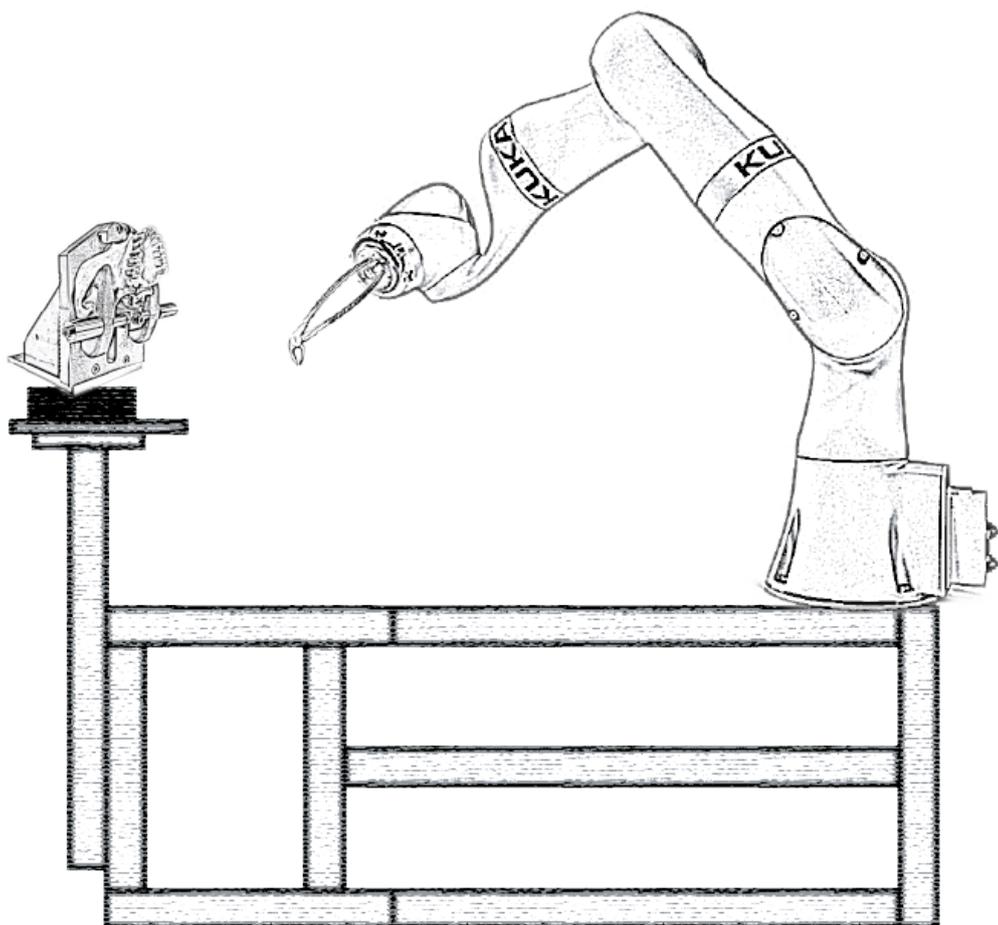


Understanding Tooth Removal Procedures using Feature Engineering and Classification Models

W.M. de Graaf

Master of Science Thesis



Understanding Tooth Removal Procedures using Feature Engineering and Classification Models

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Mechanical Engineering at Delft
University of Technology

W.M. de Graaf

November 4, 2020

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of
Technology



Copyright © Cognitive Robotics (CoR)
All rights reserved.



DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
COGNITIVE ROBOTICS (CoR)

The undersigned hereby certify that they have read and recommend to the Faculty of
Mechanical, Maritime and Materials Engineering (3mE) for acceptance a thesis
entitled

UNDERSTANDING TOOTH REMOVAL PROCEDURES USING FEATURE ENGINEERING
AND CLASSIFICATION MODELS

by

W.M. DE GRAAF

in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE MECHANICAL ENGINEERING

Dated: November 4, 2020

Supervisor(s):

dr.ing. J. Kober

drs. T.C.T. Van Riet

Reader(s):

prof.dr.ir. J. Harlaar

dr. O. Mazhar

Abstract

Tooth removal is one of the most performed surgical procedures worldwide. Despite the high amount of tooth removal procedures carried out each year, scientific understanding of these procedures is not present. Knowledge of force and torque behaviour is limited and knowledge about movements has never been subject to scientific research before. This study is an initial attempt to describe the factors that influence tooth removal in terms of forces, torques and movements. In-vitro measurements were performed that resulted in a dataset containing force, torque and movement time series of 181 human demonstrations of tooth extractions. This report showed how feature engineering and classification modelling were employed to find tooth removal explaining parameters in the dataset.

The feature engineering process led to numerical features describing the force and movement (rotation) time series. The rotation features were found to be most descriptive in describing differences in tooth removal procedures. This introduced five distinct rotation strategies that grouped the human demonstrations based on similarity of extraction strategy. These groups have been used as classification labels in the supervised learning process. A Naïve Bayes algorithm and a Logistic Regression algorithm were implemented as prediction models. These models showed that while the rotation features contributed the most to the prediction performance, there was need for additional force features to reach maximum prediction performance.

The results showed how feature engineering and classification modelling are the first steps in understanding the procedure of tooth removal.

Table of Contents

Acknowledgements	xv
1 Introduction	1
1-1 Relevance	1
1-2 Research Question	2
1-2-1 Objectives	4
1-3 Approach	4
1-3-1 Exploratory Phase	5
1-3-2 Iterative Phase	6
1-3-3 Additional Outline	6
1-4 Related Work	6
2 Theoretical Background	9
2-1 Dental Background	9
2-1-1 Anatomy	9
2-1-2 Tooth extraction	10
2-2 Machine Learning	12
2-2-1 Feature Engineering	13
2-2-2 Feature Normalization	15
2-2-3 Prediction Models	16
2-2-4 Performance Measures	18
3 Measurement Setup	21
3-1 Challenges	21
3-2 Overview	22
3-2-1 Measuring Movement	23
3-2-2 Measuring forces and torques	24

3-2-3	Calibration	24
3-3	Measurement Samples	25
3-3-1	Jaw Holding Devices	25
3-4	Software Integration	25
3-5	Results of the measurements	26
3-5-1	Resulting Dataset	26
4	Data Pre-processing	29
4-1	Reference Frame Transformation	30
4-1-1	General overview	30
4-1-2	Calibration Process	31
4-1-3	Transforming movement from Ψ_w to Ψ_t	31
4-1-4	Transforming wrench from Ψ_s to Ψ_t	33
4-2	Data cleaning	34
4-2-1	Resampling	35
4-2-2	Filtering	35
4-2-3	Data Augmentation	36
4-2-4	Trimming unnecessary data	36
4-3	Summary	37
5	Feature Engineering	39
5-1	Hypothesis	40
5-1-1	Force hypotheses	40
5-1-2	Rotation hypothesis	41
5-2	Feature engineering of force data	42
5-2-1	Extracting Strategies	43
5-2-2	Force Complexity	44
5-3	Feature engineering of rotation data	44
5-3-1	Strategies	45
5-3-2	Rotation Complexity	47
5-3-3	Overview of all features	48
5-4	Summary	52
6	Prediction Modelling	53
6-1	Feature table pre-processing	53
6-1-1	Labelling	54
6-1-2	Standardization	54
6-1-3	Train/test split	55
6-2	Model Implementation	55
6-2-1	Naïve Bayes implementation (baseline model)	56

6-2-2	Logistic Regression implementation	56
6-3	Model performance	56
6-4	Feature importance	58
6-4-1	Best subset selection	58
6-4-2	Forward Stepwise Selection	58
6-4-3	SHapley Additive exPlanations (SHAP) values	58
6-5	Summary	59
7	Results	61
7-1	Results of the Feature Engineering process	61
7-1-1	Descriptiveness in force features	62
7-1-2	Descriptiveness in rotation features	66
7-2	Results of the Supervised Prediction Modelling	71
7-2-1	Gaussian Naïve Bayes prediction Model	71
7-2-2	Logistic Regression Model	73
7-2-3	Model Interpretability	75
8	Discussion and Conclusion	77
8-1	Discussion of Feature Engineering and Prediction Model	77
8-1-1	Discussion of the force features	77
8-1-2	Discussion of the rotational features	79
8-1-3	Discussion of the prediction model	80
8-2	General Discussion	81
8-2-1	What was found in this study?	81
8-2-2	Limitations	81
8-3	Conclusion	83
8-4	Future Work	83
A	Additional Feature Engineering Results	85
A-1	Additional figures for feature engineering of forces	85
A-2	Additional figures for feature engineering of rotations	87
A-2-1	Strategy deduction and movement segmentation	87
A-2-2	Kernel Density Estimation	88
A-3	Gaussian Rotation Distributions	88
A-4	Gaussian Velocity Distributions	90
B	ATI F/T sensor information	93
B-1	Hooke's Law for Force Torque Measurement	93
B-2	Relevant information from Installation Manual	94
C	Conference Paper	99
	Glossary	111
List of Acronyms		111
List of Symbols		111

List of Figures

1-1	Simodont training simulator where the procedure of hard tissue removal is practiced with a physical rotary drilling tool.	2
1-2	Systematic hypothesis-based approach used to answer the research questions. This figure explains the approach used to analyse tooth removal, but it is also used as a guideline for the rest of the report. The steps taken to come to a justified answering of the research question are drawn in the coloured rectangles. Each step is described in a separate chapter, indicated with the abbreviation 'CH' in the corners. The inner rounded rectangles represent the sub-steps that are made, and the smaller grey rectangles are the representation of the data. Experiments are carried out with a measurement setup to obtain force, torque and movement data from human demonstrations of tooth removal procedures in 10 dimensions. This raw data is pre-processed , transformed and augmented to 14 dimensions. First insights are made together with clinical expertise, leading to a hypothesis about factors influencing tooth removal. Based on this hypothesis, feature engineering is performed. This step reduces the 14-dimensional time series to a n -column sized table, with n representing the number of features. Next, a prediction model is built on top of the feature engineering model and the extracted features are used in a Naive Bayes algorithm and a Logistic Regression algorithm to make a classification between groups of teeth. The results of the feature extraction process and prediction modelling are tested against the hypothesis, and based on the results, conclusions are drawn.	5
2-1	The left figure is showing an outline of a regular human mouth, seen from a dentist's perspective. The <i>right</i> and <i>left</i> declarations are corresponding to the perspective of the patient. The right figure shows the anatomy of an adult molar tooth, with important dental terminology visualized.	10
2-2	Visualization of the two most frequently used extraction techniques. The left figure displays a rocking motion and the right figure displays a twisting motion. ¹	11
2-3	Supervised learning principle applied on a spam filter. Each instance is labelled with a class and the model learns these pairs. In this way the model is able to classify the class of an unseen instance.	13
2-4	Typical machine learning pipeline from raw data to output predictions. The first step is the acquisition of raw data. The second step is the feature engineering step, which often also included pre-processing of the raw data. Subsequently, the features are extracted and used as an input to a machine learning model. Predictions are made with this model which leads to an output.	14

2-5	Feature engineering process on a single demonstration. The black curve represents the time series data. Two features are extracted from this time series: the amount of rotations, annotated with the circles, and the length of one back and forth movement, annotated with the arrows underneath the plot. These two features are placed in a table, as shown in the lower right of the figure. Each experiment is depicted by a row and the features are depicted by a column.	14
2-6	Sigmoid function used in Logistic Regression to calculate the output class probability. On the horizontal axis the input η is stated and the vertical axis states the output probability.	18
2-7	Confusion matrix categorizing the prediction class against the class in ground truth.	18
3-1	Overview of the setup with the following components: (1) robot arm, (2) forceps holding device, (3) video camera, (4) upper jaw holding device, (5) force-torque sensor, (6) bolts to adjust frame vertically, (7) bolts to adjust frame horizontally	22
3-2	The seven rotation joints of the KUKA iiwa robot. The schematic drawing explains the working principle of the spring damper system that is used in the compliant control mode. When a motion is applied, the spring with stiffness k and the damper with damping c resist this motion according to the value of k and c	23
3-3	Force torque sensor (FT sensor) used underneath the jaw holders.	24
3-4	Both jaw holders for the upper jaw (a) and the lower jaw (b) , rigidly connected to a rotation plate and the FT-sensor. For a more detailed description of the numbered components of the jaw-holders and the design choices made, the reader is referred to the paper in Appendix C.	25
3-5	Comparison of the removal of a central upper incisor (21) by an experienced surgeon (a) and the removal of a central upper incisor (11) by a dental student (b) . The arrow indicate the force and torque spike that occur at the instance of a tooth fracture.	26
3-6	An overview of the dataset that is used for analysis. The four executors together performed a total of 131 representative extractions. The grey stacked rectangles represent these representative extractions. A single extraction consists of time series data (TS) in green and meta-data (MD) in blue. The time series consists of 6-dimensional force and torque time series (brown) of time t and 7-dimensional movement time series (purple) of time t . The metadata consists of nine single categorical or numerical data points.	27
4-1	Pre-processing pipeline from the approach of Section 1-3. The raw data is the input for the pre-processing and the clean data is the output. The clean data results in a hypothesis, which is not stated in this chapter, but in Section 5-1. . .	29
4-2	A schematic overview of the setup with the coloured frames of reference for each component of interest. The robot is located on top of the adjustable frame, and the force-torque sensor is rigidly connected on the opposite side of the frame with respect to the robot. The black arrows a, b, c indicate how the frame widths, height or sensor rotation can be adjusted respectively. Frame Ψ_w is the robot base coordinate frame, which is fixed with respect to the world. Frame Ψ_{ee} is the coordinate frame as attached to the end-effector of the robot. The forceps with coordinate frame Ψ_f is rigidly connected with the end-effector. The location of the sensor is expressed by Ψ_s , and the location of the tooth is expressed by Ψ_t . .	30
4-3	Jaw holders with plastic jaws to show the two calibration tools and their respective reference frames. The upper jaw is placed on the left and the lower jaw is placed on the right. For the upper jaw, the luxator is used as a calibration tool. The y -axis of the luxator frame is pointing out of plane, indicated with a dot. The calibration tool for the lower jaw is a lower incisor forceps and is depicted on the right, where the y -axis is pointing into the plane. The reference frames depicted on both tools are Ψ_f . Their individual transformation to Ψ_{ee} is not identical. . .	31

4-4	The Euler angles xyz from an arbitrary measurement to show the effect of the coordinate transformation of the movement. The left figure depicts the raw Euler angles in Ψ_w and the right figure depicts the transformed Euler angles in the tooth frame Ψ_t	33
4-5	The torque data in xyz -directions from an arbitrary measurement to show the effect of the wrench coordinate transformation. In the left figure the raw torque data is depicted in the sensor frame Ψ_s and in the right figure the transformed torque data in the tooth frame Ψ_t is shown.	34
4-6	Result of the upsampling of an arbitrary f_z time series (orange) in comparison with the regular time series (blue).	35
4-7	Result of the butterworth filter for an arbitrary measurement	35
5-1	Prediction modelling pipeline from the approach of Section 1-3. The clean data results in a hypothesis that is used as a starting point for the feature engineering. The feature engineering process outputs numerical features, represented in table format. Furthermore, first results of factors influencing tooth removal are found. These results will be described in Chapter 7.	39
5-2	Force curves from an upper jaw incisor, bicuspid and a molar. The horizontal dashed line is where the force is 0N. There exists a phase where F_z is pushing into the jaw with $F_z < 0$ and a phase where the total F_z force profile is shifted above zero $F_z > 0$. Furthermore, the force complexity and magnitude during the pulling phase is increasing when the extracted tooth is located towards the back of the mouth.	40
5-3	Anatomical description of the movements as described in Table 5-1. The left figure displays the bucco-lingual/palatal movement and the right figure displays the longitudinal rotation.	41
5-4	Various rotations plotted that are encountered during the manual inspection of the rotation dataset. The left figure displays a pure z -rotation, the second figure displays a pure y -rotation. The third and the last figure displays two variations of the 8-rotation	42
5-5	Strategy deduction for all timestamps of a force experiment, where str1 is the pushing strategy and str2 is the pulling strategy, and Not Def are the non defined timestamps. The upper bar is the unfiltered strategy annotation and the lower bar is the filtered strategy annotation.	43
5-6	Figure displaying how the complexity of the force is calculated. Only the amount of force direction changes within the last part of str2 count towards the force complexity feature. The dots represent a direction change, which is annotated in all three directions.	44
5-7	Heatmaps resulting from plotting the y -rotation and the z -rotation. The kernel density distributions in y and z -directions are included on the top and the right respectively. The left figure displays a typical heatmap for an upper jaw incisor extraction. The right figure displays a typical heatmap for an upper jaw pre-molar extraction.	45
5-8	Rotation deduction strategy based on the amount of z -rotational velocity applied versus the amount of y -rotational velocity applied at time t . The resulting point on the circle determines which strategy is annotated to the data point. The resulting strategy deduction can be seen in the right plot. The blue curve is the z -velocity, and the green curve is the y -velocity. Plotting the point (dy_t, dz_t) in the circle on the left yields a strategy, indicated by the location of the point. This strategy is calculated for all timestamps in the whole experiment. The resulting strategy deduction for the whole experiment is shown by the bar underneath the right plot. The legend of the circle also indicate the colours in the bar.	46

5-9	The filtered rotation strategy deduction. The colors for the strategy are depicted in the left figure from Figure 5-8. The curve colors depict the velocity in the direction of y and z respectively.	47
5-10	Zero velocity crossing segmentation of the rotation time series based on the most dominant rotation. In this figure, the rotation in z -direction is segmented. The blue curves represent a clockwise rotation in z -direction and the red curves represent a counter-clockwise rotation in z -direction. The green curve is the rotation in y -direction. The horizontal lines are the respective segments that are introduced by the zero velocity crossing method.	48
5-11	Legend for the force curves as depicted in Table 5-2.	49
5-12	Legends for row 1-4 in Table 5-3. The left legend is the legend for the rotation curves, with the green color representing the y -rotation and the blue color representing the z -rotation. The right legend is the legend for the strategy bar, with the white color representing no strategy, the green color representing a y -rotation dominant strategy, the blue color representing a z -rotation dominant strategy and the yellow color representing a yz (both) dominant strategy.	50
5-13	Legend for the segmented rotation curves on row 5-6 in Table 5-3. The green color is representing the y -rotation, the blue color is representing a counter-clockwise z -rotation and the red color is representing a clockwise z -rotation.	50
6-1	Classification learning pipeline from the approach of Section 1-3. The feature table is used as an input the the prediction model. The prediction models are the Naive Bayes (NB) model and the Logistic Regression (LR) model. By classification, these models provide predictions, yielding results and conclusions.	53
6-2	Example of how standardization affects the distribution of two features from the dataset. In the left figure two random unscaled features are plotted against each other and in the right figure the standardized features are plotted	55
6-3	Accuracy distributions plotted for 500 randomized stratified train and test sizes, categorized per test size fraction. In blue the distribution for the the training set is plotted and in orange the distribution for the test set is plotted. The diamonds indicate single outlier samples.	56
6-4	A k -fold cross validation procedure performed on the dataset of the main surgeon ($n=68$). The respective train/test splits over 5 cross validation iterations is depicted in the first 5 rows, where a red colour represents the samples from the test set and the blue colour represents the samples from the training set. The sample division is based on the classes from Section 7-1-2 and is shown in the last row.	57
6-5	Stratified k -fold cross validation procedure performed on the dataset of the main surgeon ($n=68$). This figure shows that the training set samples and the test set samples are evenly divided over the number of samples in each class.	57
7-1	The demonstrations of the main surgeon ($n=68$) plotted for a selection of force features. The left figure shows the pushing percentage (f_str1_perc) on the vertical axis versus the pulling percentage (f_str2_perc) on the horizontal axis, both normalized from 0 to 1. The right figure shows the force complexity in F_z direction on the vertical axis versus the pulling percentage (f_str2_perc) on the horizontal axis normalized from 0 to 1. The scatter size is based on the amount of roots of the tooth, and the colour grading of the samples distinguishes based on the tooth type.	63
7-2	The pulling percentage (f_str2_perc) on the horizontal axis versus the pushing percentage (f_str1_perc) on the vertical axis of 68 demonstrations of the main surgeon. The tooth type (incisor or (pre-)molar) is described by the scatter size and the type of jaw is described by the scatter color	63

7-3	Force complexity in all <i>xyz</i> -directions plotted against each other sequentially. The size of the marker indicates the type of tooth and the color of the marker indicates the jaw type.	64
7-4	Distribution box plots of the complexity in all three force directions. The purple color is indicating the distributions related to the incisor teeth and the orange color is indicating the distributions related to the (pre-)molar teeth.	65
7-5	Distribution box plots of the normalized push and pull percentages for all executors. The number <i>n</i> is indicating the amount of samples that are spanning the distributions. The amount of push percentage <i>f_str1_perc</i> is depicted in red and the amount of pull percentage <i>f_str2_perc</i> is depicted in blue.	65
7-6	Features <i>f_str1_perc</i> and <i>f_str2_perc</i> plotted against each other for the samples of the student (<i>n</i> =15). The type of tooth is indicated by the marker size and the type of jaw is indicated with the marker colour.	66
7-7	Distribution box plots of the complexity in all three force directions for all executors. The purple color is indicating the distributions related to the incisor teeth and the orange color is indicating the distributions related to the molar teeth.	66
7-8	The normalized percentage of rotation strategy in <i>z</i> -direction (<i>z_perc</i>), <i>y</i> -direction (<i>y_perc</i>) and <i>yz</i> -direction (<i>both_perc</i>) plotted in a three fold plot. The marker color indicates the type of jaw and the marker size indicates the amount of roots of the tooth.	67
7-9	Clustering of rotation demonstrations where the percentage of <i>y</i> -rotation is plotted on the horizontal axis and the percentage of <i>z</i> -rotation is plotted on the vertical axis in the left figure. Because the 'remaining' samples in the lower jaw do not display distinctiveness based on <i>y_perc</i> and <i>z_perc</i> , they are compared amongst the features <i>dir_changes</i> and <i>mean_seg_length</i> , which can be seen in the right figure. The proposed clustering in this figure is done based on tooth type	68
7-10	Gaussian distribution of the amount of <i>y</i> -rotation against the amount of <i>z</i> -rotation in radians of all classes defined in Table 7-3, displayed in a heatmap.	69
7-11	Heatmap of the <i>y</i> -rotations against the <i>z</i> -rotations for the upper incisor class and the lower incisor class. The data of the main surgeon is plotted in (a) , the data of the other two surgeons are plotted in (b) and the data of the student is plotted in (c)	70
7-12	Confusion matrices for prediction results of the NB model. Figure (a) shows the result for the training set and figure (b) shows the result for the test set. Each instance in a column of the matrix represents the predicted label for a sample and each row represents the true label for that sample.	72
7-13	Cross validation scores when a 5-fold stratified cross validation is performed on the Naive Bayes model. The red curve shows the the training accuracy and the green line shows the cross-validation test accuracy.	72
7-14	Probability density plots from 500 randomized train/test splits for the Naive Bayes algorithm and the Logistic Regression algorithm. The blue violins show the distributions of the train data and the orange violins show the distributions of the test data.	74
7-15	Best subset selection for the Naive Bayes algorithm and Logistic Regression algorithm. The black dots represent the algorithms accuracy for a fitted models. For each number of features, the combination of possible models $\binom{k}{n}$ are fitted. Here <i>n</i> is the number of features to choose from and <i>k</i> is the total amount of features, which is 11. The red line is the subset with the highest accuracy. The green line shows the mean accuracy and the standard deviation over all $\binom{k}{n}$ models for every feature amount.	75

7-16	SHAP values for the Logistic Regression model trained on the classification classes as described in Table 7-3. The stacked bars show the impact of a feature on the output of the Logistic Regression model. The colour grading of the bars indicate the impact of the feature on the output performance of a specific class.	76
A-1	Strategy deduction results for force data for an incisor in the upper jaw (upper left), an incisor in the lower jaw (upper right), a molar in the upper jaw (lower left) and a molar in the lower jaw (lower right). In all four figures the force curves in x , y and z -direction are plotted in red, green and blue respectively in the most upper plot. The second plot shows the unfiltered strategy annotation per time sample. The third plot shows the filtered strategy annotation per time sample. The coloured bar at the bottom is a visual representation of the filtered strategy annotation.	86
A-2	Strategy deduction and segmentation results of rotation data for an incisor in the upper jaw (upper left and upper right), a molar in the upper jaw (lower left) and a molar in the lower jaw (lower right). In all four figures the rotation curves in y and z -direction are plotted. Both upper figures show experiments where the z -direction was dominant. Here, a segmentation distinction is made between counter-clockwise rotations (blue) and clockwise rotations (red). Both lower figures show experiments where the y -rotation was dominant. Here, a segmentation distinction is made between lingual/palatal rotation (green) and buccal rotation (orange). In all four figures the type of extraction strategy is plotted in the bar underneath. All time stamps are annotated with no strategy, a y -rotation strategy, a z -rotation strategy of an 8-rotation strategy (both), annotated with the white, green, blue or yellow color respectively.	87
A-3	Heatmap resulting from the plotting the Gaussian distribution of the y -rotation and the z -rotation of a demonstration of a lower jaw incisor 42 on the left and a lower jaw molar 47 on the right	88
A-4	Grouped rotation heatmaps from the rotation data of surgeon 2 and surgeon 3. The y -rotation is plotted on the horizontal axis and the z -rotation is plotted on the vertical axis.	89
A-5	Grouped rotation heatmaps from the rotation data of the student. The y -rotation is plotted on the horizontal axis and the z -rotation is plotted on the vertical axis.	89
A-6	Grouped rotation velocity heatmaps from the rotation velocity data of the main surgeon. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.	90
A-7	Grouped rotation velocity heatmaps from the rotation velocity data of surgeon 2 and surgeon 3. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.	91
A-8	Grouped rotation velocity heatmaps from the rotation velocity data the student. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.	91

List of Tables

2-1	Extraction techniques per tooth, grouped together based on similarity of extraction technique. This table is a merged and adapted version from Table 12.1 and Table 12.2 from Stegenga.	11
5-1	Comparison of basic naming conventions of the rotational movements. The description of Stegenga is shown in the first column, the anatomical description is shown in the second column and our description based on the calibration process is shown in the third column	41
5-2	Force features summarized with examples of the minimum extreme values and the maximum extreme values of the respective feature.	49
5-3	Rotational features summarized with examples of the minimum extreme values and the maximum extreme values of the respective feature.	51
6-1	Adjusted feature table where the labelled class is shown in column ' <i>label</i> '. This label is used in the supervised learning algorithm to fit the feature values in the training and testing phase.	54
7-1	Explanation of the features deduced from the force data.	62
7-2	Explanation of the features deduced from the rotation data with their descriptions.	67
7-3	Extraction techniques grouped together based on the outcomes of the feature engineering process. The tooth numbers and the group names are included.	69
7-4	Forward stepwise feature selection result for the five most important iterations. The last two columns represent the cumulative accuracy added during an iteration and the resulting accuracy of the combined features.	76

Acknowledgements

First and foremost, I would like to express my gratitude towards my supervisors Jens Kober and Tom van Riet, by proposing this research to me and guiding me during my graduation period. It felt like an adventure pioneering in the field of dentistry where I could apply my interest for the Robotic Operating System in an unknown research area of tooth removal. I will never forget the opportunities they both have given me, ranging from publishing a conference paper to performing measurements at the pathology department of the Amsterdam UMC. Even during the outbreak of COVID-19, they always kept personally engaged without losing their enthusiasm for my progress or the project.

I want to personally thank Jens Kober for the many interesting discussions during my progress meetings. I will never forget the meetings where we took a united technical front against Tom, to calm his wild ideas down. Furthermore, Jens's inconceivable eye for detail still amazes me, but foremostly, it helped me focus at times it was necessary. I benefited from his technical advice for exploring new directions of research.

I also want to thank Tom van Riet for his infinite amount of enthusiasm for my research and discovering the unknown. During the progress meetings, he always used his critical eye to keep me sharp, while constantly challenging me. I am grateful for introducing me to the field of dentistry, where I learned so much about. All the times I visited the Amsterdam UMC were inspiring and great fun.

Last but not least, I would like to thank my parents, brothers, friends and girlfriend for their love, support and encouragement during my studies. Thanks to them, I did not lose perspective and I could not have done it without them.

Thank you all, you made it happen!

Delft, University of Technology
November 4, 2020

W.M. de Graaf

Chapter 1

Introduction

Tooth removal is one of the most commonly performed surgical procedures worldwide [1]. In the United States alone, an estimated amount of 10 million wisdom teeth (only 4 out of all 32 human adult teeth) are extracted each year, accounting for an annual cost of more than \$3 billion [2].

1-1 Relevance

Despite the high amount of tooth extractions carried out each year, understanding of this procedure is limited. Knowledge about the reaction forces or torques is limited, and previous research only aimed at measuring the maximum amount of force necessary [3, 4, 5]. Furthermore, the directions in which the forces are applied, and foremost the movement the dentist performs, have never been subject to scientific research before. The lack of this technological understanding translates in the lack of a fundamental model explaining the intrinsic properties of tooth removal. In consequence, education on tooth removal also has been unchanged for 2000 years, resulting in limited technological advancement in this field. Nowadays, education only consists of reading minimal and general instructions on ‘rocking and twisting’ motions and pre-clinical training possibilities are absent or mostly inadequate. Students have to practice their skills on plastics jaws, requiring less force. Furthermore, the performed movements are not reflecting a real clinical situation. Hanson et al. showed how teaching tooth removal on cadaver jaws is more realistic than regular plastic mannequins [6]. However, most European dental schools do not have the resources to introduce such methods, so the student’s skill set has to be learned on real patients [7].

The problem of limited training facilities is in contrast to other specialisations in dentistry like restorative or endodontic dentistry. These specialisations booked more progress in implementing pre-clinical training possibilities. Virtual reality modules or simulator environments are already part of the current educational material [8]. An example of the implementation of such a training simulator in restorative dentistry is shown in Figure 1-1. This simulator uses a physical rotary drilling tool in combination with a virtual environment to teach the



Figure 1-1: Simodont training simulator where the procedure of hard tissue removal is practiced with a physical rotary drilling tool.¹

removal of hard tissue. Such methods however, are not yet implemented in the field of tooth removal [8].

This problem becomes even more evident when focusing at improvements in preventive dentistry programs. In well-developed countries, preventive dentistry gives rise to healthier teeth, contributing to a decrease in necessary extractions [9]. Dental students are thus suffering from decreased exposure to ‘learning by experience’ because fewer teeth need to be removed in general, facilitating lower confidence levels amongst recently graduate dentists [7]. A concerning development rising from this are the increasing referrals to Oral and Maxillofacial Surgery (OMFS) practices for simple tooth removal procedures [7]. This secondary line care is more expensive, so referrals due to a lack of confidence should be avoided as much as possible [9].

Summarizing, the lack of understanding of tooth removal has consequences both on educational level and on a clinical level. The need for improving this is high, and understanding the tooth removal procedure should be considered as the first step towards a scientific model describing the intrinsic properties of what makes (un)successful tooth removal. This facilitates the design of evidence-based educational material, helping dental students gaining confidence and, potentially, supplying them with reliable pre-clinical training possibilities. Next to that, it has the potential to help clinicians predict clinical outcomes (i.e. complicated treatments), leading to more (cost-)efficient referrals to oral and maxillofacial surgeons.

1-2 Research Question

In order to start to understand the tooth removal procedure, the fundamental properties of this procedure should be found. Stegenga already showed the importance of different movement patterns [10]. Ojala showed how the amount of force could be different in various teeth [11]. However, no uniform conclusion has been drawn on the actual interaction between the two and their precise value yet. Measuring these movements and forces and analysing them would be a first step towards understanding tooth removal. To date, reliable measurements of forces and movements have not been possible due to technological impossibilities. The

¹Courtesy of MOOG. Image from <https://www.moog.com/markets/medical-dental-simulation/haptic-technology-in-the-moog-simodont-dental-trainer.html>

key-hole nature of the tooth removal procedure makes it challenging to adequately measure movements. Furthermore, the forces and moments are only measured using strain gauges on the extraction tool [3, 4]. These are gripping forces, but no reaction forces of the tooth itself are measured. In addition, the rare combination of high forces and subtle movements has not been subject to research before.

However, advances in robotic technologies makes it possible to use a robot for recording human motion. The robot's precise motion tracking capabilities resolve the difficulties of measuring key-hole experiments. In addition, external force-torque sensors can be employed to measure reaction forces and reaction torques. This results in a dataset containing force, torque and movement data of human demonstrations of tooth removal procedures. Features describing differences between tooth characteristics could be deduced from the dataset and used to analyse the movement and force-torque data. In this way, extraction strategies, repetitive patterns or anomalies can be detected. Tooth specific parameters can be deduced, and information can be provided in which manner these parameters are of influence during a particular tooth extraction. With these parameters, machine learning can be employed to perform predictive modelling of the found structures and patterns. With these patterns, predictions can be made on what tooth is extracted, what the force level should be or what movement should be made.

This thesis aims to bridge the scientific gap of limited understanding of the extraction procedure. Robotic technology is employed to record human demonstrations of tooth removal procedures for modelling and analysis purposes. These human demonstrations involve the coherence of forces and movements and their interactions. Drawing conclusions from the obtained data could give insight into the explanation of tooth removal procedures which serves as the main research question and goal of this thesis:

Main Research Question (MQ)

How can the process of tooth removal be understood in terms of forces, torques and movements?

This research question gives rise to a framework which can explain tooth removal by translating human demonstrations of tooth removal procedures into understanding of the tooth removal process. This framework involves the development and validation of a reliable and safe measurement setup, which can capture the interplay of the human movement and resulting forces as precisely as possible. Furthermore, it involves the storage and analysis of this dataset, which are the first steps in bridging the scientific gap that exists in this field. If relevant features are found from the dataset that can describe similarities or differences between teeth, these features can be employed in a supervised prediction model. Analysing the behaviour and outcomes of such a prediction model provides insights into the importance of the features. To answer the main research question, the above leads to the following sub-questions:

Sub-Question 1 (SQ-1)

What are the most relevant factors in explaining tooth removal and how can we explain extraction differences between teeth?

Sub-Question 2 (SQ-2)

To what extent can a supervised prediction model be used to explain the factors that influence the process of tooth removal?

1-2-1 Objectives

The above research questions are posed in the framework of understanding the process of tooth removal, which is a broader research than only described in this thesis. Within the scope of this broader research, some overarching objectives for this thesis are defined and summarized below:

- Bridging the scientific gap in the field of tooth removal.
- Utilizing robot technology to capture clinician's extraction movement and constructing a dataset representing a real extraction as close as possible
- Providing a basis for the development of reliable educational material for dental students by providing intuitive and science based evidence for tooth removal explaining factors

1-3 Approach

A systematic approach should be followed to give a justified answer to the research questions defined above. Tooth removal is an accurate and careful procedure, influenced by numerous parameters. Therefore, it must be ensured that the movement of the extraction process and its reaction forces are captured in high detail. A hypothesis-based approach will be used while approaching (MQ), (SQ-1) and (SQ-2), which is shown in Figure 1-2. This facilitates a systematic analysis of the obtained data and helps to scope the thesis' objectives. This section will elaborate on the various steps that should be taken before the research questions can be answered. This section will also serve as an outline for the remainder of this report.

Figure 1-2 shows the approach used to draw conclusions from the complex dataset of tooth removal. Two phases are contributing to this goal, the phase before drawing the hypothesis and a phase after drawing the hypothesis. The first phase is denoted as the exploratory phase. Here the raw data is obtained from human demonstrations and it is pre-processed. Some first insights are made based on the intrinsics of the dataset. These insights are used to come up with hypotheses about what factors are explaining tooth removal. The second phase is an iterative phase where the hypotheses are used as a starting point to perform feature engineering and build a prediction model. For example, a choice for a feature is made because it looks promising. This feature is extracted from the time series and used as an input in the

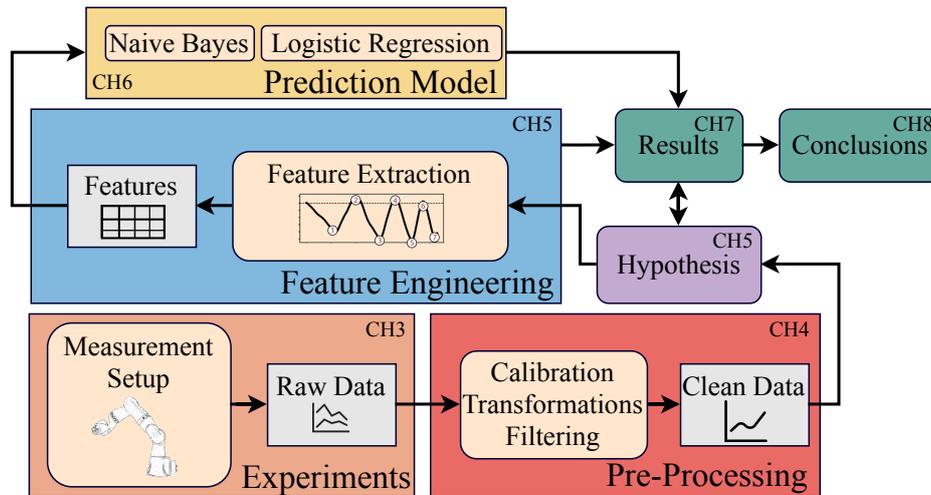


Figure 1-2: Systematic hypothesis-based approach used to answer the research questions. This figure explains the approach used to analyse tooth removal, but it is also used as a guideline for the rest of the report. The steps taken to come to a justified answering of the research question are drawn in the coloured rectangles. Each step is described in a separate chapter, indicated with the abbreviation 'CH' in the corners. The inner rounded rectangles represent the sub-steps that are made, and the smaller grey rectangles are the representation of the data. **Experiments** are carried out with a measurement setup to obtain force, torque and movement data from human demonstrations of tooth removal procedures in 10 dimensions. This raw data is **pre-processed**, transformed and augmented to 14 dimensions. First insights are made together with clinical expertise, leading to a hypothesis about factors influencing tooth removal. Based on this hypothesis, **feature engineering** is performed. This step reduces the 14-dimensional time series to a n -column sized table, with n representing the number of features. Next, a **prediction model** is built on top of the **feature engineering model** and the extracted features are used in a Naive Bayes algorithm and a Logistic Regression algorithm to make a classification between groups of teeth. The results of the feature extraction process and prediction modelling are tested against the hypothesis, and based on the results, conclusions are drawn.

prediction model. If the features do not show performance, the feature engineering process is tweaked and revised. Multiple hypotheses were adapted and reviewed before coming up with the final one. This report only describes the last iteration from hypothesis to results and which steps are taken to accomplish these results. Both feature engineering and prediction modelling leads to results. These results are tested against the hypothesis and conclusions are drawn accordingly.

1-3-1 Exploratory Phase

First, the raw data should be obtained from the human demonstrations measured with the measurement setup. The development and characteristics of the measurement setup are described in Chapter 3, where relevant design choices are explained. The design of the in-vitro extraction experiments and the characteristics of the raw dataset are described. Next, Chapter 4 describes how the raw data is pre-processed such that sensor noise is filtered and missing data is augmented. Correlations, trends and anomalies are inspected to gather

more information about the dataset. With the help of clinical expertise, a hypothesis about features explaining tooth removal is defined from these first insights. This clinical expertise is of high importance to justify if found patterns are reasonable and have a chance to survive as hypothesis. The hypotheses are used as the first building block towards the results that are obtained in the later stages of the research. It will serve as the bridge between the exploratory phase and the iterative modelling phase. As such, the hypotheses will be described at the beginning of Chapter 5.

