`. Die Methode `getAllFault()` gibt den Wert für den Gesamtfehler der evaluierten Operation zurück. Die Methode `getAllFaultRate()` hat als Ausgabe den Wert der Gesamtfehlerquote. Die berechneten Werte der evaluierten Operation als String hat die Methode `toString()` als Ausgabe. Der Wert Gesamtfehler ist die Summe alle absoluten Differenzen zwischen der vorhergesagten Dauer des Schrittes und seiner tatsächlichen Dauer. Die Gesamtfehlerquote ist der prozentuale Anteil der Gesamtfehler bezüglich der Dauer der evaluierten Operation.

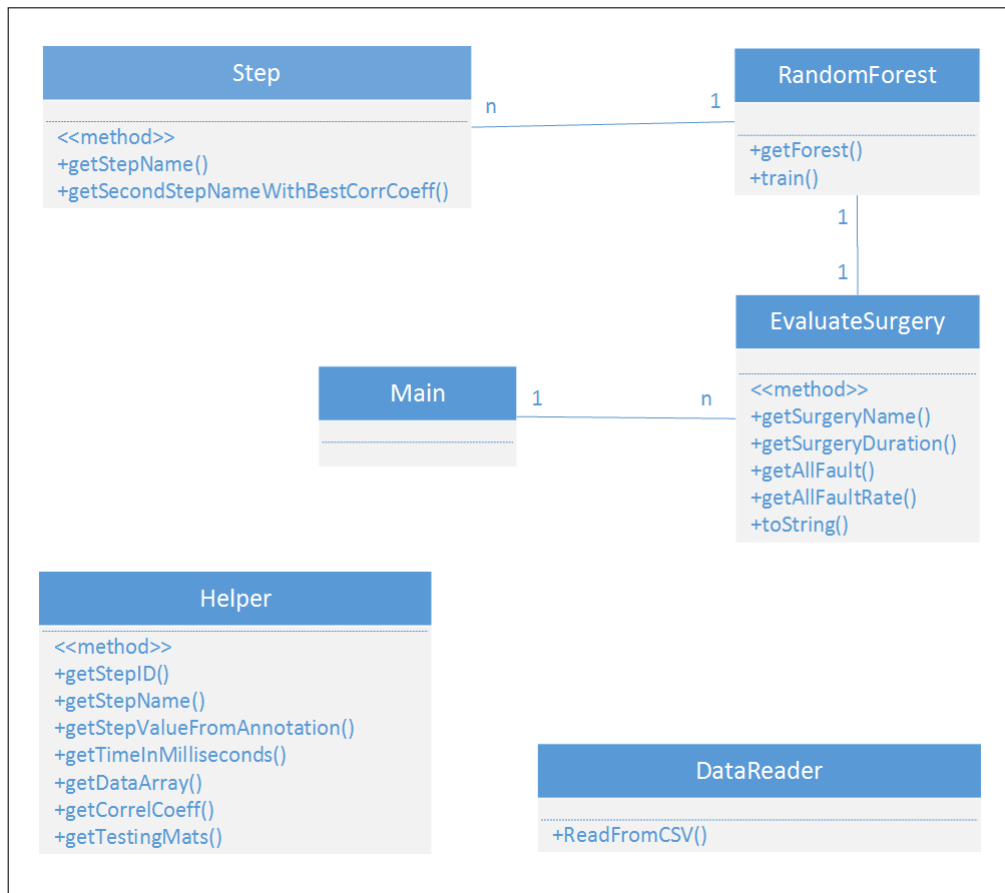


Abbildung 4.6: Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes und ihre Position in der OP verwendet.

Die Parameter des Random Forest werden mit den folgenden Werten sowohl für 13 als auch für 17 Annotationen initialisiert:

- `max_depth 12`

- min\_sample\_count 13
- nactive\_vars 2

Die Ergebnisse der Evaluation werden noch im Kapitel Evaluation im Unterkapitel 5.4.1 erläutert. In den nächsten Unterkapitel werden die Implementierungen weiterer Methoden mit Random Forest beschrieben.

#### 4.3.2 Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

Abbildung 4.7 veranschaulicht durch das UML-Klassendiagramm die Implementierung des Propheten, der zur Vorhersage der Schrittdauer einen Random Forest Algorithmus verwendet. Zum Trainieren verwendet er die Dauer von den vorherigen Schritten und ihre Reihenfolge in der Operation. Eine genauere Beschreibung, wie die Trainingsdaten aufgebaut sind, wurde im Unterkapitel 3.3.2 des Kapitels Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie beschrieben. Der Prophet besteht aus den drei Klassen PreparationCSVforRF, StepPrediction und DataReader. Die Methode makeTrainingCSV() erzeugt die Trainingsdaten im CSV-Format, mit denen StepPrediction durch die Methode train() trainiert wird. Die Trainingsdaten enthalten zeilenweise den Featurevektor und das Label. Mit Hilfe der Methode ReadFromCSV() werden die CSV-Dateien mit Trainingsdaten und Annotationsdaten geparkt und in den Methoden makeTrainingCSV() von der Klasse PreparationCSVforRF und train() von der Klasse StepPrediction verwendet. Durch die Methode predict() wird die Dauer der Schritten vorhergesagt. Um die Methode predict() zu verwenden, muss sie als Eingabe ein passendes Argument, also den Featurevektor, bekommen. Der Featurevektor wird deshalb zuvor im System in der passenden Form aufbereitet und dann an die Methode predict übergeben.

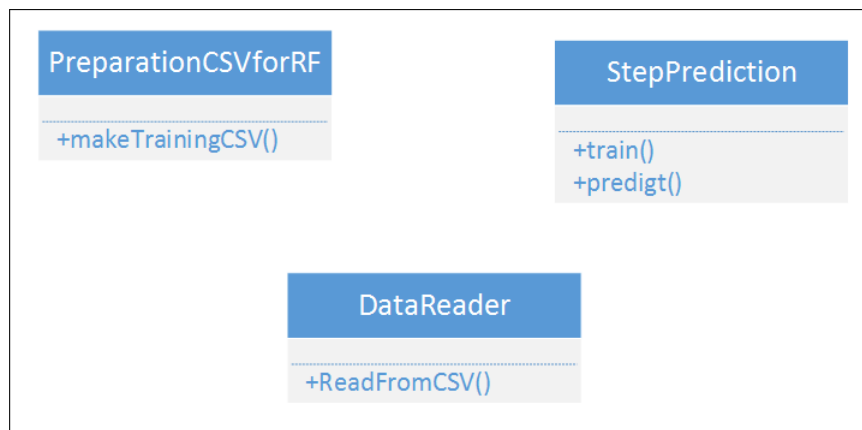


Abbildung 4.7: Klassendiagramm des Prophetn auf Basis des Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet.

Das Klassendiagramm aus der Abbildung 4.8 erklärt, wie die Leave-One-Out-Kreuzvalidierung für diese Methode implimentiert wurde. Die oben erläuterte Implementierung vom Propheten wird um die Klasse Main erweitert. In der Klasse StepPrediction werden zwei Methoden durch die Methode setup() ersetzt. In der Klasse Main in der Methode main werden zuerst die Trainingsdaten durch die Methode makeTrainingCSV() erzeugt, dann wird die Initialisierung von StepPrediction vorgenommen. Anschließend wird die Methode setup() von der Klasse StepPrediction aufgerufen, die als die Eingabe die Trainingsdaten und Featurevektoren hat. Die Parameter vom Random Forest werden mit den folgenden Werten sowohl für 13 als auch für 17 Annotationen initialisiert:

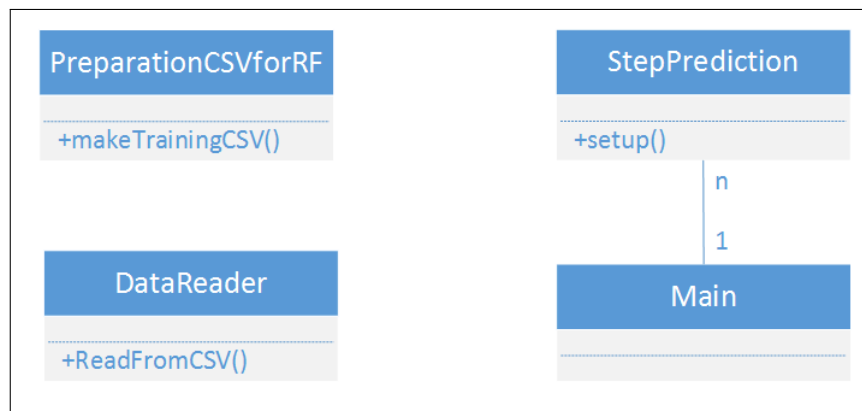


Abbildung 4.8: Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet.

- max\_depth 3
- min\_sample\_count 1
- nactive\_vars 0

Die Ergebnisse der Evaluation werden noch im Kapitel Evaluation im Unterkapitel 5.4.2 erläutert. Im nächsten Unterkapitel wird eine letzte Implementierung einer Methoden mit Random Forest beschrieben.

### 4.3.3 Korrelation zwischen Schritten und Features mit der Schrittdauer

In diesem Teil der Bachelorarbeit wird die Implementierung der Methode zur Dauervorhersage der Schritte veranschaulicht. Wie die Methode funktioniert, wurde schon im Kapitel Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie im Unterkapitel 3.3.3 beschrieben. Die Methode ist der Methode aus dem Unterkapitel 3.3.1 vom Kapitel Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie sehr ähnlich aufgebaut. Unterschiedlich sind hauptsächlich die Trainingsdaten. In dieser Methode wird zum Trainieren nur die Dauer der zusammenhängenden Schritte benutzt 3.3.3. Das Klassendiagramm 4.9 beschreibt die Implementierung des Propheten mit dem Random Forest. Der Random Forest in der Methode verwendet die Trainingsdaten, die die Dauer der zusammenhängenden Schritte in Millisekunden enthalten. Die Vorgehensweise der Implementierung ist gleich mit der Implementierung, die im Unterkapitel 4.3.1 beschrieben wurde.