1-3-2 Iterative Phase

Subsequently, there is an iterative phase where the formed hypothesis is modelled. The second part of Chapter 5 describes the feature engineering process that extracts features from the pre-processed time series data. These features are designed such that they represent the statements made in the hypothesis. These features are extracted from the time series and placed in a table. This table can be used as an input to the prediction model, which is described in Chapter 6. Multiple features combined can be used to build a Gaussian Naive Bayes classifier that uses Bayes' Theorem to find coherence amongst the extracted features. This Naive Bayes prediction model will serve as a baseline model which will be tuned and compared to a more sophisticated Logistic Regression model. These algorithms are not the most sophisticated machine learning algorithms available. In this research however, the goal is not to reach the best prediction performance. The goal is to implement an interpretable method, allowing humans to gain insights [12]. Best subset implementation and feature importance techniques will be used to gather feature specific information, giving more evidence for feature descriptiveness and model performance contribution. From the features and the modelling of the prediction model, more sophisticated insights will emerge, providing additional conclusions about the tooth removal procedure.

1-3-3 Additional Outline

In addition to the chapters described above, Chapter 2 will provide the reader with the required background knowledge about the tooth removal process and the machine learning principles used in this report. Furthermore, Chapter 7 describes the results of the feature engineering process and the prediction model analysis. Chapter 8 provides a summary of the executed research and discusses and justifies the results. Furthermore, the limitations of the research are stated and recommendations for future research are provided.

1-4 Related Work

This section shows the work done in literature regarding the application of robotics in dentistry. A small overview of robotics applied in three most essential sub-areas of dentistry is shown, involving orthodontics, implantology and restorative dentistry. This subsection is included to show what the current state of technology readiness is in dentistry, and what the possibilities are for applying robotics in the field of tooth removal.

In *orthodontics*, robots are mostly used for bending dental braces. A solution widely used in practices is the Suresmile treatment [13]. Scanning the mouth with 3D equipment is used

to achieve a Computer Aided Design (CAD) model of the mouth. With this CAD model, the researchers can plan effective treatments and reduce bending error margins. A robot is used to bend the metal wire brace precisely based on the 3D CAD model. A decrease in treatment time of 5 months (25%) is achieved compared to conventional methods. The process of bending the archwire is a crucial factor for decreasing these treatment times [14]. A bending robot not yet implemented in practices but showing potential is described by Jiang et al. They introduced a bending robot that uses a Bessel curve to model the shape of the archwire [15].

In *implantology*, the application of robotics is more focused on robotic guidance research. In the United States, a healthcare startup developed Yomi, a robot assisting surgeons in performing implant surgery. With integrated visual feedback and haptic guidance, Yomi is considered as cutting edge technology in the field of implantology and robotic integration in dentistry. However, it should be taken with caution, as results in terms of accuracy, reproducibility and repeatability of Yomi have not been published in literature so far. Additionally, other research groups have published work where the human error is being eliminated, and the accuracy is improved when performing implant surgery. Sun et al. describe a setup that generates a virtual plan to determine the ideal implant position [16]. Markers are attached in the mandibula before the mouth is scanned with a CT scan. This generates a virtual 3D model of the mouth, where the location of the implant position is known relative to the mandibula. Touching upon the markers with a guided robot will calibrate the system so the coordinates of the robot are transferred to the virtual environment. The software calculates the most efficient path and the drilling robot executes the path to drill the hole for an implant. Furthermore, Yu et al. made use of a stereo vision-based navigation system on a 3-DOF manipulator to approach the drilling holes for an implant [17]. The manipulator is moved by a clinician in a force field, causing a force feedback guidance system guiding the manipulator in the direction of the drilling hole.

In *restorative dentistry*, robotic research is mainly applied in the field of hard tissue removal. This method however, relies a lot on human accuracy [18]. With small holes, the conventional treatment strongly relies on the accuracy of the clinician. Yuan et al. proposed a fully automatic tooth preparation technique that eliminates manual operation from a clinician [19, 20]. A three-dimensional motion planning tool controlling an ultra-short pulse laser beam is used to perform the complete tooth preparation process. This research however, is only in the preliminary phase where it is not ready to be implemented in daily practice yet. Otani et al. investigated the removal of hard tissue as well, but only for aesthetic purposes and a different approach is used [21]. The hard tissue is removed with the same rotary cutting tool as in the manual removing of hard tissue, only mounted on a robotic arm which is used as guidance for this key-hole experiment. Better accuracy and precision was reached compared to the manual preparation method. No implementation was given however, so this research is still in the preliminary phase.

Concluding this subsection, most of the studies show that human intervention is still needed. Dentistry is still not at a point that robotics can be an integrated part of the daily workflow, but the existing research shows the advantages of increased precision, accuracy and more precise pre-surgical planning. Orthodontics was shown to be the only field in dentistry where other than basic research has been performed, and robotic solutions are yet implemented in the daily workflow. However, in the fields of implantology and restorative dentistry, only fundamental research and preliminary studies have been carried out.

Theoretical Background

This chapter provides the dental principles and the technical background knowledge necessary to better understand the steps made in the remainder of this report. Section 2-1 describes the background knowledge of the domain of dentistry. This section starts with explaining relevant anatomical knowledge about the mouth and the teeth inside the upper and lower jaw of a human. Furthermore, the relevant background information about the tooth removal procedure is explained. In addition to this, the various techniques currently taught to dental students are discussed.

Section 2-2 describes background information about the feature engineering process and why it is employed in the machine learning pipeline. Reasons are given why feature engineering is particularly useful for trying to understand tooth removal procedures. The features from the feature engineering process will be used in a Naïve Bayes (NB) prediction model and a Logistic Regression (LR) prediction model. The working principles of these models are discussed, and evaluation techniques to measure the performance of these algorithms are explained as well.

2-1 Dental Background

2-1-1 Anatomy

In the left figure of Figure 2-1, all 32 teeth of a healthy adult are displayed. In this research, the notation of The Fédération Dentaire Internationale (FDI) is used. The mouth is divided into four quadrants, each holding eight teeth. The first number refers to the quadrant, the second number refers to the individual tooth within this quadrant. The figure should be seen from the dentist's perspective, where the *right* and *left* annotations are corresponding to the patient's perspective. This means the notation "31" refers to an incisor in the left half of the lower jaw and the notation "17" refers to a molar in the right half of the upper jaw.

These 32 teeth can be subdivided into the following groups; 8 incisors, 4 canines, 8 premolars and 12 molars of which the four most distal ones are also known as the 'third molars' or 'wisdom teeth'. Teeth characteristics are mainly determined by the root type. A tooth can

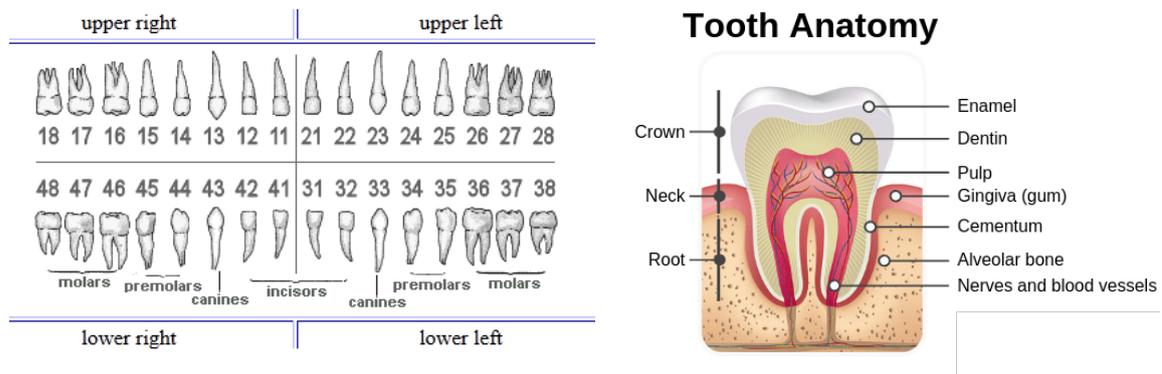


Figure 2-1: The left figure is showing an outline of a regular human mouth, seen from a dentist's perspective.¹ The *right* and *left* declarations are corresponding to the perspective of the patient. The right figure shows the anatomy of an adult molar tooth, with important dental terminology visualized.²

have single or multiple root, varying in size, length and shape (round/oval/straight/curvy). The right figure of Figure 2-1 is displaying the internal structure of a single tooth. The most important parts are the crown and the root(s), which contain the vital tissues of the tooth. The crown is the part that is visible in the mouth and it is the hardest form of tissue in a human body. In healthy patients, the root is entirely underneath the gingiva and for the largest part it is situated inside the bone. In patients with periodontal diseases, parts of the root can submerge because of retracting periodontal tissue. The cementum (outer layer of the root) is strongly connected to the surrounding bone through fibrous ligaments. The combination of loosening these fibrous ligaments and space creation between gingiva and bone is considered as 'loosening the tooth'.

2-1-2 Tooth extraction

Generally, two types of instruments are available to remove a tooth: a forceps and a luxator. A forceps is a commonly used tool while extracting teeth. Tooth crowns are varying in shape, so multiple variations of forceps' exist to facilitate a solid grip on every tooth. Based on individual preference of the clinician and varying clinical situations, a luxator (also known as an elevator) could be used as well. This is a small long tool which can be placed in between two teeth. With a very high torque, the tooth is plied out of its socket, located in the gingiva (see Figure 2-1). The constantly changing position of the luxator relative to the tooth makes reliable measurements of forces and their directions even more difficult, which was the main reason only to use forceps in this research.

Two extraction techniques are most frequently used in tooth removal: a rocking motion and a twisting motion. Both tooth motions are visualized in Figure 2-2. The chosen technique depends on the number of roots a tooth has and its shape. Single-rooted teeth are twisted around their longitudinal axis. Multiple-rooted teeth are extracted using a rocking motion, because multiple roots make a twisting motion impossible [4]. When a tooth is moved during

¹Adapted from <https://sites.google.com/site/oralhealthin/articles/tooth-numbering-systems>

²From <https://cdn.sharedentalcare.com/wp-content/uploads/2019/06/Tooth-Anatomy.png>

a rocking and twisting motion, it is believed that space is created for the physical removal of the tooth in the softer centre part of the bone, called bone marrow [10]. At the same time, the ligaments between the cementum of the tooth and its bony sockets are disrupted. Depending strongly on the position of the root(s) inside the bone relative to the strong outer layer of the bone (cortex) and softer part (marrow), a ‘preferable’ direction of movement is usually present.



Figure 2-2: Visualization of the two most frequently used extraction techniques. The left figure displays a rocking motion and the right figure displays a twisting motion.³

Currently, only these ‘rocking’ and ‘twisting’ motions are taught through a similar table as Table 2-1 to dental students, based on the book of Stegenga [10]. The extraction techniques are subdivided into groups of similarity, and for every group, only basic instructions are given. Two thousand years ago, a man named Aurelius Celcus, also described tooth removal with the same words: ‘*rocking and twisting*’ [22]. So, compared to the work of Celcus, education on these extraction techniques has not changed in 2000 years. Table 2-1 is showing how this educational material is handed to the students by Stegenga nowadays. The practical skill of extracting teeth is transferred through a table. Together with the fact that tooth removal education has not changed for over 2000 years, this table gives the impression that there is more to tooth removal than only ‘rocking’ and ‘twisting’. Other initiatives to improve tooth removal education have been described (i.e. videos, cadaver studies), but their representativeness and therefore their use is limited [7].

Table 2-1: Extraction techniques per tooth, grouped together based on similarity of extraction technique. This table is a merged and adapted version from Table 12.1 and Table 12.2 from [10].

Extraction technique	Tooth number	Root shape
Twisting	11, 21 34, 35, 44, 45	Round shaped Sometimes two roots
Rocking	12, 22	Oval shaped
Rocking + twisting	13, 23 15, 25	Long oval shaped Flat shaped
Rocking [Lingual/Palatinal] ⁴	31, 32, 33, 41, 42, 43 14, 24 16, 17, 26, 27	Long oval root Two thin roots Three roots
Rocking [Highly Lingual]	36, 37, 46, 47	Two flat shaped roots

³Adapted from <https://support.clearcorrect.com/hc/en-us/articles/203836918-Tooth-Movements>

⁴Lingual is the clinical term for *towards the tongue* as used in the lower jaw, palatinal is the clinical term for *towards the palatinum* as used in the upper jaw. However, both lingual and palatinal represent the same motion: opposite to buccal

2-2 Machine Learning

Machine Learning is a subset of the broader term Artificial Intelligence (AI). AI research focuses on techniques that enable computers to mimic human intelligence. Machine learning focuses on improving task performance from experience using statistical techniques. An algorithm develops an internal model to pair input data to output data. A formal and widely accepted definition of the term *machine learning* was proposed by Tom Mitchell [23] two decades ago:

“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 .”

— Tom Mitchell, 1997

In general, machine learning algorithms can be categorized based on the learning style an algorithm adopts to solve the problem statement. This learning style is fully dependent on the amount of supervision given on the input data. We can distinguish between supervised learning methods, unsupervised learning methods, semi-supervised learning methods and reinforcement learning. Supervised learning is the learning method where the input data is labelled. These labels are the desired solutions, and the algorithm learns from these labels. Unsupervised learning is the learning method where the input data is not labelled, and the algorithms try to learn without supervision. When providing labels for large datasets is costly, or if the labels are not available, a semi-supervised learning approach is often used [24]. The input data mostly has unlabeled instances and a few labelled instances. Reinforcement learning is the learning system that shows how agents take decisions in an environment to maximize the cumulative reward over time [25]. The system learns from these rewards while developing policies for performing actions in varying situations. Some examples of reinforcement learning applications are learning to play ball-in-a-cup [26] or learning to play table tennis [27].

In this research, we employ supervised learning for our analysis. The use of a supervised learning approach has two main reasons: First, our dataset will not be large, so unsupervised learning is expected to yield mediocre performance, while supervised learning generally performs better with smaller datasets [28]. Second, supervised learning has the advantage that the labels of the input data are known. These labels are not only used to build the model that predicts the correct output, but they can also be used to give contextual information on what happens during the learning phase. This can be used in our favour to find predictors for specific extraction strategies or various clinical situations.

Figure 2-3 shows an example of how labelling works in a spam-filter, a typical example of a supervised learning problem. Each instance (envelope) is labelled as spam or non-spam. This label can be considered as the class. The model learns a policy from the coupled class and instances in the training set, and this is applied on a new, unseen email, where the goal is to predict the correct value of the class.

Supervised learning problems can be categorized into two learning types: *classification* problems and *regression* problems [29]. **Classification** is the process where the prediction model is assigning observations to a class. The spam-filter problem in Figure 2-3 is an example of such a classification task. The classification model can be a binary, or a multi-class prediction

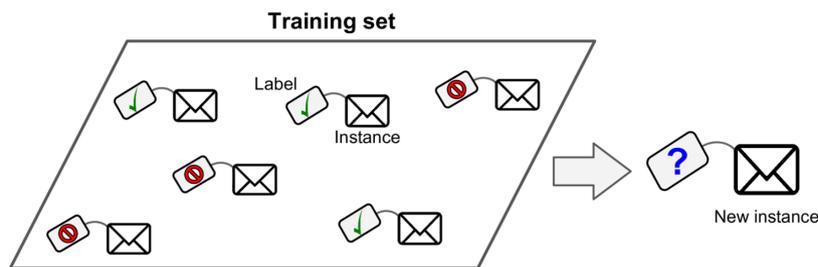


Figure 2-3: Supervised learning principle applied on a spam filter. Each instance is labelled with a class and the model learns these pairs. In this way the model is able to classify the class of an unseen instance (from [28]).

model. In multi-class prediction, the model is capable of predicting more than one class. **Regression** algorithms are predicting continuous numerical values based on the information hidden in the input information. It tries to find the relationship between response Y and input predictor X . Typically regression algorithms are used for stock value prediction or housing prices.

In this research, the classification prediction type will be used to build the prediction model. Classification will be more suitable than regression, because with classification we can check if the features make sense. Predicting continuous values with regression would only be justified if the features have successfully shown to be predictors for tooth removal, which they are not yet. The labels (classes) for classification in our model will be found based on similarity, emerging by the implementation of a feature engineering approach. Feature engineering is a common, underestimated step within the machine learning pipeline, but has proven to be decisive when domain expertise is embedded in the feature engineering process [30]. The following subsection describes how feature engineering works and how it could be applied in our research.

2-2-1 Feature Engineering

Usually, several steps in standard order have to be made when applying machine learning to real-world problems. This sequence of steps is typically called a pipeline. This includes but is not limited to the gathering of the data, removal of outlier data, filtering, applying transformations, doing feature extraction, perform model training, perform predictions and performing evaluations. A typical example of a Machine Learning pipeline can be seen in Figure 2-4.

This figure shows that there are more steps to machine learning than only picking a learning model and apply it on a dataset. As can be seen in Figure 2-4, this learning model needs features as an input and learns the mapping to the desired outputs. The general form of the last three steps of this figure can be seen in Equation (2-1). Here X is the input to the machine learning model and y is the output of the machine learning model. The function f is the mathematical model that is fitted to the input data such that it learns the corresponding outputs.

$$y = f(X) \quad (2-1)$$

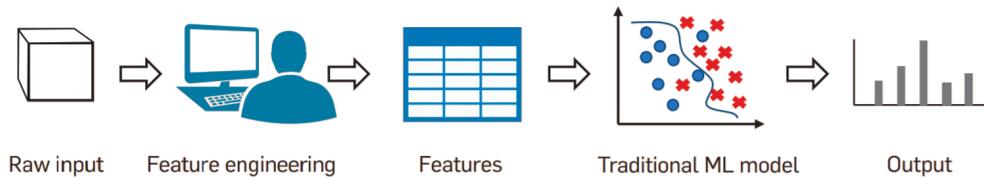


Figure 2-4: Typical machine learning pipeline from raw data to output predictions. The first step is the acquisition of raw data. The second step is the feature engineering step, which often also included pre-processing of the raw data. Subsequently, the features are extracted and used as an input to a machine learning model. Predictions are made with this model which leads to an output (figure adapted from [31]).

This shows that the features represent X and the learned classes represent y . A feature is a numeric representation of a portion of the raw data, which is the right input form for the machine learning pipeline [30]. The process of extracting these features from the raw data is called feature engineering, depicted as the second step in Figure 2-4. It is considered as an important step to make machine learning models perform well. Well-chosen features can make the modelling part easier and more informative, yet increasing the output performance of the pipeline. Harvey et al. and Bagnall et al. argued that higher model performance can be reached when the input data is transformed to feature-based representations rather than building complex prediction models in spaces that are not representative for the input data [32, 33]. Feature Engineering is widely applied in multiple domains such as text analysis (spam-filters, sentiment analysis or translation), image recognition (automatic driving, number-plate recognition or tumour detection on CT-scans) [34] or in the domain of time series analysis [35]. In the latter, the time series is transformed into a feature-based representation. This set of features can be used to find similarity amongst multiple time series [36]. An example of the feature extraction process and the resulting feature representation form can be seen in Figure 2-5. It shows how the number of rotations and the mean length of a back and forth movement of a time series can be translated into a feature-based representation.

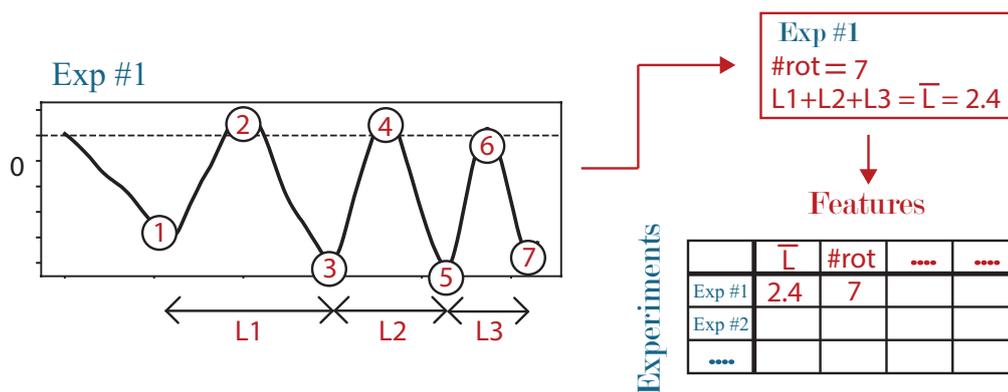


Figure 2-5: Feature engineering process on a single demonstration. The black curve represents the time series data. Two features are extracted from this time series: the amount of rotations, annotated with the circles, and the length of one back and forth movement, annotated with the arrows underneath the plot. These two features are placed in a table, as shown in the lower right of the figure. Each experiment is depicted by a row and the features are depicted by a column.

The resulting representation is a table where the columns are the feature vectors and the rows are the experiments. Referring back to Equation (2-1) and step 3 in Figure 2-4, this table is the input format X for the machine learning algorithm.

With the variety of applications of feature engineering to real-world problems, the manual selection of features from a dataset is dependent on the knowledge of the researcher. This automatically means that the performance of the resulting prediction model is dependent on the knowledge of the researcher. If the domain expertise of the researcher is put into practice in the right way, the features will describe the underlying mechanisms of the data in an efficient way and critical information can be isolated. This can help the machine learning model to focus on what the significant predictors for the specific problem are. In our our the implementation of feature engineering is mainly chosen because the information retrieval from the feature engineering process is high. This gives insight in what factors influence tooth removal and how features can explain differences between (groups of) teeth. Furthermore, a time series representation is not a suited input format for the prediction algorithm, so the time series have to be represented by numerical values: the features. The importance of domain expertise and model performance is well summarized by Timmer et al. [37]:

“The crucial problem is not the classifier function (linear or nonlinear), but the selection of well-discriminating features. In addition, the features should contribute to an understanding [...]”

— J. Timmer, 1993

2-2-2 Feature Normalization

Usually, the table in the lower right of Figure 2-5 is the correct input format for the features to work in a machine learning algorithm. However, the dataset can contain features which are highly varying in ranges and magnitudes. As machine learning algorithms make use of the Euclidean distance to calculate the distance between samples, this is problematic. Higher feature magnitudes will contribute significantly more to the distance calculation than features that have lower magnitudes, yielding a skewed contribution to the output performance. Scaling can be performed to let all features contribute evenly to the output performance. A selection of scaling methods will be explained below.

Standardization

Standardization, also referred to as variance scaling, is used to subtract the mean off the individual feature value, while dividing by the standard deviation. This will center the feature distribution around 0 mean ($\bar{x} = 0$) and scale the variance of the distribution to 1 ($\sigma_x = 1$).

Min-Max Scaling

The next scaling method is Min-Max scaling. The values of the feature vector are scaled to a fixed range between 0 and 1. Because this lower and upper bound are often smaller than the bounds of the unscaled dataset, the standard deviations are smaller, resulting in a smaller outlier error.

l^2 Normalization

The feature vector can also be scaled using the Euclidean norm, also referred to as the l^2 norm. This is the distance from the origin to the vector coordinate in the vector space.

$$\tilde{x} = \frac{x}{\|\mathbf{x}\|_2} \quad (2-2)$$

where x is the unscaled feature value, $\|\mathbf{x}\|_2$ is the Euclidean distance of the feature vector, denoted as $\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ and \tilde{x} is the scaled feature value.

2-2-3 Prediction Models

After the features are scaled, they can be used as an input for the supervised classification model. Some examples of supervised classification algorithms are Decision Tree (DT), (Deep) Neural Network (NN), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Logistic Regression (LR). Within the scope of this project, two main criteria are considered the most important for choosing a suitable prediction algorithm: the amount of data (which is limited) and the amount of information inferred from the model (which should be high).

A strong property of a NB algorithm is that even with limited data, it could be successfully applied in practice [38, 39]. SVMs and KNN have proven to be robust classifiers, but with limited data and strong linear calculation methods or solvers, they are prone to overfit [40].

Considering the amount of information inferred from an algorithm, DTs are most insightful in terms of absolute feature importance. The output of the prediction is directly correlated with the values of the features, giving a good insight into what features are descriptive and could potentially be predictors for a specific tooth group [30]. On the other hand, the LR algorithm outputs probabilities, making the classification decision more informative than only a yes/no [41].

(D)NNs are very advanced and could discover high nonlinear trends in datasets [41]. They consist of activation layers where weights and biases are randomly adapted to find patterns in the input data. With a low amount of data, the weights and biases are tuned such that high nonlinear patterns are found within the data. With a small dataset, this presumably results a strong overfit [28] on the training data. Furthermore, because of the perceptron structure in the NN, it is not possible to infer information on the specific feature importance.

Based on the proven performance of NB models on small datasets and the amount of information that can be inferred from both NB and LR models, a Naïve Bayes algorithm and a Logistic Regression algorithm will be used in this research. DTs have proven a to be most insightful, but are not used because they are prone to overfit on small datasets. Below the main working principle of both the NB algorithm and the LR algorithm is explained:

Naïve Bayes Algorithm

NB classification algorithms are algorithms based on the probabilistic Bayes Theorem. A NB algorithm can discriminate between features and predict a class based on the relationship

between features. However, these features should be independent, which is the core principle of the Naive Bayes classifiers. All features contribute evenly to the probability of the class prediction. This class prediction is called the *posterior probability* and is calculated based on Bayes' rule as follows:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (2-3)$$

Here $P(y|X)$ is called posterior probability. It is the probability of class y being predicted, given event X has occurred. $P(X|y)$ is called the likelihood, which is the probability of a predictor being true given class y . $P(y)$ is the priori of y , the probability of y before the evidence $P(X)$ is seen. Here X is representing the features which are used to predict y and it is given as $X = (x_1, x_2, \dots, x_n)$. Substituting this in Equation (2-3) gives:

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_n)} \quad (2-4)$$

For any given input the denominator stays constant. Hence, a proportionality is introduced, and with some rewriting the following equation remains:

$$P(y|x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i|y) \quad (2-5)$$

with n the amount of features representing the dataset. From this equation, a classifier model can be created, extracting the maximum probability of class variable y with given set of predictors X :

$$y = \arg \max_y P(y) \prod_{i=1}^n P(x_i|y) \quad (2-6)$$

The predictors of Equation (2-6) are still discrete. To make predictions with continuous predictors, a Gaussian distribution can be used to model the likelihood of the features X continuously. This is called a Gaussian Naïve Bayes classifier:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (2-7)$$

with μ_y and σ_y^2 being the mean and the variance of class x_i respectively.

Logistic Regression

Logistic Regression is a different type of classification algorithm. It is based on a type of regression, but the usage is to classify independent variables by means of a probability. The foundation of the algorithm is the sigmoid function (Figure 2-6):

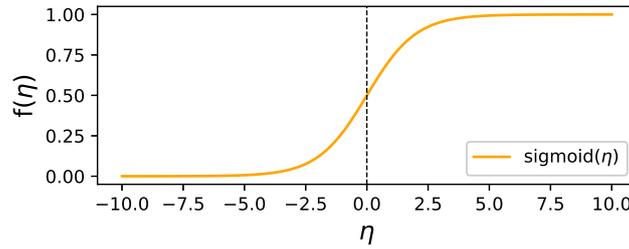


Figure 2-6: Sigmoid function used in Logistic Regression to calculate the output class probability. On the horizontal axis the input η is stated and the vertical axis states the output probability.

$$f(\eta) = \frac{1}{1 + \exp(-\eta)} \quad (2-8)$$

We can substitute the linear relationship from linear regression $f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ into Equation (2-8) to force the output to be a probability between 0 and 1:

$$P(x) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p))} \quad (2-9)$$

In this way, the model can make classifications based on the probability measure P . Although Logistic Regression depends on the sigmoid function, the coefficients β are still linear dependent, making the model a linear model. The model learns by adapting and estimating the coefficients β with the help of Maximum-likelihood Estimation. Based on a Gaussian distribution, the values for β will be found such that they maximise the likelihood of the model producing the output data as observed.

2-2-4 Performance Measures

The performance of an algorithm can be calculated by comparing the predicted values with the ground truth. This is possible because a supervised algorithm is used, so the labels are known. The most common technique is the use of a confusion matrix [42]. An example can be seen in Fig. 2-7. This figure shows how to compare the predicted class with the ground truth. It defines the True Positive (TP), True Negative (TN), False Positive (Type I error, FP) and False Negative (Type II error, FN) variables. TN indicates we predicted a class and this is correct. FP indicates we predicted a class while this is not correct, and FN indicates we did not predict a class while it is that class, respectively representing Type I and Type II errors. From these errors we can derive the accuracy metric, which is a ratio between the correct predictions (diagonal) and the total amount of predictions made:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-10)$$

		Ground Truth	
		Yes	No
Prediction	Yes	TP	FP
	No	FN	TN

Figure 2-7: Confusion matrix categorizing the prediction class against the class in ground truth.

However, accuracy can be a distorted metric. It assumes both classes have the same presence. If the classes are not well balanced, the prediction can be biased and will favour one class above the other. From this perspective, we need to be more specific. Recall is a metric that measures how much we predicted correctly, from all classes that have a positive ground truth.

$$Re = \frac{TP}{TP + FN} \quad (2-11)$$

The precision metric gives us the ratio out of all positive classes we predicted correct, how many actually are positive:

$$Pr = \frac{TP}{TP + FP} \quad (2-12)$$

A high recall means minimizing the FN rate. An example is the prediction of cancer. Cancer should always be detected, so predicting that a patient does not has cancer while cancer is present (FN) is worse then predicting the patient has cancer while cancer is not present (FP). On the other hand, high precision means minimizing the FP rate. This is mainly used is spam-filters. An important e-mail should not be classified as spam (FP), while it matters less if spam is not correctly classified as spam (FN).

A comparison metric is used to compare models where the precision metric and the recall metric are combined. Extreme values are punished more which gives a proper balance between Equation (2-11) and (2-12), which is called the F1 score:

$$F1 = \frac{2 \cdot Pr \cdot Re}{Pr + Re} \quad (2-13)$$

Chapter 3

Measurement Setup

This chapter describes the setup that is designed and used for measuring the forces, torques and movements from human demonstrations of tooth removal procedures. These measurements are performed on fresh frozen cadaver jaws. It will give a brief description of the challenges involved in measuring tooth removal, leading to design choices and compromises in the setup design. Furthermore, an overview of the components involved in the measurement setup is given. The mechanical components are discussed briefly and the robotic components, including the software, are discussed in more detail. A selection of the resulting dataset is shown to ensure the reader becomes familiar with the type of data acquired, adding to the readability of the rest of the report.¹

3-1 Challenges

Several challenges should be overcome to design a setup that can measure the forces-torques and rotations involved in tooth removal. This section describes the four most significant challenges and considerations that had an impact on the design process.

The first challenge is to capture these measurements in an in-vivo setting, so on ‘real’ patients. This would be the ideal setting, because the measurements would represent reality as close as possible. An alternative would be measuring in an in-vitro setting, where fresh frozen or conserved jaws are used. A significant drawback with this in-vitro approach is the representativeness of the results for tooth extractions in the clinic.

Second, all the parts of the setup should be attached rigidly to each other. Movement of parts should be minimized when applying forces during an experiment since sub-millimetre motion is recorded. Furthermore, the anatomy of the upper jaw and the lower jaw varies, causing a difference in the orientation of the extraction direction. This means two separate jaw holders

¹The general design choices of the hardware (jaw holders, robot and frame) were already determined by the Department of Biomechanical Engineering before the author started this project. The author is responsible for the software and control of the robot and the data storage.

have to be developed, both having several extraction orientations. The height, distance and the orientation of both jaw holders with respect to the robot should be able to be adjusted to compensate for variations in jaw anatomy.

The third challenge is the measurement of high forces and subtle movements simultaneously. The setup should capture movement and force-torque interactions in high detail, such that differences between various teeth are visible and reproducible.

At last, during an experiment, the presence of the setup must not lead to any restriction, otherwise the results will be biased. Handling the forceps attached to the robot should feel as natural as possible, even if extreme movements are made. Additionally, the jaw holders should be designed in such that the clinician's hand movements are not blocked in any way.

3-2 Overview

The proposed solution is shown in Figure 3-1, where the components are numbered from 1-7. A significant concession to the ideal setup is the use of an in-vitro experimental setting. Compared to in-vitro measurements, accurate sub-millimetre movement tracking and registration of forces and torques, and their directions in an in-vivo setup are questionable. One of the main issues is the mobility of the patient, which is difficult to compensate for. This is especially true for the lower jaw, being not rigidly fixated to the human skull. The setup consists of a robotic arm, a camera, an adjustable frame, a force-torque sensor on a rotation plate and a holding device for cadaver jaws (separate holding devices for upper and lower jaw).

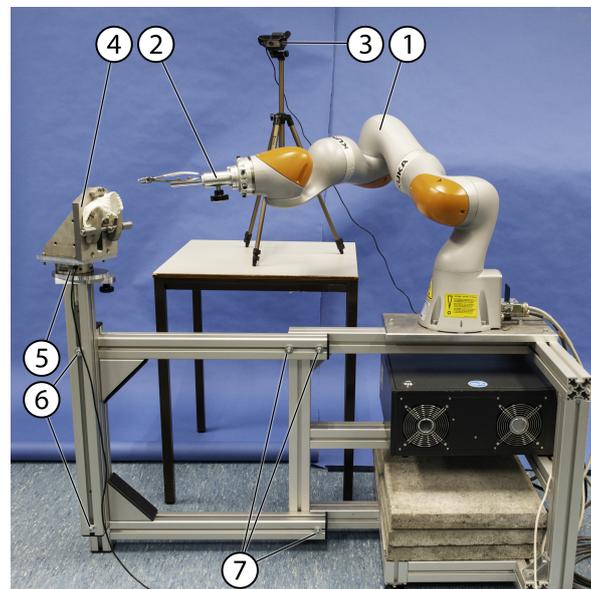


Figure 3-1: Overview of the setup with the following components: (1) robot arm, (2) forceps holding device, (3) video camera, (4) upper jaw holding device, (5) force-torque sensor, (6) bolts to adjust frame vertically, (7) bolts to adjust frame horizontally

With a dental forceps connected to the robot arm, the movement of a dentist is measured. The procedure of measuring movement is described in Section 3-2-1. The movement of the

dentist results in an application of forces and torques on the tooth. The forces and torques are measured with a force-torque sensor. This sensor is connected to the robot by means of an adjustable frame. The procedure of measuring forces and torques is described in Section 3-2-2. Because an in-vitro setting is used, the tooth extractions are performed on cadaver jaws. Two holding devices for upper and lower jaw respectively, have been developed to fixate the cadaver jaws rigidly. The position of the holding devices for the upper and lower jaws can be changed relative to the robot and placed at different heights. This is necessary to mimic clinical circumstances in which the position of the upper and lower jaw are, respectively, vertical and horizontal. Furthermore, the jaw holders can rotate along the vertical axis. This is implemented to mimic the position of the dentist in practice as close as possible.

3-2-1 Measuring Movement

A KUKA LBR iiwa7 R800 [43] is the heart of the setup, being a collaborative robot with seven rotational joints. Position and orientation data is needed for analysis, and this robot is capable of recording position and orientation data of the end-effector at 100Hz in 7-DOF (number 1 in Figure 3-1). The robot's integrated torque and rotational sensors enable for detection of external forces applied on the robot, making this robot highly suitable for collaborative purposes.

Where most robots are controlled by a simple position controller, where the desired position is reached by moving the rotational joints with a pre-defined velocity. The KUKA has an additional *compliance control* mode, where all joints are acting as separate rotational spring damper systems. In this mode, not the rotation of all joints is controlled, but the virtual rotational stiffness k and virtual damping c are controlled. This k and c of all seven joints can be adjusted separately. If there is no stiffness and no damping in all joints, the robot can be moved freely around, by exerting a minimum of external force on the robot's body. Attaching a dental forceps to the end-effector allows the robot to follow its movement while logging the position and orientation of the dental forceps. This principle can be used to measure the movement of a clinician performing a tooth removal procedure.

Figure 3-2 displays the location of all seven rotational joints. Additionally, a schematic drawing of the rotational mechanism is drawn. By incidence of the spring with stiffness k and the damper with damping c , the rotation θ is brought back to original state 0. The stiffness k and damping c are the rates at which this rotation returns to the original state. The higher k and c , the stiffer the rotation of that particular joint. If all values are set to 0, the robot will collapse under its own weight.

While performing an experiment, the clinician should feel a minimum of resistance from the robot. The clinician is free to move the forceps around, but certain extreme movements are

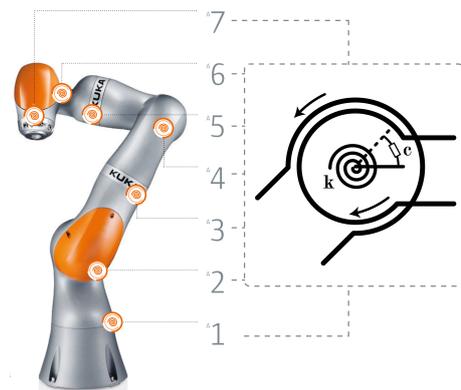


Figure 3-2: The seven rotation joints of the KUKA iiwa robot. The schematic drawing explains the working principle of the spring damper system that is used in the compliant control mode. When a motion is applied, the spring with stiffness k and the damper with damping c resist this motion according to the value of k and c .

causing the robot to resist this movement because the robot is not able to provide the desired position or orientation from the current joint configuration. The kinematic chain of the robot cannot provide the desired end-effector motion, so it results in high resistance of the robot at the end-effector. It can also happen that the robot is able to provide for the desired end-effector motion, but it suddenly reaches its joint limits and automatically shuts down for safety precautions. The stiffnesses of joints 2 and 5 are set to 5 N/m to prevent these pitfalls from happening. Movement in joint 2 and 5 is thus becoming limited, ensuring no joint limits overshooting. But most importantly, it allows the clinician to perform extractions with almost no resistance from the robot. Compliance control is based on the starting configuration of the robot, and because extractions of upper jaw teeth require other movements than extractions of lower jaw teeth, multiple starting configurations are pre-programmed accordingly for both the upper jaw and the lower jaw experiments. An example of the upper jaw starting configuration is shown in Figure 3-1.

3-2-2 Measuring forces and torques

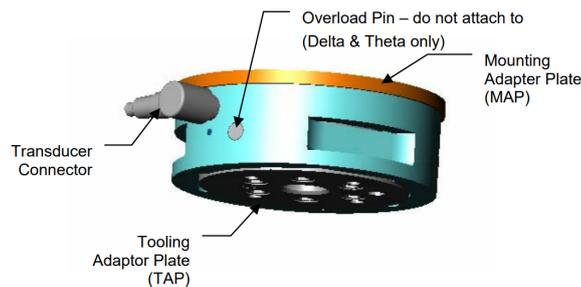


Figure 3-3: Force torque sensor (FT sensor) used to measure force and torque time series. The sensor is placed underneath the jaw holders [44].

An ATI 16 bit Delta transducer is used for recording the force and torque data in six axes at a rate of 20Hz (number 4 in Figure 3-1). Figure 3-3 shows the force-torque (FT) sensor used in the setup. It consists of a transducer that converts an applied load to force and torque components. Three symmetrically placed metal beams are located inside the transducer with semiconductor strain gauges attached to them. When applying a force to the transducer, Hooke's Law can be used to calculate the force and torque components by measuring the strain applied to the beam. A detailed explanation can be found in Appendix B.

3-2-3 Calibration

The measurements of the forces, torques and movements are all with respect to the location and orientation of the tooth. By means of calibration the location and orientation are determined. A summary of this calibration process can be read in Appendix C, and a detailed analysis can be read in Section 4-1-2.