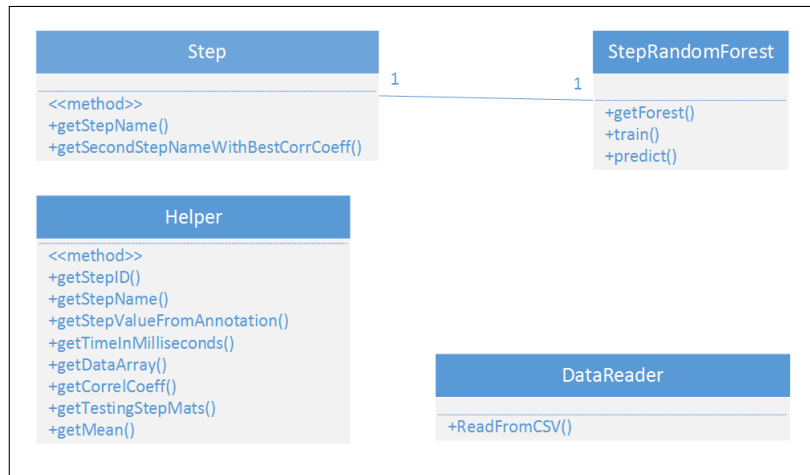


Abbildung 4.9: Klassendiagramm des Propheten auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes verwendet wird.

Die Abbildung 4.10 beleuchtet die Implementierung der Leave-One-Out-Kreuzvalidierung. Diese Implementierung verwendet wie der Prophet einen Random Forest, der auf Trainingsdaten arbeitet, die die Dauer der zusammenhängenden Schritte in Millisekunden enthalten. Die Vorgehensweise der Implementierung ist gleich mit der Implementierung aus Unterkapitel 4.3.1. Die Parameter vom Random Forest werden mit den folgenden Werten für 13

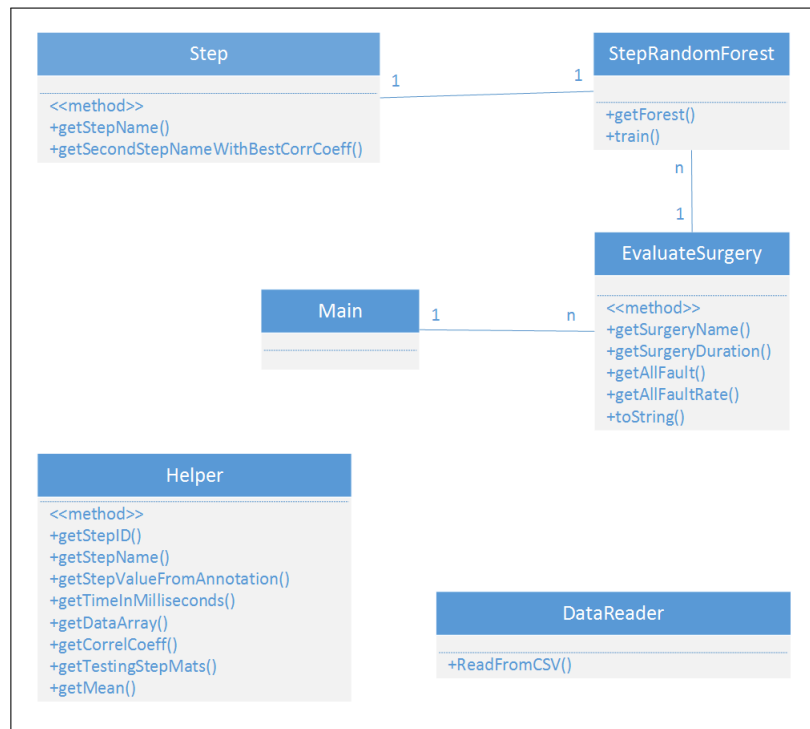


Abbildung 4.10: Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes verwendet.

Annotationen initialisiert:

- `max_depth` 12
- `min_sample_count` 13

- `nactive_vars` 2

und die Parameter des Random Forest werden mit den folgenden Werten für 17 Annotationen initialisiert:

- `max_depth` 10
- `min_sample_count` 20
- `nactive_vars` 5

Die Ergebnisse der Evaluation werden noch im Kapitel Evaluation im Unterkapitel 5.4.3 beschrieben.





## 5. Evaluation

Die Evaluation der in dieser Bachelorarbeit vorgestellten Methoden, die im Kapitel 3 beschrieben sind, findet durch die Leave-One-Out-Kreuzvalidierung statt. Mit der Leave-One-Out-Kreuzvalidierung [23] ist ein Verfahren zur Evaluierung gemeint, das analysiert, welche Methode bessere Ergebnisse liefert. Das Leave-One-Out-Kreuzvalidierungsverfahren ist eine einfache k-fache Kreuzvalidierung [23]. Dabei werden  $n$  Elemente genommen. Man nimmt man  $n-1$  Elemente zum Trainieren und ein  $n$ -tes zum Testen. Es werden bei der Analyse  $n$  Durchläufe gemacht, bis kein Element zum Evaluieren mehr übrig bleibt. Für jedes evaluierte Element wird ein Fehlerwert gebildet. Das Ergebnis ist das arithmetische Mittel, die sogenannte Gesamtfehlerquote. Der Fehlerwert in der Bachelorarbeit ergibt sich als absoluten Wert der Differenz zwischen der vorhergesagten Operationsschrittdauer und der evaluierten Dauer des Schrittes. Der Nachteil einer solchen Analyse besteht darin, dass die Methode einen sehr hohen Rechenaufwand bei großen Datensätzen hat und die Aufteilung der Teilmengen zum Testen und zum Trainieren während der Evaluation sehr schwer ist [23] [16]. Die Evaluation wird zuerst mit 13 Operationen und dann mit 17 Operationen durchgeführt. Wobei die erste 13 Operationen in den 17 Operationen enthalten sind. In den nächsten Abschnitten der Arbeit werden die Evaluationsmethoden näher erläutert. Der verwendete Datensatz wird im nächsten Unterkapitel 5.1 erläutert. Durch die Klassendiagramme im Kapitel Implementierung 4 und die genauere Beschreibung dieser Klassen wird die Leave-One-Out-Kreuzvalidierung deutlicher und verständlicher.

### 5.1 Verwendeter Datensatz

Am Anfang des Kapitels Vorhersage der Dauer chirurgischer Schritte in der Laparoskopie 3 wurde schon auf die benötigten Daten eingegangen und kurz erläutert. Bevor auf den verwendeten Datensatz eingegangen wird, bietet sich an dieser Stelle eine kurze Zusammenfassung an. Von Medizinern aus dem Uni-Klinikum Heidelberg wurde ein Rektumresektion auf einem menschlichen Phantom durchgeführt und annotiert. 21 Operationen wurden protokolliert. Die ersten vier Operationen wurden sicherheitshalber nicht bei der Evaluation berücksichtigt, da die Mediziner dies zu Trainingszwecken genutzt hatten. Die Daten liegen im CSV-Datenformat vor. Die Rektumresektion besteht aus 14 Schritten. Die Daten dieser Schritte aus den CSV-Dateien sind von Interesse in dieser Bachelorarbeit, da man die Dauer dieser Schritte prognostizieren will.

Um sich einen besseren Überblick über die Daten zu verschaffen, wurden zuerst das arithmetische Mittel [7], die Standardabweichung [24] und die Varianz [25] berechnet. Die Ergebnisse werden in der Tabelle 5.1 dargestellt. Aus der Tabelle kann man entnehmen, dass

einige Schritte kleine Abweichungen und andere Schritte wiederum große Abweichungen vom arithmetischen Mittel aufweisen. Der Wert der Standardabweichung lässt eine solche Schlussfolgerung zu. Durch eine solche Analyse kann man, sogar ohne Evaluation der implementierten Methoden, einige Schlussfolgerungen machen. Zum Beispiel könnte man die Schritte mit kleiner Standardabweichung könnte man durch die Durchschnittsmethode 3.1 4.1 gut vorhersagen. Der Schritt *check for metastasis and adhesions* weist ca. 10 Sekunden Abweichung vom Mittelwert auf. Im schlechtesten Fall wird der Schritt 10 Sekunden früher beendet oder wird noch 10 Sekunden dauern. Der Schritt *dissection rectum* weist mehr als 70 Sekunden Abweichung auf. Daher wird es bei der Verwendung der Durchschnittsmethode zur Prognose dieses Schrittes bei der Evaluation und auch im Echtzeitsystem zu großen Fehler kommen. Die Abbildung 5.2 stellt die Abweichungen der Schritte vom arithmetischen Mittel von zwei Operationen dar. Um die Abweichung für alle Schritte in einem Diagramm zu zeigen, wurde das arithmetische Mittel auf den Wert 0 normiert. Bei der Betrachtung der annotierten Operationen wurde festgestellt, dass die Abweichung vom arithmetischen Mittel sogar für gleiche Schritte willkürlich ist. Die Abweichung schwankt in negative und positive Richtung vom Mittelwert. In der Abbildung 5.2 wird das veranschaulicht. Zum Beispiel hat der Schritt *dissection rectum* in der OP 6 eine negative Abweichung und in der OP 13 eine positive Abweichung vom Mittelwert.

Die Schritte werden auch auf einen Zusammenhang hin analysiert. Da die Reihenfolge der Schritte eine Rolle spielt, stellt sich die Frage, ob der aktuelle Schritt in der OP mit den vorherigen Schritten zusammenhängt. In der Tabelle "Die OP-Schritte mit dem besten linearen Zusammenhang und ihre Korrelationskoeffizienten" 5.2 sind die Schritte mit dem besten Korrelationskoeffizient [20] aufgelistet. Der Korrelationskoeffizient ist ein Maß, das einen linearen Zusammenhang zwischen zwei Merkmalen beschreibt [20]. Durch Analyse auf einen linearen Zusammenhang wurde festgestellt, dass die Dauer von Schritten einen linearen Zusammenhang hat. Nur beim ersten Schritt *check for metastasis and adhesions* gibt es ein Problem mit dem Zusammenhang, weil der Schritt keinen vorherigen Schritt hat. Daher kam die Idee, sich so einen Zusammenhang bei der Prognose der Schrittdauer zu Nutze zu machen. Wenn es einen Zusammenhang zwischen Schritten gibt, dann sollte es so sein, dass es zu einem Schritt, der länger dauert, einen Nachfolger gibt, der auch länger dauert. Andersrum sollte es auch gehen. Zur Prognose sollte zuerst für jeden Schritt ein Schritt mit dem größten Korrelationskoeffizient gefunden werden. Die Tabelle "Die OP-Schritte mit bestem linearem Zusammenhang und ihre Korrelationskoeffizienten" 5.2 und die Abbildung "Der beste lineare Zusammenhang von zwei Schritten" 5.1 veranschaulichen das. In der Abbildung 5.1 wird ein Zusammenhang der Dauer vom Schritt *mobilization sigmoid* und dem Schritt *division of vein* dargestellt. Die Schritte haben einen Korrelationskoeffizient in Höhe von 0,75. Die Gleichung der Regressionsgerade [19], die in der Abbildung 5.1 gezeigt wird, wird in dieser Arbeit auch zur Prognose verwendet. Näheres dazu wird im Unterkapitel "Lineare Regression" 5.3 und im Kapitel "Implimentierung" im Unterkapitel "Lineare Regression" 4.2 beschrieben. Folglich wurde als Datensatz zum Trainieren die Dauer der Schritte verwendet, um zu Analysieren, Evaluieren und eine gute Methode zur Prediktion zu entwickeln. Im Folgenden werden die Methoden evaluiert und der Evaluationsvorgang näher beleuchtet.

	Mittelwert	Standardabweichung	Varianz
check for metastasis and adhesions	59769,23077	10430,38347	108792899,4
delineating of vessels	95461,53846	31543,90197	995017751,5
dissection rectum	196384,6154	72359,71942	5235928994
division of artery	132000	61374,01104	3766769231
division of vein	113000	52336,19133	2739076923
inspection colon	43923,07692	19605,57824	384378698,2
lancing retroperitoneum	60461,53846	20243,19595	409786982,2
mobilization descending colon	113000	41572,92018	1728307692
mobilization sigmoid	178538,4615	58972,76475	3477786982
mobilization splenic flexure	44461,53846	15370,37803	236248520,7
opening of lesser pelvic peritoneum	333076,9231	67594,33469	4568994083
salvage rectum	54615,38462	18623,96496	346852071
transect rectum	95769,23077	27124,0848	735715976,3
visual inspection lesser pelvis	37307,69231	15157,55519	229751479,3

Tabelle 5.1: Mittelwert, Standardabweichung und Varianz der Schrittdauer

Schritt 1	Schritt 2	Korrelationskoeffizient
check for metastasis and adhesions	mobilization sigmoid	-0,43
check for metastasis and adhesions	mobilization descending colon	0,41
mobilization sigmoid	mobilization splenic flexure	-0,28
mobilization sigmoid	inspection colon	0,69
mobilization descending colon	lancing retroperitoneum	-0,47
check for metastasis and adhesions	delineating of vessels	0,53
mobilization descending colon	division of artery	-0,39
mobilization sigmoid	division of vein	0,75
delineating of vessels	opening of lesser pelvic peritoneum	0,46
division of vein	dissection rectum	0,71
check for metastasis and adhesions	transect rectum	0,3
dissection rectum	salvage rectum	0,55
mobilization splenic flexure	visual inspection lesser pelvis	-0,51

Tabelle 5.2: Die OP-Schritte mit bestem linearen Zusammenhang und ihre Korrelationskoeffizienten

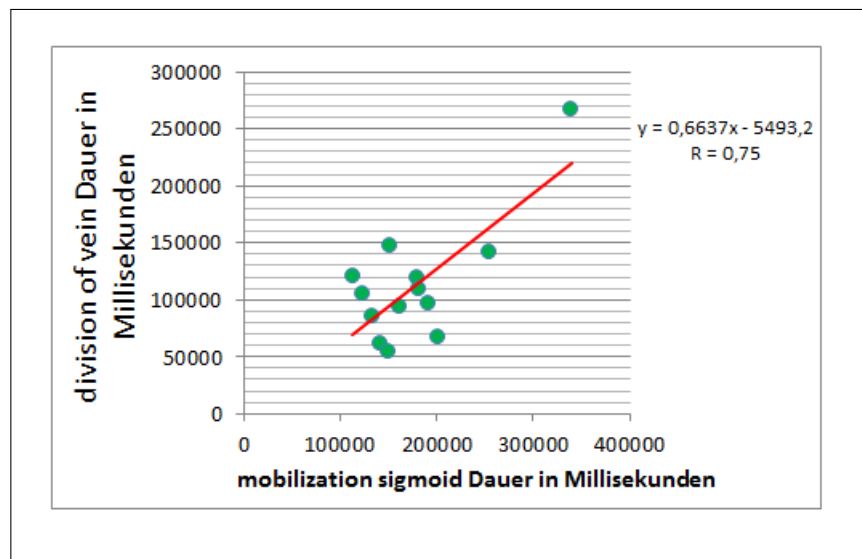


Abbildung 5.1: Der besten linearen Zusammenhang von zwei Schritten

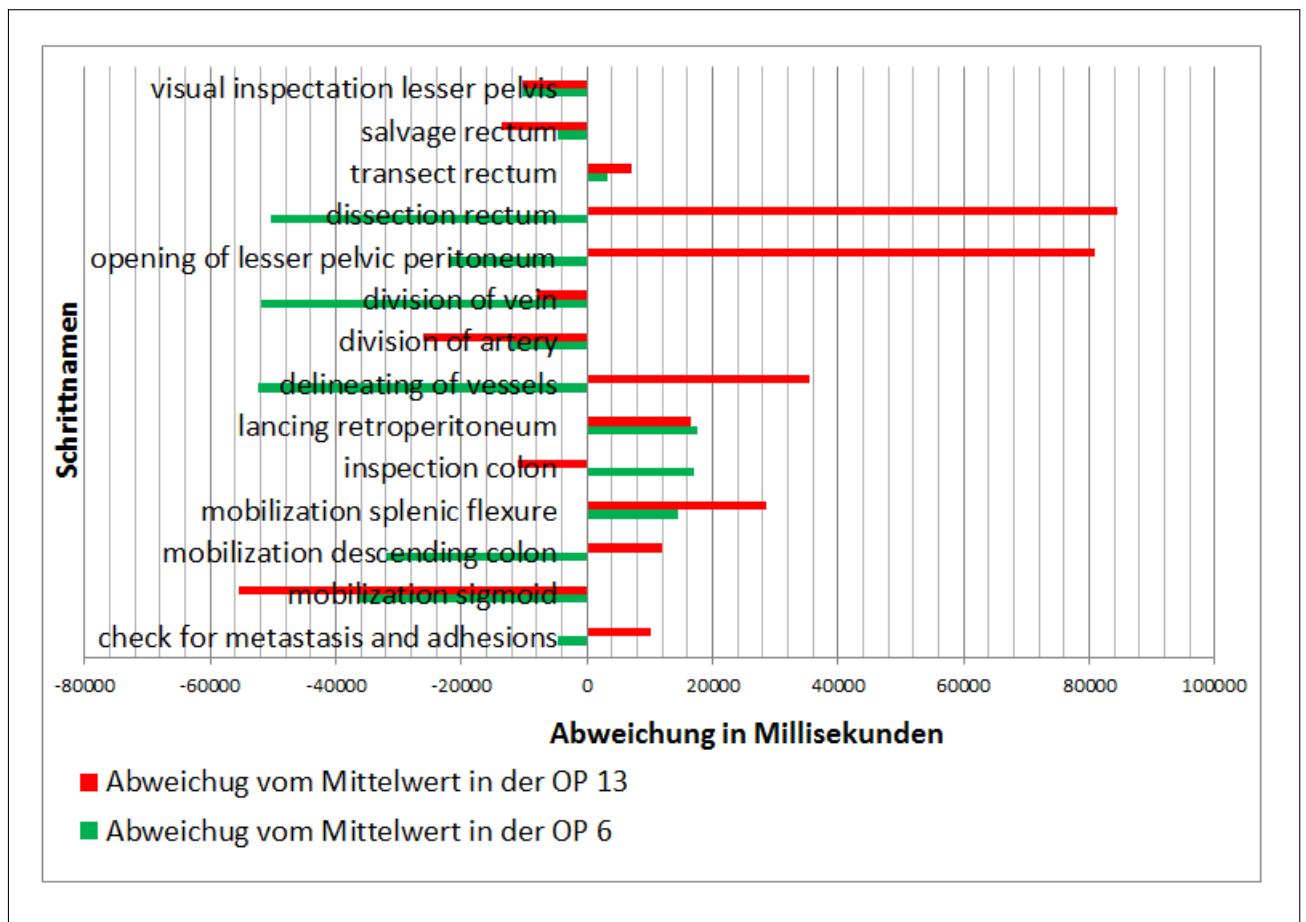


Abbildung 5.2: Abweichung vom Mittelwert in der OP 6 und OP 13 in Millisekunden

## 5.2 Durchschnittsmethode

Die Evaluation der Durchschnittsmethode wurde, wie schon oben erwähnt, mit der Lieve-One-Out-Kreuzvalidierung [23] gemacht. Die Evaluation wurde in zwei Etappen durchgeführt. Zuerst wurden 13 Operationen und dann noch 17 Operationen evaluiert. Wobei die

ersten 13 Operationen ein Teil der 17 Operationen bei der Evaluation sind. Die Tabellen 5.3, 5.4 und die Abbildung 5.5 geben einen gesamten Überblick über die Evaluation dieses Verfahrens. In den Tabellen in der ersten Spalte steht die Operation (Annotation), die evaluiert wurde. Die Spalten Gesamtfehler, Dauer der Test OP und Gesamtfehlerquote stellen die Ergebnisse der Evaluation dar. Der Gesamtfehler einer Operation ist die Summe von allen absoluten Abweichungen zwischen dem tatsächlichen und dem vorhergesagten Wert für die Dauer eines Schrittes. Die Dauer der Schritte wird in Millisekunden berechnet. Die Gesamtfehlerquote ist der Prozentsatz der Gesamtfehler zur Dauer der evaluierten Operation. Zur Repräsentation der Ergebnisse wurde die Evaluation von zwei Operationen ausgewählt, die das schlechteste und das beste Ergebnis liefert 5.3 5.4. Das durchschnittliche Ergebnis kann man in der letzten Zeile "Durchschnitt" der Tabellen 5.3 und 5.4 betrachten.