3-3 Measurement Samples

The experiments took place in an in-hospital anatomy laboratory. Because an in-vitro measurement setting is used, fresh frozen cadaver jaws are used to extract teeth from. These jaws are prepared to a pre-defined and standardized form to fit the holding devices optimally as described in Section 3-3-1. Any remaining soft tissue is removed by using standard surgical blades. Care was taken not to remove any of the attached gingivae as periodontal health was one of the clinical parameters. The bodies from which the samples were taken were donated to science in accordance with Dutch legislation and the regulations of the medical ethical committee of the Amsterdam UMC at the location Academic Medical Center.

3-3-1 Jaw Holding Devices

Figure 3-4 show how the fresh-frozen jaws are placed in the jaw holders. Essential for reproducible, accurate and thus meaningful measurements is a completely rigid fixation of both upper and lower jaw. Two separate holding devices had to be designed. The most important reason for this is that the anatomical differences between the two jaws do not facilitate the design of a single device to fit both. In general, non-corrosive and smooth surface materials were used to facilitate cleaning, which is especially necessary when working with (fresh) human material.

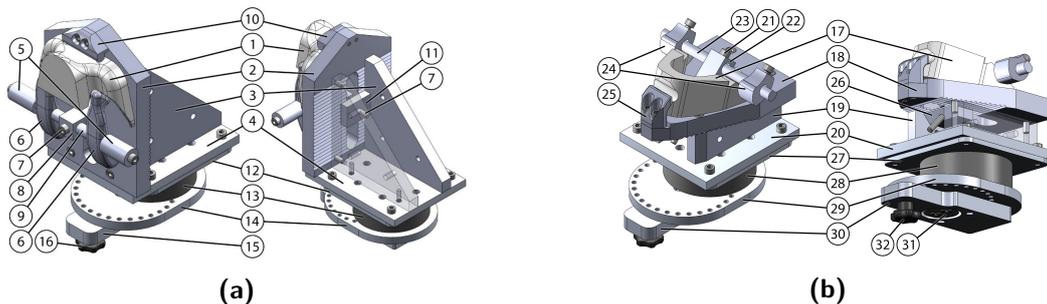


Figure 3-4: Both jaw holders for the upper jaw (a) and the lower jaw (b), rigidly connected to a rotation plate and the FT-sensor. For a more detailed description of the numbered components of the jaw-holders and the design choices made, the reader is referred to the paper in Appendix C.

3-4 Software Integration

The platform Robotic Operating System (ROS) is used to develop a control centre where all the components of the measurement setup are managed [45]. With a Graphical User Interface (GUI) the robot and the force-torque sensor can be switched on or off. At the same time, the position, orientation, force and torque data are logged and saved to the computer. The GUI handles and saves this experimental data and clinically essential parameters such as periodontal health, amount of roots and root size can be linked to the respective experiment. Additionally, a video stream of the experiment is saved to facilitate the analysis and interpretation of the data later on. For controlling the KUKA, the `iiwa_stack` is used, which first appeared in the work of Hennersperger et al. [46].

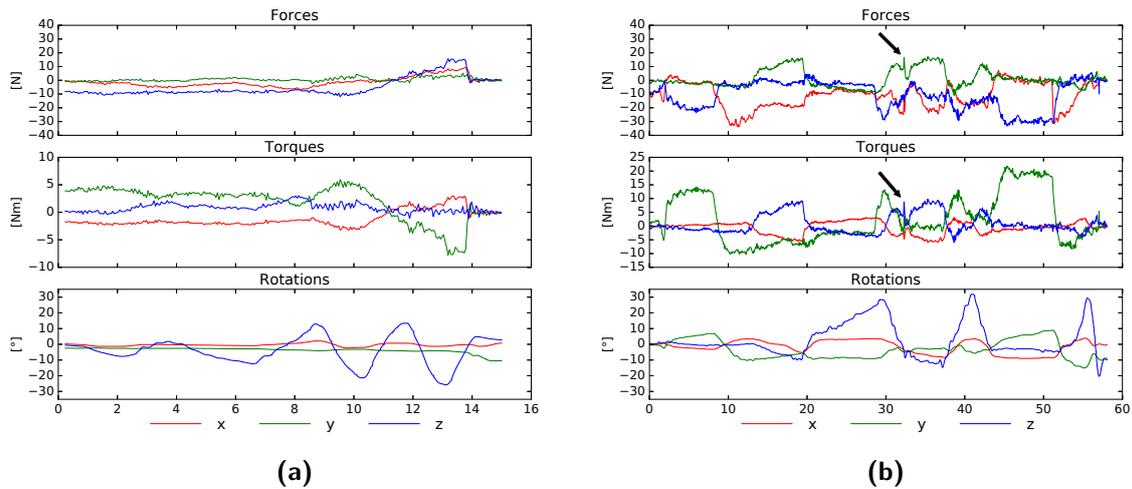


Figure 3-5: Comparison of the removal of a central upper incisor (21) by an experienced surgeon **(a)** and the removal of a central upper incisor (11) by a dental student **(b)**. The arrow indicate the force and torque spike that occur at the instance of a tooth fracture.

3-5 Results of the measurements

In total, an amount of 150 teeth are extracted from seven upper jaws and seven lower jaws. The jaws were fresh frozen cadaver jaws as described in Section 3-3. Three experienced oral and maxillofacial surgeons extracted a total of 128 teeth. Most experiments have been performed by the same experienced oral and maxillofacial surgeon ($n=79$), to gain a representative dataset. Additionally, a dental intern was asked to perform experiments as well ($n=22$). This was done to test if the differences between an experienced and an inexperienced clinician can be visualized.

Figure 3-5 show the comparison of an experiment carried out by both an experienced surgeon and a dental intern. The force, torque and rotations in xyz -directions are plotted on the vertical axis. This is raw data that is transformed from the end-effector frame Ψ_{ee} to the tooth frame Ψ_t (a detailed explanation of this transformation can be found in Section 4-1-3). It can be seen that the measurement setup is able to measure anomalies (arrows in **(b)**) and accurate force, torque and rotation changes. Interesting to see is the more than 200% increase of force the dental student uses compared to the surgeon. The dental student also shows a less recognizable plan in terms of movement. In contrast, the surgeon manages to keep the forces and torques at a relatively low and stable amount whilst increasing the movement.

This is just a small portion of the results showing the potential of this setup for measuring force, torque and movement data in in-vitro tooth removal procedures. For additional results and a more thorough explanation on design choices and the experiments conducted, the reader is referred to Appendix C.

3-5-1 Resulting Dataset

The dataset that originates from the experiments will contain non-representative experiments. This can be due to no successful tooth removal, slippage of the forceps, tooth breakage or

endodontic treatments that were present in the tooth. These faulty experiments are filtered out before the analysis starts. After removing the samples that were not representative, the dataset is as follows:

Study Population

Three experienced oral and maxillofacial surgeons performed 116 successful and representative tooth extraction from a total of 127 extractions. The student performed 15 successful and representative tooth extractions from a total of 22 extractions. This means, a total of 131 successful representative extractions will be used for the analysis. From these, 61 teeth are extracted from the lower jaw, and 70 teeth are extracted from the upper jaw. All extractions yield 6-dimensional force-torque data and 7-dimensional cartesian (position and orientation quaternion) movement data, in the format of time series. In addition to the time series, each experiment is annotated with metadata. This includes the jaw type, the tooth number, the race, the extraction instrument, the state of the tooth, the state of the periodontium, the complications, the number of roots and the root length, which adds another nine dimensions. Figure 3-6 is showing how this dataset is built up.

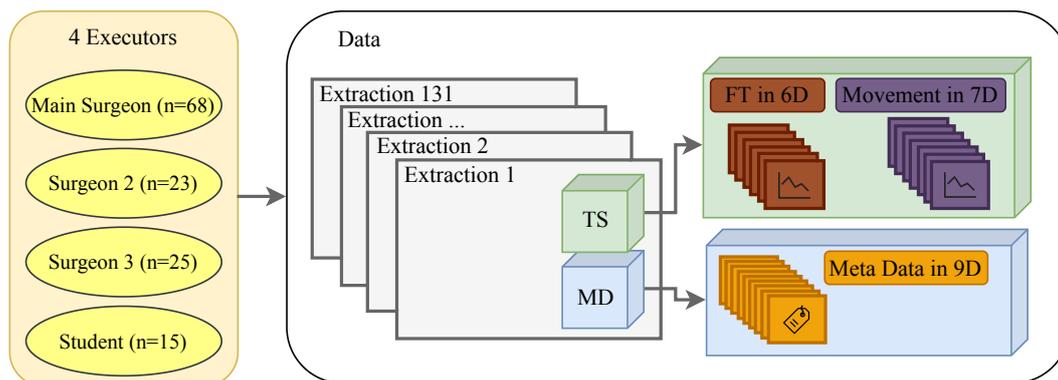


Figure 3-6: An overview of the dataset that is used for analysis. The four executors together performed a total of 131 representative extractions. The grey stacked rectangles represent these representative extractions. A single extraction consists of time series data (TS) in green and meta-data (MD) in blue. The time series consists of 6-dimensional force and torque time series (brown) of time t and 7-dimensional movement time series (purple) of time t . The metadata consists of nine single categorical or numerical data points.

This means a total of $131 \times 13 = 1703$ time series are included and $131 \times 9 = 1179$ single metadata points are included.

Data Pre-processing

This chapter describes how the raw data pre-processed into a format suited for feature engineering. Figure 4-1 displays where the pre-processing step fits in our approach. It can be seen that we get the raw data as an input to this process, and we get the clean data as an output of this process. The clean data will be used to propose hypotheses. These hypotheses are described in Section 5-1.

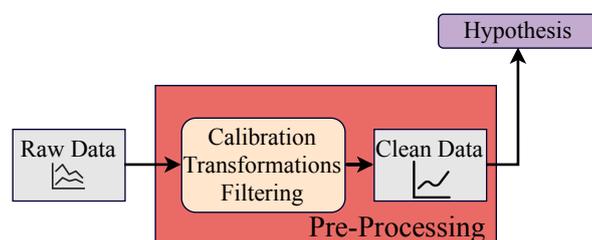


Figure 4-1: Pre-processing pipeline from the approach of Section 1-3. The raw data is the input for the pre-processing and the clean data is the output. The clean data results in a hypothesis, which is not stated in this chapter, but in Section 5-1.

The raw data is oriented in the local reference frames. This means the force and torque time series are expressed in the sensor frame, and the movement data of the end-effector is expressed in the robot's frame. Section 4-1 discusses the transformations of the local reference frames to a single reference frame that is situated on the tooth to be extracted. Section 4-1-1 describes the general overview of the current dataset and how the data formats can be interpreted. Also an overview of the local reference frames with respect to the world frame is given. The transformation from a local frame to the tooth frame can only be performed when the location and orientation of the tooth frame are known. We can obtain this location and orientation by calibrating the individual tooth. This calibration method is described in Section 4-1-2. Section 4-1-3 and Section 4-1-4 uses this calibration to describe the transformation of the rotations and the forces and torques respectively from their local reference frames to the tooth frame. When the transformation is complete, the data can be cleaned. This is described in Section 4-2. The effect of resampling, filtering and data augmentation is discussed and applied

on the transformed data, yielding a clean dataset with minimum noise which can be used for feature extraction.

4-1 Reference Frame Transformation

4-1-1 General overview

The raw data from the measurement setup consists of 13 dimensions. The force in sensor xyz -dimensions grouped together with the torque in sensor xyz -dimensions gives a 6-dimensional wrench. The position of the robot's end-effector in xyz -dimension and the orientation of the robot end-effector in four dimensions gives seven dimensions. The position of the end-effector is considered as a 3-dimensional vector. The orientation is represented as a quaternion, which is the 4-dimensional representation of any rotation in a 3-dimensional space [47]. Usually, quaternions are favoured in robotic applications above Euler angles or rotation matrices. Although quaternions are less intuitive and more complex to understand, they avoid the problem of 'gimbal lock' [48]. This gimbal lock occurs when the pitch angle of a mechanism approaches 90° and one degree of freedom is lost.

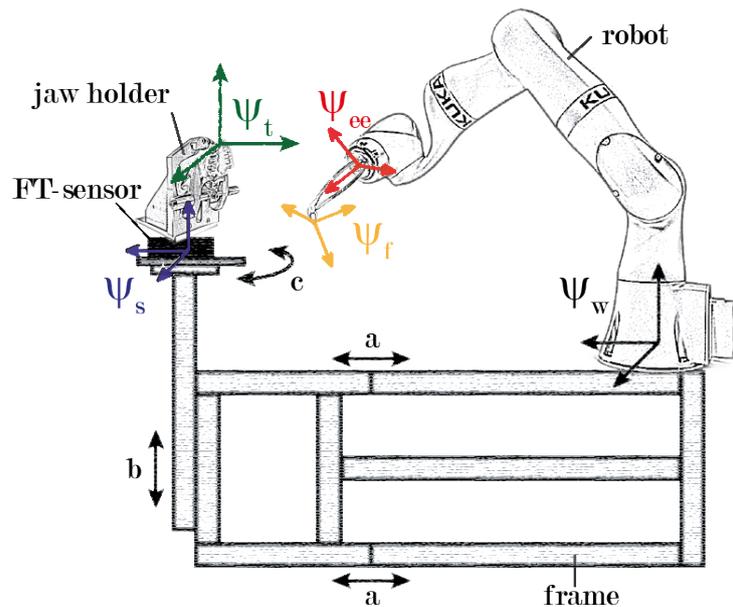


Figure 4-2: A schematic overview of the setup with the coloured frames of reference for each component of interest. The robot is located on top of the adjustable frame, and the force-torque sensor is rigidly connected on the opposite side of the frame with respect to the robot. The black arrows a, b, c indicate how the frame widths, height or sensor rotation can be adjusted respectively. Frame Ψ_w is the robot base coordinate frame, which is fixed with respect to the world. Frame Ψ_{ee} is the coordinate frame as attached to the end-effector of the robot. The forceps with coordinate frame Ψ_f is rigidly connected with the end-effector. The location of the sensor is expressed by Ψ_s , and the location of the tooth is expressed by Ψ_t .

Figure 4-2 is showing the schematic overview of the measurement setup of Chapter 3. The Kuka robot is located at the right on top of the adjustable frame. The force-torque sensor is

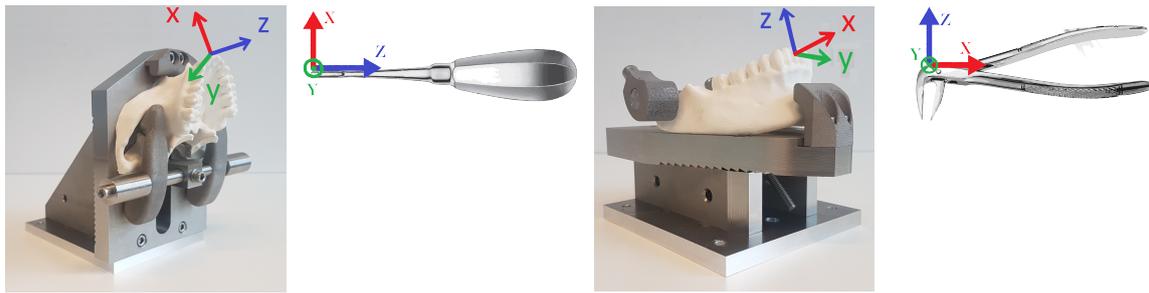


Figure 4-3: Jaw holders with plastic jaws to show the two calibration tools and their respective reference frames. The upper jaw is placed on the left and the lower jaw is placed on the right. For the upper jaw, the luxator is used as a calibration tool. The y -axis of the luxator frame is pointing out of plane, indicated with a dot. The calibration tool for the lower jaw is a lower incisor forceps and is depicted on the right, where the y -axis is pointing into the plane. The reference frames depicted on both tools are Ψ_f . Their individual transformation to Ψ_{ee} is not identical.

located opposite of the robot and is depicted in black. The jaw holder is located on top of the force-torque sensor. The black arrows next to the frame indicate the type of adjustment that can be made to the frame: the horizontal arrows **a** indicate a change of the distance between the robot and the sensor, the vertical arrow **b** indicates the change of height of the sensor and the rotational arrow **c** indicates the orientation change of the sensor with respect to the robot. Moreover, it shows the reference frames of the Kuka robot (Ψ_w), the force-torque sensor (Ψ_s), the end effector (Ψ_{ee}), the forceps (Ψ_f) and the tooth (Ψ_t). The tooth frame is the frame in which the forces, torques and movements will be expressed. Frame Ψ_s is determined by the dimensions of the adjustable frame. Frame Ψ_{ee} and frame Ψ_f are determined by the movement of the robot and frame Ψ_t is determined by the calibration process.

4-1-2 Calibration Process

With the help of the robot, the position and orientation of every tooth can be found. Because of the orientation difference between the upper and lower jaw (vertical/horizontal), two calibration tools were necessary. A lower incisor dental forceps is used for calibration in the lower jaw, due to the 90-degree angle and its straight design. For the upper jaw, a straight dental elevator (Usto-Lux, Ustomed, Germany) is used for calibration. This is done for the sake of convenience; any other tool with a flat pointy end could be used as well. Both tools and their orientation frames with respect to the tooth can be seen in Figure 4-3. The calibration is done by touching the centre of the crown. The tip of the tool is oriented in line with the z -axis of the tooth. For the elevator tool, this is along the length of the tool. For the forceps tool, this is the fixed part of the tip that is perpendicular to the handle. The calibration is done successfully if the reference frames on the tools are matched up with the reference frames depicted on the tooth in Figure 4-3. The tool's position and orientation are now registered as Ψ_t , which is expressed in Ψ_w .

4-1-3 Transforming movement from Ψ_w to Ψ_t

The movement performed by the clinician should be understandable. The format of the movement data allows for two ways of describing it: by the use of end-effector position data

or by the use of the forceps' tip rotation data. To compare against the motions described by Stegenga, the tip of the forceps rotation representation is used. This is most easy to interpret, and if the results of this study will be used later on, rotation is a better interpretable metric from a clinician's point of view. Furthermore, it is assumed that the tip of the forceps that grabs the tooth does not change in position during an extraction. A rotation representation thus captures the same movement as a position representation, only better interpretable.

To express the movement data in an interpretable format, the obtained orientation quaternion should be converted to Euler angles. Euler angles are chosen because their representation of movement is easier to interpret than rotation matrices are. To show this conversion, we start with a representation of a quaternion q :

$$q = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \quad (4-1)$$

The first component w is referred to as the 'real' part of the quaternion. The remaining three components x, y and z are the imaginary part, with \mathbf{i}, \mathbf{j} and \mathbf{k} being the mutually orthogonal imaginary unit vectors. In this way, x, y and z is the vector part of the quaternion about which the rotation should be constructed. The real part w is the scalar which determines the amount of rotation around the vector part.

From the calibration process, we obtained the orientation quaternion of the tooth. We call this quaternion q_t , which is the orientation of the reference frame Ψ_t . This quaternion is fixed and does not change during an experiment. The orientation of the end effector, and thus the orientation of the tip of the forceps, changes during an experiment. This means the orientation of the forceps, expressed as quaternion q_f with frame Ψ_f also changes over time. With quaternion multiplication, we can find the relative quaternion between q_t and q_f , describing the change of orientation of Ψ_f in Ψ_t :

$$q_{rel} = q_t^{-1} \cdot q_f \quad (4-2)$$

This quaternion in the form of Equation (4-1), consists of the terms $a_{rel}, b_{rel}, c_{rel}$ and d_{rel} . These terms can be used to calculate the corresponding Euler angles, based on the yaw (ϕ), pitch (θ) and roll (ψ) conversion, being the zyx -axis in frame Ψ_t :

$$\phi = \arctan \left(\frac{2(w_{rel}x_{rel} + y_{rel}z_{rel})}{w_{rel}^2 - x_{rel}^2 - y_{rel}^2 + z_{rel}^2} \right) \quad (4-3)$$

$$\theta = -\arcsin(2(x_{rel}z_{rel} - w_{rel}y_{rel})) \quad (4-4)$$

$$\psi = \arctan \left(\frac{2(w_{rel}z_{rel} + x_{rel}y_{rel})}{w_{rel}^2 + x_{rel}^2 - y_{rel}^2 - z_{rel}^2} \right) \quad (4-5)$$

An example of an arbitrary raw time series as measured in Ψ_w and converted via Ψ_{ee} to Ψ_t can be seen in Figure 4-4. The quaternions from the raw measurements are also converted to Euler angles zyx using Equations (4-3 - 4-5). The combined rotation around the xz -axis in Ψ_w results in almost a pure rotation around the z -axis in frame Ψ_t . Using Figure 4-3 we can see that this represents a rotation around the longitudinal axis of an upper incisor.

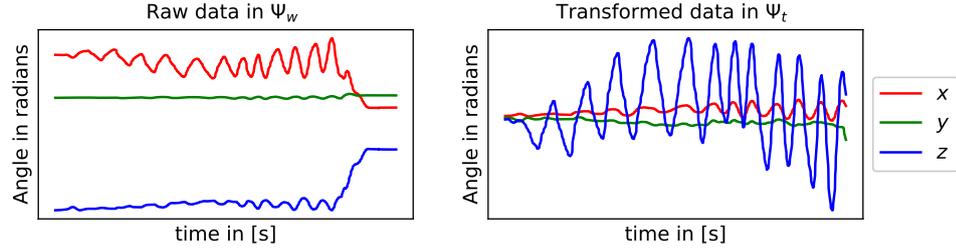


Figure 4-4: The Euler angles xyz from an arbitrary measurement to show the effect of the coordinate transformation of the movement. The left figure depicts the raw Euler angles in Ψ_w and the right figure depicts the transformed Euler angles in the tooth frame Ψ_t .

4-1-4 Transforming wrench from Ψ_s to Ψ_t

Frame Ψ_s is the sensor frame. This frame is fixed to the sensor, but because the sensor is mounted on a rotation plate, frame Ψ_s can rotate around its vertical axis. In Figure 4-2 this is indicated with the rotational arrow next to the sensor. The location of Ψ_s with respect to the world frame can be determined by measuring the height adjustments and width adjustments of the frame, indicated with the black arrows in Figure 4-2. With frame width \mathbf{a} , height \mathbf{b} and sensor rotation \mathbf{c} in mind, the position and orientation of Ψ_s with respect to the world frame Ψ_w is determined.

To find the transformation between Ψ_s and Ψ_t , we can make use of homogeneous matrices. The transformation between the sensor frame Ψ_s and the calibrated tooth frame Ψ_t can be found by calculating H_s^t :

$$H_t^s = H_w^s H_{ee}^w H_t^{ee} = \begin{bmatrix} R_t^s & p_t^s \\ \mathbf{0} & 1 \end{bmatrix} \quad (4-6)$$

It starts with the calculation of H_w^s , the transformation between the sensor and the world. The principal axis of the sensor and thus the orientation with respect to the robot is determined by the mechanical description of the sensor and denoted as R_w^{s0} . A detailed explanation can be read in Section 3.4 of the sensor manual, included in Appendix B. The location of the sensor with respect to the world is determined by the height \mathbf{b} and width \mathbf{a} dimensions of the adjustable frame and denoted as p_w^{s0} . This completes the transformation matrix H_w^{s0} . In addition, the transformation of the rotation plate (H_{rp}) should also be taken into account. This involves a pure rotation ψ around the z -axis, and because the plate rotates the sensor around its centerline, no translation is involved. Thus H_{rp} consist of $R_{rp}(\psi)$ and $p_{rp} = [0, 0, 0]^T$. This yields the following transformation:

$$H_w^s = H_{rp} H_w^{s0} = \begin{bmatrix} R_{rp} & p_{rp} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} R_w^{s0} & p_w^{s0} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \text{with } R_{rp}(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4-7)$$

Next, the transformation H_{ee}^w is defined, which is the transformation from the world frame to the end-effector frame. This is the logged position and orientation of the end-effector which would be H_w^{ee} . Instead, we can use the inverse of H_w^{ee} to calculate the transformation H_{ee}^w :

$$H_{ee}^w = (H_w^{ee})^{-1} = \begin{bmatrix} (R_w^{ee})^T & -(R_w^{ee})^T p_w^{ee} \\ \mathbf{0} & 1 \end{bmatrix} \quad (4-8)$$

At last, the transformation H_t^{ee} is defined. This is the transformation from the end-effector to the tip of the calibration tools in Figure 4-3. The precise transformations are not included in this report, but they are taken into account during the calculations. This transformation matrix thus is two-fold: one for the luxator calibration tool and one for the incisor calibration tool. The reference frame Ψ_f is not taken into account here, because during the calibration process this frame coincides with Ψ_t .

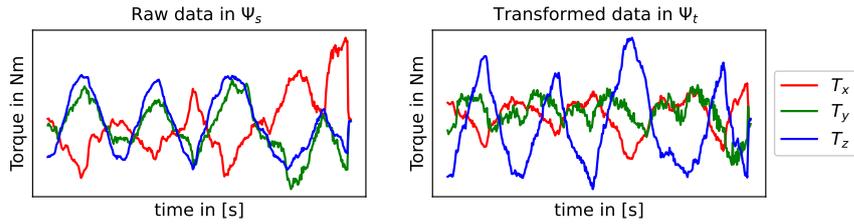


Figure 4-5: The torque data in xyz -directions from an arbitrary measurement to show the effect of the wrench coordinate transformation. In the left figure the raw torque data is depicted in the sensor frame Ψ_s and in the right figure the transformed torque data in the tooth frame Ψ_t is shown.

Now the transformation between Ψ_s and Ψ_t is found, we can use this to calculate the force and torque transformation. The forces and torques in Ψ_s can be represented by a wrench, which is a 6-dimensional generalized force containing a pure linear force and a pure moment around it. It is expressed in the following form:

$$W = \begin{bmatrix} f \\ \tau \end{bmatrix} \quad \begin{array}{l} f \in \mathbb{R}^3 \text{ pure linear component} \\ \tau \in \mathbb{R}^3 \text{ pure rotation component} \end{array} \quad (4-9)$$

Using screw theory we can perform the adjoint transformation of H_t^s to transform the applied wrench from Ψ_s to Ψ_t :

$$W^t = \begin{bmatrix} f_t \\ \tau_t \end{bmatrix} = \text{Ad}_{H_t^s}^T W^s = \begin{bmatrix} (R_t^s)^T & \mathbf{0} \\ -(R_t^s)^T \tilde{p}_t^s & (R_t^s)^T \end{bmatrix} \begin{bmatrix} f_s \\ \tau_s \end{bmatrix} \quad (4-10)$$

where \tilde{p}_t^s is the skew-symmetric form of p_t^s . Equation (4-10) also takes the generated torque that results from the translation from the sensor to the tooth at a distance $-p_t^s$ into account by the term $-p_t^s \times f_s$. An example of the transformation of a 3D torque measurement is shown in Figure 4-5.

4-2 Data cleaning

The transformed data can be processed further so it can be used in the feature engineering process. To ease the comparison in the `python` environment, the force-torque data is resampled from the obtained 20Hz to 100Hz. This is done to match the time series length of the

movement data. Next, the unnecessary data at the beginning and end of an experiment is deleted where the forces and torques are zero and the movement stops. Next, the noise is reduced by the use of filtering. This is needed to let the feature extraction process become robust, and the predictive features more stable. At last, data augmentation takes place where additional time series such as velocity and magnitudes are obtained.

4-2-1 Resampling

When resampling the wrench data (20Hz) and the movement data (100Hz) to the same sample length, three choices can be made; the wrench data can be upsampled to the movement data, the movement data can be downsampled to the wrench data, and both wrench data and movement data can be upsampled and downsampled respectively to a third frequency. To lose the least amount of information, the choice has been made to upsample the wrench data. In this way, no information is lost from the movement data. Due to upsampling however, noise will be introduced in the wrench data. This will be filtered out later when the dataset is filtered as a whole.

The wrench data is upsampled from 20Hz to 100Hz using a standard Fast Fourier Transformation (FFT). The result can be seen in Figure 4-6.

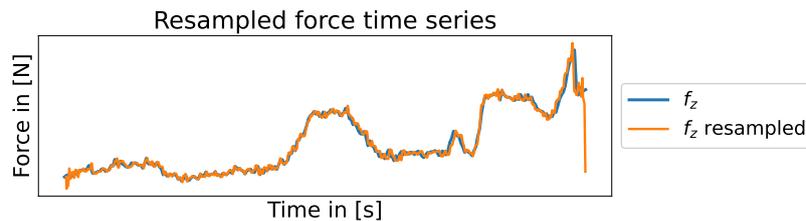


Figure 4-6: Result of the upsampling of an arbitrary f_z time series (orange) in comparison with the regular time series (blue).

4-2-2 Filtering

The feature extraction methods we aim to use will be sensitive to noise. We can filter out the high frequencies of the data by applying a causal low-pass Butterworth filter. With sampling frequency $f_s = 100\text{Hz}$ and cutoff frequency of 1Hz, we can filter out the higher frequencies resulting in the following smoothing of the data:

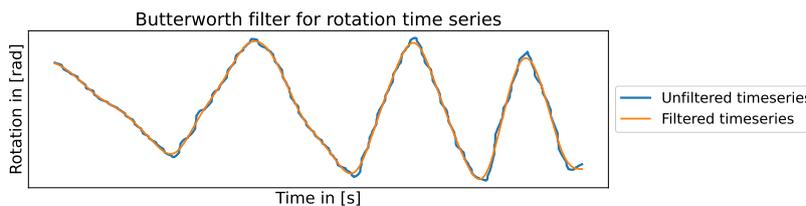


Figure 4-7: Result of the butterworth filter for an arbitrary measurement

This technique is applied to the 6-dimensional wrench data and the 3-dimensional rotational data. It can be seen that the filtered data is able to follow the shape of the demonstration and at the same time is able to filter out the high frequencies.

4-2-3 Data Augmentation

For feature extraction, we can include additional data, other than the data obtained from the measurement setup. The rotational velocity for instance, is needed to perform zero velocity crossings analysis. This can be used to define segments of movements. Furthermore, the magnitude of the forces and torques can be used to find differences between teeth in terms of total amount of applied force.

Velocity estimation

Because the velocity of the movement is not directly obtained from the measurement setup, we can estimate it using a finite difference method of the rotation xyz at time t :

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\mathbf{x}_{t+1} - \mathbf{x}_t}{\Delta t} \quad (4-11)$$

Where \mathbf{x} is the vector containing $[x, y, z]^T$ rotation, t is the instance in time and Δt is the time interval between two samples, which is equal to 0.01 seconds.

Magnitude estimation

The magnitude of the 3-dimensional forces and 3-dimensional torques can be used to express the total amount of force applied during an instance in time. We can use the Euclidean norm to calculate this magnitude as follows:

$$\|\mathbf{x}_t\| = \sqrt{x_{x,t}^2 + x_{y,t}^2 + x_{z,t}^2} \quad (4-12)$$

Here \mathbf{x}_t is the 3-dimensional force vector with xyz -components or the 3-dimensional torque vector with xyz -components at timestamp t . This Euclidean norm is the length of the vector so it is always a positive scalar.

4-2-4 Trimming unnecessary data

Usually, the obtained data stream has some meaningless data at the beginning and the end of an experiment. To ease the pre-processing work, the trimming of this meaningless data was done right at the end of an experiment, using the GUI described in Section 3-4. Trimming is done manually right after the demonstration because the whole dataset is quite diverse, and no general metric can be defined to trim all data accurately. If it was not done right after the demonstration during the experiments, it should have been done manually either way, only later in the process.

4-3 Summary

In this chapter, we applied several pre-processing steps to improve the quality of the raw data obtained from the measurement setup. We started with specifying the calibration method and the tools needed to perform this calibration. This calibration method is needed to specify the tooth location and orientation, which is used for the transformation of the raw data in a specific unified format.

We continued by transforming the data from the sensor frame and the robot frame to the unified tooth frame. In this way, the force, torque and rotational data for every tooth is expressed in a unified reference frame. This makes it possible to compare different extraction behaviours of teeth amongst each other, but also simplifies the feature engineering process. Because the 7-dimensional position and orientation data was reduced to only three rotation dimensions, the dataset consists of 9 dimensions after transformation.

Furthermore, we described the process of cleaning the data by the use of resampling, filtering, data augmentation and trimming. In this way, we obtained a clean and consistent dataset without meaningless data, which is ready to be used in the feature engineering process. With the augmentation of three velocity time series, one force magnitude time series and one torque magnitude time series, the 9-dimensional time series is increased to a final amount of 14 dimensions. A single tooth extraction measurement thus has 14 time series representing it.

Feature Engineering

This chapter describes the feature engineering process from clean data to a feature table. From manually inspecting the dataset, multiple hypotheses about what factors are influencing tooth removal are formed. This process is highly dependent on the expertise of the researcher, as such, a clinician of the Amsterdam UMC helped in forming the hypotheses. In this case, both technical expertise and clinical expertise were combined in forming the hypotheses. Based on the hypotheses, the 14-dimensional time series is reduced in dimension to a feature table through the feature extraction process. This table is the input format that is needed for the prediction model. Figure 5-1 shows the steps needed to transform the clean data into a feature table.

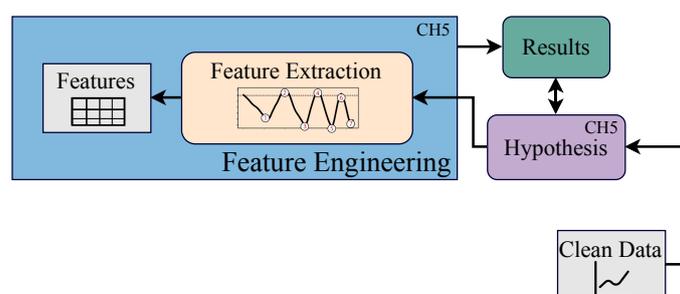


Figure 5-1: Prediction modelling pipeline from the approach of Section 1-3. The clean data results in a hypothesis that is used as a starting point for the feature engineering. The feature engineering process outputs numerical features, represented in table format. Furthermore, first results of factors influencing tooth removal are found. These results will be described in Chapter 7.

Within this chapter, a division will be retained between the analysis of force data and the analysis of rotational data. First, the hypotheses from the forces and rotations are described in Section 5-1. Section 5-2 discusses the method of extracting features from the force data. Chapter 5-3 discusses the method of extracting features from the rotational data. This chapter ends with summarizing the force and rotational features that will be used as an input to the prediction model in Chapter 6. Examples of extreme values of the features are given.

5-1 Hypothesis

To form the hypothesis, we can analyse the cleaned data that is obtained up to this point. The cleaned data consists of force, torque and movement data. The next subsections describe how the hypotheses of the forces and the rotations are constructed. The torque data was inconsistent but had some resemblance with the rotations. After multiple iterations of analysis, no sufficient proof was found that the torques could render good feature descriptiveness for the feature engineering model and the prediction model, so the choice was made to not propose a hypothesis about the torque data. The analysis leading to hypotheses about force and rotation, is based on the dataset of 68 experiments from the main surgeon. For readability of this report, not all figures are included, but an interesting selection is discussed.

5-1-1 Force hypotheses

Figure 5-2 is displaying the force plots of three different experiments. The left figure shows the reaction forces of an upper incisor, the central figure shows the reaction forces of a bicuspid (pre-molar), and the right figure displays the reaction forces of a molar. First, it can be seen that the extraction can be subdivided into two phases: a phase where the force F_z is negative, indicating a pushing action of the tooth in the jaw and a phase where the force F_z is positive, indicating a pulling action. In addition to these phases, it can be seen that during the pulling phase (and in limited amount during the pushing phase), the complexity of all three the forces increases when the tooth is located more to the back of the mouth. With the term ‘complexity’ the amount of zero velocity points and the magnitude of the forces is meant.¹ From manual inspection of the data it can be seen that these statements are true for teeth in both the upper and lower jaw. The expertise of the clinician supports these statements.

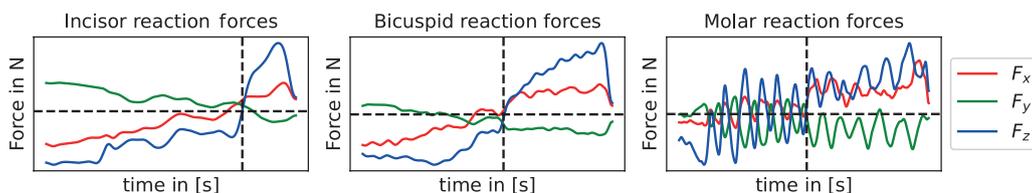


Figure 5-2: Force curves from an upper jaw incisor, bicuspid and a molar. The horizontal dashed line is where the force is 0N. There exists a phase where F_z is pushing into the jaw with $F_z < 0$ and a phase where the total F_z force profile is shifted above zero $F_z > 0$. Furthermore, the force complexity and magnitude during the pulling phase is increasing when the extracted tooth is located towards the back of the mouth.

It is hypothesised that the above mentioned can be of influence on the prediction model. With the human eye, it is possible to see differences in various force curves. Based on the statements about force complexity and the vertical force, features could be extracted that contain these differences. This leads to the hypotheses about the force data, which is summarised as follows:

¹With the term complexity it is not claimed that the actual force complexity of tooth removal is found. The term ‘complexity’ is used for convenience sake, since the researchers are unsure what complexity in terms of tooth removal means

Hypotheses about the Forces

1. An extraction can be subdivided by at least two strategies, which are based on the reaction force F_z . These strategies are the pushing and pulling strategies respectively.
2. Forces increase with complexity when teeth are located more towards the back of the mouth.

5-1-2 Rotation hypothesis

When manually inspecting the rotational data, it can be seen that the time series are more varied than the ‘rocking and twisting’ movements described by Stegenga in Table 2-1. Based on his explanations, there are two main rotation strategies: a twisting motion and a rocking motion. Because Stegenga uses a different naming convention than our calibrated reference frames, Table 5-1 includes the comparison of the naming conventions of Stegenga, the anatomical movement and ours.

Table 5-1: Comparison of basic naming conventions of the rotational movements. The description of Stegenga is shown in the first column, the anatomical description is shown in the second column and our description based on the calibration process is shown in the third column

Stegenga description	Anatomical description	Our description
Rocking	Bucco-Lingual / bucco-palatal	y -rotation
Twisting	Longitudinal rotation	z -rotation

For reference Figure 5-3 includes the rotational movements described in Table 5-1. This figure show the same movements as Figure 2-2, only the anatomical description is used here.



Figure 5-3: Anatomical description of the movements as described in Table 5-1. The left figure displays the bucco-lingual/palatal movement and the right figure displays the longitudinal rotation.²

The inspected data showed more rotations than only the rocking and twisting rotations. A selection of the groups covering all the different rotations in our dataset are shown in Figure 5-4. The left figure shows a pure rotation around the z -axis, and the second figure shows a

²Adapted from <https://support.clearcorrect.com/hc/en-us/articles/203836918-Tooth-Movements>

pure rotation around the y -axis up until the end of the experiment. The last part of this experiment includes a pure z -rotation as well. This second figure thus captures the rocking and the combination of rocking and twisting movement as described by Stegenga. The third and fourth figure show a rotation that is not described by Stegenga. It is a rotation where all three axes are active and play a role in the extraction movement. The clinician refers to this movement as the movement of ‘8’. This type of movement is not described in the tables of Stegenga but is applied in practice now and then. We will refer to this movement as the 8-rotation in the remainder of this report. The third and fourth plot of Figure 5-4 displays multiple variations of this movement, which has to be investigated further.