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	753333.33	2163000	34.83%
OP 2	329416.67	1728000	19.06%
OP 3	380916.67	1646000	23.14%
OP 4	379250	1674000	22.66%
OP 5	478083.33	1614000	29.62%
OP 6	357000	1333000	26.78%
OP 7	407666.67	1497000	27.23%
OP 8	438500	1371000	31.98%
OP 9	340000	1501000	22.65%
OP 10	407500	1359000	29.99%
OP 11	373333.33	1363000	27.39%
OP 12	435666.67	1293000	33.69%
OP 13	433333.33	1709000	25.36%
Durchschnitt	424153.85	1557769.23	27.23%

Tabelle 5.3: Evaluation mit 13 Annotationen Durchschnittsmethode

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	745750	2163000	34.48%
OP 2	338437.5	1728000	19.59%
OP 3	367562.5	1646000	22.33%
OP 4	404937.5	1674000	24.19%
OP 5	466687.5	1614000	28.91%
OP 6	314375	1333000	23.58%
OP 7	398250	1497000	26.6%
OP 8	410625	1371000	29.95%
OP 9	337500	1501000	22.49%
OP 10	387375	1359000	28.5%
OP 11	366750	1363000	26.91%
OP 12	407375	1293000	31.51%
OP 13	430250	1709000	25.18%
OP 14	407687.5	1784000	22.85%
OP 15	423812.5	1270000	33.37%
OP 16	378812.5	1320000	28.7%
OP 17	360437.5	1258000	28.65%
Durchschnitt	408625	1522529.41	26.84%

Tabelle 5.4: Evaluation mit 17 Annotationen Durchschnittsmethode

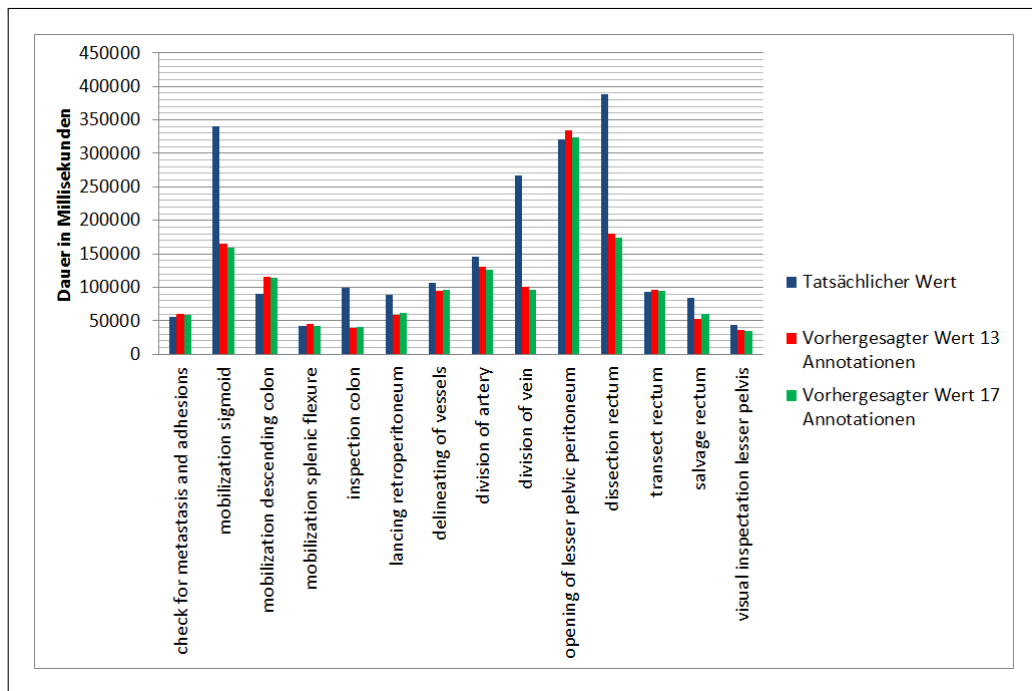


Abbildung 5.3: Evaluation der OP 1 mit der Durchschnittsmethode

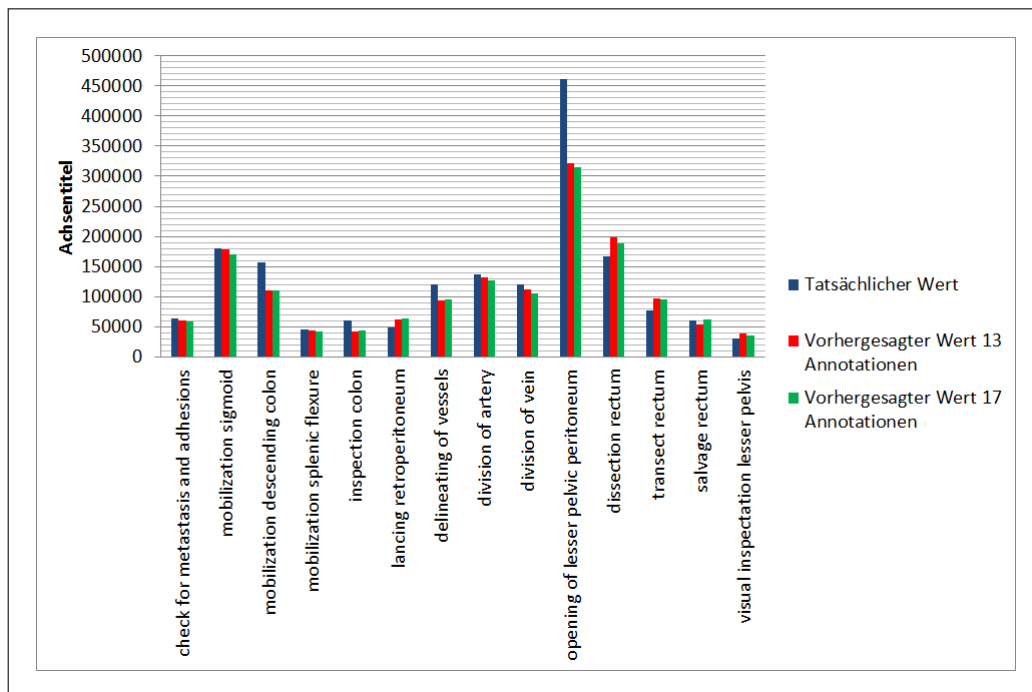


Abbildung 5.4: Evaluation der OP 2 mit der Durchschnittsmethode

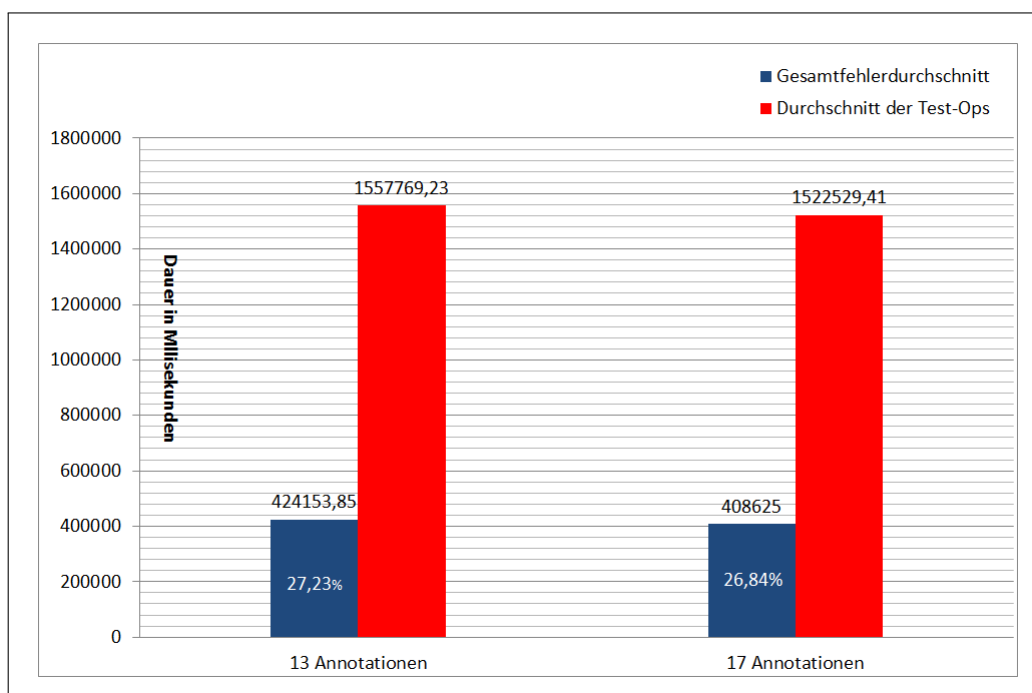


Abbildung 5.5: Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerquote mit der Durchschnittsmethode

### 5.3 Lineare Regression

Die Linearregressionsmethode wird auch wie die andere Methoden in dieser Bachelorarbeit durch Lieve-One-Out-Kreuzvalidierung [23] evaluiert. Die Evaluierung wurde auf 13 und dann auf 17 Annotationen durchgeführt. Die Tabellen 5.5–5.6 enthalten die Evaluationsergebnisse sowie den Gesamtfehler und die Gesamtfehlerquote. Zur besseren Übersicht enthalten diese Tabellen noch eine Spalte, um die Dauer der evaluierten Operation zu zeigen. Das schlechteste Ergebnis liefert die Operation 1 sowohl bei der Evaluation auf 13 als auch auf 17 Annotationen. Das schlechteste Ergebnis bei der Evaluation auf 13 Annotation beträgt 53.24% und auf 17 Annotationen 37.84% (Gesamtfehlerquote). Das beste Ergebnis wurde auch bei der Evaluation der Operation OP 2 sowohl bei der Evaluation auf 13 als auch auf 17 Annotationen erzielt. Das Ergebnis beträgt 21.82% für die Evaluation auf 13 Annotationen und 21.08% für die Evaluation auf 17 Annotationen.

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	1151492.64	2163000	53.24%
OP 2	377014.97	1728000	21.82%
OP 3	452351.33	1646000	27.48%
OP 4	526225.03	1674000	31.44%
OP 5	574159.44	1614000	35.57%
OP 6	503696.35	1333000	37.79%
OP 7	435057.22	1497000	29.06%
OP 8	539698.88	1371000	39.37%
OP 9	443868.08	1501000	29.57%
OP 10	360171	1359000	26.5%
OP 11	372008.94	1363000	27.29%
OP 12	570164.2	1293000	44.1%
OP 13	409602.19	1709000	23.97%
Durchschnitt	516577.71	1557769.23	33.16%

Tabelle 5.5: Evaluation mit 13 Annotationen Lineare Regression

Diese Ergebnisse werden in der Abbildung 5.6 und 5.7 grafisch veranschaulicht. Man sieht, dass die OP 1 bei mehreren Schritten große Abweichungen vom tatsächlichen Wert aufweist und die OP 2 wiederum nur bei einem Schritt eine große Abweichung hat.

Durch die Abbildungen werden die schlechtesten und die besten Ergebnisse erklärt. Die Verallgemeinerung der Ergebnisse dieser Methode erfolgt durch die letzte Zeile der Tabellen 5.5–5.6 und die Abbildung 5.8. Der Durchschnitt der Gesamtfehlerquote wird im Unterkapitel 5.5 sehr wichtig. In dieser Bachelorarbeit ist es ein sinnvolles Maß, um die Methoden zu bewerten. Der Durchschnitt der Gesamtfehlerquote beträgt bei der Evaluation auf 13 Annotationen 33.16% und auf 17 Annotationen 28.79%. Der Durchschnitt des Gesamtfehlers beträgt 516577.71 ms.(13 Annotationen) und 438381.42 ms.(17 Annotationen). Abbildung 5.8 veranschaulicht, dass bei mehreren Trainingsdaten(Annotationen) ein besseres Ergebnis erzielt wird.



	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	818499.3	2163000	37.84%
OP 2	364268.8	1728000	21.08%
OP 3	347312.43	1646000	21.1%
OP 4	495041.97	1674000	29.57%
OP 5	568868.45	1614000	35.25%
OP 6	349678.28	1333000	26.23%
OP 7	445968.97	1497000	29.79%
OP 8	444530.45	1371000	32.42%
OP 9	437009.77	1501000	29.11%
OP 10	352955.03	1359000	25.97%
OP 11	403748.75	1363000	29.62%
OP 12	364688.32	1293000	28.2%
OP 13	405708.09	1709000	23.74%
OP 14	414265.61	1784000	23.22%
OP 15	367530	1270000	28.94%
OP 16	475054.59	1320000	35.99%
OP 17	397355.34	1258000	31.59%
Durchschnitt	438381.42	1522529.41	28.79%

Tabelle 5.6: Evaluation mit 17 Annotationen Lineare Regression

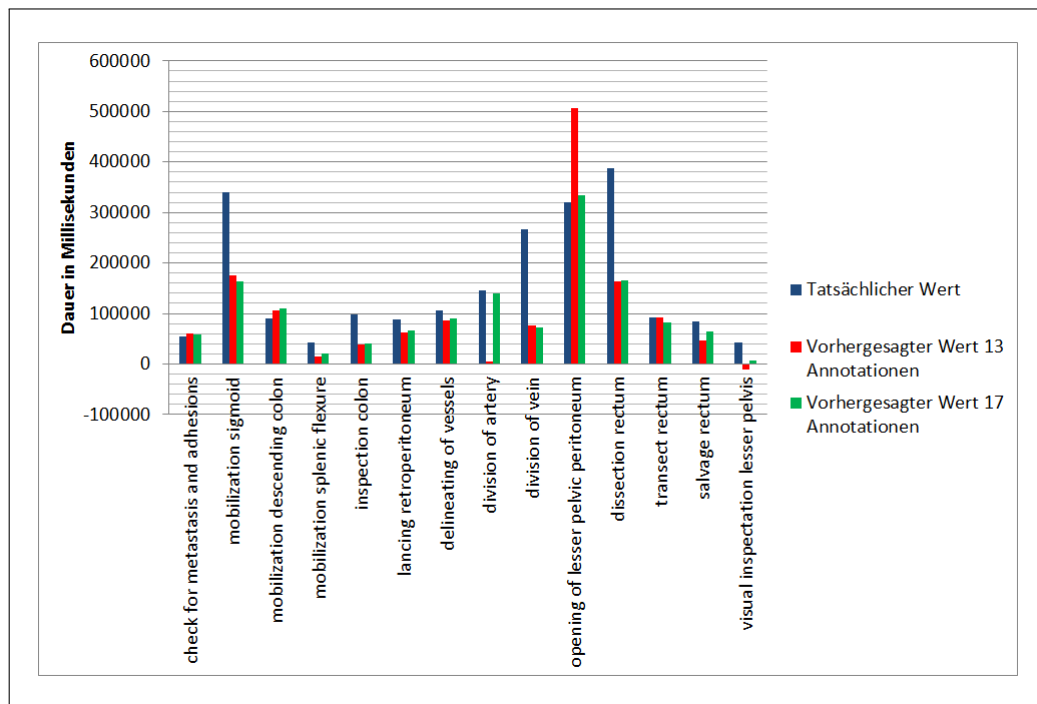


Abbildung 5.6: Evaluation der OP 1 mit der Linearregressionsmethode

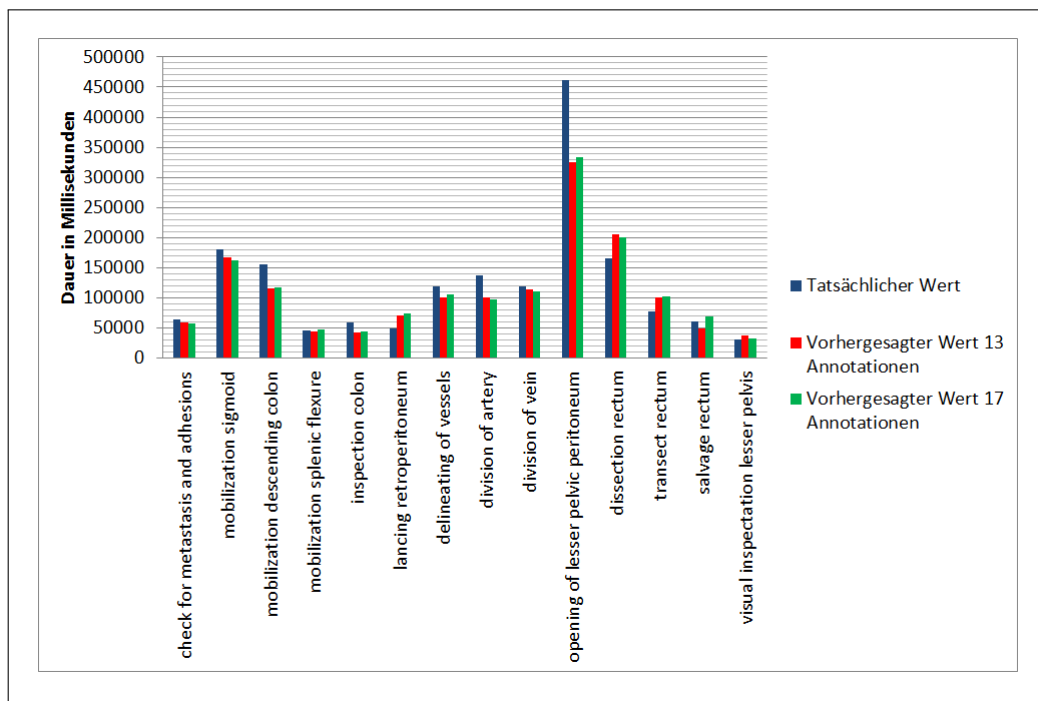


Abbildung 5.7: Evaluation der OP 2 mit der Linearregressionsmethode

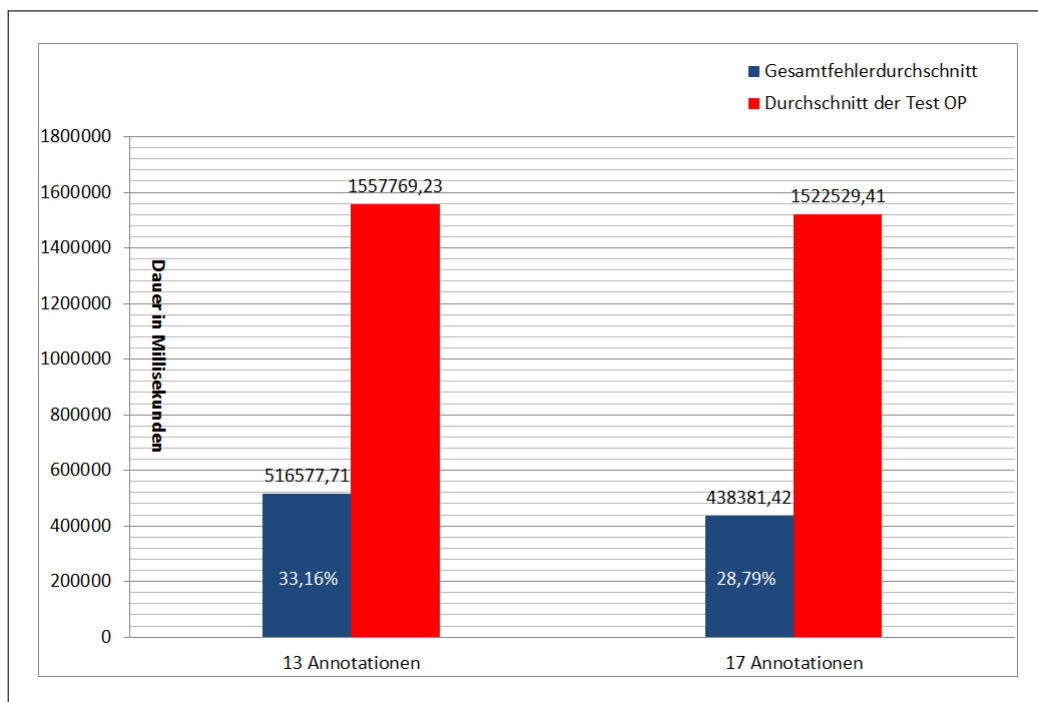


Abbildung 5.8: Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerquote mit der Linearregressionsmethode

## 5.4 Random Forest Methoden

In diesem Unterkapitel werden die Evaluationsergebnisse der Methoden, die den Random Forest verwenden, repräsentiert.