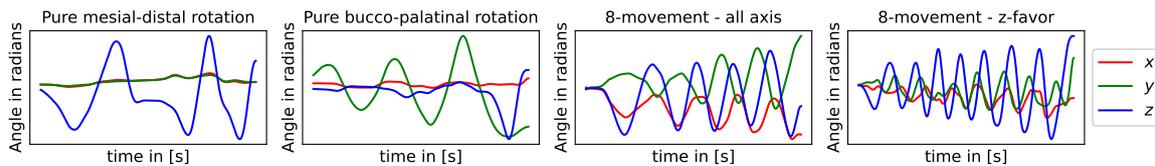


Figure 5-4: Various rotations plotted that are encountered during the manual inspection of the rotation dataset. The left figure displays a pure z -rotation, the second figure displays a pure y -rotation. The third and the last figure displays two variations of the 8-rotation

We want to find a descriptor for the found groups, and it is hypothesized that the combination of rotations amongst multiple axes can be a predictor for these tooth groups. The dataset showed that the demonstrations were either performed by performing a dominant y -rotation, a dominant z -rotation or a combination of yz -rotation. The rotation around x was present but was not dominant. This is probably due to adjacent crowns that do not allow for rotation along the x -axis. The experience of the clinician in practice supports this.

Furthermore, it was argued that the more the tooth was located to the back, the larger number of rotations is needed to extract the tooth. Figure 5-4 is showing this, where the left figure is an incisor and the most right figure is a molar. An increase in the number of rotations was seen, which could be a predictor for the type of tooth being extracted. The two hypotheses are summarized as follows:

Hypotheses about the Rotations

1. Rotation differences around the y -axis, z -axis or a combination of both can be used to distinguish extraction strategies between groups of teeth.
2. The more the tooth is located towards the back of the mouth, the larger number of rotations are needed to perform a successful extraction.

5-2 Feature engineering of force data

The hypotheses of the forces is stated in Section 5-1-1. The first hypothesis is based on the behaviour of the reaction force F_z . This force can have a pushing or a pulling character. It showed the characteristics from the cross-over point, where the strategy switches from a

pushing form to a pulling form. The second hypothesis is based on the force curves after this cross-over point. It seemed that the forces are increasing in complexity when teeth are located towards the back of the mouth. Section 5-2-1 describes how the strategies are deduced from the force data and how this leads to features describing the strategy of applying force. Section 5-2-2 describes how the complexity is modelled and how this leads to features describing the complexity of the applied force.

5-2-1 Extracting Strategies

The extraction of the strategy is done via the algorithm stated below. It is based on the first hypothesis, where we use the force F_z to make a distinction between the strategies.

Algorithm 1: Strategy deduction algorithm for the force time series

Result: Strategy at time t

```

for time  $t$  do
  if  $F_z < 0$  then
    | S = phase push;
  else if  $F_z > 0$  &  $F_z > F_y$  then
    | S = phase pull;
  else
    | S = not defined;
  end
end
end

```

Here S is the strategy at time t . The strategy is assigned to a data point based on the value of F_z , but also based on the value F_y . By manual inspection, it was seen that the value of F_y was usually centred around zero, which gives a good benchmark for F_z . It is also possible that the value of F_z does not comply with the given rules, then an unknown (not defined) strategy will be assigned to the datapoint. An example of the strategy deduction can be seen in Figure 5-5. The upper figure displays the force curves, and the first horizontal bar displays the deduced strategy.

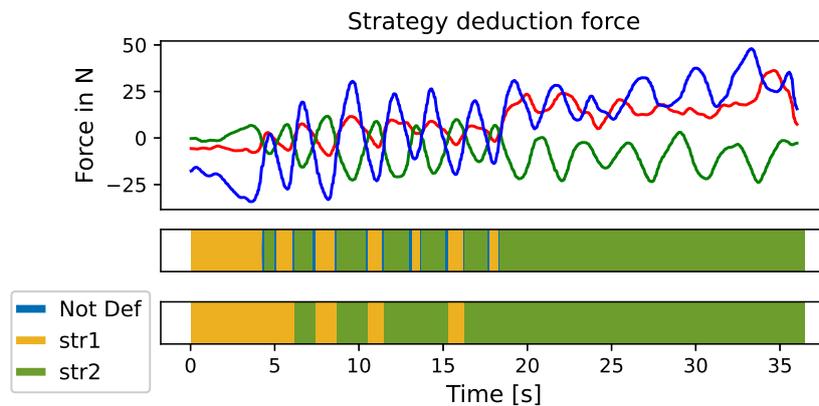


Figure 5-5: Strategy deduction for all timestamps of a force experiment, where str1 is the pushing strategy and str2 is the pulling strategy, and Not Def are the non defined timestamps. The upper bar is the unfiltered strategy annotation and the lower bar is the filtered strategy annotation.

It can be seen that until timestamp $t = 18s$, the strategy is swapping and changing relatively fast from `str1` to `str2`. Based on empirical conclusions from the experiments, advised by clinicians, we assume that it is not possible to change from strategy within a time span of 1 second deliberately. We can use this to filter out the strategies that have a total duration of less than a second. This can be seen in the lower bar of Figure 5-5.

The strategy algorithm is executed on all successful extractions. This yields an additional time series where every sample is annotated with a strategy for force application. This extra time series can be used to extract numerical features, describing the total extraction in a generic way. We can calculate the time a strategy is performed as a percentage of the total time of the extraction. This gives us two features: `f_str1_perc`, which is the percentage of the usage of `str1` during the extraction, and `f_str2_perc`, which is the percentage of the usage of `str2` during the extraction.

5-2-2 Force Complexity

The strategy conversion can also be used to define the cross over point where pushing of the tooth passes over to pulling of the tooth. This cross over point can be used to calculate the complexity of the extraction in the final pulling phase. Figure 5-6 is showing the same force curves and filtered strategies as plotted in Figure 5-5. In addition, the cross over point is plotted, and the force direction changes are marked. The total amount of points in the final pulling phase is a measure of complexity, which is hypothesized to increase when teeth are located towards the back of the mouth. This complexity of the force in xyz -directions is thus extracted from the data as a feature, yielding three complexity features: `fx_complexity`, `fy_complexity` and `fz_complexity`.

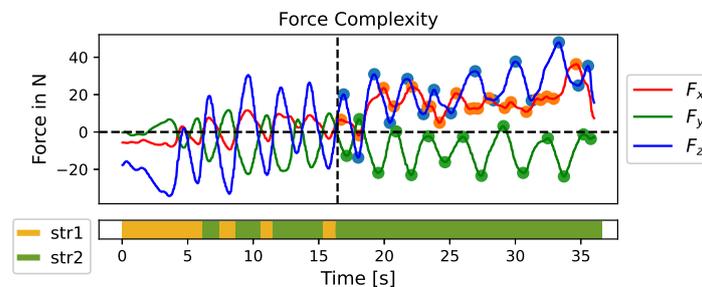


Figure 5-6: Figure displaying how the complexity of the force is calculated. Only the amount of force direction changes within the last part of `str2` count towards the force complexity feature. The dots represent a direction change, which is annotated in all three directions.

5-3 Feature engineering of rotation data

For the rotation data, the hypotheses are stated in Section 5-1-2. It is hypothesized that there is more to tooth removal than only a rocking and twisting motion. With the formed hypotheses, we try to validate this statement by finding extraction strategies in the rotations. As such, this section will focus on modelling the rotations in terms of the y -rotation and the z -rotation. Section 5-3-1 describes how these rotations are modelled and how we can extract

features from this model. Section 5-3-2 is describing the method where the rotation curves are segmented based on direction changes, such that the number of back and forth movements can be modelled. These can be used to extract features that could be predictors for complexity, because more changes in movement direction can indicate an increase in complexity.

5-3-1 Strategies

Currently, the rotation data is in three dimensions (xyz). To make use of a comprehensive type of modelling which is easy to interpret, we exclude the use of the x -direction of the rotations. The data had shown that the rotation in the x -direction is the least present (Section 5-1-2) and the anatomical position of neighbouring crowns prevent the tooth from rotating around the x -axis. In general, the x -axis's contribution to the total amount of rotation is less or equal compared to the y -axis's contribution in 94% of the experiments (64 out of 68 experiments). This was true for both the upper and lower jaw. As such, we assume that the rotation strategy depends on the relation between the yz -rotation. The differences in extraction techniques become apparent when the heatmaps and the corresponding kernel density estimation are plotted. Figure 5-7 show the plots of a typical z -rotation extraction and a typical y -rotation extraction.

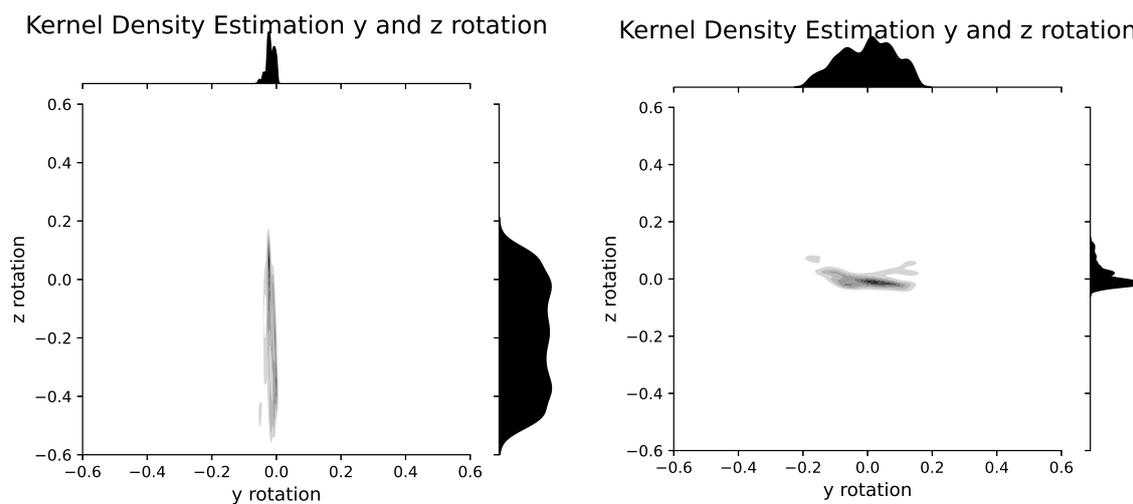


Figure 5-7: Heatmaps resulting from plotting the y -rotation and the z -rotation. The kernel density distributions in y and z -directions are included on the top and the right respectively. The left figure displays a typical heatmap for an upper jaw incisor extraction. The right figure displays a typical heatmap for an upper jaw pre-molar extraction.

The darker the region, the more frequent the corresponding yz -rotation is performed. A clear difference between the overall extraction strategy between an upper incisor on the left and an upper molar on the right can be seen. The incisor extraction is purely z -oriented and the pre-molar extraction is purely y -oriented. Also the amount of rotation is larger for the incisor than for the pre-molar. Appendix A is showing more kernel density graphs of various strategies found in the time series. However, Figure 5-7 is not showing whether and at which time the extraction strategy is changing. From the inspection of the data, it was seen that there are also strategy changes during an extraction. This behaviour is not captured by these

figures and is thus investigated further. Capturing these interchanging behaviour between strategies could be of high importance because the number of strategies can give information about the difficulty of an extraction.

To capture these strategy changes, we first need to find the strategy at every instance in time. We can use the velocity of the rotation curves to see how many rotation in what direction is applied per time stamp. The velocity at an instance in time can be thought of a velocity vector in the yz -plane. In this way, it is possible to build a feature from the coupled behaviour of the single y and z -rotation, capturing the influence of both directions in one metric. We can calculate the angle that this velocity vector makes with the use of the tangens rule:

$$\theta_{yz} = \tan\left(\frac{dz}{dy}\right) \quad (5-1)$$

From this angle, we can build a ruleset where we assign strategies to a certain angle. The left figure of Figure 5-8 is showing the ruleset for the specific angles chosen in a pie chart. The vertical axis is the amount of z -rotational velocity and the horizontal axis is the amount of y -rotational velocity. When plotting the y -component and the z -component, points as p_1 , p_2 and p_3 emerge. Based on its location with respect to the circle, the points fall within the coloured regions. The region is telling which strategy is used for that particular data point. The yellow region is where the y -velocity and z -velocity contribute evenly (20° - 70° and 110° - 160°), the green region is where a pure y -velocity occurs (0° - 15° and 165° - 180°) and the blue region is where a pure z -velocity occurs (75° - 105°). An example of this strategy deduction performed on a whole experiment can be seen in the right figure of Figure 5-8.

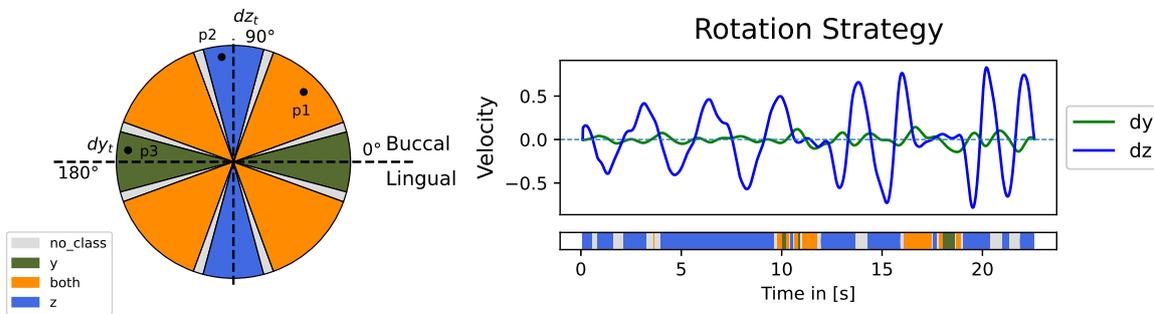


Figure 5-8: Rotation deduction strategy based on the amount of z -rotational velocity applied versus the amount of y -rotational velocity applied at time t . The resulting point on the circle determines which strategy is annotated to the data point. The resulting strategy deduction can be seen in the right plot. The blue curve is the z -velocity, and the green curve is the y -velocity. Plotting the point (dy_t, dz_t) in the circle on the left yields a strategy, indicated by the location of the point. This strategy is calculated for all timestamps in the whole experiment. The resulting strategy deduction for the whole experiment is shown by the bar underneath the right plot. The legend of the circle also indicate the colours in the bar.

It can be seen that the strategy is changing relatively fast, so it should be filtered to be more smooth. We can use a rolling window filter to smooth out the strategy to be more consistent and logical. The size of the window is one second, and it iterates over the whole length of the time series. The results of this filtering are shown in Figure 5-9.

From this filtered strategy, we can extract our features. We assume the majority of the strategy in an experiment is representative for the type of tooth extracted. As such, we can use the percentage of the respective strategies of a whole experiment as a feature. This yields the following features: `y_perc`, `z_perc` and `both_perc`. Furthermore, the number of strategies can be used to discriminate between teeth. The more changes in strategy, the more complex the applied rotation is. A feature is constructed from this, which it is called `#_rot_strategies`.

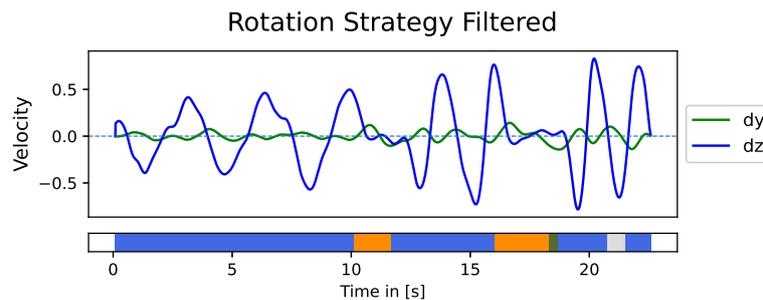


Figure 5-9: The filtered rotation strategy deduction. The colors for the strategy are depicted in the left figure from Figure 5-8. The curve colors depict the velocity in the direction of y and z respectively.

Figure 5-9 shows how the model captures the change in strategy. Around $t = 10s$ the executor is changing from a pure z -rotation to a combination of yz -rotation: an 8-rotation. It can be seen that this is repeated two times, indicating that only a z -rotation is not sufficient enough to extract this tooth.

5-3-2 Rotation Complexity

Knowledge from manual inspection of the data combined with clinical expertise led to a hypothesis that the number of rotations increases when the tooth is located to the back of the mouth. The rotational data is mostly composed of oscillating motions. This is due to the nature of the rotational movement, where back and forth and left and right movements are made periodically. We are interested in counting these amounts of periodic movement, and we can use the velocity curves to segment them out. From Section 5-3-1 we know that the strategy features are build from two dimensions, the y -velocity and the z -velocity. Zero velocity crossing segmentation can only segment a single time series. This means the most dominant rotation direction is chosen to be segmented. The feature values of `y_perc` and `z_perc` are compared and the velocity of the most dominant direction is chosen to be segmented. The locations of the zero velocity crossings are the local maxima or minima of the rotations, representing a change in the rotation direction. The result of the segmentation of a z -rotation can be seen in Figure 5-10 where a clockwise z -rotation is shown in blue and a counter clockwise z -rotation in red. The rotation of the y -direction is shown in green.

From this calculation, we can extract features that are a measure for the rotation complexity. In this way, the segment length is used as a feature, because this varies a lot amongst the demonstrations. This is supported by the clinician's expertise, who agreed that in practice a single longitudinal rotation movement for an incisor takes less time than a bucco-lingual

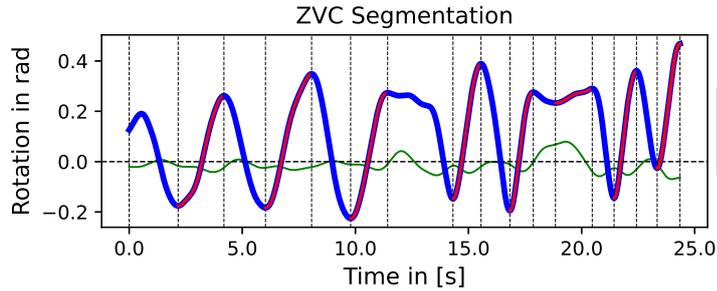


Figure 5-10: Zero velocity crossing segmentation of the rotation time series based on the most dominant rotation. In this figure, the rotation in z -direction is segmented. The blue curves represent a clockwise rotation in z -direction and the red curves represent a counter-clockwise rotation in z -direction. The green curve is the rotation in y -direction. The horizontal lines are the respective segments that are introduced by the zero velocity crossing method.

rotation for a molar. Hence we can use the `mean_seg_length` as a feature. In addition to this, we can count the number of segments. This is a measure of rotation complexity, where more direction changes can indicate a higher level of complexity. As such, the feature is defined as the number of direction changes: `dir_changes`.

5-3-3 Overview of all features

A visual overview of the features and their extremes is summarized in Table 5-2 and Table 5-3. For convenience, the features and its textual interpretations are restated in Section 7-1-1 and Section 7-1-2 as well.

Table 5-2 is showing the extreme values of the features deduced from the force data. The first column shows the feature name, the second and third columns show the minimum extreme value of the feature and the maximum extreme value of the feature.

The plotted curves are the transformed and filtered force curves in x , y and z -direction respectively. The legend in Figure 5-11 show the corresponding colors for the force directions. Underneath the force curves, a coloured bar can be seen. This is the strategy bar, showing the annotation of the pull strategy (`str1`) or push strategy (`str2`) per time stamp.

For `f_str1_perc`, the minimum extreme value has a small `str1` percentage, as indicated by the yellow colour in the strategy bar. The maximum extreme value has a high `str1` percentage, shown by a large yellow area in the strategy bar.

For `f_str2_perc`, the opposite is shown. The green area (`str2`) in the strategy bar is small for the minimum extreme value, where it is large for the maximum extreme value.

The last row shows the complexity features (`f_xyz_complexity`). For all three force directions, this complexity is based on the amount of force direction changes, as indicated by the coloured dots. The strategy bar is included because the complexity is only measured from the last phase of the total extraction, indicated by the dotted vertical line.

Table 5-3 is showing the extreme values of the features deduced from the rotation data. The first column shows the feature name, the second and third columns show the minimum extreme value of the feature and the maximum extreme value of the feature. The plotted curves in row 1-4 of Table 5-3 are the transformed and filtered rotation curves in y -direction

Table 5-2: Force features summarized with examples of the minimum extreme values and the maximum extreme values of the respective feature.

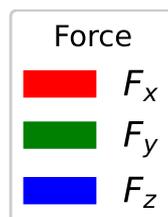
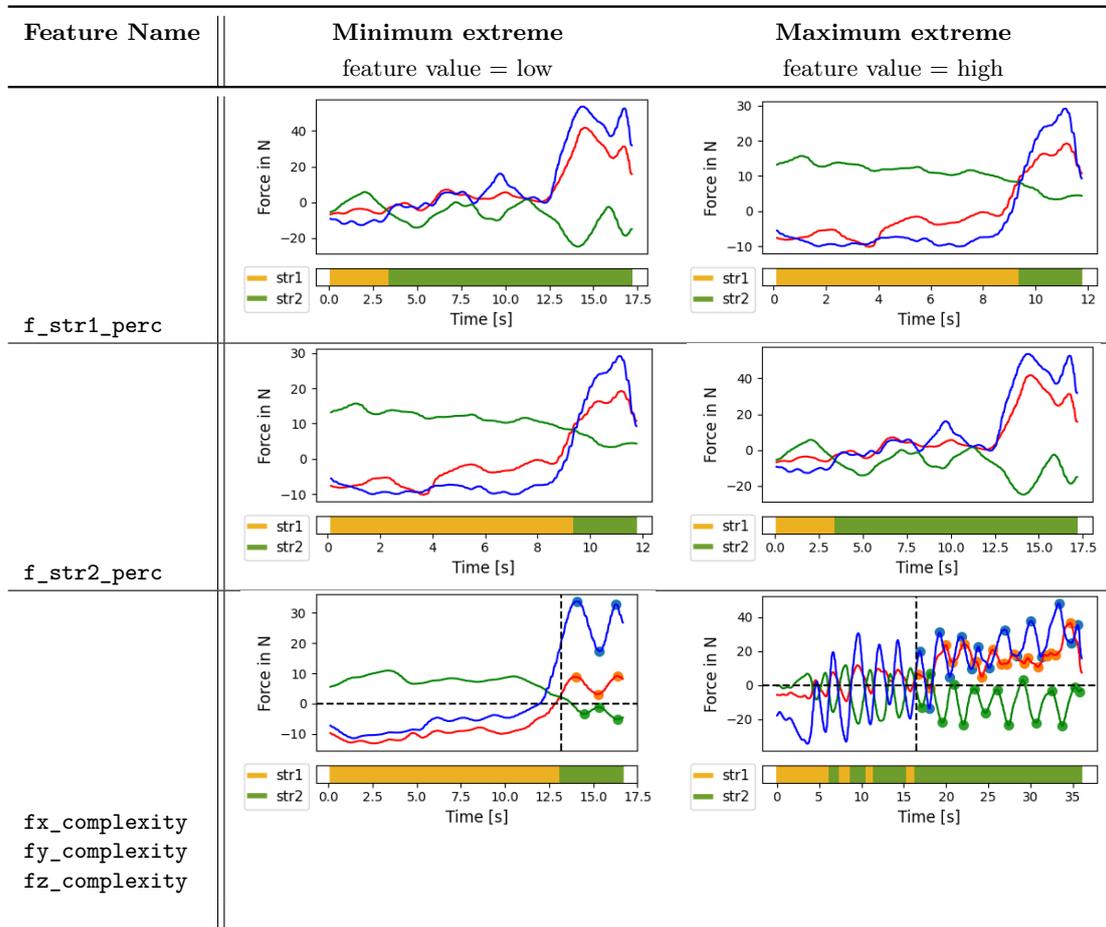


Figure 5-11: Legend for the force curves as depicted in Table 5-2.

(green color) and in z -direction (blue color). The left figure of Figure 5-12 shows the legend with the corresponding colors. Underneath the rotation curves, a coloured bar can be seen. This is the strategy bar, showing the annotation of the rotation strategy used per time stamp. The right figure of Figure 5-12 shows the legend of the strategy bar with the corresponding colors. The colour of this strategy bar indicates in which direction (unknown, y , z , both yz) the rotation at time t is performed, based on the ruleset of Section 5-3-2.

For y_perc , z_perc and $both_perc$, the minimum and maximum feature values are based



Figure 5-12: Legends for row 1-4 in Table 5-3. The left legend is the legend for the rotation curves, with the green color representing the y -rotation and the blue color representing the z -rotation. The right legend is the legend for the strategy bar, with the white color representing no strategy, the green color representing a y -rotation dominant strategy, the blue color representing a z -rotation dominant strategy and the yellow color representing a yz (both) dominant strategy.

on the percentage of the rotation in y , z or yz -direction respectively. This is indicated by a small or large covered area of the respective colour in the strategy bar.

For `#_rot_strategies`, the amount of strategy changes in the strategy bar is counted. The minimum value shows no change (one single colour), while the maximum value shows a lot of strategy changes (multiple changing colours).

Row 5-6 of Table 5-3 show the minimum and maximum values for `mean_seg_length` and `dir_changes`. These features are based on the segmentation of the dominant rotation, as indicated by the vertical dotted lines. Figure 5-13 shows the legend with the corresponding colors of the segmented rotation curves. In all four figures of row 5-6 of Table 5-3 the z -rotation was the dominant rotation, so this line is segmented. The red parts indicate a clockwise rotation around z , while the blue parts represent a counter-clockwise rotation around z . The green line is the y rotation. For `mean_seg_length`, the left figure displays a minimal distance between the segments and the right figure displays the maximum mean distance between the segments. For the `dir_changes`, the left figure displays the least amount of segments. The right figure displays the most amount of segments.

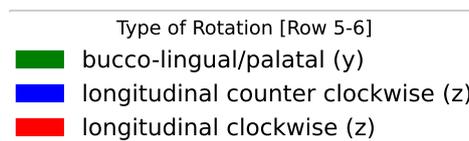
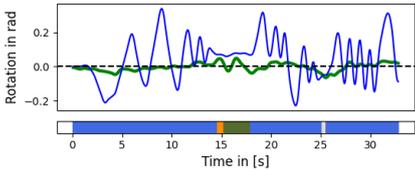
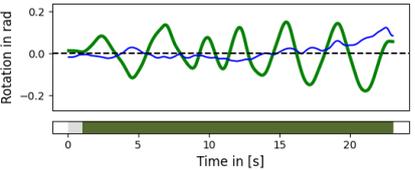
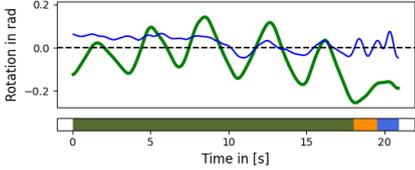
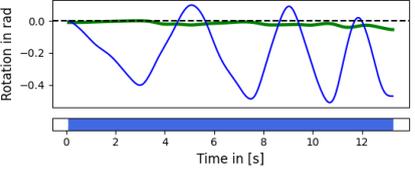
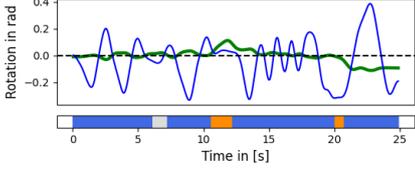
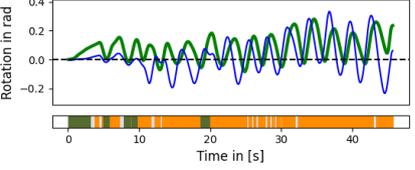
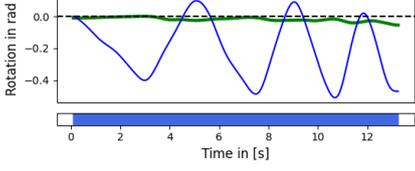
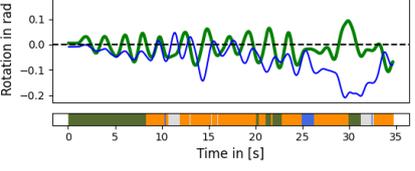
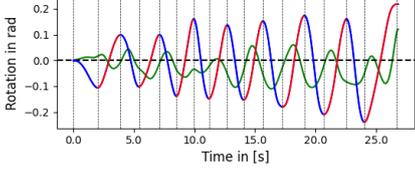
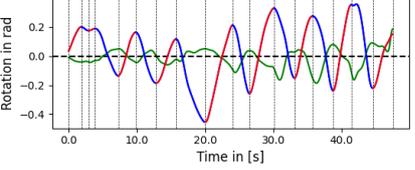
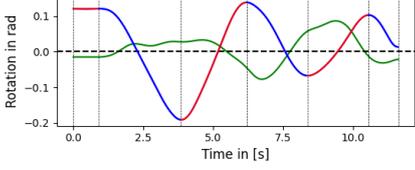
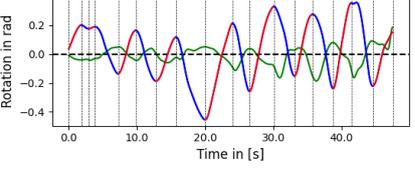


Figure 5-13: Legend for the segmented rotation curves on row 5-6 in Table 5-3. The green color is representing the y -rotation, the blue color is representing a counter-clockwise z -rotation and the red color is representing a clockwise z -rotation.

Table 5-3: Rotational features summarized with examples of the minimum extreme values and the maximum extreme values of the respective feature.

Feature Name	Minimum extreme feature value = low	Maximum extreme feature value = high
y_perc		
z_perc		
both_perc		
#_rot_strategies		
mean_seg_length		
dir_changes		

5-4 Summary

This chapter described the feature engineering process, where relevant numerical features are extracted from the pre-processed time series. This chapter started with the formation of four hypotheses: two hypotheses for the force data and two hypotheses for the rotational data. We started the feature extraction based on these hypotheses, guiding the process towards clinically interpretable models.

From the force data, it was possible to deduce a strategy per time sample based on the vertical force F_z . We found a pushing and pulling strategy, which we were able to express in percentages of the total experiment time. This resulted in the `f_str1_perc` and `f_str2_perc` features. Furthermore, we found that we could express the force complexity as the number of directional changes during the pulling phase after the cross over point. We did this for the force in x , y and z direction, resulting in the `fx_complexity`, `fy_complexity` and `fz_complexity` features.

From the rotational data, we deduced a strategy measure and a complexity measure. The strategy is based on the type of yz -rotation performed and can be calculated using the rotational velocity. At every instance in time, the magnitude of the dy -component versus the magnitude of the dz -component yields a strategy based on the direction of the resultant. This strategy deduction can be used to find the dominant rotation in an experiment, while also capturing changes of strategies within this experiment. The resulting features from this strategy model are the percentage of rotational velocity in y direction, the percentage of rotational velocity in z direction and the percentage of rotation in both directions; `y_perc`, `z_perc` and `both_perc` respectively. The amount of changes in strategy is also used as a feature, named `#_rot_strategies`. Furthermore, the complexity of a rotation is investigated. With Zero Velocity Crossing (ZVC) segmentation, the dominant rotation time series is segmented and the changes in rotation direction are captured. The number of segments and the length of these segments are the resulting features that are a measure of complexity: `dir_changes` and `mean_seg_length`.

Prediction Modelling

This chapter describes the prediction modelling process that is performed to make predictions between different groups of teeth. This prediction model builds on top of the feature engineering model that is described in Chapter 5. It takes features in table format as an input, and by applying a classification method, it outputs a prediction of a specific class. The pipeline that will be used as reference can be seen in Figure 6-1.

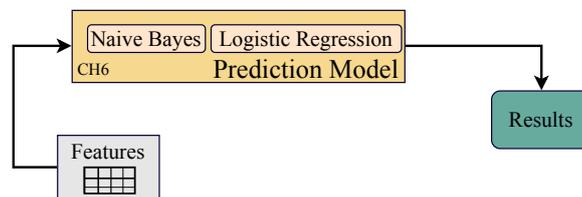


Figure 6-1: Classification learning pipeline from the approach of Section 1-3. The feature table is used as an input the the prediction model. The prediction models are the Naïve Bayes (NB) model and the Logistic Regression (LR) model. By classification, these models provide predictions, yielding results and conclusions.

This chapter starts with the description of the data structure and how the instances of the table are labelled such that a supervised learning algorithm can be applied. Next, the learning methods that will be used to analyse the feature table are discussed. We will make use of a NB algorithm that serves as a baseline model. A LR algorithm will be used to compare performance. At last, the techniques used to assess the performance of the models during the implementation and comparison are discussed, as well as the methods that describe feature importance.

6-1 Feature table pre-processing

This section describes how the input feature table could be modified such that it could be used in a supervised prediction model. First the labelling is discussed in Section 6-1-1, where

classification groups are used as classes. These classes used are described in the results, Section 7-1. Section 6-1-2 shows how the features can be scaled such that they are represented to the algorithms on the same scale. This makes sure all feature distributions have zero mean and a standard deviation of one so that the prediction algorithm weighs their importances equally. At last, prediction modelling consists of a training phase and a testing phase, each with separate parts of the feature table. Section 6-1-3 describes how the feature table is split such that the test set is a representative part of the data.

6-1-1 Labelling

As can be seen in Figure 6-1, the input data is the feature table as stated in Section 5-4. The rows of this table represent the experiments, the columns represent the features derived with the feature engineering process. In order to use a prediction model successfully, the experiments should be labelled with a class. Our goal is to find a set of features that is descriptive for a specific group of teeth. When we are able to develop a model that renders high performance, we can argue that the features are good predictors for the specific groups that we made. As such, these features render descriptiveness between teeth.

Table 6-1: Adjusted feature table where the labelled class is shown in column '*label*'. This label is used in the supervised learning algorithm to fit the feature values in the training and testing phase.

Exp #	label	f_str1_perc	f_str2_perc	fx_complexity	...	y_perc	dir_changes
Exp 1	Class 1						
Exp 2	Class 2						
⋮	Class 3						
Exp <i>n</i>	Class 2						

Table 6-1 is displaying the input format of the table for the prediction model. The features from Section 5-3-3 are placed in the columns, the experiments are placed in the rows. A column displaying the *label* is added. This label is the desired output and can be of any kind: a group of teeth, an extractor, the number of roots or the root size. As long as it is a categorical value, we can use it as our class in the prediction algorithm. Section 7-1-2 is showing the results from feature engineering, which includes the classes that are used in the prediction algorithm.

6-1-2 Standardization

Furthermore, some pre-processing needs to be done on feature level. Most of the machine learning algorithms are using Euclidean distance measures in determining similarity [28]. When features are differing drastically in scale (thus distance), the algorithm can give more weight to the feature with a higher magnitude. In our case, the force complexity can range from [0:25], and the rotational velocity percentage ranges from [0:1]. This will result in an unwanted bias towards the feature with the smallest distance, which we want to avoid. To make the features more comparable, we can use a scaling method to scale the features to the same range so the prediction algorithm can compare them evenly. The standardization

method as described in Section 2-2-2 is used in favour of the MinMax scaling option. The MinMax scaler is sensitive to outliers and is mostly used for small standard deviations. The values are scaled between a range of 0 and 1. So the occurrence of one outlier can reduce the performance drastically. Our dataset has some outliers which should be taken into account more smoothly than done with MinMax scaling. They should not be scaled to the same bounding range. Figure 6-2 is showing the effect of standardization on the raw feature data for two arbitrary features for the dataset of the main surgeon ($n=68$).

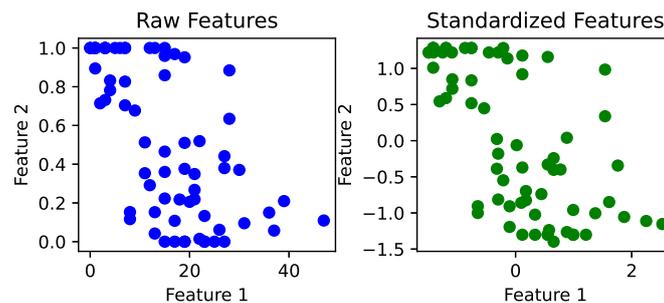


Figure 6-2: Example of how standardization affects the distribution of two features from the dataset. In the left figure two random unscaled features are plotted against each other and in the right figure the standardized features are plotted

6-1-3 Train/test split

The development of the classification model consists of a training phase and a testing phase. Both phases use different portions of the available feature table, which is called the train/test split. During training, the model uses the labels of the training data to learn decision making. Next, this decision making is applied to the test set to validate the predictions. The splitting of this data is done randomly, where an x percentage of the data is chosen to represent the test set, and the rest is chosen to be the training set. With the splitting of this dataset, a significant tradeoff has to be made. When the amount of training data is high, the model improves on training accuracy, because the model has more samples to learn from. On the other side, the amount of test data needs to be representative enough to assess the performance of samples that are not seen yet. For instance, if a class in the test set has only 1 sample and the training set has 99 samples, this is considered as a poor train/test split. In our models, a percentage of 35% for the test set is used. Figure 6-3 shows how this percentage is determined. A Gaussian Naive Bayes model is used to assess the training and testing accuracy for 500 random train/test splits between 20% and 50% test set size. A test set size of 45% is performing slightly better than a test set size of 35%, while yielding slightly more outliers than a test set size of 35%. Our dataset is relatively small with 68 samples, so to keep the training accuracy high while having a representative test set, a test set size of 35% is chosen.

6-2 Model Implementation

This section describes what the characteristics of the implemented Naïve Bayes and Logistic Regression models are.

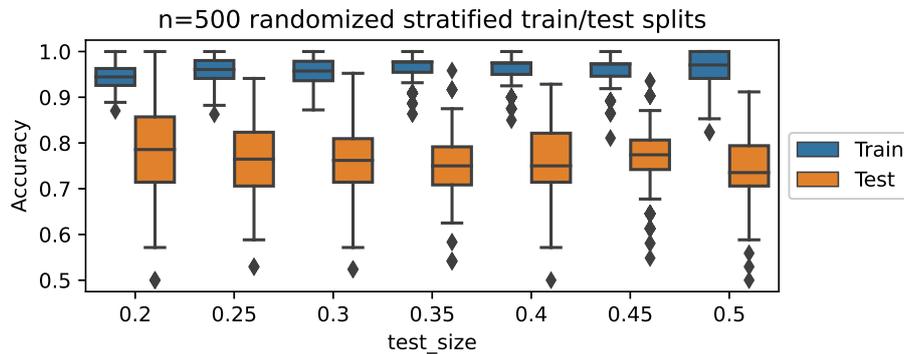


Figure 6-3: Accuracy distributions plotted for 500 randomized stratified train and test sizes, categorized per test size fraction. In blue the distribution for the the training set is plotted and in orange the distribution for the test set is plotted. The diamonds indicate single outlier samples.

6-2-1 Naïve Bayes implementation (baseline model)

The training data can now be used to make predictions. First, the training set is fitted to the Naive Bayes model, which serves as a baseline model. A Gaussian implementation is chosen, where the classes are fitted based on the Gaussian distribution as described in Equation (2-7). The distribution consists of μ_y and σ_y^2 , being the mean and the variance of class x_i respectively. The Naive Bayes model is easy to implement, having no parameters to tune. The only implementation to make is the choice for a Gaussian model because the feature values are continuous.

6-2-2 Logistic Regression implementation

The second algorithm used is a Logistic Regression algorithm. Section 2-2-3 describes how the main principle of Logistic Regression is based on the sigmoid function of Equation (2-8). The sigmoid function of the Logistic Regression model only allow for binary classification.

For our application however, a multiclass classifier is needed. The use of a One-Versus-Rest (OVR) solving strategy can make this happen. The binary classifier will be trained for each class, and the class that has the largest probability returned by Equation (2-8) is chosen as the respective class. In this way, the binary Logistic Regression model is transformed into a multiclass classifier.