### 5.4.1 Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

In diesem Unterkapitel werden in den Tabellen 5.7 und 5.8 die Ergebnisse der Evaluation der Methode, die den Random Forest vorgstellt. Zum Trainieren des Random Forest werden Informationen der Schritten verwendet, wie z.B. die Korrelation zwischen den Schritten, ihre Dauer und ihre Position in der Operation. Es wird wieder die Lieve-One-Out-Kreuzvalidierung [23] benutzt. Der Random Forest wird zuerst mit 13 Annotationen und dann mit 17 Annotationen trainiert. Die Tabellen enthalten den Gesamtfehler, die Dauer der evaluierten Operation und die Gesamtfehlerquote. Die letzte Zeile der Tabellen enthält den jeweiligen Durchschnitt aller Werte. Der Durchschnitt der Gesamtfehlerquote wird in dem Unterkapitel Vergleich aller Methoden benutzt, um die in dieser Arbeit implementierte Methoden zu bewerten. Die schlechtesten Ergebnisse liefert sowohl von 13 Annotationen als auch von 17 Annotationen die Operation 1. Die besten Ergebnisse liefert sowohl von 13 Annotationen als auch von 17 Annotationen die Operation 2. Die schlechteste Gesamtfehlerquote für die Evaluation mit 13 Annotationen beträgt 37.24% und 36.54% mit 17 Annotationen. Die beste Gesamtfehlerquote für die Evaluation mit 13 Annotationen beträgt 21.04% und 20.46% mit 17 Annotationen. Der Durchschnitt der Gesamtfehlerquote für die 13 Annotationen ist 28.34% und 27.59% für die 17 Annotationen.

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	805410.59	2163000	37.24%
OP 2	363638.2	1728000	21.04%
OP 3	363044.63	1646000	22.06%
OP 4	462144.69	1674000	27.61%
OP 5	527257.59	1614000	32.67%
OP 6	391969.76	1333000	29.41%
OP 7	411880.76	1497000	27.51%
OP 8	456451.68	1371000	33.29%
OP 9	332861.61	1501000	22.18%
OP 10	380435.25	1359000	27.99%
OP 11	368476.93	1363000	27.03%
OP 12	470875.21	1293000	36.42%
OP 13	409224.14	1709000	23.95%
Durchschnitt	441820.85	1557769.23	28.34%

Tabelle 5.7: Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet.

Abbildung 5.9 veranschaulicht die Evaluation der Methode, bei der schlechteste Ergebnisse erzielt wurde. Die besten Ergebnisse werden durch die Abbildung 5.10 gezeigt. Aus den Abbildungen sieht man, warum die Evaluation beste oder schlechteste Ergebnisse liefert. Die Evaluation der Methode, die die schlechtesten Ergebnisse liefert, weist bei mehreren Schritten die größte Abweichungen von den tatsächlichen Werten auf.

Die Abbildung 5.11 zeigt, dass die Ergebnisse bei der Verwendung mehrere Annotationen um fast 1% verbessert werden. Die Verbesserung wird schon bei der Erhöhung der Trainingsdaten um vier Annotationen erzielt.

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	790284.75	2163000	36.54%
OP 2	353546.68	1728000	20.46%
OP 3	338661.16	1646000	20.57%
OP 4	505270.19	1674000	30.18%
OP 5	511478.91	1614000	31.69%
OP 6	323346.16	1333000	24.26%
OP 7	403435.2	1497000	26.95%
OP 8	443671.03	1371000	32.36%
OP 9	342699.87	1501000	22.83%
OP 10	344314.18	1359000	25.34%
OP 11	417392.16	1363000	30.62%
OP 12	376815.77	1293000	29.14%
OP 13	387658.83	1709000	22.68%
OP 14	523627.62	1967000	26.62%
OP 15	348313.35	1299000	26.81%
OP 16	456410.05	1450000	31.48%
OP 17	393226.44	1286000	30.58%
Durchschnitt	427067.79	1544294.12	27.59%

Tabelle 5.8: Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet.

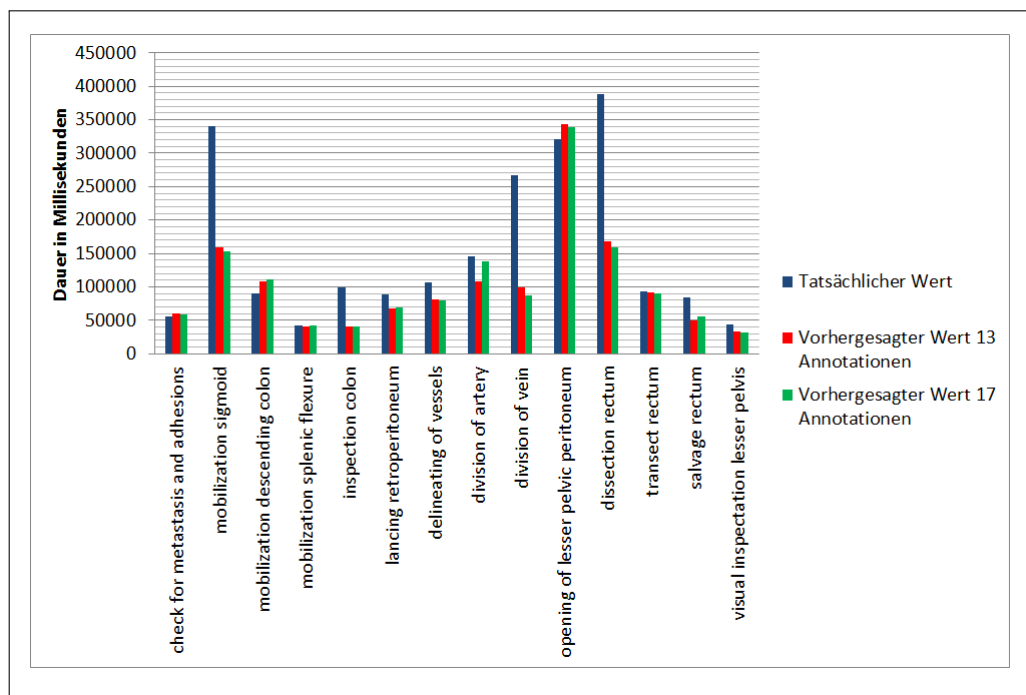


Abbildung 5.9: Evaluation der OP 1 mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet.

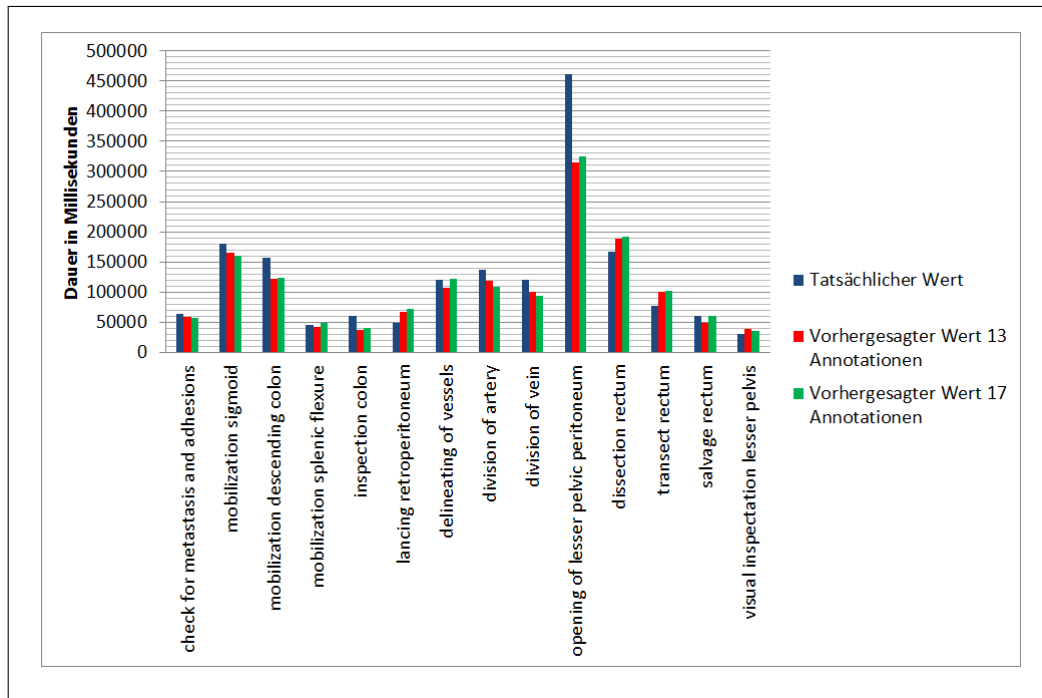


Abbildung 5.10: Evaluation der OP 2 mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet.