6-3 Model performance

This section describes how the performances of the supervised prediction models of Section 6-2 are assessed. In addition to the confusion matrix, where techniques of accuracy, precision and recall are used to calculate performance, we can also make use of the cross-validation technique. Cross-validation makes use of multiple (not necessarily random) train/test splits to assess the performance of the algorithm on unseen data. It splits the data in k -folds, repeating the train/test splitting method multiple times. It is different from the simple train/test

splitting method because in cross-validation, the algorithm ensures that all samples from the original dataset end up in both the training set and the test set. An example of the regular k -fold cross-validation that is usually performed in machine learning is shown in Figure 6-4. This figure is constructed with the dataset samples of the main surgeon and the classes from Section 7-1-2.

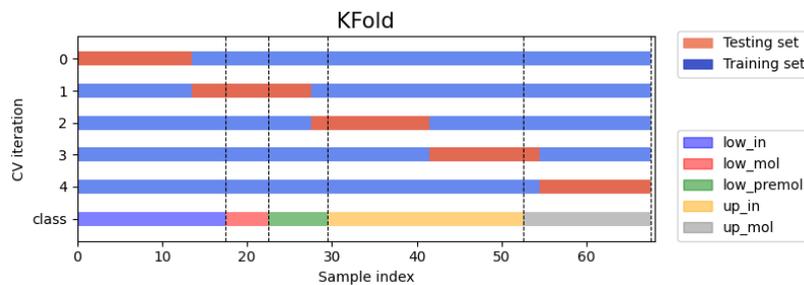


Figure 6-4: A k -fold cross validation procedure performed on the dataset of the main surgeon ($n=68$). The respective train/test splits over 5 cross validation iterations is depicted in the first 5 rows, where a red colour represents the samples from the test set and the blue colour represents the samples from the training set. The sample division is based on the classes from Section 7-1-2 and is shown in the last row.

This figure shows how the test set is 1/5 of the total amount of samples. In the second iteration (annotated with 1) however, it can be seen that the test set spans the whole `low_mol` class and almost the whole `low_premol` class. This means that the algorithm could not train on the characteristics of these classes because none or limited data is available. The performance will drop because the test set is not 'representative' for the training set. A more representative split on the test set is needed, so the division of samples in the test set should be performed differently than in regular k -fold cross-validation. This can be accomplished by introducing a stratified k -fold cross-validation. Figure 6-5 is displaying this cross-validation where 5 iterations are performed.

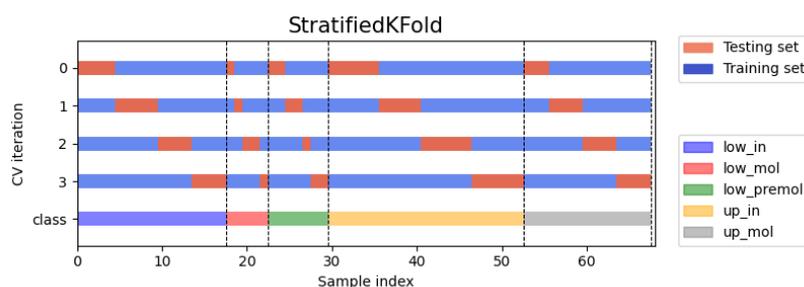


Figure 6-5: Stratified k -fold cross validation procedure performed on the dataset of the main surgeon ($n=68$). This figure shows that the training set samples and the test set samples are evenly divided over the number of samples in each class.

This figure shows that the train/test splits are evenly divided (stratified) over the number of samples that exists in a class. In this way, all five classes are represented in the test set, so that the test set becomes a good representation of the classes in the original dataset. The stratified k -fold cross-validation will be used in analysing the performance of the models on

unseen data.

6-4 Feature importance

This section describes how the impact of the features on the model performance can be assessed. Section 6-4-1 describes best subset selection, Section 6-4-2 describes forward stepwise selection and Section 6-4-3 describes the SHapley Additive exPlanations (SHAP) implementation.

6-4-1 Best subset selection

Best subset selection will be employed to demonstrate the differences between the predictive power of the NB algorithm and the LR algorithm. Best subset selection tries to find the best combination of features for n amount of features [49]. It iterates through the amount of features $n = 1, \dots, k$, with k being the maximum amount of features. At every iteration it fits $\binom{k}{n}$ models, yielding a total amount of 2^k models. At every iteration, the model with the highest accuracy is assigned the best subset. Our model contains 11 features, so a total of $2^{11} - 1 = 2047$ models have to be fitted.

6-4-2 Forward Stepwise Selection

If the amount of features n increases, the computation time of the best subset selection algorithm described above increases exponentially. Additionally, the best subset selection does not provide information about which features are responsible for the model to perform in this way.

The forward stepwise selection algorithm has a different approach [50]. It starts with fitting a model where 0 predictors are present, and it iterates through all n amount of features. Every iteration, the best improvement on the model for the remaining features is calculated, and the feature that gives the best additional performance on the current model is added. This continues until all features are included in the model.

6-4-3 SHapley Additive exPlanations (SHAP) values

Additionally, SHAP values can be used to define the individual contribution of a feature on the model's performance [51]. It is based on the Shapley value that is mainly used in game theory. Game theory describes how two or more players are involved in a strategy to achieve a desired outcome or payoff. In this way, Shaply values could determine a payoff for the involved players and their contribution. In machine learning, this is used to determine the contribution of a feature to the model output. SHAP values are computed for single samples from the dataset, so they are a local interpretation of the supervised learning algorithm [51]. They do not explain the working principle of the algorithm: they only provide interpretations of the features.

6-5 Summary

This chapter described the classification learning process, where the input features in the form of a table are managed to be used in a prediction model that yields good and reliable accuracy.

First, the process of class labelling was explained. Next, the features were standardized to eliminate high scaling differences amongst the features. Next, the type of train/test set splitting is discussed where evidence was given for the split amount. Furthermore, the implementation of two prediction models was discussed. An implementation of a NB algorithm was given to serve as a baseline model. An implementation of a LR algorithm was given to compare the baseline model to. At last, we showed how cross-validation could be used to assess the performance of the prediction model on new, unseen data, and we showed how best subset selection, forward subset selection and SHAP values could be used to identify feature importance.

Chapter 7

Results

This chapter describes the results of the modelling of tooth removal. Section 7-1 states the results of the feature engineering process that leads to features that can be distinguished in tooth removal procedures. The deduction of these features results in key insights. The results of the force analysis and the results of the rotation analysis are discussed separately.

Section 7-2 discusses the outcomes of the classification learning process. The features found from the feature engineering are fed into multiple prediction models. First, the results of a Naïve Bayes (NB) algorithm are discussed. Limited number of parameters can be altered in a NB model, therefore the NB model serves as a baseline model. The results of the Logistic Regression (LR) model are compared with the results of the NB model. Subsequently, significant differences in accuracy and stability with respect to the baseline model are given. At last, the results of the feature importance analysis are given. The results of forward stepwise selection and SHapley Additive exPlanations (SHAP) values are discussed, supporting the interpretations of the predictions made by the LR model.

An overview of the total dataset and the study population is included in Section 3-5-1. It describes the properties of the dataset including, but not limited to, the number of successful extractions, the number of executors and the number of complications.

This chapter show graphs containing datapoints for incisors and (pre-)molars. The teeth [11, 12, 13, 21, 22, 23] and [31, 32, 33, 41, 42, 43] are included as incisors. The teeth [14, 15, 16, 17, 24, 25, 26, 27] and [34, 35, 36, 37, 44, 45, 46, 47] are included as (pre-)molars

7-1 Results of the Feature Engineering process

The output of the feature engineering process are features that seemed useful for the prediction model. This section describes how these features describe differences between tooth extractions. Section 7-1-1 shows how F_z and the complexity of the force curves contribute to differences between teeth. Section 7-1-2 shows how the features derived from the rotation can indicate differences between teeth.

7-1-1 Descriptiveness in force features

The process of feature extraction of the forces described in Section 5-2 led to the following features: `f_str1_perc`, `f_str2_perc`, `fx_complexity`, `fy_complexity` and `fz_complexity`. Table 7-1 states these features and their descriptions.

Table 7-1: Explanation of the features deduced from the force data.

Feature Name	Feature Description
<code>f_str1_perc</code>	Percentage of the total extraction where $F_z < 0$, indicating vertical pushing
<code>f_str2_perc</code>	Percentage of the total extraction where $F_z > 0$ and $F_z > F_y$, indicating vertical pulling
<code>fx_complexity</code>	Amount of force direction changes in x -direction during the final force application strategy
<code>fy_complexity</code>	Amount of force direction changes in y -direction during the final force application strategy
<code>fz_complexity</code>	Amount of force direction changes in z -direction during the final force application strategy

The features of the force data can be analysed on two levels. First, on an anatomical level. Anatomical data such as type of tooth (incisor or molar), jaw type (upper or lower) or amount of roots are used to compare feature values amongst different teeth. This should be done with consistent data, preferably from a single person. As such, the data from the main surgeon (68 samples) will be used for this analysis. Second, features are looked for that can distinguish between different surgeons. This is done for two main reasons: validating if the found model generalizes to unseen datasets and comparing the techniques of multiple executors amongst each other. The analysis on executor level is carried out with the data of three surgeons (116 samples) and one student (15 samples).

Descriptiveness on anatomical level

The features `f_str1_perc` and `f_str2_perc` are both expressed in percentages of the same extraction. This means they are strongly related. When we plot both percentages against each other and add information about the complexity of F_z , the tooth type and root amounts, we get the plot in Figure 7-1. In this way, we can investigate the influence of the type of tooth or the number of roots on the strategy of force application.

The left figure displays the pushing percentage (`f_str1_perc`) versus the pulling percentage (`f_str2_perc`). The datapoints of the molar teeth are situated more in the lower right quadrant, and the datapoints of the incisor teeth are situated in both upper left and lower right quadrant, indicating either a higher than 50% of pulling percentage or a higher than 50% of pushing percentage. The multi-rooted teeth are mainly extracted using a majority of pulling force. Furthermore, no distinction for single rooted teeth can be argued for these features based on Figure 7-1.

The right figure is displaying the complexity in F_z direction on the vertical axis versus the amount of pulling percentage normalized from 0 to 1 on the horizontal axis. In this dataset, multi-rooted teeth always require more than 12 F_z complexity, which is indicated with the black horizontal line. In general, most incisors are situated under this line of 12 F_z complexity, but still the distribution of the incisors in terms of F_z complexity is overlapping with the distribution of the molars.

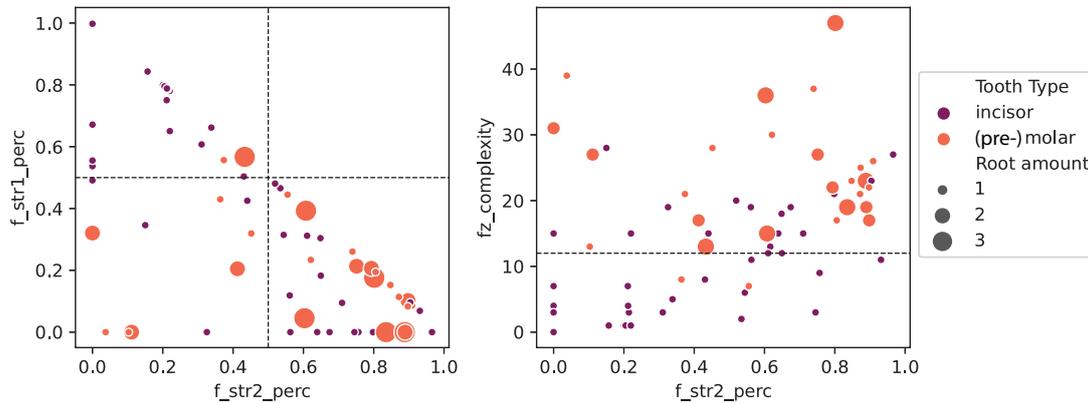


Figure 7-1: The demonstrations of the main surgeon ($n=68$) plotted for a selection of force features. The left figure shows the pushing percentage (f_{str1_perc}) on the vertical axis versus the pulling percentage (f_{str2_perc}) on the horizontal axis, both normalized from 0 to 1. The right figure shows the force complexity in F_z direction on the vertical axis versus the pulling percentage (f_{str2_perc}) on the horizontal axis normalized from 0 to 1. The scatter size is based on the amount of roots of the tooth, and the colour grading of the samples distinguishes based on the tooth type.

Figure 7-1 does not discriminate between jaw type. Figure 7-2 displays the same axis and datapoint values as Figure 7-1, only the colour grading and marker size is adapted. In this new figure, the colour indicates the jaw type (upper jaw, lower jaw), and the marker size indicates the tooth type (incisor, (pre-)molar).

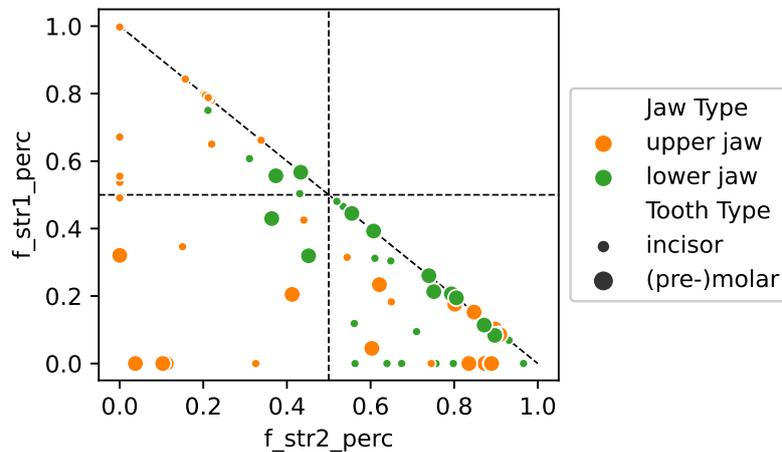


Figure 7-2: The pulling percentage (f_{str2_perc}) on the horizontal axis versus the pushing percentage (f_{str1_perc}) on the vertical axis of 68 demonstrations of the main surgeon. The tooth type (incisor or (pre-)molar) is described by the scatter size and the type of jaw is described by the scatter color

From this figure, a higher level of descriptiveness occurs. The variance in the distribution of the lower jaw samples is much smaller than the variance in the upper jaw samples. The lower jaw samples are mostly situated in the lower right quadrant. Most lower jaw molars are

situated on the diagonal line that is plotted, whereas the incisors are not. This diagonal line represents the situation that the addition of the pushing percentage and the pulling percentage gives 100%. When a point is situated on this line, all timestamps in an experiment are thus annotated with either a pushing strategy or a pulling strategy, and no datapoint is annotated with an unknown strategy. The more datapoints are situated on this line, the better the pushing/pulling strategy deduction system works. Especially the lower molars are situated on this line, indicating the force strategy deduction algorithm performs best for lower molars.

In addition to analysis of the F_z strategy features, also the degree of descriptiveness in the complexity features is investigated. Figure 7-3 shows two combinations of plotting the complexity in all force directions sequentially. Both plots show the amount of complexity for incisor teeth is distributed towards the lower left of the plot, while also overlapping with the (pre-)molar distribution significantly. Furthermore, only the left plot shows a significant level of descriptiveness. The upper jaw samples are almost fully separated by a 45 degree line from the lower jaw samples in plotting the xy complexity.

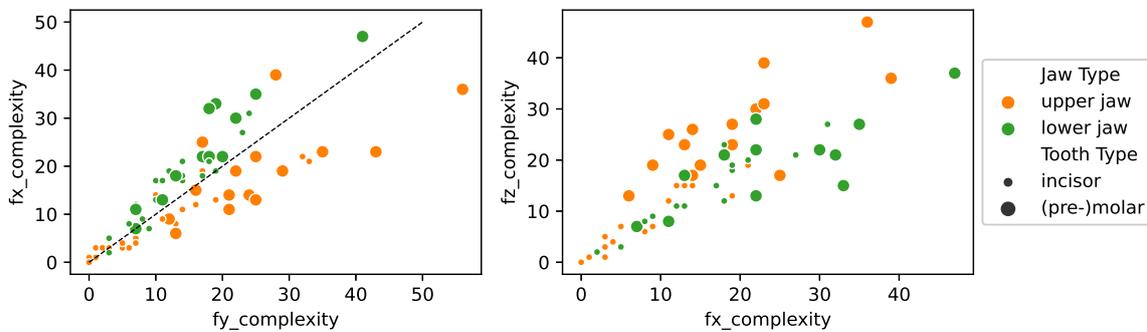


Figure 7-3: Force complexity in all xyz -directions plotted against each other sequentially. The size of the marker indicates the type of tooth and the color of the marker indicates the jaw type.

Except for a clear separation between upper jaw teeth and lower jaw teeth in F_x complexity and F_y complexity, plotting the jaw type did not give satisfactory results to argue for descriptiveness in tooth type. To quantify the suspicion that the distributions for incisor teeth differ from the distributions of molar teeth, the boxplots in Figure 7-4 are included.

The horizontal axis is displaying the three directions in which the force can be applied, and the vertical axis is showing the respective complexity. Comparing the force direction amongst each other, the median complexity of the incisors show the same complexity and the median complexity of the molars show the same median complexity. Furthermore, the lowest 25% of incisor observations have a significant smaller variance than the biggest 25% of incisor observations. Comparing tooth type wise, the medians of the molar complexity are higher than the medians of the incisor complexity. Furthermore, there is a separation between the interquartile ranges of the incisor and molar distributions in the F_y and the F_z direction.

Interoperability descriptiveness

We can further investigate the descriptiveness of the force strategy deduction system when looking at differences between executors. The boxplot in Figure 7-5 displays the push

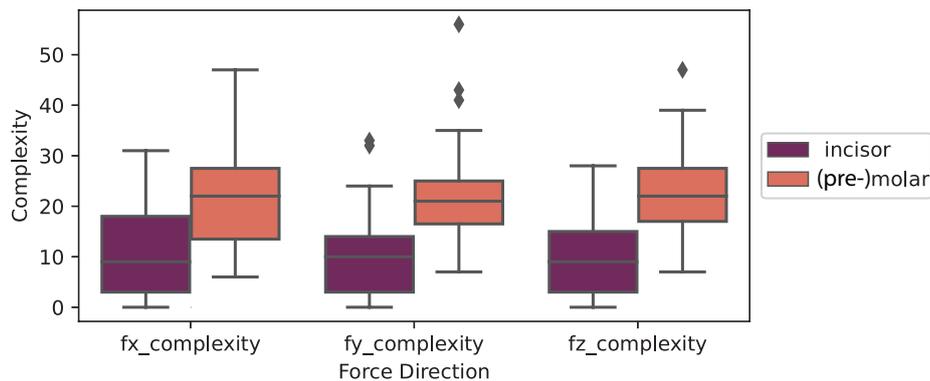


Figure 7-4: Distribution box plots of the complexity in all three force directions. The purple color is indicating the distributions related to the incisor teeth and the orange color is indicating the distributions related to the (pre-)molar teeth.

(`f_str1_perc`) and pull (`f_str2_perc`) strategy distribution of the four executors. The first column displays the distribution of the main surgeon. The second and third column displays the distribution of the two other surgeons, and the last column displays the distribution of the student. The distributions of the main surgeon and surgeon 2 are showing the most similarity in terms of median push and pull percentage. The distributions for the student are deviating the most from all other distributions in terms of pulling strategy variance. Furthermore, a clear distinction between the pushing and pulling distribution is seen. It shows no overlap between push and pull distributions whatsoever.

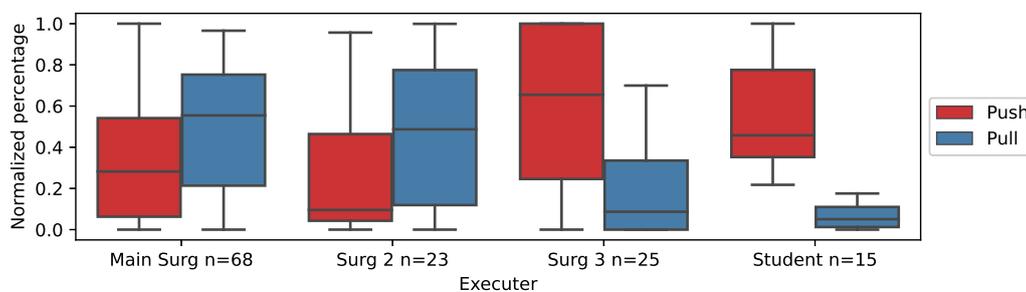


Figure 7-5: Distribution box plots of the normalized push and pull percentages for all executors. The number n is indicating the amount of samples that are spanning the distributions. The amount of push percentage `f_str1_perc` is depicted in red and the amount of pull percentage `f_str2_perc` is depicted in blue.

We can take a look at how the actual distribution of the student is constructed by plotting the pushing percentage (`f_str1_perc`) versus the pulling percentage (`f_str2_perc`). Figure 7-6 is showing this plot. It stands out that all datapoints are situated at the left of this plot (which was expected based on the student's small and low pull distributions in Figure 7-5). It can be seen however, that only one sample is situated on the diagonal line, indicating the strategy deduction system returns high percentage of not defined timestamps.

In addition to the strategy deduction system, we can also compare the distributions of the

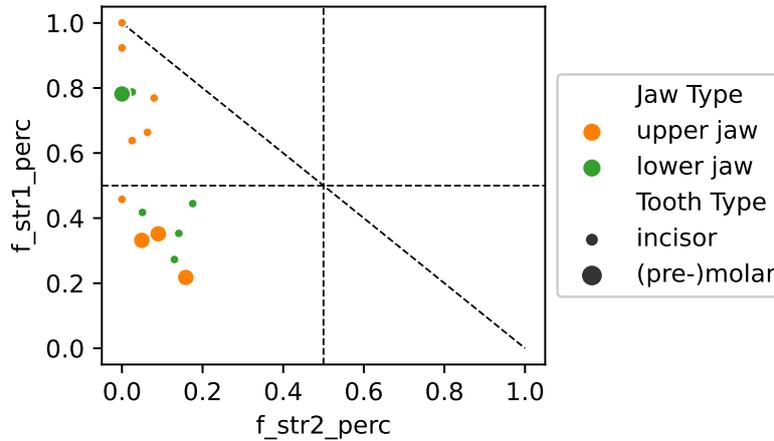


Figure 7-6: Features f_str1_perc and f_str2_perc plotted against each other for the samples of the student ($n=15$). The type of tooth is indicated by the marker size and the type of jaw is indicated with the marker colour.

force complexity amongst all four executors. This can be seen in Figure 7-7.

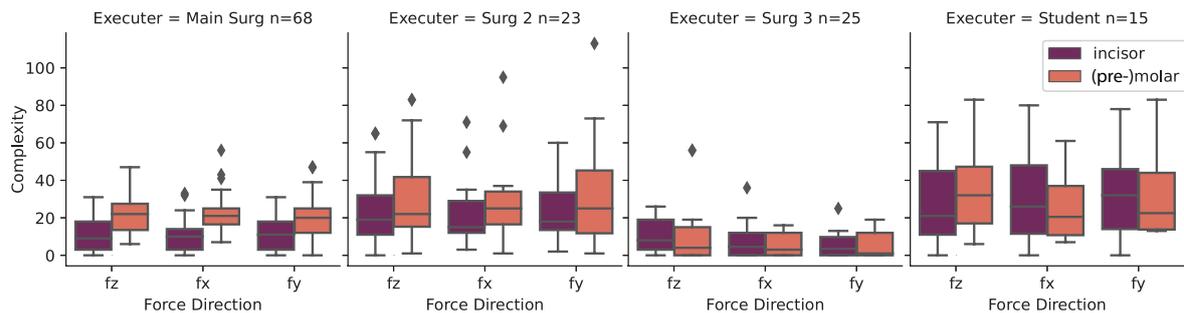


Figure 7-7: Distribution box plots of the complexity in all three force directions for all executors. The purple color is indicating the distributions related to the incisor teeth and the orange color is indicating the distributions related to the molar teeth.

From this figure, we can see that the expected behaviour of descriptiveness between median complexity between the incisor distribution and the molar distribution can only be found at the main surgeon. All other three executors have overlapping distributions for incisors and molars in all three force directions. Furthermore, the distributions of the student are standing out the most in terms of variance. The distributions of the main surgeon and surgeon three are the most narrow in general. Furthermore, the distributions of surgeon 2 and the student are most alike. This is probably related to the long execution times both executors adopted, influencing the complexity feature.

7-1-2 Descriptiveness in rotation features

The process of feature extraction of the rotational data described in Section 5-3 led to the following features: y_perc , z_perc , $both_perc$, $mean_seg_length$, $dir_changes$ and

#_rot_strategies. Table 7-2 states these features and their description.

Table 7-2: Explanation of the features deduced from the rotation data with their descriptions.

Feature Name	Feature Description
y_perc	Percentage of the total tooth extraction the forceps is rotating around the y -axis (bucco-lingual in lower jaw or bucco-palatal in upper jaw)
z_perc	Percentage of the total tooth extraction the forceps is rotating around the z -axis (longitudinal rotation)
both_perc	Percentage of the total tooth extraction the forceps is rotating the 8-movement (combination of yz -rotation)
mean_seg_length	Mean length of a segment (one back/forth movement) of the main axis of rotation
dir_changes	Amount of direction changes in the main axis of rotation (amount of segments)
#_rot_strategies	Total number of strategies (y_perc, z_perc, both_perc) used to remove a tooth

In analysing the features deduced from the rotational data we can, just as with the force features, make a distinction between descriptiveness on anatomical level or executor level. For the sake of consistency, also the results of the rotational feature extraction will be discussed by means of this same structure, with the dataset of the main surgeon (n=68 samples) and the dataset of all executors (n=131 samples) respectively.

Descriptiveness on anatomical level

Using the same plotting technique as for the forces, we can plot the percentage of rotation in y -direction (y_perc), the rotational velocity in z -direction (z_perc) and the rotational velocity in yz -direction (both_perc) for all the experiments of the main surgeon in a three-fold figure. This is shown in Figure 7-8.

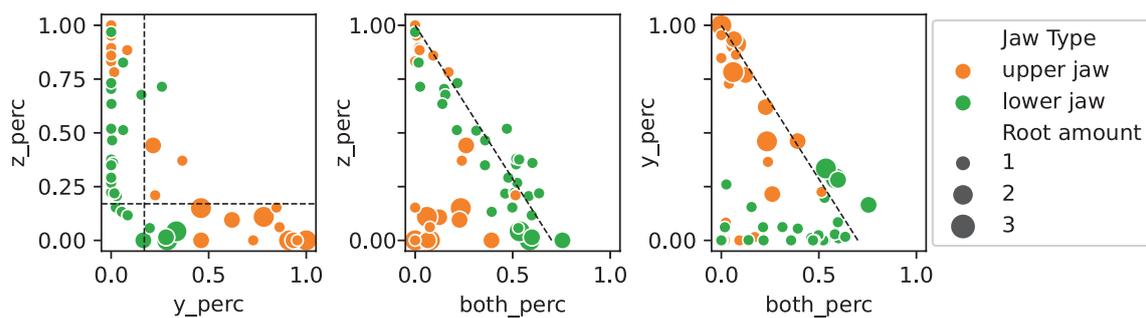


Figure 7-8: The normalized percentage of rotation strategy in z -direction (z_perc), y -direction (y_perc) and yz -direction (both_perc) plotted in a three fold plot. The marker color indicates the type of jaw and the marker size indicates the amount of roots of the tooth.

The left figure of Figure 7-8 displays a clear separation between the single-rooted teeth (1 root) and the multi-rooted teeth (2-3 roots). For multi-rooted teeth (big markers), no more than 20% of a pure z -rotation is present, except for one outlier. The percentage of y -rotation is spread from 15% to 100%, indicating the extraction technique is either a pure y -rotation

or a combination of an 8-rotation and a y -rotation. Subsequently, the left plot also displays a clear area of interest where the single-rooted teeth are situated. Especially the single-rooted teeth in the lower jaw show less than 20% of y -rotation. This indicates the single-rooted teeth are mostly extracted using a z -rotation or by a combination of z -rotation and an 8-rotation. The centre plot and the right plot show an apparent linear trend. For z_perc and $both_perc$ mostly single-rooted lower jaw samples meet this linear relation, while for y_perc and $both_perc$ mostly multi-rooted upper jaw samples meet this linear relation.

Zooming in on the left plot of Figure 7-8, the amount of roots metric can be replaced by the type of tooth to allow for a more tooth specific comparison. The result is shown in the left plot of Figure 7-9.

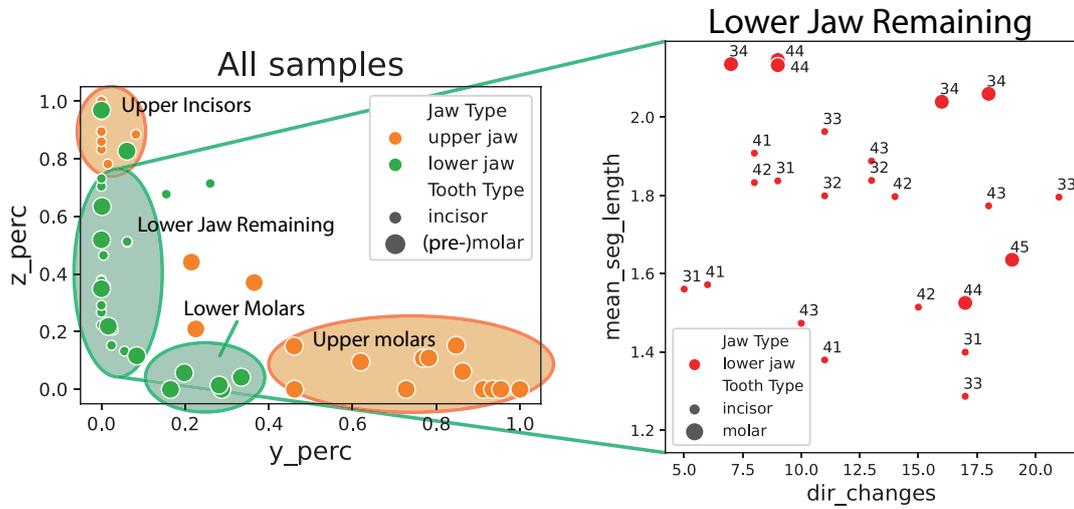


Figure 7-9: Clustering of rotation demonstrations where the percentage of y -rotation is plotted on the horizontal axis and the percentage of z -rotation is plotted on the vertical axis in the left figure. Because the ‘remaining’ samples in the lower jaw do not display distinctiveness based on y_perc and z_perc , they are compared amongst the features $dir_changes$ and $mean_seg_length$, which can be seen in the right figure. The proposed clustering in this figure is done based on tooth type

This figure shows the extraction strategy of molars in the upper jaw mainly include y -rotation. The extraction strategy of incisors in the upper jaw mainly include z -rotation. The molars in the lower jaw have a varying y -rotation between 0% and 40% of the total demonstration, whilst having no z -rotation. We are left with a group of teeth in the upper jaw where no uniform distinction can be made y_perc or z_perc . To see if the complexity of the rotation can give descriptiveness in this group, the right plot of Figure 7-9 is included. It shows the samples plotted with the number of segments on the horizontal axis and the mean segment length on the vertical axis. When a segment length is small and/or the amount of segments is small, a distinguishable group emerges with lower jaw incisors. When a segment length is big and/or the amount of segments is big, the last group with lower pre-molars appears. In this way, both the features representing rotation strategy and the features representing rotation complexity are yielding descriptiveness between groups of teeth.

Because the groups in Figure 7-9 show significant similarities amongst groups of teeth, groups are formed based on these similarities of extraction technique. The various extraction tech-

niques and the corresponding tooth numbers are described in Table 7-3. These groups will also be used later on as classification labels in the prediction algorithm. The corresponding names as used in the remainder of this report are added in the last column.

Table 7-3: Extraction techniques grouped together based on the outcomes of the feature engineering process. The tooth numbers and the group names are included.

Extraction technique	Tooth number	Group name
z_rot	11, 21, 12, 22, 13, 23	up_in
y_rot	14, 24, 15, 25, 16, 26, 17, 27	up_mol
8-rot [high rotation velocity]	31, 41, 32, 42, 33, 43	low_in
8-rot [low rotation velocity]	34, 44, 35, 45	low_premol
8-rot [buccal oriented]	36, 46, 37, 47	low_mol

To justify that the found groups of Table 7-3 indeed yield differences in extraction techniques, the kernel density analysis as performed in Figure 5-7 can be carried out on all the extraction techniques belonging to the respective group. This yields a heatmap of all the rotations in y and z -directions of that particular group. The results of grouping the data of the main surgeon are shown in Figure 7-10.

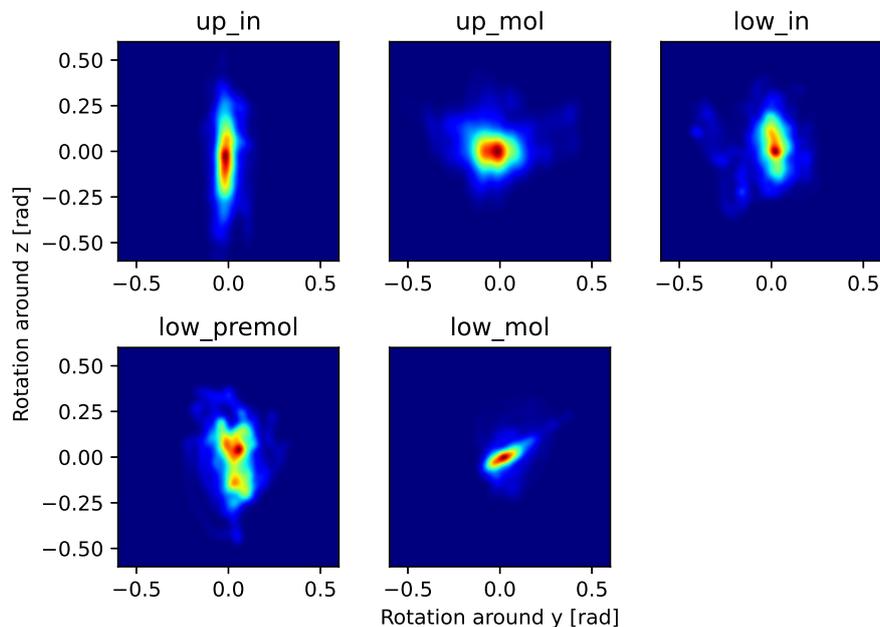


Figure 7-10: Gaussian distribution of the amount of y -rotation against the amount of z -rotation in radians of all classes defined in Table 7-3, displayed in a heatmap.

Various extraction techniques can be seen in this figure. The most obvious technique is applied on upper incisors. It is a dominant rotation around the z -axis, whilst almost no rotation around the y -axis is present. The upper molars have a dominant y -rotation, but the rotation in z is still present. The lower incisors and the lower premolars are mainly oriented

towards the rotation around z , but the y -rotations are present. The lower premolar plot shows two distinct distributions, indicating multiple techniques are used to extract this type of tooth. The lower molar group show an apparent 45 degree angle between the y -rotation and the z -rotation. This indicates the application of a pure 8-rotation as extraction technique for lower molars.

The heatmaps in Figure 7-10 are the result of the heatmaps that are based on the rotation data. To compare amongst executors in an organised manner, the most interesting results from their heatmaps are included in the following results. The remaining heatmaps not included in this report can be found in Appendix A-4.

Descriptiveness on executor level

Figure 7-11 is showing the comparison of the heatmaps of all the executors. The technique for the upper incisors is roughly the same for all executors. The main surgeon however, performs the extractions with less variance in y -rotation. The variances in y -rotation of surgeon 2 and surgeon 3 are higher and the variance in y -rotation of the student is large. The main surgeon also extracts the lower incisors with a small favour in z -rotation, but compared to the upper incisors, an increase in y -rotation is shown. This is an indication for the ‘8’ extraction technique. The lower incisors heatmaps of surgeon 2 and surgeon 3 are shown to be oriented to a y -rotation, where the spread along the z -rotation is less. It can be seen however that the heatmap lights up a bit towards the z -rotation, indicating the presence of a z -rotation as well. This difference can be due to the fact that this group consists out of two persons, both adopting their own different extracting techniques.

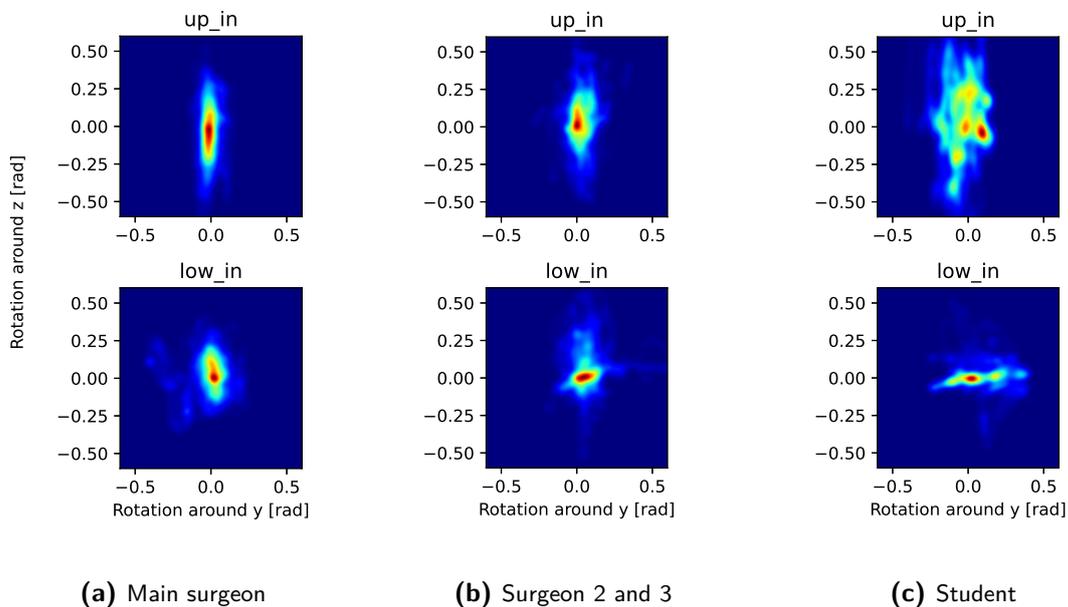


Figure 7-11: Heatmap of the y -rotations against the z -rotations for the upper incisor class and the lower incisor class. The data of the main surgeon is plotted in (a), the data of the other two surgeons are plotted in (b) and the data of the student is plotted in (c).

The last column shows the heatmap of the student. It can be seen that for the upper incisors,

there is no clear strategy. The movements are all over the place. This is in high contrast with the clean and subtle movements from the clinicians. The lower incisor group also deviates from the findings of the clinicians. It shows a pure y -rotation, indicating a buccal-lingual movement. This is interesting to see because this type of movement is exactly the extraction movement how Stegenga describes the procedure on extracting lower incisor teeth (Table 2-1).

In addition to the heatmaps shown above, the velocity heatmaps of all the four executors are calculated. These are included in Appendix A-4. The most interesting finding from these velocity heatmaps is that the student is using less rotation velocity than the surgeons.

7-2 Results of the Supervised Prediction Modelling

Section 7-1 showed how the feature engineering process led to features that describe tooth removal on anatomical level or executor level. Furthermore, tooth groups are formed, mainly based on features of rotation. These groups will serve as the classification groups for the supervised prediction algorithm. Section 7-2-1 describes how a Naïve Bayes algorithm serves as a baseline model taking the features as input to fit them based on the classification groups. This baseline model yields a certain performance and stability. These are compared with the outcomes of a LR model by means of accuracy and bet subset selection. At last, the interpretability of the features used in the LR is discussed in Section 7-2-3, where the results of forward stepwise selection and SHapley Additive exPlanations values are stated.

7-2-1 Gaussian Naïve Bayes prediction Model

The NB algorithm is implemented on the dataset of the main surgeon ($n=68$). The train/test split as described in Section 6-1-3 yields a training set size of 47 samples and a test set size of 21 samples. The respective train class sizes are: `low_in`: 18, `low_mol`: 5, `low_premol`: 7, `up_in`: 23 and `up_mol`: 15. The respective test class sizes are: `low_in`: 6, `low_mol`: 1, `low_premol`: 2, `up_in`: 7 and `up_mol`: 5. When the NB model learns from the training data and predicts output classes from the unseen samples of the test set, a confusion matrix can be employed to give insight in the prediction performance of the model on the respective classes. The confusion matrix for the training set is shown in Figure 7-12a and the confusion matrix for the test set is shown in Figure 7-12b.

Looking at individual classes, it can be seen that for both testing and training the classes `low_in`, `up_in` and `up_mol` render a high accuracy. The amount of samples in these classes is higher than in the other classes, which contributes to better training accuracy. However, the other classes perform reasonably well, despite the low amount of samples.

Furthermore, a mixup can be seen between the classes `low_in` and `low_premol` in the training set. Two `low_premol` predictions are wrongly predicted as `low_in` and two `low_mol` predictions are wrongly predicted as `low_premol`. This can be an indication that the groups `low_in` and `low_premol` are more similar than thought. This confusion is also visible in the test set whilst being less present. It can also be seen that from both the training set and the test set, only one classification mistake is made between the upper and lower jaw, indicating the prediction model is good at predicting differences between upper and lower jaw.

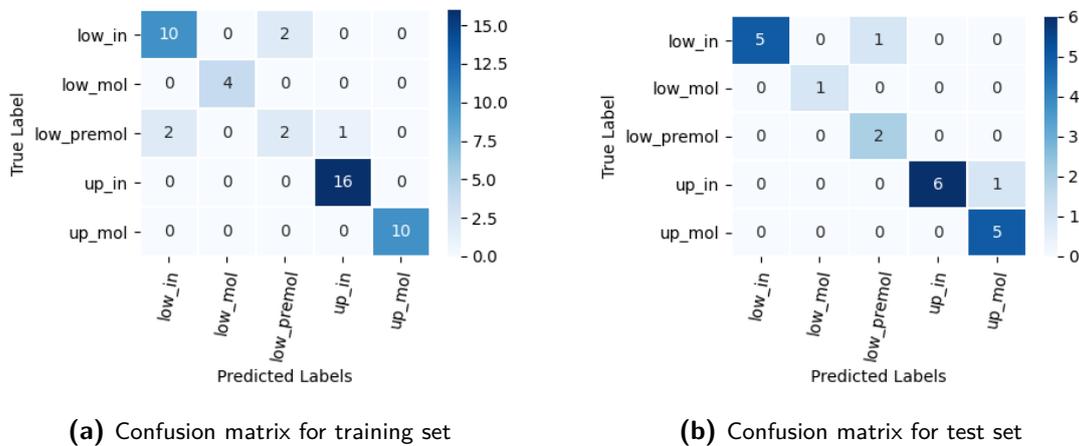


Figure 7-12: Confusion matrices for prediction results of the NB model. Figure (a) shows the result for the training set and figure (b) shows the result for the test set. Each instance in a column of the matrix represents the predicted label for a sample and each row represents the true label for that sample.

Cross-validation

Figure 7-12 is showing the confusion matrices of the NB prediction model. This is the result of a random stratified split of the training and test data. Because various splits render different results, cross-validation can be performed. The full process is described in Section 6-3. The result in terms of accuracy implementation of a 5-fold stratified cross-validation is shown in Figure 7-13. The training and the test accuracy are plotted with respect to the fold.

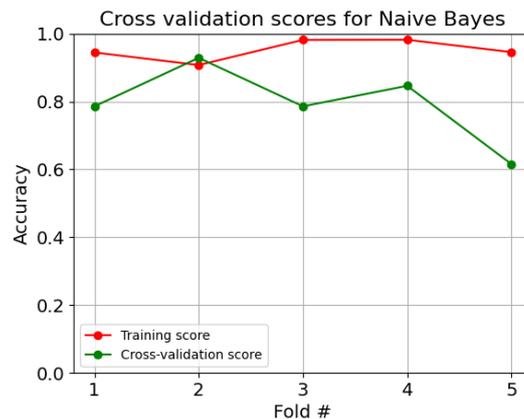


Figure 7-13: Cross validation scores when a 5-fold stratified cross validation is performed on the Naive Bayes model. The red curve shows the the training accuracy and the green line shows the cross-validation test accuracy.

The training accuracies show good performance over all folds. The cross-validation accuracies however, show instability. The accuracies of fold 1-4 are going up and down, and the accuracy of the last fold is dropping significantly. Comparing the cross-validation accuracies with the training accuracies, it is argued the training set is overfitting. This causes the cross-validation

scores to drop when they contain outliers (last fold).

Improvements can be made based on the obtained model so far. Because of the small dataset, it is hard to gain improvements in accuracy, precision and recall. The Naive Bayes is good at using the existing features for its predictions. The model however, is unstable. The cross-validation shows a instability and a fairly big drop in accuracy, probably caused by an outlier. In the next section, the results of the implementation of a Logistic Regression algorithm are discussed, which should render more stability and less variance in prediction accuracy. The current algorithm is compared with this new algorithm.

7-2-2 Logistic Regression Model

To allow for multiclass prediction a Logistic Regression model with an One-Versus-Rest (OVR) structure is introduced. The maximum iterations are set to 1000, and the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) solver is used. First, the performance of the LR algorithm is compared with the performance of the NB algorithm. Then the best feature set selection is compared for both algorithms. This indicates how much features are necessary to gain good performance.

Naïve Bayes and Logistic Regression comparison on accuracy

We can compare the performance of this model in terms of stability with the Naive Bayes model described above. Showing plain confusion matrices of the Logistic Regression model and comparing them with the confusion matrices of Figure 7-12 does not make sense to compare in terms of stability. From the Naive Bayes cross-validation it was shown that the way the training and test data was split had much influence on the prediction outcome, so to compare the two algorithms, 500 randomized train/test splits for both the Naive Bayes algorithm and the Logistic Regression algorithm are performed and plotted. This is shown in Figure 7-14.

First, the training sets are discussed. For NB and LR, both training medians are nearly equal. However, the interquartile range for Naive Bayes is centred around the mean, indicating the mean and the median are identical. The median in the LR model is more centred towards the third quartile, meaning half of the data has higher accuracy than 90%. Furthermore, the LR training distribution show a discrete character of frequency peaks. This can be caused due to the small number of samples in the dataset, where the randomized train/test splits render the same accuracy by wrongly predicting the same amount of samples over these various train/test splits. At last, the 95% confidence interval of the Naive Bayes almost spans across the whole distribution, while the 95% confidence interval of the Logistic regression is way more centred towards the median, which indicates higher stability for the LR model.

Next, the test set distributions are inspected. For both NB and LR, the test set shape varies from the training set shape. There is more variety in accuracy and the spread is higher, leading to a narrower distribution plot. The ‘discretization’ behaviour we saw for the training set in LR can be seen for the test set as well. The NB testing accuracy shows a wide 95% confidence interval, spanning almost the whole distribution. This was seen in the training set as well. For the LR, the 95% confidence interval is again smaller than for NB, which can be caused by the more stable training set in Logistic Regression.

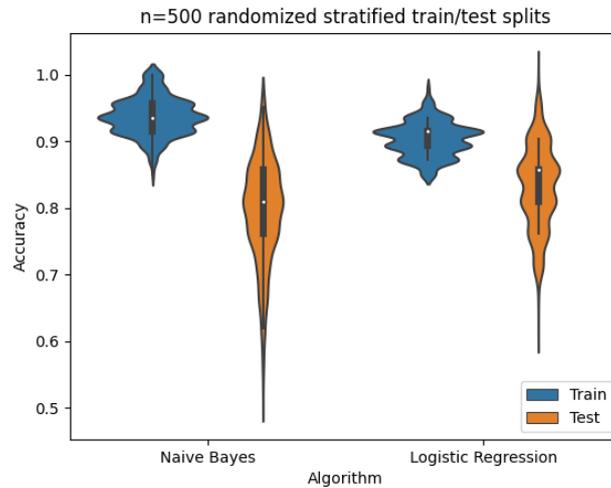


Figure 7-14: Probability density plots from 500 randomized train/test splits for the Naive Bayes algorithm and the Logistic Regression algorithm. The blue violins show the distributions of the train data and the orange violins show the distributions of the test data.

Finally, the median training accuracy and the median testing accuracy are more similar for the LR model than for the NB model. The NB distribution of the test set is more shifted downwards with respect to the training set compared to the distributions of the LR model.

Best subset selection

The best subset selection is performed to show how the accuracy changes in relation with the number of features. This is done for the NB model and the LR model. The result of implementing best subset selection on both models is shown in Figure 7-15.

The left figure shows the best subset selection for the NB model, and the right figure shows the best subset selection for the LR model. The average accuracy of the NB model is indicated by the green line and the red line indicates the maximum accuracy. This shows that the NB model has significantly higher accuracy than the LR model, especially when the amount of features is small (94% over 89%). However, the mean accuracies for both models do not differ significantly. The green line shows the NB model has higher mean accuracy in general, but a bigger standard deviation than the LR model. The standard deviation of the LR model is almost half the size of the standard deviation of the NB model. Also, if we take a look at the minimum accuracies per feature amount, the slope for the LR algorithm is almost linear, yielding a much smaller spread between the maximum score and the minimum score for the accuracy, which is likely causing the smaller standard deviation.

From this plot it looks like that the NB algorithm outperforms the LR algorithm. These subset calculations however, are based on the total dataset, so no split between a training dataset and the testing dataset is done. It can be argued that the NB model is better at fitting complex relationships when small amount of features are present than the LR model. This is not always desirable, because generalization to unseen samples is important. Taking the train/test performance results of Figure 7-14 into account, it is argued the LR model performs

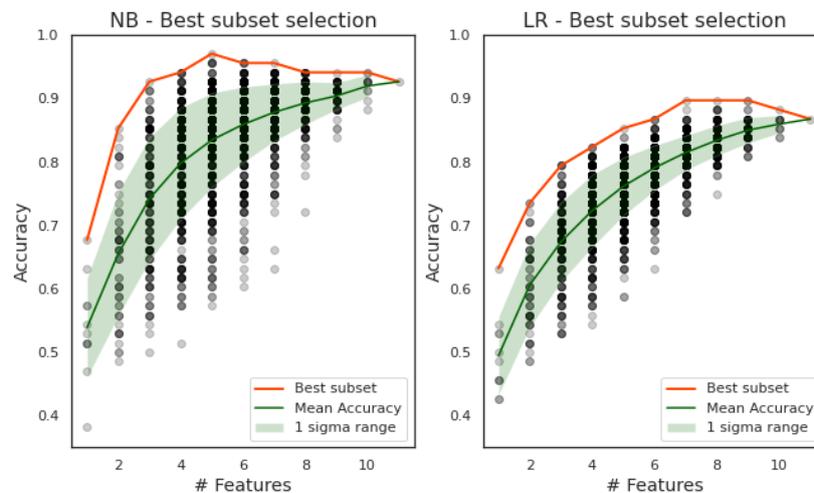


Figure 7-15: Best subset selection for the Naive Bayes algorithm and Logistic Regression algorithm. The black dots represent the algorithms accuracy for a fitted models. For each number of features, the combination of possible models $\binom{k}{n}$ are fitted. Here n is the number of features to choose from and k is the total amount of features, which is 11. The red line is the subset with the highest accuracy. The green line shows the mean accuracy and the standard deviation over all $\binom{k}{n}$ models for every feature amount.

more stable while being less prone to overfit compared to the NB model. As such, the LR algorithm is used for performing the forward stepwise selection and SHAP interpretability tests.

7-2-3 Model Interpretability

The best subset selection from Section 7-2-1 and Section 7-2-2 can be of help when interpreting the impact of the features on the model accuracy. When a feature has a high impact on the model accuracy, it means this feature is descriptive and defining for the intrinsic properties of this class. We can use two methods to investigate the feature importance in our Logistic Regression model. First, we employed a forward stepwise selection model. This model starts with fitting a model with 0 features and at every iteration, it adds the feature that gives the best additional improvement. Furthermore, we used the SHAP interpretability optimization to find the parameters with the most predictive power.

Forward stepwise selection

The results of the forward stepwise selection on the LR model can be seen in Table 7-4. This table shows the cumulative improvements on the model accuracy when using the forward stepwise feature selection method. The seven most essential iterations are plotted, yielding a training accuracy of 86.8%. Interesting to see is that the rotation features `z_perc` and `y_perc` together account for 73.5% of the accuracy, after only two iterations. After 6 iterations the performance of the model stagnates, yielding no additional performance increase at iteration 7. The ultimate performance is thus reached by combining three features of rotation and three features of force complexity.

Table 7-4: Forward stepwise feature selection result for the five most important iterations. The last two columns represent the cumulative accuracy added during an iteration and the resulting accuracy of the combined features.

It.	Features					Cum Acc. %	Acc. %
1	z_perc					63.2	63.2
2	z_perc	y_perc				10.3	73.5
3	z_perc	y_perc	fy_complexity			3	76.5
4	z_perc	y_perc	fy_complexity	fz_complexity		1.4	77.9
5	z_perc	y_perc	fy_complexity	fz_complexity	#_rot_strategies	4.5	82.4
6	z_perc	y_perc	fy_complexity	fz_complexity	#_rot_strategies	4.2	86.8
	fx_complexity						
7	z_perc	y_perc	fy_complexity	fz_complexity	#_rot_strategies	0	86.8
	fx_complexity		mean_seg_length				

SHAP Interpretability

In addition to the forward stepwise selection, we can use the SHAP interpretability that is basing the feature importance on game theory. Figure 7-16 show that again, the `z_perc` feature and the `y_perc` feature are impacting the model performance the most.

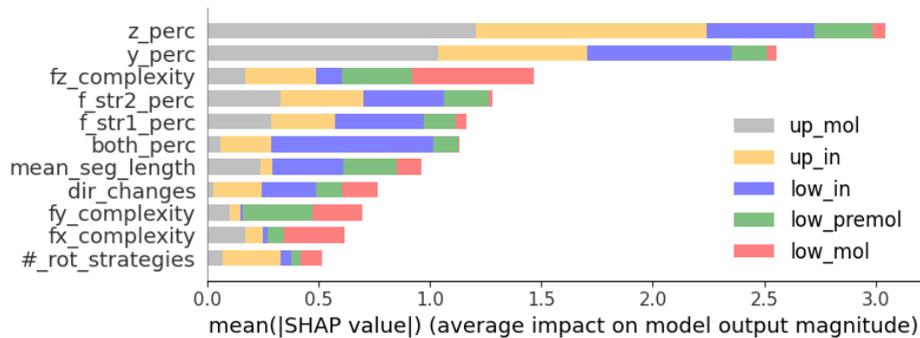


Figure 7-16: SHAP values for the Logistic Regression model trained on the classification classes as described in Table 7-3. The stacked bars show the impact of a feature on the output of the Logistic Regression model. The colour grading of the bars indicate the impact of the feature on the output performance of a specific class.

Especially for the classes in the upper jaw, the impact of the `z_perc` feature and the `y_perc` feature on the model output is significant. Furthermore, it can be seen that the `Fz_complexity` has the most impact on the lower molar class. The feature `both_perc` has the most impact on the lower incisor class.

Discussion and Conclusion

This chapter intends to discuss the results of Chapter 7. First, the results of the feature engineering of the force data, the feature engineering of the rotation data and the supervised prediction model are discussed in Section 8-1. Next, a general discussion is provided in Section 8-2, responding to the research question and the sub-questions posed in Section 1-2. Section 8-3 states the conclusion and Section 8-4 gives recommendations for future work.

8-1 Discussion of Feature Engineering and Prediction Model

This section provides an in-depth discussion of the results of the feature engineering of the force data, the feature engineering of the rotation data, and the supervised prediction model.

8-1-1 Discussion of the force features

Section 7-1-1 stated how the most essential features deduced from the force data tried to describe descriptiveness of tooth removal on anatomical level and interoperability level. In general, on anatomical level, only a descriptive trend could be seen for features `fx_complexity` and `fy_complexity` when comparing on jaw type (upper, lower). The samples from the upper jaw showed an increase in `fy_complexity` accompanied with a decrease in `fx_complexity`, while the samples for the lower jaw showed the opposite (Figure 7-3). With the results of the rotations in mind, an explanation in retrospect for this clear separation could be the fact that the movements for the upper jaw (especially the upper jaw molars) are more y -rotation oriented. This results in an application of force in the y -direction, increasing the `fy_complexity`. On the other hand, the samples of the lower jaw (especially the lower jaw molars) require more of an 8-rotation. This results in an increase of force application in x -direction, increasing the `fx_complexity`. Apparently, the force application direction changes in y -direction comes forward while applying bucco-lingual rotations, while the force application direction changes in x -direction comes forward while applying an 8-rotation. Future work could investigate the coupling of these movements and their resulting force directions.

Furthermore on interoperability level, the distributions of the force features did not show significant descriptiveness amongst the executors, except for the student. The student used a different strategy than the other three surgeons, especially compared with the main surgeon. This becomes evident when comparing Figure 7-2 and Figure 7-6. The location of the samples of the student is clustered and oriented to the left of the plot, while the majority of the samples of the main surgeon is oriented to the lower right of the plot. This indicates the student is applying a different force strategy than the surgeon. This statement is strengthened by the amount of datapoints of the student situating on the ‘perfect line’. Only a single datapoint is situated here, indicating the student’s force data differs from the main surgeon’s force data. Furthermore, the student needed more complexity than the other three surgeons, while showing most similarities with surgeon 2 (Figure 7-7). The latter was probably due to the amount of time needed to extract. For both surgeon 2 and the student, this was higher compared to the extraction times of the main surgeon and surgeon 3. Right away, this could be considered a downside of the complexity features. They are still dependent on time: the more time needed, the more complexity. An improvement on the current complexity features would be the introduction of a time-independent force feature, where the complexity of the extraction could be expressed as ‘*complexity per second*’.

The torques were not used in the analysis. During the brainstorm sessions, the torque data did not show apparent descriptiveness, and the data looked random. First iterations of prediction performance of the force features in the prediction model showed disappointing results in terms of accuracy. In this way, the torques were not analysed and no features are based on them. However, with the current knowledge about the rotations and forces, a hypothesis about the torques is proposed to show how features from the torque data could have been deducted in the feature engineering process. The torques operate around the same axis of rotation as the rotation data, lending itself for an introduction of a ‘*resistance*’ measure. In this way, when a small rotation value is measured and a high torque value, the resistance is high. When the rotation is high and the torque is small, the resistance is small. It could be that the clinicians are rotating in the direction of least resistance and therefore avoiding strong forces and torques in specific directions to minimize the risk of fractures. For further research, the working principle of the torques could be investigated in combination with this ‘*resistance*’ measure. This direction of research could be performed in combination with rotations, or it could be more focused towards the detection of anomalies, such as tooth fracture or complications.

In literature, limited research on the topic of force interactions showed strong contradictory results. This shows the complexity of analysing forces during tooth removal. Comparing the results with the hypothesis stated about the forces, the first hypothesis about the pushing and pulling strategies is rejected. This strategy deduction could only be used to show differences between a student and a surgeon, but it did not provide enough information about descriptiveness between teeth. The second hypothesis could be assumed with caution. In general, a slight increase of complexity was seen for molars compared to incisors, but the distributions of feature values show much overlap between incisors and molars. From this, the question ‘*Why was the hypothesis proposed in the first place?*’ can be asked. This was done because at earlier iterations of the modelling phase, predictions of the force features (and especially the complexity features) showed reasonable prediction performance. The feature engineering analysis was done after the prediction modelling. This should have been the other way around, so the limited descriptiveness was noticed in an earlier stage of the process. Besides,

the researchers were too much fixated at finding a metric for complexity, above just using basic features such as maximum force.

In summary, the feature extraction of the force data did not show the desired results in terms of descriptiveness. Feature Engineering remains a process that is highly dependent on human intervention and the expertise of the human concerned [30]. The researchers did not find the best working force features yet, but further analysis could change this. This statement is strengthened by the fact that the most basic characteristics such as maximum force, total force applied or peak force are not included as features in this research. At first iterations of the modelling, they did not look promising and not descriptive. Further research could look into their descriptiveness.

8-1-2 Discussion of the rotational features

Section 7-1-2 stated how the force features resulted in descriptiveness between groups of teeth and descriptiveness between executors. The features `z_perc`, `y_perc`, `dir_changes` and `mean_seg_length` were able to discriminate five different groups of rotation. In line with textbook instructions, for example in Stegenga [10], we found apparent differences between certain movement strategies, most important being the pure rotation around the z -axis and a pure rotation around the y -axis. However, it seems that an extra group of descriptive movement patterns exists as well (8-rotation), which has three variants: buccal oriented, high rotation velocity and low rotation velocity. The results of the feature engineering process of the rotation data seems to be well in line with our hypotheses about rotations. The features `z_perc`, `y_perc` showed they can be used to find descriptiveness in extraction strategy. These features showed interoperability coherence amongst surgeons, which was especially high for the `up_in` class and the `low_premol` class. The extraction strategy of the student differed significantly, which was most apparent in the `up_mol` class and the `low_in` class.

The rotational features are found to be descriptive but do have their limitations. Still, the rotation of the extraction procedure is measured with a robot attached to a forceps. Although the learning curve for handling the robot is pretty steep, the compliant movement of the robot could still impose movements to the surgeon that were not intended by the surgeon in the first place. This has a larger effect on the smaller samples sizes ($n=23$, $n=26$, $n=22$) for respectively surgeon 2, surgeon 3 and the student because they did not have time to learn how to handle the robot. This in contrary to the main surgeon, who significantly performed more experiments ($n=79$). Furthermore, the main surgeon removed all teeth during a single day. Possibly, this has influence on the resulting data. On one hand, it can introduce the effect of consistency when performing the same experiment over and over again. On the other hand, it can introduce the effect of increased (physical) workload, where fatigue could be increased and concentration could be decreased in time. Currently, these factors are assumed constant, and their influence on the data is unknown.

Furthermore, it is assumed the position of the tip of the forceps does not change during an extraction. However, when a tooth comes loose, the center of rotation of the tooth is changing. As such, the position of the forceps changes as well. The assumption that the tip does not move was initially made because in dental literature there is no consensus about what the location of this center of rotation is (if any). In hindsight, we could argue that, especially for more significant deviations from the original centre of rotation (generally bucco-lingual and

8-movements), Ψ_t is continuously moving, while the transformation of the forces, torques and rotations assume a fixed frame. The rotation is less dependent on this fixed frame because it is calculated as a relative movement, while not being dependent on position. The wrench transformation, on the other hand, is dependent on the position. This could be a reason that the rotation features worked out better than the force-torque features.

8-1-3 Discussion of the prediction model

Section 7-2 showed the results of the modelling of tooth removal features with a supervised machine learning model. The Naïve Bayes (NB) baseline model showed it is fairly good at distinguishing between the upper jaw and the lower jaw. Figure 7-12b showed only one classification mistake on the training data was made between upper and lower jaw (97% accuracy). No classification mistakes between upper and lower jaw were made in the test set, yielding 100%. The absolute accuracies are not the most important however. Apparently, the impact of the combination of force and rotation features differs significantly between the upper and lower jaw, yielding high prediction accuracy between the jaws. The specific impact of the force features or rotation features on the jaw type is not investigated. In hindsight, a separate prediction model classifying the jaw type should have been integrated. This could have given more proof for the decisiveness of the force features.

The Logistic Regression (LR) model outperformed the NB model based on stability and mean test accuracy, when the same 500 train/test splits are used. The median accuracies of the training set and test set were closer for LR compared to NB. These results show that with such a small dataset, the accuracies and performance of the prediction model are dependent on the class sizes. Here the train-test split ratio is critical and influences the performance of the predictions. Some test set classes showed only one sample, which in general is not representative for the training set. Additional data is needed to make the samples in the test set more reliable. On the other hand, when more data is gathered, the need for a more sophisticated machine learning algorithm such as neural networks emerge [12]. The downside of 'just gathering an enormous amount of data' is two-fold: It is costly, especially the acquisition of fresh frozen cadaver jaws is expensive. Furthermore, with the application of a more sophisticated algorithm you lose upon interpretability. The amount of data should lie in a sweet spot where applied algorithms still find descriptiveness amongst classes and coherence within classes, without losing upon techniques to gain information from the learning model.

The interpretation analysis of the model showed how the rotation features account for most of the descriptiveness in the model. This was expected, because the most descriptive features emerged from the rotation data and the labelling was based on the descriptiveness of the rotation features. However, the prediction algorithm still needs information from the force features to yield high prediction performance.

8-2 General Discussion

8-2-1 What was found in this study?

This study explored whether feature engineering and supervised prediction modelling are able to describe tooth removal in terms of forces and movements. A scientific gap exists in the field of tooth removal and this research provides an innovative first step to bridge this gap. With the use of robotic technology, 131 successful human demonstrations of tooth removal procedures were recorded from four different executors. Feature engineering was used to extract relevant features from the demonstrations. These features provide information about the tooth removal process but were also used as an input to a supervised classification model. This classification model was used to classify tooth groups that were formed during the feature engineering process.

The feature engineering process showed that we were able to discriminate between teeth mainly based on the features of the rotation rather than the features of the force. We were able to show how the percentage of y -rotation and the percentage of z -rotation could be used to form groups of similar extraction strategy. The formed groups proved how different extraction techniques are used for different teeth. Furthermore, the groups were used to highlight strategy differences between a student and surgeons. The strategy of the student in terms of rotation direction and velocity, significantly differs from the strategies of the surgeons.

The prediction modelling confirmed that indeed the features from rotation have high predictive importance. The rotation features alone have shown to account for more than 73.5% of the model's performance. In this way, the prediction model helped to justify the descriptive factor of the rotation features. Even with a small dataset, the supervised prediction algorithm yielded significant performance.

The analysis of the feature engineering model and the prediction model are based on finding differences between feature values. Hence, the process adopted in this thesis is of a discriminative nature. Differences in the dataset are used to our advantage in trying to explain anatomical variations or variations between executors. However, a discriminative model does not necessarily describe the total process of tooth removal. If a surgeon is using the same technique for all extractions, a discriminative model is not able to find this. A generative model on the other hand, is able to extract this specific behaviour from the data. Our discriminative model thus does not describe the most important factors for tooth removal, but describes how differences in terms of forces, rotations and movements between tooth groups could be explained. This means, factors influencing differences in tooth removal procedures are found, but factors influencing all teeth in the same manner still have to be found.

8-2-2 Limitations

However, some limitations still exist for this research. Despite this research showed promising results into a first insight in tooth removal procedures, the main pitfall seems to be the limited amount of data. This was mainly seen in the results of the prediction model. The NB baseline model did show significant accuracy with 84% on the test set (2 out of 21 wrongly predicted samples). However, the lower molar class still contains only one sample, indicating

a limited representativeness of this class in the test set. The current model predict this test set sample accurate, but with a different train/test split, this will not always be the case. Classes with with larger test sizes (i.e. class `up_in`: upper incisors) show much stronger accuracy performance. In the training set, all 16 samples are predicted correctly and in the test set only one sample is predicted as false positive. Extending the dataset with more data would justify the development of i.e. scientific backed educational material. It was valuable to show that even with small amount of data, predictions on the formed groups differed significantly and significant differences between the main surgeon and the student emerged. The datasets of the other two surgeons however, were too small to draw justified conclusions. For further research, especially because the area of forces needs more research, we should gather more data that focuses on consistency within extractions of a single surgeon to ensure representativeness of all classes in the test set.

Secondly, as described in Chapter 3, an in-vitro measurement approach is used in this research with fresh frozen cadaver jaws. Research of Cicciù et al. and Ahel et al. has pointed out that in-vivo experiments will not lead to reliable data [3, 52]. But without reliable in-vivo data to compare our dataset to, it should be questioned whether our dataset is representative in clinic. It is possible that the extraction force is significantly lower than in real jaws because the gingiva in cadaver jaws is softer. Furthermore, the state of the jaws is assumed constant over the total amount of experiment day. All extractions are performed over a total time span of four days. It is possible that the jaws used at the first day were more rigid than the jaws used on the last day because the gingiva was not fully thawed at the start of the experiments. Additionally, the movements with the robot and the way force is applied, differs from an in-vivo setting. The robot is not fully resistance-free and it requires a steep learning curve to adjust to the compliant dynamics of the robot. This could introduce fatigue of the executor after many extractions. The rotation velocity could decrease after a certain amount of extractions, which could affect the chosen strategy. A perceived workload questionnaire for every full jaw extracted could be used to measure this. An example could be a NASA Task Load Index (NASA-TLX) [53] or a Shortened Fatigue Questionnaire (SFQ) [54].

Knowing the relation between the in-vitro measurements and real extractions in terms of periodontal state, application of movement and reaction forces, could make this research more justified. Not much is known yet about how the frozen jaws anatomically relate to real jaws, so this could be investigated in future work. Also the relation of in-vitro movements and in-vitro forces to real movements and forces is not known. This could also be investigated for future research.

Thirdly, we did not use a trustworthy and reliable calibration tool. The calibration process is still based on human interpretation of the location and orientation of the tooth. Determining the orientation of the tooth is difficult since roots are situated inside the bone. Furthermore, it is well known that the root's tip might deviate in shape and orientation, which seems to be a good reason to consider using radio-graphical images for a more precise calibration process. Also, two separate calibration tools were used for the upper and lower jaw, respectively. The design of a solid and reliable calibration tool that is generic for both upper and lower jaw would help the calibration process enormous. It will lead to higher repeatability, which relies less on the interpretation of the human.

At last, we will reflect on the process that led to the results. This project involved the application of technical knowledge in the domain of dentistry, that was first unknown to the

author. The innovative character of this project and contributing to educational improvements fed the author's enthusiasm to work on this project. However, this enthusiasm also caused the author to dive too deep into the research and modelling, without taking a step back and reflecting on the progress and approach so far. For example, a step back should have been taken after the feature engineering process to reflect on the feasibility of the force features. In this way, the force features could have been more descriptive in describing differences between teeth. In retrospect, the process of finding the factors that influence tooth removal might have been dedicated too much to 'finding the best features as possible' for the prediction model. On forehand, we should have been less optimistic about the prediction model and should have focused more on incorporating the more obvious features in combination with clinical data such as treatments and executors expectations. Currently, these factors are left out of the equation, while they could have add importance in clarifying predictions or outliers. This is a consequence of the approach taken. At the point of defining the features, the focus was too much on finding features for complexity that where intuitive to understand. At first, these features looked promising in the prediction model step, but in the end, they did not show descriptiveness in the feature engineering step.

8-3 Conclusion

In conclusion, with a relatively small amount of data, we gained new insights into the process of tooth removal. Feature engineering was employed to find the most important factors influencing tooth removal procedures. Despite not all data was used and some data might be under-evaluated, feature engineering proved to be well suited for deducing features from the dataset. The rotation features were shown to be most descriptive, distinguishing between five groups of teeth. The force features did not prove to give significant descriptiveness between the extractions, but it showed potential. Furthermore, apparent extraction differences between the surgeons and student are found. In general, the student's extraction technique differs in strategy and complexity, illustrating the need for evaluation of current educational methods.

At this stage of the research, it can be concluded that the outcomes of the NB model and the LR mainly justify the descriptiveness of the rotation features. Supervised prediction showed that even with a small amount of data, the models are able to distinguish between different groups of teeth. However, the outcomes also show the limitations of the small dataset, confirming the need for more consistent data from a single executor.

An all-encompassing explanation of the tooth removal process is not found yet. However, this research showed that the process of feature engineering contributes to the understanding of various extraction strategies. These strategies are found to be highly influenced by the imposed rotation on the tooth. The prediction model highlighted these differences but showed how rotation data alone is not sufficient enough to fully explain differences in tooth removal procedures.

8-4 Future Work

Based on the discussion and the conclusion, the following recommendations for research directions and future work are proposed:

- The methodology mainly affected the force features and thus the information gain of the force features. Ahel et al. and Ciccì et al. already showed the difficulty of modelling the force in tooth removal procedures [3, 52], but it is inevitable that our dataset cannot lead to a conclusion about forces and force differences. So the deduction of the force features should be improved for future work, which could lead to clinical understanding of the applied force. A start can be made with deducing the total applied force over the time span of an extraction. This can be an informative descriptor for differences between molars and incisors because in general, more time is needed to remove molars. Furthermore, the maximum of applied force could be a feature that is worth investigating.
- It was argued that the rotation and the torque combined would yield a kind of resistance. A high torque measurement with a low rotation measurement indicates a tooth that is fixated in the gum and a low torque measurement with a high rotation measurement indicates a loose tooth. A feature could be proposed that combines both rotation and torque into a single metric defining resistance. When this resistance is modelled, the amount of torque could be coupled with the strategy performed. In this way, resistance could be used to explain why and how clinicians are swapping to the next strategy.
- If the force features are more stable, and a resistance metric is introduced, another prediction model (NB or LR) could be used to build a new classification model. We can hypothesize that with better features and more data, the accuracy will increase. If this classification model performs stable and has accurate predictions, a regression model can be introduced, performing supervised learning from classes to numerical values. This regression can be employed to predict the feature values for specific teeth; a tooth number is given as input to the model, and the model predicts the corresponding feature values.
- Especially with the use of rotational features, new educational material could be developed. The strategies can be used to serve as a basis for movement training of dental interns. The rotational directions for the different strategies are known, so these can be used as haptic guiding movements in a simulator. A force field could be generated, which guides the movement of the intern in only z -rotation or only θ -rotation. Based on the deviation of the prescribed movement patterns, the guiding force increases. In this way, the dental intern is forced to learn the movements that correspond to the specific teeth groups. Furthermore, the measurement setup could be of significant value to evaluate the current educational methods. The validity of plastic jaws and movements performed on these jaws could be tested to show the representativeness of the current educational methods.
- At last, a study could be performed that investigates the transferability of the movements made and the forces applied in the clinic to our in-vitro measurement setting. Yet it is unknown how representative our measurements are for real-life extractions. Knowing the representativeness of our measurements would justify the feasibility of the measurement setup and our analysis performed. The more representative the measurements, the better the measurement setup could be used for educational purposes.

Additional Feature Engineering Results

This appendix shows additional information on the feature engineering process as described in Chapter 5. Not all figures could be included in the report, so a selection covering the total diversity of the dataset is given in this chapter. Section A-1 covers the additional graphs on how the force strategy is deduced. Particularly, the filtering method is shown that makes filters irrelevant strategies out. Section A-2 shows additional graphs for the segmentation process of the rotation data, where the regular rotations and the velocities are plotted in the same figure. This is done to see how the velocity impacts the segmentation and the strategy filtering. Furthermore, additional kernel density plots and Gaussian heatmaps are given that were not included in the report.

A-1 Additional figures for feature engineering of forces

Figure A-1 includes the results of the pushing and pulling strategy deduction algorithm for four measurements. In all four figures (upper left, upper right, lower left and lower right), the upper plot shows the force curves in x , y and z -direction (red, green, blue colors respectively). The second plot the annotation per time stamp on which strategy is currently used. The three options are `str1`, `str2` and `Not Def`. These are a pushing strategy, a pulling strategy and an unknown strategy respectively. Because it is hypothesised that a strategy is always applied for at least one second, the strategies yielding a smaller time span than one second are filtered out. The result of the filtered annotations are plotted in the third figure. The bar at the bottom is a visual representation of this third graph, making the strategies assignment somewhat easier to interpret.

The upper two figures show a comparison between the strategy deduction of an incisor in the upper jaw (left figure) and an incisor in the lower jaw (right figure). It can be seen that the pushing strategy (`str1`) in the left figure is the main type of extraction, where in the right figure the pulling strategy (`str2`) is more present. It looks like the two strategies are swapped in place.

The lower two figures show a comparison between the strategy deduction of a molar in the upper jaw (left figure) and a molar in the lower jaw (right figure). Here it can be seen that for the upper molar only a pulling strategy (str2) is used and no pushing strategy at all. This is in contrast with the pushing strategy of the molar in the lower jaw (right figure) that occupies half of the extraction strategy.

At last we can compare the molar strategy deduction (both lower figures) with the incisor strategy deduction (both upper figures). The force profiles of the molar figures show an increase in complexity compared to the force profiles of the incisors. The back and forth appliance of the force is more present in the lower two figures than in the upper two figures.

Based on these differences shown in both pushing and pulling strategy and the differences shown in complexity, the hypothesis about F_z and complexity as stated in Section 5-1 is proposed.

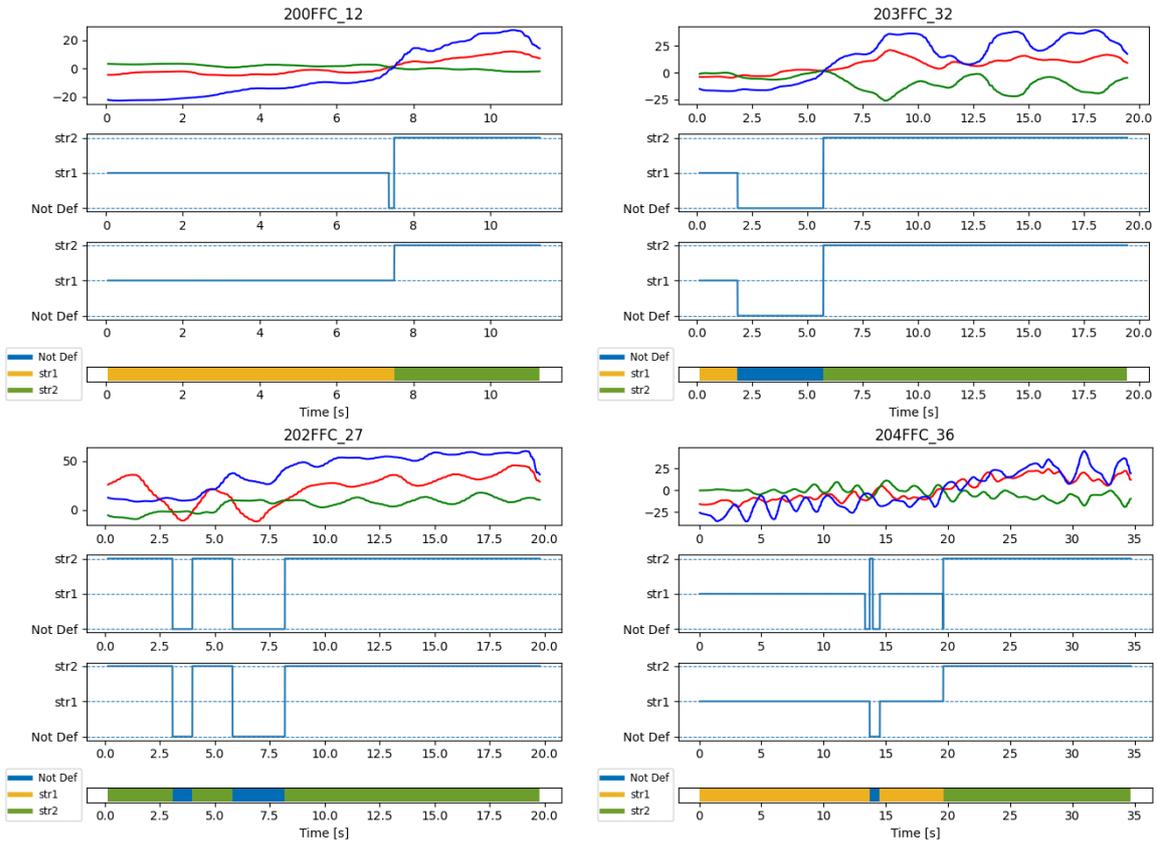


Figure A-1: Strategy deduction results for force data for an incisor in the upper jaw (upper left), an incisor in the lower jaw (upper right), a molar in the upper jaw (lower left) and a molar in the lower jaw (lower right). In all four figures the force curves in x , y and z -direction are plotted in red, green and blue respectively in the most upper plot. The second plot shows the unfiltered strategy annotation per time sample. The third plot shows the filtered strategy annotation per time sample. The coloured bar at the bottom is a visual representation of the filtered strategy annotation.