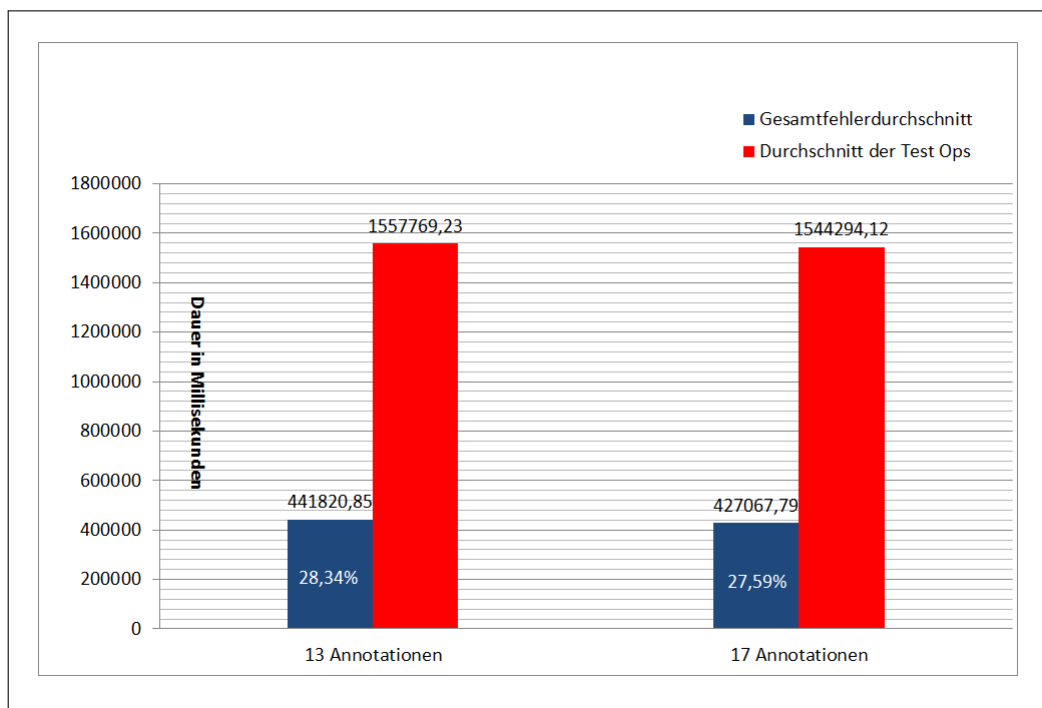


Abbildung 5.11: Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerquote mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet.

### 5.4.2 Historiegröße und Features mit der Schrittdauer und den Schrittpositionswerten in der Operation

Das Unterkapitel veranschaulicht die Ergebnisse der Evaluation der Methode, die im Teil 3.3.2 der Bachelorarbeit erklärt und erläutert wurde. Es wird wieder die Lieve-One-Out-Kreuzvalidierung [23] benutzt. Die Tabelle 5.9 enthält die Ergebnisse der Methode, für die 13 Annotationen zum Trainieren verwendet wurde. Tabelle 5.10 enthält die Ergebnisse der Evaluation durch das Trainieren des Propheten mit 17 Annotationen. Die Namen der zweiten und der dritten Spalten der Tabellen haben gleiche Bedeutung wie in den vorherigen Teilen der Bachelorarbeit. Die Spalte Historiegröße enthält die Anzahl der Schritte, die die Dauer für eine Zeile aus den Trainingsdaten verwenden.

Die beste Gesamtfehlerquote beträgt 26.06% mit den 13 Annotationen und 26.36% mit den 17 Annotationen. Die schlechteste Gesamtfehlerquote hat 29.03% mit 13 Annotationen und 28.31% mit 17 Annotation.

Historiegröße	Gesamtfehlerdurchschnitt	Gesamtfehlerquote
1	423264.5385	27.1711965%
2	452246.1538	29.03165277%
3	429260.6154	27.55611082%
4	427612.9231	27.45033827%
5	421604.7692	27.06464867%
6	429225.8462	27.55387883%
7	431506.8462	27.70030617%
8	440910.3077	28.30395537%
9	433406.0769	27.82222608%
10	437048.4615	28.05604663%
11	437277.8462	28.07077183%
12	441573	28.34649648%
13	438667.9231	28.16000693%
14	438970.0769	28.1794035%

Tabelle 5.9: Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet wird.

Historiegröße	Gesamtfehlerdurchschnitt	Gesamtfehlerquote
1	410030.9412	26.93090449%
2	431067.0588	28.3125604%
3	408041.0588	26.80020866%
4	407418.5294	26.75932082%
5	401361.4118	26.36148827%
6	411372.8235	27.01903956%
7	412975.7647	27.12432101%
8	417659.7647	27.43196696%
9	411111.6471	27.00188544%
10	418844.6471	27.50979024%
11	416607.1176	27.36282892%
12	422303.8824	27.73699343%
13	418662.2353	27.4978094%
14	422401.5882	27.74341077%

Tabelle 5.10: Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet wird.

### 5.4.3 Korrelation zwischen Schritten und Features mit der Schrittdauer

In diesem Unterkapitel werden in den Tabellen 5.11 und 5.12 die Ergebnisse der Evaluation der Methode, die den Random Forest repräsentiert. Zum Trainieren vom Random Forest werden Informationen von den Schritten über die Korrelation zwischen den Schritten und ihre Dauer verwendet. Es wird wieder die Lieve-One-Out-Kreuzvalidierung [23] benutzt. Der Random Forest wird zuerst mit 13 Annotationen und dann mit 17 Annotationen trainiert. Die Tabellen enthalten den Gesamtfehler, die Dauer der evaluierten Operation und die Gesamtfehlerquote. Die letzten Zeilen der Tabellen enthalten den jeweiligen Durchschnitt aller Werte. Der Durchschnitt der Gesamtfehlerquote wird in dem Unterkapitel Vergleich aller Methoden benutzt, um die in dieser Arbeit implementierten Methoden zu bewerten. Die schlechtesten Ergebnisse liefert die Evaluation für Operation 1 mit 13 Annotationen und 17 Annotationen. Die besten Ergebnisse liefert die Evaluation für Operation 2 mit 13 Annotationen und 17 Annotationen. Die schlechteste Gesamtfehlerquote für die Evaluation mit 13 Annotationen beträgt 34.73% und 34.49% mit 17 Annotationen. Die beste Gesamtfehlerquote für die Evaluation mit 13 Annotationen beträgt 19.52% und 19.5% mit 17 Annotationen. Der Durchschnitt der Gesamtfehlerquote für die 13 Annotationen ist 27.36% und 26.91% für die 17 Annotationen.

	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	751293.34	2163000	34.73%
OP 2	337321.67	1728000	19.52%
OP 3	383480.84	1646000	23.3%
OP 4	388243.33	1674000	23.19%
OP 5	478183.32	1614000	29.63%
OP 6	357863.33	1333000	26.85%
OP 7	415242.5	1497000	27.74%
OP 8	439555.83	1371000	32.06%
OP 9	338130.83	1501000	22.53%
OP 10	405775.83	1359000	29.86%
OP 11	372662.51	1363000	27.34%
OP 12	434009.16	1293000	33.57%
OP 13	432485.85	1709000	25.31%
Durchschnitt	425711.41	1557769.23	27.36%

Tabelle 5.11: Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer in der OP verwendet.



	Gesamtfehler	Dauer der Test OP	Gesamtfehlerquote
OP 1	746103.8125	2163000.0	34.49%
OP 2	336920.0	1728000.0	19.5%
OP 3	367285.25	1646000.0	22.31%
OP 4	405085.5625	1674000.0	24.2%
OP 5	467140.9375	1614000.0	28.94%
OP 6	313108.9375	1333000.0	23.5%
OP 7	396337.4375	1497000.0	26.48%
OP 8	410780.9375	1371000.0	29.96%
OP 9	337718.25	1501000.0	22.5%
OP 10	386796.0	1359000.0	28.46%
OP 11	368141.0625	1363000.0	27.01%
OP 12	407111.3125	1293000.0	31.49%
OP 13	427720.375	1709000.0	25.03%
OP 14	407274.4375	1784000.0	22.83%
OP 15	424901.5	1270000.0	33.46%
OP 16	378294.6875	1320000.0	28.66%
OP 17	359650.625	1258000.0	28.59%
Durchschnitt	46315941.18	1522529.41	26.91%

Tabelle 5.12: Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer in der OP verwendet.

## 5.5 Vergleich aller Methoden

In der Bachelorarbeit wurden folgende Methoden zur Schrittdauervorhersage in der Laparoskopie entwickelt und evaluiert:

- Durchschnittsmethode
- Linearregressionsmethode
- die Methode mit Random Forest, der zum Trainieren die Korrelation zwischen den Schritten, ihre Dauer und ihre Position in der Operation verwendet.
- die Methode mit Random Forest, der zum Trainieren die Dauer der vorherigen Schritte und ihre Reihenfolge in der Operation verwendet.
- die Methode mit Random Forest, der zum Trainieren die Informationen der Korrelation zwischen den Schritten und ihre Dauer verwendet.

Das beste Ergebnis wurde mit der Methode erreicht, die den Random Forest, der zum Trainieren die Dauer der vorherigen Schritte und ihre Reihenfolge in der Operation verwendet hat 3.3.2. Das Ergebnis beträgt 26.36% (Durchschnitt der Gesamtfehlerquote pro Historie) und wurde mit 17 Annotationen und mit der Historiegröße 5 erreicht.

Das schlechteste Ergebnis wurde mit der Linearregressionsmethode 3.2 mit 13 Annotationen erreicht und der Durchschnitt der Gesamtfehlerquote beträgt 33.16%.

## 6. Zusammenfassung und Ausblick

Im Rahmen dieser Bachelorarbeit wurden folgende Methoden entwickelt und evaluiert:

- Durchschnittsmethode
- Linearregressionsmethode
- die Methode mit Random Forest, der zum Trainieren die Korrelation zwischen den Schritten, ihre Dauer und ihre Position in der Operation verwendet.
- die Methode mit Random Forest, der zum Trainieren die Dauer der vorherigen Schritte und ihre Reihenfolge in der Operation verwendet.
- die Methode mit Random Forest, der zum Trainieren die Informationen der Korrelation zwischen den Schritten und ihre Dauer verwendet.