A-2 Additional figures for feature engineering of rotations

A-2-1 Strategy deduction and movement segmentation

Figure A-2 includes the results of the strategy deduction algorithm and the segmentation algorithm for four measurements. In all four figures (upper left, upper right, lower left and lower right), the upper plot shows the rotations curves in y and z -direction (green, blue colors respectively). The bar underneath the rotation curves show the strategy that is annotated to a specific timestamp. These are an unknown strategy, a strategy of y -rotation, a strategy of z -rotation or a strategy of yz -rotation (both) with colors white, green, blue and yellow respectively. Below the figure the legend is shown, with on the left the legend for the rotation movements and on the right the legend for the strategy bar.

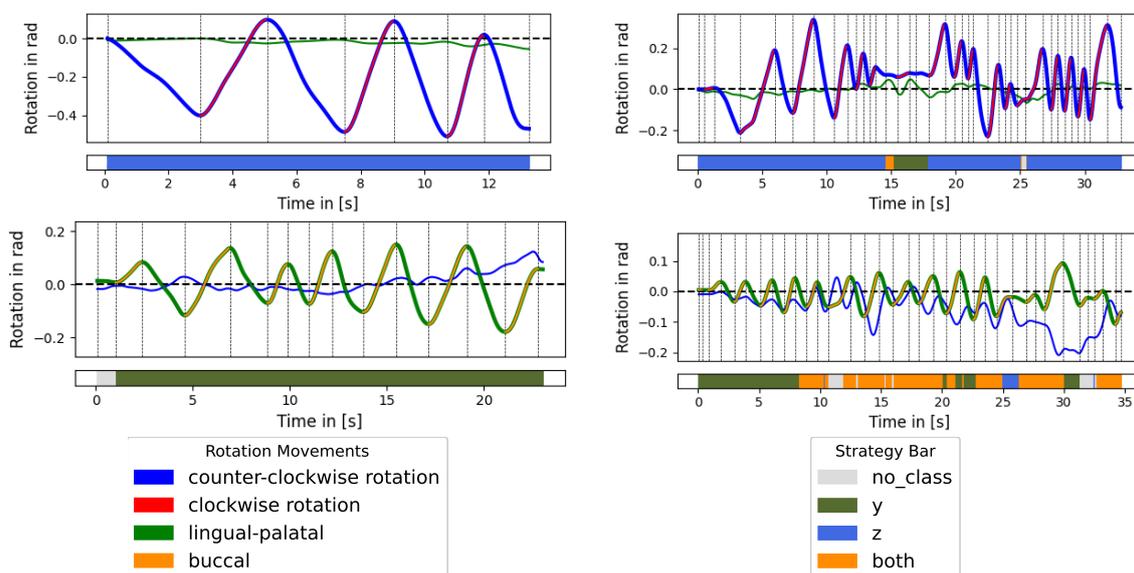


Figure A-2: Strategy deduction and segmentation results of rotation data for an incisor in the upper jaw (upper left and upper right), a molar in the upper jaw (lower left) and a molar in the lower jaw (lower right). In all four figures the rotation curves in y and z -direction are plotted. Both upper figures show experiments where the z -direction was dominant. Here, a segmentation distinction is made between counter-clockwise rotations (blue) and clockwise rotations (red). Both lower figures show experiments where the y -rotation was dominant. Here, a segmentation distinction is made between lingual/palatal rotation (green) and buccal rotation (orange). In all four figures the type of extraction strategy is plotted in the bar underneath. All time stamps are annotated with no strategy, a y -rotation strategy, a z -rotation strategy or an 8-rotation strategy (both), annotated with the white, green, blue or yellow color respectively.

The upper two figures show the segmentation of the rotations in experiments where the z -direction was dominant. A clear difference between the left figure and the right figure can be seen in amount of rotation changes needed. The direction changes (indicated with the color change from blue to red) is more present in the right figure than in the left figure. In the upper right figure it is interesting to see that around $t = 15s$, the algorithm detects a change from a pure z -rotation to a pure y -rotation.

The lower two figures show the segmentation of the rotations in experiments where the y -

direction was dominant, again indicated with the color change (green, orange). Comparing the lower left and the lower right figure, again the amount of segments is higher in the right figure. The strategy in the left figure is a pure y -rotation (bucco-lingual/palatal), while the strategy in the right figure is mostly using both yz -directions (8-rotation).

These figures are included to show the versatile results of the strategy deduction algorithm and the segmentation algorithm on the rotation data.

A-2-2 Kernel Density Estimation

Two extra figure of the Kernel Density Estimation as performed in Section 5-3-1 are included. Figure A-3 shows the differences between a lower jaw incisor and a lower jaw molar. The round shaped heatmap in the left plot indicates the strategy is changing quite a lot and a combination of separate y and z rotations is used. The thin line shaped curve which has an angle of 45° in the right plot indicates a pure simultaneous y - z rotation.

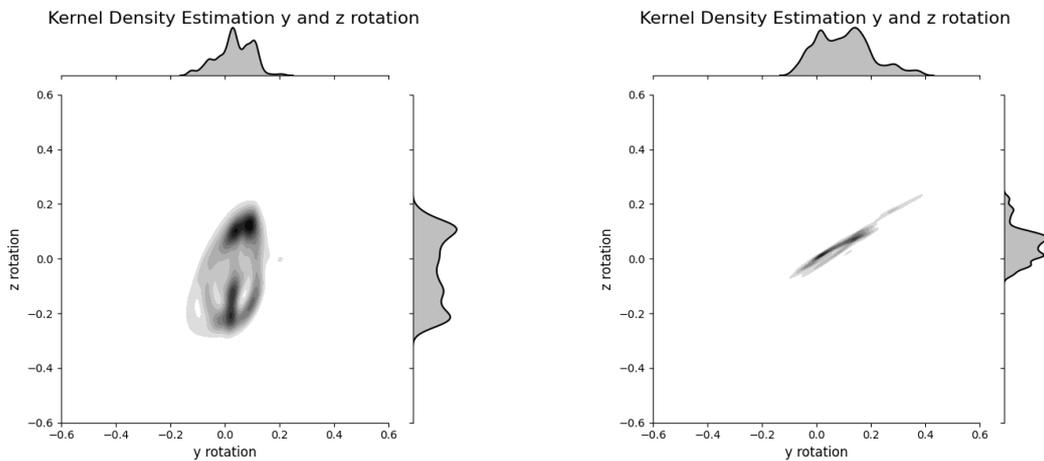


Figure A-3: Heatmap resulting from the plotting the Gaussian distribution of the y -rotation and the z -rotation of a demonstration of a lower jaw incisor 42 on the left and a lower jaw molar 47 on the right

A-3 Gaussian Rotation Distributions

Section 7-1-2 show the heatmap results of the main surgeon for the rotational data. In this section, the heatmaps for the other three extractors are provided. Figure A-4 provides the heatmaps of the rotation in y and z -direction for surgeon 2 and surgeon 3. Figure A-5 provides the heatmaps of the rotation in y and z -direction for the student.

Interestingly, the rotations of the surgeons do not show a coherent strategy along the classes. The heatmap does not show a pure y or a pure z -rotation. Instead, the heatmaps are more round shaped and a lot of deviation can be seen away from the centre point (0,0). The heatmaps of the student show do show interesting behaviour for the upper incisor class (`up_in`). The movement is all over the place, so no clear strategy is applied. This could indicate a lack of confidence and experience. Furthermore, the most applied strategy (red

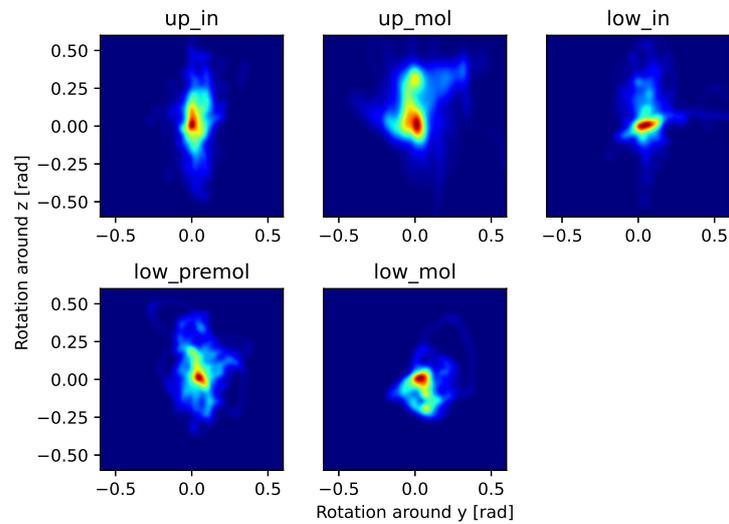


Figure A-4: Grouped rotation heatmaps from the rotation data of surgeon 2 and surgeon 3. The y -rotation is plotted on the horizontal axis and the z -rotation is plotted on the vertical axis.

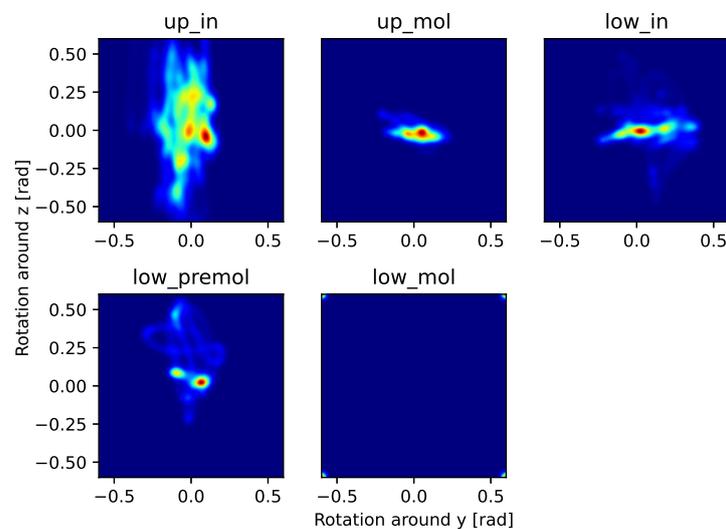


Figure A-5: Grouped rotation heatmaps from the rotation data of the student. The y -rotation is plotted on the horizontal axis and the z -rotation is plotted on the vertical axis.

dot) is located at a z -rotation of 0, and a positive y -rotation of 0.1 radians. This indicates a palatal strategy, which is a different strategy than the longitudinal rotation (z -rotation) the main surgeon performs.

A-4 Gaussian Velocity Distributions

In addition to the heatmaps based on the y -rotation and z -rotation data, the heatmaps of the velocity data in y and z -direction are included in this appendix. These were not included in the report, but yield interesting findings, which could be of interest for the reader. Figure A-6 displays the heatmaps for the tooth groups of the velocity data from the main surgeon.

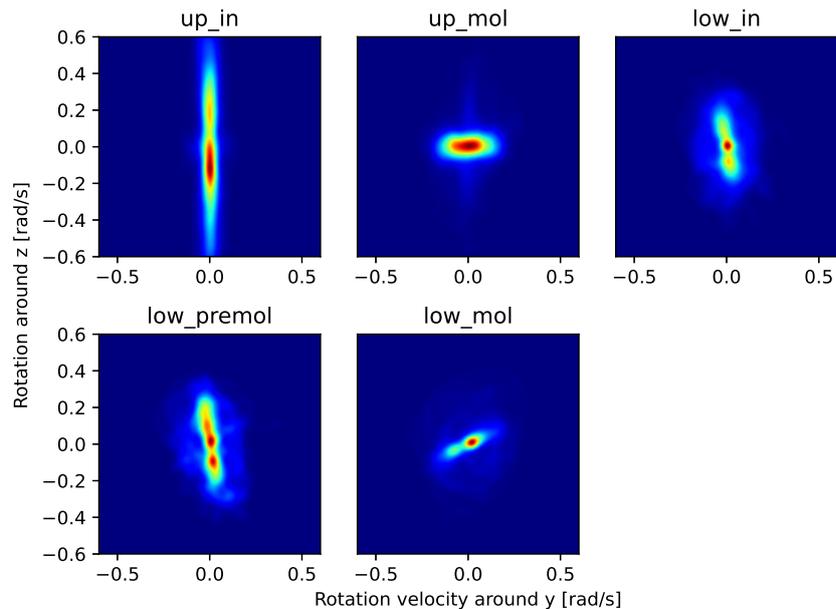


Figure A-6: Grouped rotation velocity heatmaps from the rotation velocity data of the main surgeon. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.

Interesting to see, is that the velocity in extracting the upper incisors (`up_in`), reaches twice the values of velocities in the other classes. Also the direction of the velocity is quite distinguishable. For the upper incisors (`up_in`), the lower incisors (`low_in`) and the lower premolars (`low_premol`), the main direction of velocity is in the z -direction. For the upper molars (`up_mol`), this is clearly in y -direction and for the lower molars (`low_mol`) this is clearly in the yz -direction.

The next page show the heatmaps of the other executors. Figure A-7 shows the combined velocity heatmaps of surgeon 2 and surgeon 3. Figure A-8 shows the velocity heatmaps of the student. Most remarkable is that the rotation velocity around the z -axis is most distinguishable. Furthermore, the heatmaps of the surgeons show a lot of variation, especially a little further away from the centre. At last, it can be seen that all the rotation velocities of the student are low, much lower than the velocities of the surgeons. This can be due to the student's inexperience and low confidence level.

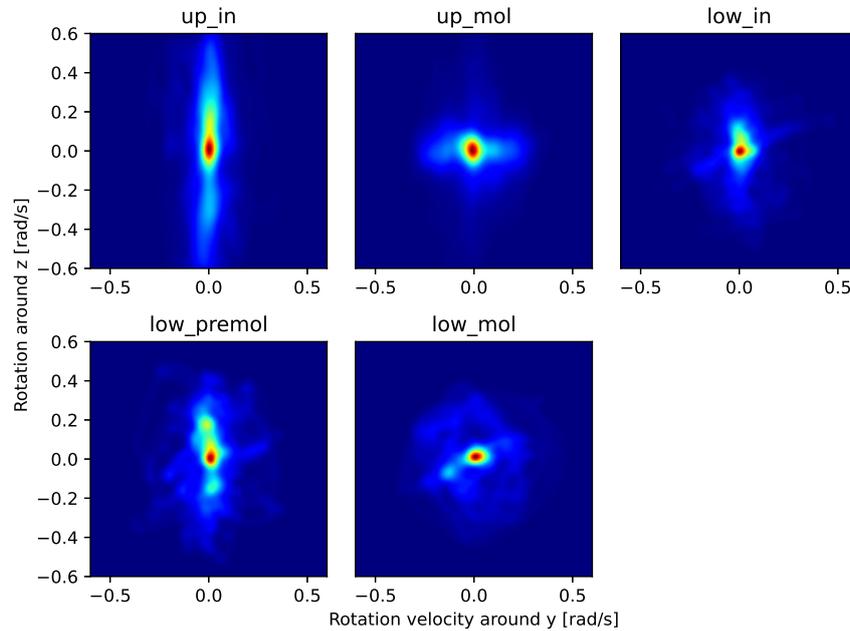


Figure A-7: Grouped rotation velocity heatmaps from the rotation velocity data of surgeon 2 and surgeon 3. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.

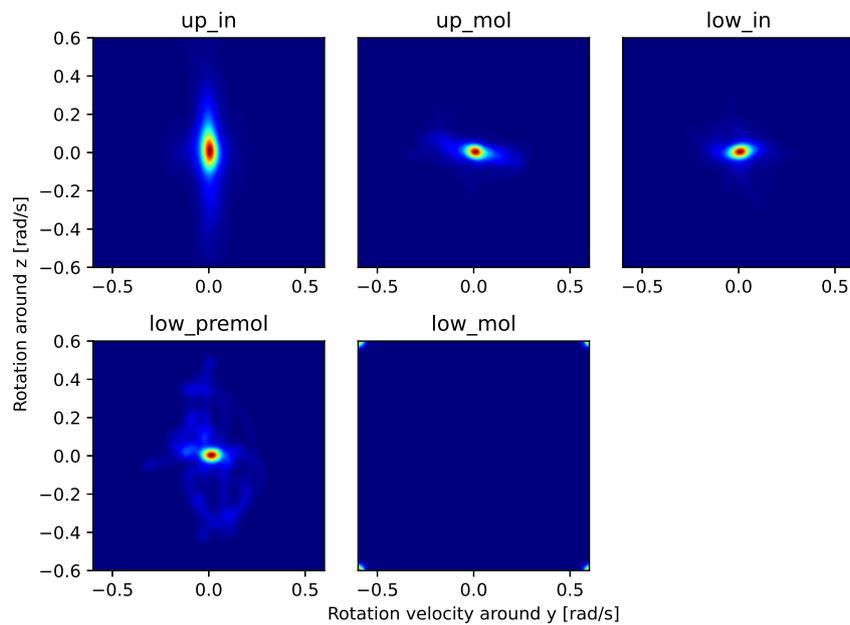


Figure A-8: Grouped rotation velocity heatmaps from the rotation velocity data the student. The y -rotation velocity is plotted on the horizontal axis and the z -rotation velocity is plotted on the vertical axis.

Appendix B

ATI F/T sensor information

This Appendix covers the relevant information regarding the working principle of the force torque sensor. Section B-1 describes how Hooke's law is applied to measure the forces and torques applied on the transducer. Section B-2 provides the functional outline of the electrical, software and mechanical parts of the force torque system as stated in the installation and operation manual as provided by ATI industrial Automation [44]. Additionally, the schematic drawings of the transducer and the mounting plate are added for reference.

B-1 Hooke's Law for Force Torque Measurement

Equation (B-1) can be used to calculate the force and torque components by measuring the strain on the beam when a force is applied to the transducer:

$$\sigma = E \cdot \epsilon \quad (\text{B-1})$$

Here σ is the stress applied to the beam, which is proportional to the applied force. E is the elasticity modulus of the beam and ϵ is the strain applied to the beam. The strain ϵ is the unknown variable, but it can be measured by measuring the change in resistance of the strain gauges attached to the beam. The change in resistance ΔR changes as a function of the strain ϵ in the strain gauge and the strain gauge constant:

$$\Delta R = GF \cdot R_0 \cdot \epsilon \quad (\text{B-2})$$

R_0 is the resistance of the strain gauge in an unstrained state and GF is the ratio between the relative change in electrical resistance and the mechanical strain; the gauge factor. The change in resistance ΔR is measured and with the use of Equation (B-1) and Equation (B-2) this resistance is converted to force and torque components σ .

B-2 Relevant information from Installation Manual

F/T Controller Installation and Operation Manual
 Document #9610-05-1001-22

3. How It Works

3.1 Introduction

This section provides a functional outline of the F/T system. The F/T system is broken into three areas: electrical, controlling software, and mechanical. A graphical representation of the electronics is presented in *Section 3.2*. A controlling software flow chart is shown in *Section 3.3*. A mechanical description is shown in *Section 3.4*.

3.2 Electronic Hardware

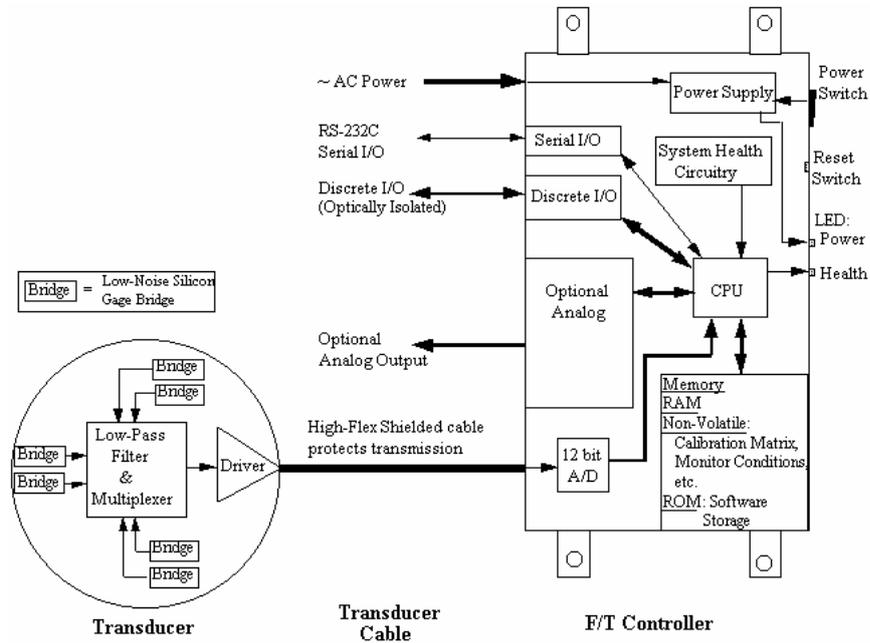


Figure 3.1—Electronic hardware outline

3.3 Software Outline

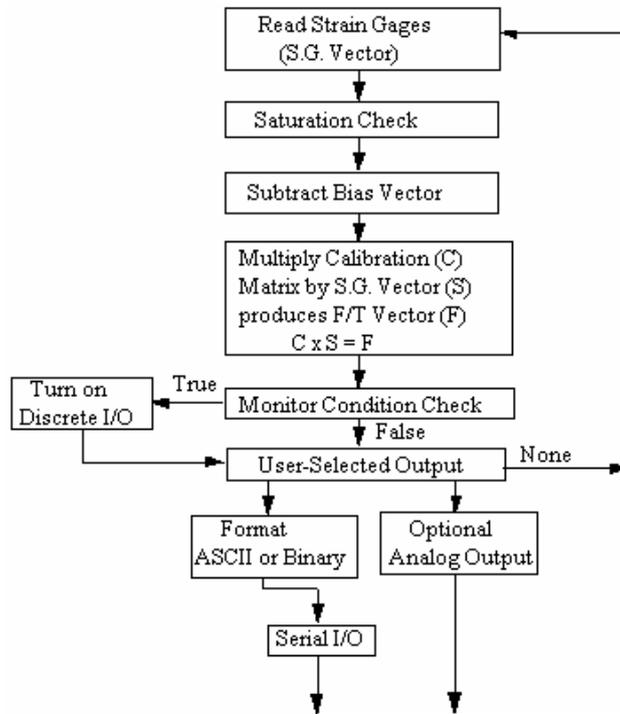


Figure 3.2—F/T Controller Data flowchart

3.4 Mechanical Description

The property of forces was first stated by Newton in his third law of motion: “To every action there is always opposed an equal reaction; or, the mutual action of two bodies upon each other are always equal, and directed to contrary parts.” The transducer reacts to applied forces and torques using Newton’s third law.

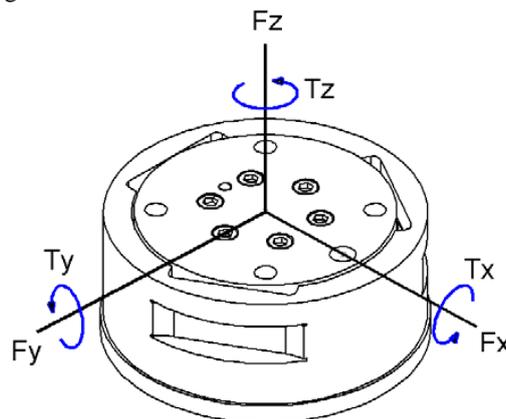


Figure 3.3—Applied force and torque vector on transducer

F/T Controller Installation and Operation Manual
 Document #9610-05-1001-22

The force applied to the transducer flexes three symmetrically placed beams using Hooke's law:

$$\begin{aligned} \sigma &= E \cdot \varepsilon \\ \sigma &= \text{Stress applied to the beam } (\sigma \text{ is proportional to force}) \\ E &= \text{Elasticity modulus of the beam} \\ \varepsilon &= \text{Strain applied to the beam} \end{aligned}$$

Aside:

The transducer is monolithic structure. The beams are machined from a solid piece of metal. This decreases hysteresis and increases the strength and repeatability of the structure.

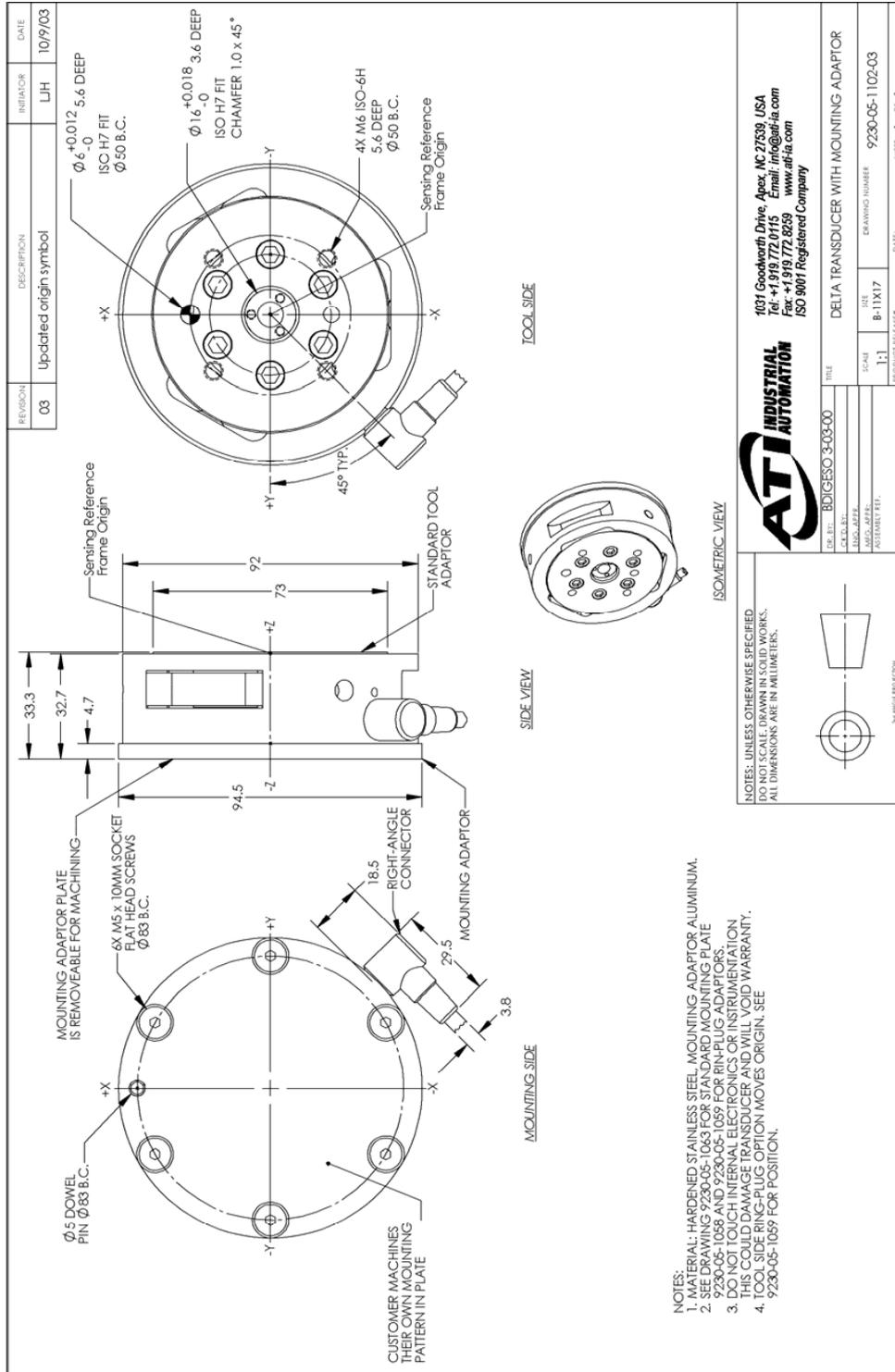
Semiconductor strain gages are attached to the beams and are considered strain-sensitive resistors. The resistance of the strain gage changes as a function of the applied strain as follows:

$$\begin{aligned} \Delta R &= S_a \cdot R_0 \cdot \varepsilon \\ \Delta R &= \text{Change in resistance of strain gage} \\ S_a &= \text{Gage factor of strain gage} \\ R_0 &= \text{Resistance of strain gage unstrained} \\ \varepsilon &= \text{Strain applied to strain gage} \end{aligned}$$

The electronic hardware, described in *Section 3.2*, measures the change in resistance and the software, described in *Section 3.3*, converts this change to force and torque components.

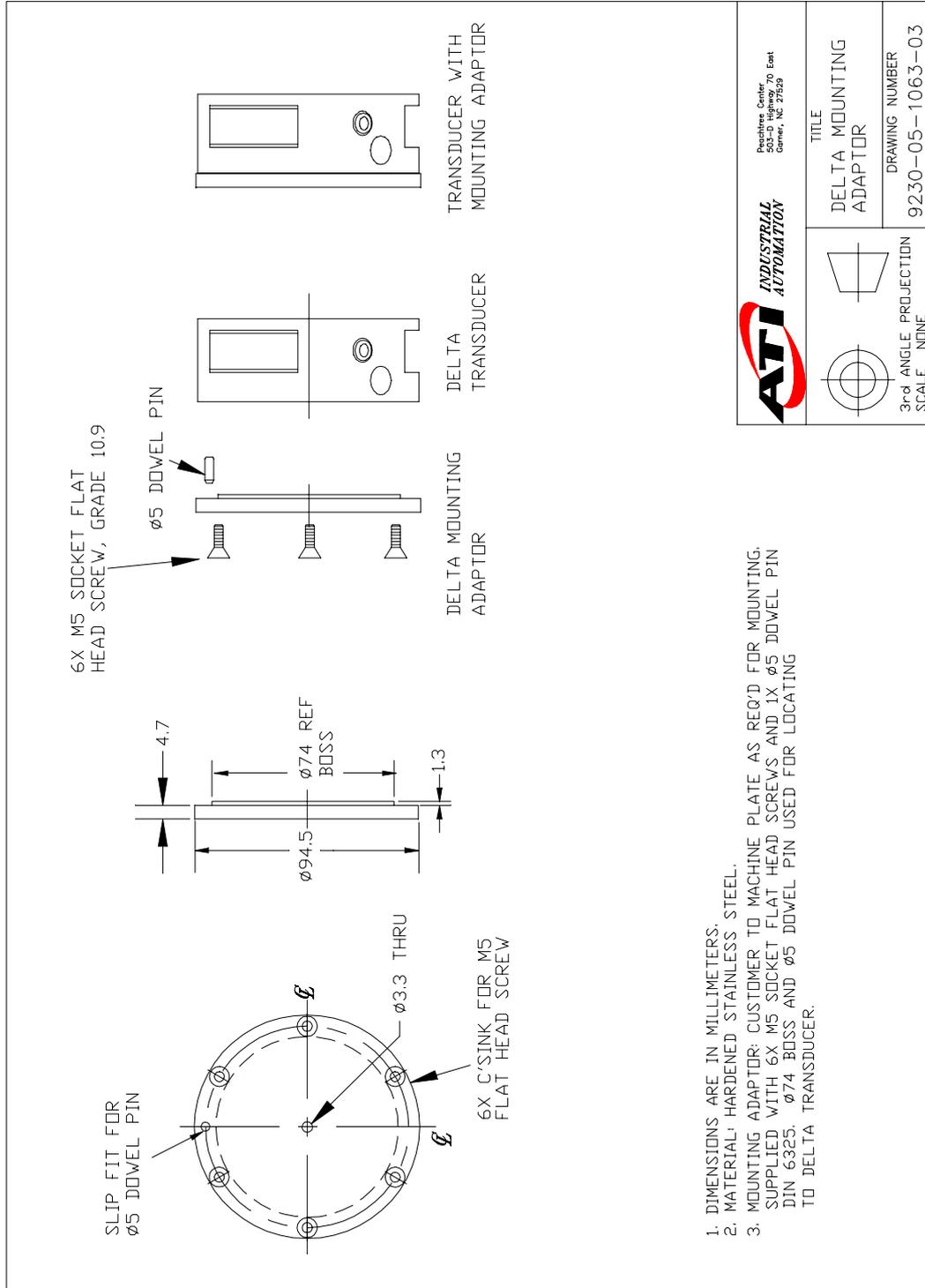
F/T Controller Installation and Operation Manual
Document #9610-05-1001-22

10.7 Delta Transducer with Mounting Adapter



F/T Controller Installation and Operation Manual
 Document #9610-05-1001-22

10.16 Delta Mounting Adapter Plate



Appendix C

Conference Paper

Parts of this report are submitted as a conference paper to the 42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society 2020 and is written in collaboration with Dr. Ing. J. Kober (Cognitive Robotics, Delft University of Technology) and Drs. T.C.T. Van Riet (OMFS, Amsterdam UMC). The paper is a detailed and extended explanation of the material described in Chapter 3 and is titled '*Robot Technology in Analyzing Tooth Removal - a Proof of Concept*'. Chapter 3 explains is focused on the robotic and software parts of the measurements setup. For readability of the report some parts of the paper are left out of the report. As a result a more thorough explanation of the design considerations and concessions made in the design process are discussed in the paper, as added in this appendix.

Robot Technology in Analyzing Tooth Removal - a Proof of Concept

Tom C.T. van Riet¹, Willem M. de Graaf², Reinier van Antwerpen³, Jan van Frankenhuyzen³,
Jan de Lange¹, Jens Kober²

Abstract—a measurement setup is proposed that, for the first time, is capable of capturing the combination of high forces and subtle movements exerted during tooth removal procedures in high detail and in a reproducible manner by using robot technology. The outcomes of a design process from a collaboration between clinicians, mechanical and software engineers together with first results are presented in this proof of concept.

Clinical relevance— by measuring all aspects of tooth removal in a single setup a strong database can be built that will deliver the data needed to gain scientific understanding of what makes (un)successful tooth removal. It gives a unique opportunity to model the procedure, evaluate techniques, understand and predict adverse events as well as to create new evidence-based teaching methods.

I. INTRODUCTION

Tooth removal, or exodontia, is one of the most commonly performed surgical procedures on our planet. Despite its high prevalence, surprisingly little is known about this procedure. During these procedures dental surgeons use a combination of subtle movements and strong forces to free a tooth from its surrounding bony socket. Previous (very limited) research aimed at measuring just the total amount of forces necessary for exodontia [1]–[5]. The precise direction (in 3 dimensions) of the involved forces and the movements of the dental surgeon were, to the authors' knowledge, never before subject to research. The latter is probably due to the limitations of available instruments to precisely measure these parameters in a “key-hole” environment. It has led to a large scientific gap which becomes more evident when looking at the education of dental students. Tooth removal is the most invasive procedure dental students need to learn during their training but it is also the single procedure for which adequate preclinical training possibilities are absent or largely inadequate [6], [7]. Up until today students mostly

*Research supported by the Dutch Research Council (NWO) by means of the Open Mind Grant (2018, project number 17394)

¹Tom C.T. van Riet and Jan de Lange are with the department of Oral and Maxillofacial Surgery of the Amsterdam University Medical Center and the Academic Centre for Dentistry Amsterdam (ACTA), Meibergdreef 9, 1105 AZ, Amsterdam {t.c.vanriet, j.delange}@amsterdamumc.nl

²Willem M. de Graaf and Jens Kober are with the department of Cognitive Robotics of the 3mE faculty of the Delft University of Technology, Mekelweg 2, 2628 CD, Delft {w.m.degraaf@student.tudelft.nl and j.kober@tudelft.nl}

³Reinier van Antwerpen and Jan van Frankenhuyzen are with the Department of Biomechanical Engineering of the 3mE faculty of the Delft University of Technology, Mekelweg 2, 2628 CD, Delft {r.vanantwerpen, j.vanfrankenhuyzen}@tudelft.nl

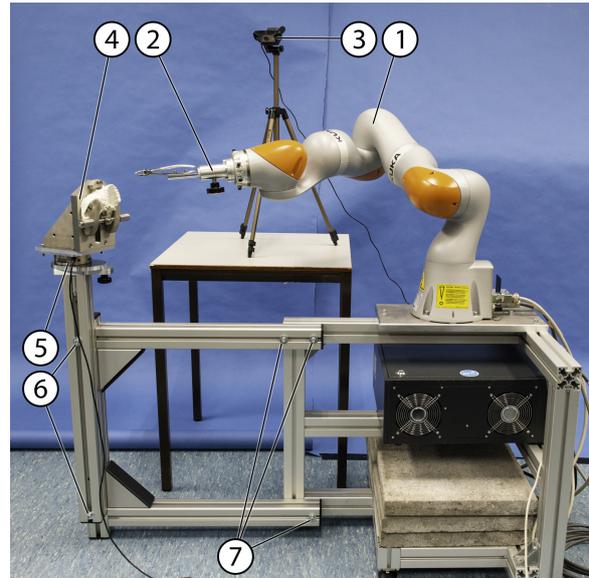


Fig. 1: Overview of the setup. (1) robot arm, (2) forceps holding device, (3) video camera, (4) upper jaw holding device, (5) force torque sensor, (6) bolts to adjust frame vertically, (7) bolts to adjust frame horizontally

learn their skills from textbooks with only minor instructions and train their skillset on actual patients [7]. Students in well-developed countries, where extensive preventive dentistry programs are present, are suffering from decreased exposure to ‘learning by experience’ because less teeth need to be removed in general. This contributes to low confidence levels in tooth removal procedures of young dentists and an increase in referrals to (more expensive) oral and maxillofacial surgery practices [7], [8]. Complete data on every aspect of these procedures is needed to be able to understand what makes (un)successful tooth removal and to scientifically describe and model the procedure. This dataset should additionally contain clinical parameters and perioperative data to be able to find relevant parameters in successful tooth removal. It would facilitate the design of evidence-based educational instruments but, next to that, it has the potential to help clinicians predict clinical outcomes (i.e. complicated treatments) and could lead to more (cost-) efficient referrals to oral and maxillofacial surgeons.

The goal of this project is to design a measurement setup that captures the high forces and subtle movements involved

in tooth removal procedures in detail. The design of the setup and integration of, amongst others, a collaborative robot and 6-axis force-torque sensor are shown in this article together with first results as a proof of concept.

II. MATERIAL AND METHODS

A. Challenges in detailed measuring of tooth removal

Several challenges had to be overcome during the design of the measurement setup. Dental surgeons use a combination of high forces and subtle motions to loosen a tooth from its bony socket. It is necessary to measure these sub-millimeter movements in 3 dimensions and at a high rate to be able to analyze movements in full detail and, for example, enable analysis of adverse events like tooth fracture. These measurements should take place without restricting dental surgeons in their movements in any way. Forces and torques should be measured in 3 dimensions in the center of rotation of the tooth, simultaneously with the movements. Clinically important parameters such as periodontal health, amount of roots, root size, age of the patient, and restorative state should be easily integrated into the measurements. Preferably these measurements should all be performed on patients in an *in vivo* setup.