Es wurden auch Trainingsdaten aus den Annotationen analysiert. Zur Evaluierung wurde die Leave-One-Out-Kreuzvalidierung [23] benutzt. Das beste Ergebnis wurde mit der Methode erreicht, die den Random Forest, der zum Trainieren die Dauer der vorherigen Schritte und ihre Reihenfolge in der Operation verwendet hat 3.3.2. Das Ergebnis beträgt 26.36% (Durchschnitt der Gesamtfehlerquote pro Historie) und wurde mit 17 Annotationen und mit der Historiegröße 5 erreicht. Obwohl das beste Ergebnis mit dieser Methode erreicht wurde, könnte man auch die Durchschnittsmethode verwenden, weil das beste Ergebnis der Durchschnittsmethode 26.84% beträgt und die Methode die intuitive Verständlichkeit und der geringe Implementierungsaufwand hat.

Für weitere Verbesserungen des Verfahrens könnte man in der Zukunft, die in dieser Arbeit entwickelten Methoden, kombinieren. Die Linearregressionsmethode könnte man für die multiple Regression [19] implementieren, um den Zusammenhang von mehreren Variablen in Betracht zu ziehen. Als Feature könnte weiter interessant sein, ob der Chirurg unter Stress war, ob er müde war und ob er ein erfahrener Chirurg ist. Die Werte für den Stress und die Müdigkeit könnten durch die Analyse des Sauerstoffs im Blut gewonnen werden. Der Grad der Erfahrung könnte von Chirurgen eingeschätzt und in die Annotationen eingetragen werden. In Zukunft könnte man die Annotationen von echten Operationen und nicht nur Rektumresektion, sondern auch anderen laparoskopischen Operationen benutzen.



# Abbildungsverzeichnis

3.1	Ablauf der Dauervorhersage . . . . .	6
4.1	Klassendiagramm des Propheten auf Basis der Durchschnittsmethode . . .	12
4.2	Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis der Durchschnittsmethode . . . . .	12
4.3	Klassendiagramm des Propheten auf Basis der linearen Regressionsmethode .	13
4.4	Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis der linearen Regressionsmethode . . . . .	14
4.5	Klassendiagramm des Propheten auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes und ihre Position in der OP verwendet. . . . .	16
4.6	Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes und ihre Position in der OP verwendet. . . . .	17
4.7	Klassendiagramm des Propheten auf Basis des Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet. .	18
4.8	Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet. . . . .	19
4.9	Klassendiagramm des Propheten auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes verwendet wird. . . . .	20
4.10	Klassendiagramm der Leave-One-Out-Kreuzvalidierung auf Basis des Random Forest, der zum Trainieren die Dauer des bestkorrelierten Schrittes verwendet. . . . .	20
5.1	Der besten linearen Zusammenhang von zwei Schritten . . . . .	26
5.2	Abweichung vom Mittelwert in der OP 6 und OP 13 in Millisekunden . . .	26
5.3	Evaluation der OP 1 mit der Durchschnittsmethode . . . . .	28
5.4	Evaluation der OP 2 mit der Durchschnittsmethode . . . . .	29
5.5	Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerquote mit der Durchschnittsmethode . . . . .	29
5.6	Evaluation der OP 1 mit der Linearregressionsmethode . . . . .	31
5.7	Evaluation der OP 2 mit der Linearregressionsmethode . . . . .	32
5.8	Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerquote mit der Linearregressionsmethode . . . . .	32
5.9	Evaluation der OP 1 mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet. . . . .	34
5.10	Evaluation der OP 2 mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet. . . . .	35

5.11 Gesamtfehlerdurchschnitt, Durchschnitt der evaluierten OPs und der Durchschnitt der Gesamtfehlerroute mit Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet. . . . .	35
---	----

# Tabellenverzeichnis

5.1	Mittelwert, Standardabweichung und Varianz der Schrittdauer . . . . .	25
5.2	Die OP-Schritte mit bestem linearen Zusammenhang und ihre Korrelationskoeffizienten . . . . .	25
5.3	Evaluation mit 13 Annotationen Durchschnittsmethode . . . . .	27
5.4	Evaluation mit 17 Annotationen Durchschnittsmethode . . . . .	28
5.5	Evaluation mit 13 Annotationen Lineare Regression . . . . .	30
5.6	Evaluation mit 17 Annotationen Lineare Regression . . . . .	31
5.7	Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet. . . . .	33
5.8	Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer und den Schrittpositionswerten in der OP verwendet. . . . .	34
5.9	Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet wird. . . .	36
5.10	Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Dauer vorheriger Schritte und ihre Position in der OP verwendet wird. . . .	37
5.11	Evaluation mit 13 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer in der OP verwendet. . . . .	38
5.12	Evaluation mit 17 Annotationen Random Forest, der zum Trainieren die Korrelation zwischen Schritten und Features mit der Schrittdauer in der OP verwendet. . . . .	39





# Literaturverzeichnis

- [1] Laparoskopische Chirurgie. Abgerufen am 3.04.16. [Online]. Available: [https://de.wikipedia.org/wiki/Laparoskopische\\_Chirurgie](https://de.wikipedia.org/wiki/Laparoskopische_Chirurgie).
- [2] M. H. Unerbein, S. Gretsche, B. Rau, and P. Schlag, "Traumareduktion durch minimal invasive Chirurgie," *Der Chirurg*, vol. 74, no. 4, pp. 282-289, 2003. [Online]. Available: <http://link.springer.com/article/10.1007/s00104-003-0646-2>.
- [3] R. Schult, P. Matuszyk, and M. Spiliopoulou, "Prediction of surgery duration using empirical anesthesia protocols," in *Proc. of the KD-HCM 2011 at the ECML PKDD Int. Conf*, 2011.
- [4] G. Schmidberger and E. Frank, "Unsupervised discretization using tree-based density estimation," in *Knowledge Discovery in Databases: PKDD 2005*. Springer, 2005, pp. 240-251.
- [5] J. Dougherty, R. Kohavi, M. Sahami *et al.*, "Supervised and unsupervised discretization of continuous features," in *Machine learning: proceedings of the twelfth international conference*, vol. 12, 1995, pp. 194-202.
- [6] J. K. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach, chapter Discretization Methods*. Springer Verlag, 2007.
- [7] Arithmetisches Mittel. Abgerufen am 20.03.16. [Online]. Available: [https://de.wikipedia.org/wiki/Arithmetisches\\_Mittel](https://de.wikipedia.org/wiki/Arithmetisches_Mittel).
- [8] Naiver Bayes-Klassifikator. Abgerufen am 20.03.16. [Online]. Available: [https://de.wikipedia.org/wiki/Bayes-Klassifikator#Naiver\\_Bayes-Klassifikator](https://de.wikipedia.org/wiki/Bayes-Klassifikator#Naiver_Bayes-Klassifikator).
- [9] ID3 Algorithmus. Abgerufen am 20.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/ID3>.
- [10] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [11] Random Forest. Abgerufen am 20.03.16. [Online]. Available: [https://de.wikipedia.org/wiki/Random\\_Forest](https://de.wikipedia.org/wiki/Random_Forest).
- [12] E. Kayis, H. Wang, M. Patel, T. Gonzalez, S. Jain, R. Ramamurthi, C. Santos, S. Singhal, J. Suermondt, and K. Sylvester, "Improving prediction of surgery duration using operational and temporal factors," in *AMIA Annual Symposium Proceedings*, vol. 2012. American Medical Informatics Association, 2012, p. 456.
- [13] A. Macario and F. Dexter, "Estimating the duration of a case when the surgeon has not recently scheduled the procedure at the surgical suite," *Anesthesia & Analgesia*, vol. 89, no. 5, pp. 1241-1245, 1999.
- [14] Systematische Messabweichung. Abgerufen am 22.03.16. [Online]. Available: [https://de.wikipedia.org/wiki/Messabweichung#Systematische\\_Messabweichung](https://de.wikipedia.org/wiki/Messabweichung#Systematische_Messabweichung).

- [15] Mean absolute deviation. Abgerufen am 22.03.16. [Online]. Available: [https://en.wikipedia.org/wiki/Average\\_absolute\\_deviation#Mean\\_absolute\\_deviation\\_around\\_the\\_mean](https://en.wikipedia.org/wiki/Average_absolute_deviation#Mean_absolute_deviation_around_the_mean).
- [16] Kreuzvalidierungsverfahren. Abgerufen am 08.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/Kreuzvalidierungsverfahren>.
- [17] Maschinelles Lernen. Abgerufen am 15.03.16. [Online]. Available: [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning).
- [18] S. Dua, "Machine learning in healthcare informatics," Berlin, Heidelberg, 2014. [Online]. Available: <http://swbplus.bsz-bw.de/bsz399525416cov.htmhttp://dx.doi.org/10.1007/978-3-642-40017-9>
- [19] Lineare Regression. Abgerufen am 04.04.16. [Online]. Available: [https://de.wikipedia.org/wiki/Lineare\\_Regression](https://de.wikipedia.org/wiki/Lineare_Regression).
- [20] Korrelationskoeffizient. Abgerufen am 04.04.16. [Online]. Available: <https://de.wikipedia.org/wiki/Korrelationskoeffizient>.
- [21] Klassifikationsverfahren. Abgerufen am 22.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/Klassifikationsverfahren>.
- [22] Entscheidungsbaum. Abgerufen am 20.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/Entscheidungsbaum>.
- [23] Leave-One-Out-Kreuzvalidierung. Abgerufen am 08.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/Kreuzvalidierungsverfahren#Leave-One-Out-Kreuzvalidierung>.
- [24] Standardabweichung. Abgerufen am 22.03.16. [Online]. Available: <https://de.wikipedia.org/wiki/Standardabweichung>.
- [25] Varianz. Abgerufen am 22.03.16. [Online]. Available: [https://de.wikipedia.org/wiki/Varianz\\_\(Stochastik\)](https://de.wikipedia.org/wiki/Varianz_(Stochastik)).