Multiple sessions with a team of clinicians, mechanical engineers and computer scientists led to inevitable compromises in the setup. One of the major concessions to the ideal setup was the use of an *in vitro* measurement setup. Simultaneous and reproducible recordings of position/orientation/force/torque measurements are essential in this fundamental research. Compared to *in vitro* measurements, accurate sub-millimeter movement tracking and registration of forces/torques and their directions *in vivo* is questionable. One of the main issues is that the mobility of the patient is difficult to compensate for, which is especially true for the lower jaw, which is not rigidly fixated to the human skull. The force/torque sensor would need to be integrated in the forceps between the surgeon's hand and the tooth, which is unrealistic due to very limited space and high forces. Next to that, *in vivo* tooth removal requires considerable counterforce from the surgeons' second hand which would interfere with the force measurements. Finally, we made the assumption that the forceps and the tooth are rigidly connected once the tooth is grabbed. Therefore, we do not need to measure the movement of the tooth itself and can place the force/torque sensor under the jaw. To capture the clinicians' movement, several techniques were proposed of which optical tracking (infrared) and robot technology were the most promising. Robot assisted motion capture was preferred due to the high accuracy associated with robotic positional measurements. Next to that, by rigidly fixating the standard dental forceps to the end-effector, the surgeon can hold the forceps as they would do in clinical circumstances. Compared to optical trackers it prevents the need for markers and it avoids visibility issues of the tracking system during these 'key hole' surgical procedures.

B. An overview of the measurement setup

The measurement setup, see Fig. 1, consists of:

- a holding device for the upper- and lower jaw in an adjustable frame (Section II-C)
- 7 dental forceps (Section II-C)
- a six-axis force/torque (FT-) sensor (Section II-D)
- a compliant robot arm (Section II-D)
- a video camera (Section II-D)
- the Robot Operating System (ROS) (Section II-D)
- a graphical user interface (GUI) (Section II-E)

C. The adjustable frame and holding devices

To add to the readability of this subsection, numbers put between parentheses are referring to Fig. 2 (numbers 1 to 16) and Fig. 3 (numbers 17 to 32). A framework of a 60 by 60 millimeter aluminium profile (Item Industrietechnik, Solingen, Germany) was designed to mechanically integrate the different components (Fig. 1). The framework is adjustable, meaning the position of the holding devices for the upper and lower jaws can be changed relative to the robot and placed at different heights. This is necessary to mimic clinical circumstances in which the position of the upper and lower jaw are, respectively, vertical and horizontal. For ergonomic reasons, the patient is positioned higher when removing teeth from the upper jaw. The addition of a rotational plate (14,29) between the frame and the holding devices mimics the turning of the patients head and leads to a more clinical representative situation in which the clinician can maintain an ergonomic pose during the extraction procedure. The plate is located just below the FT-sensor (13,28) and can be rotated by pulling a locking bolt (16,32) on the bottom plate (15,30). The locking bolt falls into one of the position holes upon its release and can be further tightened to eliminate any slack. The position holes allow a 137.5 degree rotation in 11 steps of 12.5 degree increment in either direction (a total range of 275 degrees). Next to the ergonomic advantages, the usage of an adjustable frame largely overcomes an important issue of working with a robot arm. When any of the robot's joints reaches a joint limit, it needs to adjust other joints to enable the end-effector to reach the desired position. This can involve a rigorous movement of the robot which inevitably leads to some resistance for the clinician. By placing the most relevant joints in a neutral position just before starting the experiment, reaching joint limits can be avoided. This is enhanced by placing the upper and lower jaw in a favorable position relative to the robot arm. The frame was provided with a scale (millimeter) to measure the exact position of the holding devices for calibration purposes, see Section II-D.

Essential for reproducible, accurate and thus meaningful measurements is a completely rigid fixation of both upper and lower jaw. Two separate holding devices had to be designed. First because the above mentioned difference in ergonomic position (horizontal/vertical) of both jaws. Second because the anatomical differences between the two jaws do not facilitate the design of a single device to fit both. In general, non-corrosive and smooth surface materials were

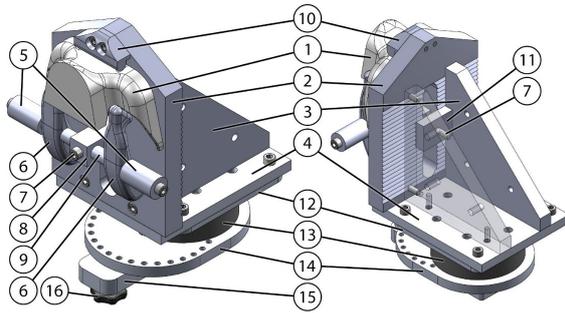


Fig. 2: Holding device for upper jaw: (1) upper jaw, (2) surface plate, (3) support plate, (4) ground plate, (5) axle boxes, (6) clamping arms, (7) clamping bolt, (8) sliding block, (9) clamp axis, (10) front block, (11) clamping nut, (12) top plate of sensor build-up, (13) force/torque sensor, (14) rotation plate, (15) bottom plate, (16) locking bolt

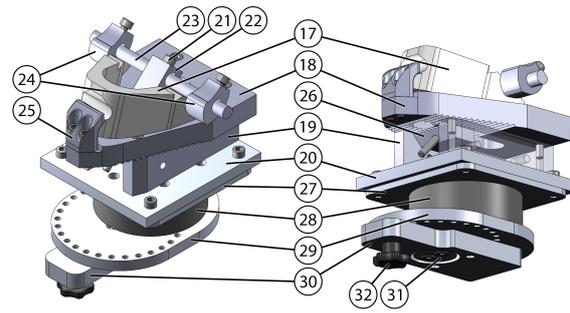


Fig. 3: Holding device for lower jaw: (17) lower jaw, (18) surface plate, (19) supporting plate, (20) ground plate, (21) clamping bolt, (22) sliding block, (23) clamp axis, (24) side blocks, (25) front block, (26) clamping nut, (27) top plate of sensor build-up, (28) force-torque sensor, (29) rotation plate, (30) bottom plate, (31) rotation axis, (32) locking bolt

used to facilitate cleaning which is especially necessary when working with (fresh) human material.

The shape of the upper jaw is geometrically unsuitable to fixate (inverted trapezoid shape) and can be very thin at certain points. As is known from facial trauma surgery, other parts of the midface (located just above the upper jaw) have better properties in terms of fixation because of the strength and shape of the bone. This counts for both the paranasal region (besides the nose) and, more lateral, the connection between upper jaw and zygomatic bone ('zygomatic buttress'). For holding the upper jaw, see Fig. 2, a clamping nut (7) was placed in an angular position relative to grooves on the main plate (4). Tightening the clamping bolt will force the 3D-printed titanium clamping arms (6), which were manufactured through selective laser melting (material: Ti6Al4V-ELI), to push the maxilla (1) downwards and forwards into a 45-degree angle. This way the frontal part of the maxilla, with its strong paranasal zones is fixated underneath a ridge (10). The ridge's geometry allows the upper jaw to slide slightly under it and prevents it from tilting upwards. Vertical grooves in this ridge minimizes translation from left to right. Sideward motion is further limited by tightening the axle boxes on the clamp axis (5) against the clamping arms. The arms push the strong zygomatic buttresses downwards and inwards. The rough surface of the clamps ensure grip even when remnants of muscle attachments are not completely removed during preparation of the skull. The shape of the clamp's head is designed to fit the natural shape of the zygomatic buttress which reduces the risk of iatrogenic fractures during any of the experiments.

Compared to the upper jaw, the lower jaw can be geometrically adjusted to make it more suitable for fixation. Its thick and strong cortical lining lends itself for fixation, even when the bone is reduced in size, see Fig. 3. Similar to the fixation of the upper jaw a clamping nut (21) is placed in an angular position to the grooves of the surface plate (20). By tightening the clamping bolt the clamp axis will



Fig. 4: Holding device for dental forceps.

force the jaw in a 45 degree angle downwards and forwards against the front block (25). The design of the front block ensures that the jaw can slide slightly under it to prevent the jaw from tilting upwards, while vertical grooves prevent translation sideways. Further translation is limited by sliding the side blocks (24) on the clamp axis against the sides of the jaw and locking them on the axis with a bolt. The design of the blocks is lean to facilitate the movement of the clinician, even when removing dorsally located molars.

To remove teeth, dental surgeons have a large variety of forceps at their disposal. To enhance grip on the tooth, the forceps are designed to specifically fit a certain type of tooth. For these experiments, seven dental forceps (Aesculap, B.Braun, Melsungen, Germany) are used: the left upper molar, right upper molar, upper premolar, upper incisor, lower molar, lower premolar and lower incisor forceps. They are fixated to the end-effector through a custom aluminum holder with two bolts (5mm), see Fig. 4. The aluminum holder is fixated in the end-effector by tightening one clamping bolt. The partially flat design of the custom aluminum frame ensured a reproducible position of the dental forceps in the end-effector.

D. The robot and force-torque measurements

To obtain sub-millimeter precision and accurate repeatability of movements during the procedure, the KUKA LBR iiwa 7 R800 is used [9]. This robot is a 7 degree of freedom collaborative robot with 7 rotational joints and recording position and orientation data of the dental forceps at 100Hz. The integrated torque and rotational sensors enable the robot to detect external forces which makes this robot collaborative and highly suitable for integration in this measurement setup. An ATI 16 bit Delta transducer is used for recording the force and torque data in 6 axis at a speed of 20Hz. A Logitech C920 Pro HD webcam is used to record a video stream of the experiment. The latter will facilitate the interpretation of data patterns when analyzing the data later on.

The platform Robotic Operating System (ROS) is used for software integration of the force/torque sensor, the video camera, and the collaborative robot [10]. ROS is an open source framework that allows for easy integration of several hardware sensors with robotic control and simulation. It provides hardware abstraction, device drivers, and libraries. The `image_pipeline` repository is used to convert the image data from the video camera to the ROS framework. For controlling the KUKA, the `iiwa_stack` repository is used which contains high level commands to collaborate with the robot through the ROS framework [11]. A custom ROS driver was written to read out the serial data from the FT-sensor and enable its usage in the ROS environment.

To enable the clinician to freely move the forceps, the robot mode is switched to a passive mode (impedance control). Impedance control enables a dynamic collaboration between the clinician and the robot. In this mode all 7 joints are acting as separate spring-damper systems. The stiffness and damping constants can be tuned by the user for each individual joint. High values will result in rigid joint motion, whereas lower values will result in more compliant/floating motion. To prevent joints drifting into joint limits and to facilitate the clinician during the experiments, joints numbers a_2 and a_5 are set to a higher stiffness and damping value compared to the other joints (Fig. 1). It results in a more compliant motion of the dental forceps.

Both the FT-sensor and robot need to be calibrated before each experiment to register the position and orientation of the teeth. The robot is used for calibration of the position and orientation of the teeth. Because of the orientation difference of the upper and lower jaw (vertical/horizontal) two calibration tools were necessary. A lower incisor dental forceps is used for calibration in the lower jaw, due to the 90 degree angle and its straight design. For the upper jaw a straight dental elevator (Usto-Lux, Ustomed, Germany) is used for calibration. The calibration is done by touching the center of the crown holding the tool in line with the z -axis of the tooth (see Fig. 5). The tool's position and orientation was then registered using the graphical user interface (see below, Section II-E). By combining the exact position of the holding device (using the scale provided on the setup's frame) and the positional information of the robot,

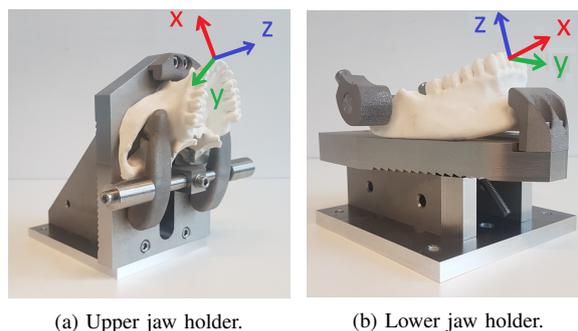


Fig. 5: Representation of the anatomical preparation of the upper- and lower jaw to fit the holding devices. The reference frames for upper- and lower teeth are shown.

a mathematical conversion can be made to determine the position and orientation of the teeth. Because the teeth in the upper jaw are positioned horizontally and the teeth in the lower jaw are positioned vertically, the z -axis of the teeth in the upper jaw is oriented along the x -axis of the robot's world frame, as opposed to the lower jaw in which the z -axis is aligned with the z -axis of the robot's world frame. Therefore, teeth in the upper jaw need a different transformation to the world frame than teeth in the lower jaw. The calibration method, as described above, enables the forces, torques and rotations of all teeth in both upper and lower jaw to be expressed in exactly the same reference frame, easing data analysis.

E. Graphical User Interface

To improve the workflow during the experiments, a Graphical User Interface (GUI) is designed as a platform where all components of the setup as well as the experiments can be managed simultaneously. The GUI allows for meta-information to be added to the experiments. It consists of a pre-operative, perioperative and post-operative window in which data are shown and can be edited, if necessary. In the pre-operative screen clinical data such as periodontal or restorative state can be filed. To optimize the flow of the experiments, predefined joint positions are determined in which most relevant joints are in their neutral status (Section II-C). These predefined starting positions are different for upper and lower jaw because of their different positions relative to the robot. They can be requested and executed from within the preoperative part of the GUI. During the experiments the GUI shows graphical information on actual measurements to enable live monitoring of the experiment. A summary of the experiment is shown and certain 'events' can be added to the experiment in the postoperative section. As an example, a marking can be added at a point in time where a complication has happened. The postoperative part also offers the opportunity to trim unuseful data, for example the time between the tooth being removed and the moment where the experiment is actually stopped in the GUI (usually a few seconds later).

The experiments took place in an in-hospital anatomy laboratory. Samples were obtained through the body donation program from the Department of Medical Biology, Section Clinical Anatomy and Embryology, of the Amsterdam UMC at the location Academic Medical Center in The Netherlands. The bodies from which the samples were taken were donated to science in accordance with Dutch legislation and the regulations of the medical ethical committee of the Amsterdam UMC at the location Academic Medical Center. The setup was tested with experiments on both conserved and fresh frozen cadaver jaws. A band saw was used to reduce the cadaver heads to the proportions as necessary to fit the holding devices. For the lower jaw this meant an oblique 45 degree bone cut from the gonial angle of the mandible towards the retromolar area. For the upper jaw a horizontal cut starting at the level of the infra-orbital rim was made. The cut was continued dorsally to the level of the articular tubercle and then connected to the oropharynx. See Fig. 5a and 5b. Soft tissue was largely removed by using standard surgical blades. Care was taken not to remove any of the attached gingiva as periodontal health was one of clinical parameters. As dental notation system the ISO system is used (International Standards Organization number 3950, Fédération Dentaire Internationale).

III. RESULTS

In order to provide a comprehensive overview of the data that can be obtained using this measurement setup, while also safeguarding the readability of this article, representative examples of data on movements, forces, and clinical data are shown. One of the main goals of this setup was to visualize what movements happen during tooth removal. To the authors best of knowledge, this has never been done before. In textbooks on oral surgery usually a short and basic movement pattern is advised for successful tooth removal [12]. Which movement pattern to choose is largely based on tooth root morphology. For example, a central upper incisor, which has only 1 root that usually has a round shape, is advised to ‘rotate’ out of the bony socket. For an upper molar with 3 roots a movement from buccal to the palatal side is advised, largely luxating towards the buccal side. Fig. 6 shows the movements recorded during removal of an upper central incisor (tooth number 21). In this figure the described pattern from the textbook can be clearly recognized. Rotations around x and y -axis are absent whilst a recurrent rotation around the tooth’s axis is evident. The data shows both a clockwise and counterclockwise rotation around the tooth’s axis that increases towards the clockwise side before the tooth is taken out. At the end of the movement a slight increase in movements around the x and y -axis shows a wiggle to release the tooth.

When compared to the movements during removal of a first upper molar (tooth number 16) on the right side a difference in movement pattern can be found. This first molar had, as usual, 3 roots. This means that rotation of the tooth is geometrically unfavorable. In Fig. 6 this can be recognized by the flat character of z -axis meaning no rotation takes

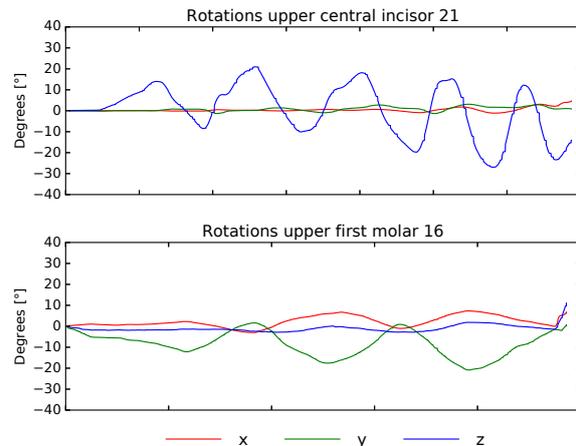


Fig. 6: Comparison of rotations of an upper incisor (21) and upper first molar (16)

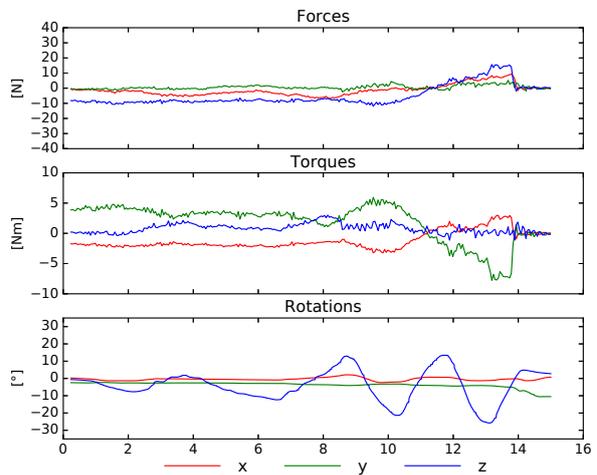


Fig. 7: Removal of a central upper incisor (21) by an experienced surgeon

place throughout the entire procedure. Rotation around the y -axis shows a buccal movement which increases over time. Movement around the x -axis (mesiodistal movement) shows a slight movement towards the mesial side during this buccal movement which means the tooth is moved in the direction of the opening of the mouth.

A. Forces and torques

When explaining tooth removal to dental students, usually one of the first things that is explained is that the idea of ‘pulling’ a tooth is incorrect. A tooth needs to be ‘pushed’ out. In terms of forces one could expect a negative force in the tooth’s root axis (z -axis). Fig. 7 shows the forces exerted during removal of a central upper incisor. It can be appreciated that, during the first phase of the treatment, the tooth is actually pushed into its socket. During this phase

only a little movement (rotation) can be distinguished. Later during the treatment we can see a clear turnaround in terms of forces. Pushing into the socket becomes pulling whilst movements are increasing, meaning the tooth is coming loose.

B. Clinical data

To gain a representative dataset, most experiments during the testing phase have been performed by the same experienced oral and maxillofacial surgeon. To test if the differences between an experienced and an inexperienced clinician can be visualized, a dental intern was asked to perform experiments as well. In total the surgeon removed 76 teeth of fresh frozen cadavers of which in 5 (7%) cases a fracture of a root occurred. The dental intern removed 21 teeth, also of fresh frozen cadaver head of which in 9 (43%) cases a fracture of a root occurred.

To see if the data can deliver us further insight in what the differences between the two clinicians are, a comparison of a removal of the same type of tooth between the dental intern and the experienced oral and maxillofacial surgeon can be made. Without the necessity of an in-depth analysis, we can see major differences between the removal a central upper incisor when this procedure is performed by a dental intern (Fig. 8) and an experienced oral and maxillofacial surgeon (Fig. 7). Both teeth were central upper incisor with a composite restoration, a healthy periodontium and a root length of 14mm. The dental student:

- exerts more than twice the amount of forces in the beginning of the procedure
- shows a less recognizable plan in terms of movements consisting of a mixture of rotational and buccopalatal movements
- fractures the root of the tooth. This was clinically noted to happen at $T(\text{seconds}) = 33$. Here a small spike in the forces and torques can be observed

The surgeon manages to keep forces and torques at a relative low and stable amount whilst increasing the movements (loosen the tooth).

IV. DISCUSSION

In this study a measurement setup is proposed that is the result of a strong collaboration between clinicians, mechanical and software engineers. It is capable of, for the first time, capturing the combination of high forces and subtle movements exerted during tooth removal procedures in high detail by using, amongst others, robot technology. First outcomes of experiments are used as a proof of the concept and show promising results. The dataset which can be built with this setup offers a unique insight in one of the oldest and most performed surgical procedures worldwide.

It is remarkable how underdeveloped the scientific understanding of tooth removal is. Only a few attempts have been undertaken in which moments were measured in an in vivo setting, in contrast to this study where an in vitro setup is proposed [1]–[5], [13]. The studies that have been performed thus far used either a strain gauge or manometer attached to,

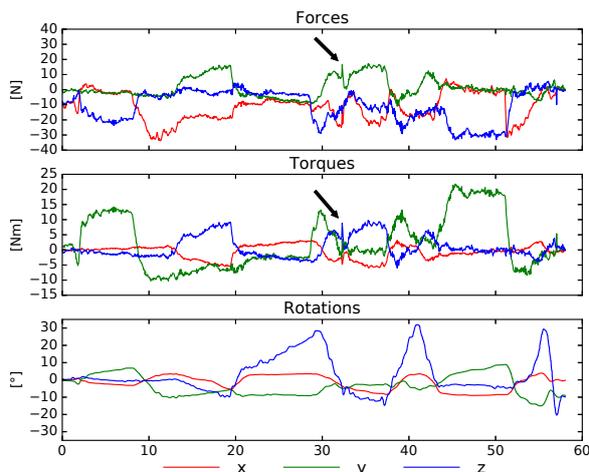


Fig. 8: Removal of a central upper incisor (11) by a dental student. The arrows indicate the spikes that occur at the instance the tooth fractures.

or integrated in, a dental forceps. They were therefore limited to measuring forces and moments, not the movements of the clinician. The outcomes are very limited and heterogeneous which shows the difficulty of analyzing tooth removal in vivo conditions. For example, Cicciu et al. [1] found a 25 fold increase in forces used in upper premolar removal compared to lower premolar removal whilst Lehtinen [2] and Ojala [5] found the forces between upper and lower canines to be indifferent. This shows that a benchmark to compare our results to is unfortunately not available.

The lack of technical possibilities to measure subtle (sub-millimeter) movements and high forces in all directions in an in vivo condition is the main reason why an in vitro setup was chosen to study tooth removal. Its design for in vitro measurements is also one of the major drawbacks of this setup. It will be unsure how data can be translated into in vivo circumstances. This is even more true since there is very limited in vivo data available to correlate the outcomes to. Next to that the setup is limited to the use of dental forceps. The elevator is also frequently used in tooth removal procedures, but its usage is much more diverse (different positions relative to the tooth for example) and we would need to measure the movement of the teeth themselves, which made it unsuitable for a first proof of concept. Finally the setup does not provide the possibility to measure clamping forces between the tooth and dental forceps. This would require mechanical changes to the dental forceps itself and might interfere with the normal usage of a dental forceps by the clinician. Despite its disadvantages the authors believe that, especially when using fresh frozen cadavers, the setup can be used to gain a unique and relevant new insight into tooth removal techniques.

Mechanically the development of the rigid fixation method for a human upper and, to a lesser extent, lower jaw was most challenging. Several designs were 3D-printed in plastic

and tested on conserved cadaver jaws on ease-of-fixation and rigidity of the fixation method before the final design was chosen and manufactured in stainless steel. When first testing the stainless steel setup a slight mobility of the jaw holders was noted due to the locking bolt in the rotational plate which was a prefabricated and gave some slack. It was later customized to a locking pin that could be tightened by rotation which resulted in a strong and complete rigid fixation of the jaws. During the experiments with fresh frozen jaws, out of 146 experiments, only 2 times an experiment failed because of loosening of the jaw within the holding device. Both times it involved an upper jaw and loosening was due to improper tightening of the holding device at the start of the experiment.

For the measurement of movements a robot was added to the setup. One of the major concerns when using the robot in a 'compliant' mode was the robot not being fully passive at all times. Especially when joint limits are approached with some pace, the robot showed resistance when adapting its joint position to enable certain positions or movements. To overcome this problem a 'best fit' starting position of the end-effector of the robot was to be found where most (relevant) joints were in a neutral position to ensure as little resistance as possible. Although it is difficult to measure the exact value of the resistance, it seems relatively small in comparison with the large amounts of forces exerted. The upper jaw was fixated with the occlusal plane in a vertical way and the lower jaw with its occlusal plane horizontal to mimic the clinical situation which required different "preset" joint positions for upper and lower jaws. These positions, that were optimized based on preference from the surgeons, were programmed starting position for all experiments. The combination of an adjustable frame and a rotational plate ensured roughly the same starting position for all experiments in upper and lower jaw. Pre-programming the same joint positions at the start of each experiment also added to the reproducibility of the experiments. Despite all efforts on creating a setup that comes as close to a clinical setting as possible, it must be noted that some resistance seems inevitable and this should be taken into account when interpreting results of these experiments. Despite a slight learning curve was noted when it comes to working with a passive robot arm, the feedback the authors received on clinical representativeness in general was very positive.

To calibrate the position of the tooth and its orientation relative to the FT-sensor and the robot a different dental instrument was used for both upper and lower jaw. It was aligned with the tooth axis by the clinician based on the orientation of the crown of the tooth. Despite efforts made to be as precise as possible some comments should be made. Firstly, even in an in vitro setting, it can be quite challenging to align a tool in all axis at the same time. Secondly, the crown forms only a small portion of the tooth. It is common knowledge in the field of dentistry that roots tend to divert to some extent (usually distally). To add to the precision of the measurements in future experiments it can be considered

to use CT-data to calibrate the position of the entire jaw by using anatomical landmarks rather than calibrating each tooth separately. This could also reduce duration of the experiments.

V. CONCLUSIONS AND FUTURE WORK

It is the goal of this research group to acquire data on every aspect of tooth removal. With this setup a dataset can be build that contains high quality data on every aspect of tooth removal. Data driven modelling will be used to analyze the large amount of data. A model is necessary to be able to understand what makes tooth removal (un)successful. Clinicians could learn from a model what parameters are essential to look for in clinic and to help predict the level of difficulty of an upcoming procedure. It could help them to decide when referral is necessary based on their own competence. The setup allows for different teaching instruments, i.e., plastic models or conserved cadavers, to be tested on representativeness. The derived dataset will be used to create new and evidence based learning material for dental students and young dentists. In a later phase some parts of the setup can be transformed for the use in an in vivo experiment to enable a correlation to clinical data.

REFERENCES

- [1] M. Ciccù, E. Bramanti, F. Signorino, A. Ciccù, and F. Sortino, "Experimental study on strength evaluation applied for teeth extraction: an in vivo study," *The Open Dentistry Journal*, vol. 7, p. 20, 2013.
- [2] R. Lehtinen and T. Ojala, "Rocking and twisting moments in extraction of teeth in the upper jaw," *International Journal of Oral Surgery*, vol. 9, no. 5, pp. 377–382, 1980.
- [3] A. MacGregor and G. Tomlinson, "An apparatus for measuring the forces of dental extraction," *British Journal of Oral Surgery*, vol. 17, no. 1, pp. 71–76, 1979.
- [4] V. Ahel, I. Brekalo, J. Ahel, and G. Brumini, "Measurement of tooth extraction forces in upper incisors," *Collegium Antropologicum*, vol. 30, no. 1, pp. 31–35, 2006.
- [5] T. Ojala, "Rocking moments in extraction of teeth in the lower jaw," *International Journal of Oral Surgery*, vol. 9, no. 5, pp. 367–372, 1980.
- [6] C. Hanson, T. Wilkinson, and M. Macluskey, "Do dental undergraduates think that Thiel-embalmed cadavers are a more realistic model for teaching exodontia?" *European Journal of Dental Education*, vol. 22, no. 1, pp. 14–18, 2018.
- [7] H. S. Brand, C. C. Van Der Cammen, S. M. Roorda, and J. A. Baart, "Tooth extraction education at dental schools across Europe," *BDJ Open*, vol. 1, p. 15002, 2015.
- [8] P. Onclin, van Minnen B, and B. Stegenga, "Recent increase in simple extractions in OMF-surgeries in the Netherlands studied," *Nederlands Tijdschrift voor Tandheelkunde*, vol. 123, no. 7-8, pp. 365–372, 2016.
- [9] KUKA Roboter GmbH, *LBR iiwa 7 R800, LBR iiwa 14 R820 specification*, 5th ed., Zugspitzstrasse 140, Augsburg, Germany, Jan 2015.
- [10] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng, "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, vol. 3, 2009.
- [11] C. Hennersperger, B. Fuerst, S. Virga, O. Zetting, B. Frisch, T. Neff, and N. Navab, "Towards MRI-based autonomous robotic US acquisitions: a first feasibility study," *IEEE Transactions on Medical Imaging*, vol. 36, no. 2, pp. 538–548, 2017.
- [12] B. Stegenga, A. Vissink, L. de Bont, F. Spijkervet, and E. Ommen, *MKA-chirurgie: Handboek voor mondziekten, kaak-en aangezichtschirurgie*. Koninklijke van Gorcum B.V., 2013.
- [13] V. Ahel, T. Čabov, S. Špalj, B. Perić, D. Jelušić, and M. Dmitrašinović, "Forces that fracture teeth during extraction with mandibular premolar and maxillary incisor forceps," *British Journal of Oral and Maxillofacial Surgery*, vol. 53, no. 10, pp. 982–987, 2015.

Bibliography

- [1] K. Gaballah, “Avoiding common problems in tooth extractions,” *DentalTribuneUK Edition*, 2016.
- [2] J. W. Friedman, “The prophylactic extraction of third molars: A public health hazard,” *American Journal of Public Health*, vol. 97, no. 9, pp. 1554–1559, 2007.
- [3] M. Cicciù, E. Bramanti, F. Signorino, A. Cicciù, and F. Sortino, “Experimental Study on Strength Evaluation Applied for Teeth Extraction: An In Vivo Study,” *The Open Dentistry Journal*, vol. 7, no. 1, pp. 20–26, 2013.
- [4] R. Lehtinen and T. Ojala, “Rocking and twisting moments in extraction of teeth in the upper jaw,” *International Journal of Oral Surgery*, vol. 9, no. 5, pp. 377–382, 1980.
- [5] T. Ojala, “Rocking moments in extraction of teeth in the lower jaw,” *International Journal of Oral Surgery*, vol. 9, no. 5, pp. 367–372, 1980.
- [6] C. Hanson, T. Wilkinson, and M. Macluskey, “Do dental undergraduates think that Thiel-embalmed cadavers are a more realistic model for teaching exodontia?,” *European Journal of Dental Education*, vol. 22, no. 1, pp. e14–e18, 2018.
- [7] H. S. Brand, C. C. J. van der Cammen, S. M. E. Roorda, and J. A. Baart, “Tooth extraction education at dental schools across Europe,” *BDJ Open*, vol. 1, no. 1, pp. 1–6, 2015.
- [8] I. R. de Boer, D. R. Bakker, P. R. Wesselink, and J. M. Vervoorn, “[The Simodont in dental education].,” *Nederlands tijdschrift voor tandheelkunde*, vol. 119, no. 6, pp. 294–300, 2012.
- [9] P. Onclin, B. van Minnen, and B. Stegenga, “Recente toename van reguliere extracties in mka-chirurgie onderzocht,” *Nederlands tijdschrift voor tandheelkunde*, vol. 123, no. 7-8, pp. 365–372, 2016.
- [10] B. Stegenga, A. Vissink, L. de Bont, and F. Spijkervet, *MKA-chirurgie Handboek voor Mondziekte, Kaak- en Aangezichtschirurgie*. Gorcum B.V., Koninklijke van, 2013.

- [11] T. Ojala, "Dental extraction movements and their action times. Strain gauge measurements of rocking and twisting movements in relation to tooth extraction, including observations on the length of the action time," *Proc Finn Dent Soc*, vol. 80, pp. 145–148, 1984.
- [12] W. M. De Graaf, "Towards the Modern Analysis of Tooth Removal Procedures - A Comparative Study," tech. rep., TUDelft, 2020.
- [13] A. K. Saxe, L. J. Louie, and J. Mah, "Efficiency and effectiveness of SureSmile.," *World journal of orthodontics*, vol. 11, no. 1, pp. 16–22, 2010.
- [14] H. Deng, Z. Xia, S. Weng, Y. Gan, J. Xiong, Y. Ou, and J. Zhang, "Motion planning and control of a robotic system for orthodontic archwire bending," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3729–3734, 2015.
- [15] J. Jiang, X. Ma, Y. Zhang, B. Huo, and Y. Liu, "Study on Three-Dimensional Digital Expression and Robot Bending Method of Orthodontic Archwire.," *Applied bionics and biomechanics*, vol. 2018, p. 2176478, 2018.
- [16] X. Sun, F. D. McKenzie, S. Bawab, J. Li, Y. Yoon, and J.-K. Huang, "Automated dental implantation using image-guided robotics: registration results.," *International journal of computer assisted radiology and surgery*, vol. 6, pp. 627–634, 9 2011.
- [17] K. Yu, S. Uozumi, K. Ohnishi, S. Usuda, H. Kawana, and T. Nakagawa, "Stereo vision based robot navigation system using modulated potential field for implant surgery," *2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 493–498, 2015.
- [18] A. Banerjee, T. F. Watson, and E. A. M. Kidd, "Dentine caries excavation: a review of current clinical techniques," *British Dental Journal*, vol. 188, no. 9, pp. 476–482, 2000.
- [19] F. Yuan, Y. Wang, Y. Zhang, Y. Sun, D. Wang, and P. Lyu, "An automatic tooth preparation technique: A preliminary study.," *Scientific reports*, vol. 6, p. 25281, 4 2016.
- [20] F. Yuan and P. Lyu, "A preliminary study on a tooth preparation robot," *Advances in Applied Ceramics*, pp. 1–6, 9 2019.
- [21] T. Otani, A. J. Raigrodski, L. Mancl, I. Kanuma, and J. Rosen, "In vitro evaluation of accuracy and precision of automated robotic tooth preparation system for porcelain laminate veneers.," *The Journal of prosthetic dentistry*, vol. 114, pp. 229–235, 8 2015.
- [22] A. C. Celsus, "Celsus: De Medicina. Volume I," *Journal of the American Medical Association*, vol. 105, p. 1632, 1935.
- [23] T. Mitchell, *Machine Learning*. McGraw-Hill, 1997.
- [24] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*. The MIT Press, 2006.
- [25] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An Introduction to Deep Reinforcement Learning," *Foundations and Trends® in Machine Learning*, vol. 11, no. 3-4, pp. 219–354, 2018.
- [26] J. Kober and J. Peter, "Policy search for motor primitives in robotics," *Springer Tracts in Advanced Robotics*, vol. 97, pp. 83–117, 2014.

-
- [27] J. Kober, A. Wilhelm, E. Oztop, and J. Peters, “Reinforcement learning to adjust parametrized motor primitives to new situations,” *Springer Tracts in Advanced Robotics*, vol. 97, pp. 119–147, 2014.
- [28] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, Inc., 2019.
- [29] D. W. Gareth James Trevor Hastie, Robert Tibshirani, *An introduction to statistical learning : with applications in R*. New York : Springer, [2013] ©2013, 2013.
- [30] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O’Reilly Media, Inc., 2018.
- [31] M. Du, N. Liu, and X. Hu, “Techniques for interpretable machine learning,” *Communications of the ACM*, vol. 63, pp. 68–77, 12 2019.
- [32] D. Y. Harvey and M. D. Todd, “Automated Feature Design for Numeric Sequence Classification by Genetic Programming,” *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 4, pp. 474–489, 2015.
- [33] A. Bagnall, L. Davis, J. Hills, and J. Lines, “Transformation based ensembles for time series classification,” *Proceedings of the 12th SIAM International Conference on Data Mining, SDM 2012*, pp. 307–318, 2012.
- [34] H. Liu and G. Dong, *Feature Engineering for Machine Learning and Data Analytics*. CRC Press, 3 2018.
- [35] B. D. Fulcher, “Feature-Based Time-Series Analysis,” *Feature Engineering for Machine Learning and Data Analytics*, pp. 87–116, 2018.
- [36] B. D. Fulcher and N. S. Jones, “Highly comparative feature-based time-series classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3026–3037, 2014.
- [37] J. Timmer, C. Gantert, G. Deuschl, and J. Honerkamp, “Characteristics of hand tremor time series,” *Biological cybernetics*, vol. 70, no. 1, p. 75–80, 1993.
- [38] P. Domingos and M. Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss,” *Machine Learning*, vol. 29, no. 2, pp. 103–130, 1997.
- [39] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian Network Classifiers,” *Machine Learning*, vol. 29, no. 2, pp. 131–163, 1997.
- [40] I. Rish, “An Empirical Study of the Naïve Bayes Classifier,” *IJCAI 2001 Work Empir Methods Artif Intell*, vol. 3, 1 2001.
- [41] A. Singh, N. Thakur, and A. Sharma, “A review of supervised machine learning algorithms,” in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315, 2016.
- [42] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

- [43] KUKA Roboter GmbH, “LBR iiwa 7 R800, LBR iiwa 14 R820 specification,” tech. rep., KUKA Roboter GmbH, 2015.
- [44] ATI Industrial Automation, “Intelligent Multi-axis Force / Torque Sensor System Installation and Operation Manual,” Tech. Rep. August, ATI Industrial Automation, 2004.
- [45] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng, “ROS: an open-source Robot Operating System,” in *ICRA Workshop on Open Source Software*, 2009.
- [46] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, and N. Navab, “Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 2, pp. 538–548, 2017.
- [47] W. Hamilton, *Elements of Quaternions*. Chelsea Pub Co; 3rd Edition (March 1, 1969), 1866.
- [48] D. Hoag, “Consideration of Apollo IMU Gimbal Lock,” tech. rep., MIT Instrumentation Laboratory Document E-1344, 1963.
- [49] Narendra and Fukunaga, “A Branch and Bound Algorithm for Feature Subset Selection,” *IEEE Transactions on Computers*, vol. C-26, no. 9, pp. 917–922, 1977.
- [50] D. Ververidis and C. Kotropoulos, “Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition,” *Signal Processing*, vol. 88, no. 12, pp. 2956–2970, 2008.
- [51] K.-g. Zhang, Y.-d. Zhang, and M. Wang, “A Unified Approach to Interpreting Model Predictions Scott,” *Nips*, vol. 16, no. 3, pp. 426–430, 2012.
- [52] V. Ahel, T. Čabov, S. Špalj, B. Perić, D. Jelušić, and M. Dmitrašinić, “Forces that fracture teeth during extraction with mandibular premolar and maxillary incisor forceps,” *British Journal of Oral and Maxillofacial Surgery*, vol. 53, no. 10, pp. 982–987, 2015.
- [53] Human Performance Research Group, “NASA Task Load Index User Manual v. 1.0,” 1986.
- [54] M. Alberts, E. M. Smets, J. H. Vercoulen, B. Garssen, and G. Bleijenberg, “[’Abbreviated fatigue questionnaire’: a practical tool in the classification of fatigue].,” 1997.

Glossary

List of Acronyms

CAD	Computer Aided Design
ROS	Robotic Operating System
GUI	Graphical User Interface
NB	Naïve Bayes
KNN	K-Nearest Neighbor
SHAP	SHapley Additive exPlanations
LR	Logistic Regression
OVR	One-Versus-Rest
NASA-TLX	NASA Task Load Index
SFQ	Shortened Fatigue Questionnaire
LBFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno
DT	Decision Tree
NN	Neural Network
SVM	Support Vector Machine
ZVC	Zero Velocity Crossing

