Follow us @BoneJointOpen

Freely available online OPEN ACCESS

BJO

GENERAL ORTHOPAEDICS

The application of machine learning to balance a total knee arthroplasty

M. A. Verstraete,
R. E. Moore,
M. Roche,
M. A. Conditt

St. Helena Hospital, Saint Helena, California, USA

Aims

The use of technology to assess balance and alignment during total knee surgery can provide an overload of numerical data to the surgeon. Meanwhile, this quantification holds the potential to clarify and guide the surgeon through the surgical decision process when selecting the appropriate bone recut or soft tissue adjustment when balancing a total knee. Therefore, this paper evaluates the potential of deploying supervised machine learning (ML) models to select a surgical correction based on patient-specific intra-operative assessments.

Methods

Based on a clinical series of 479 primary total knees and 1,305 associated surgical decisions, various ML models were developed. These models identified the indicated surgical decision based on available, intra-operative alignment, and tibiofemoral load data.

Results

With an associated area under the receiver-operator curve ranging between 0.75 and 0.98, the optimized ML models resulted in good to excellent predictions. The best performing model used a random forest approach while considering both alignment and intra-articular load readings.

Conclusion

The presented model has the potential to make experience available to surgeons adopting new technology, bringing expert opinion in their operating theatre, but also provides insight in the surgical decision process. More specifically, these promising outcomes indicated the relevance of considering the overall limb alignment in the coronal and sagittal plane to identify the appropriate surgical decision.

Cite this article: *Bone Joint Open* 2020;1-6:236–244.

Keywords: Balancing, Machine Learning, Total Knee Arthroplasty

Introduction

With the advance of technology in the field of joint arthroplasty surgery, quantification is quickly becoming the new normal for both component alignment and soft-tissue balancing. Various studies have shown a direct clinical benefit to the patient when using advanced technologies during surgery. Examples include a reduction in pain when using robotic technology to assist the implant component positioning in total knee arthroplasty (TKA)[1] and improvement of patient satisfaction when assuring a quantitatively balanced knee.[2,3] However, an often-overlooked aspect of these technologies is the surgeons' learning experience that accompanies adoption and implementation of new technology. As the introduction of new technology in the operating theatre often comes with an overload of numbers representing newly quantified metrics, it can take some time for the surgeon to understand how to interpret the numbers and make subsequent intra-operative decisions. This early learning curve may therefore result in some early technical challenges resulting in outcomes such as extended surgical times.[4-6] Though cadaver training labs have their value in lowering the entry barrier, they seldomly replicate pathology and deformity frequently encountered in the operating room in arthritic knees. The experience built by early adaptors of technology often gets lost or at best inadequately transferred when subsequent users adopt these new

Correspondence should be sent to Matthias A Verstraete; email: Matthias.Verstraete@OrthoSensor.com

doi: 10.1302/2633-1462.16.BJO-2020-0056.R1

Bone Joint Open 2020;1-6:236–244.

technologies using the traditional mentor learning model prevalent in medicine.

With advances in data science and machine learning (ML) now demonstrating the ability to assist patient selection for joint arthroplasty surgery,[7] the possibility arises of also using such techniques to bring expert decision making to the novice technology user. More specifically, this paper aims to provide an ML framework that helps navigate a multi-dimensional space such as the one encountered when balancing the soft-tissue envelope during TKA surgery. While balancing a total knee, the leg's alignment, range of movement as well as soft-tissue tensions need to be considered during the surgical decision-making. This creates a multi-dimensional problem as both soft tissue corrections and bone recuts can be considered as successful maneuvers to achieve a balanced knee. It is thus not always clear which surgical corrections should be performed, in what order and at what phase during surgery, presenting a classical classification problem in the world of ML.

Whereas the skill of balancing a total knee has historically been mystified as it is built around expert opinion and surgeon feel,[8,9] the introduction of intra-operative sensor technology has brought quantification to the knees' state of balance.[10,11] Given the multi-dimensional nature of such decisions as well as the variability in pathological conditions encountered during surgery, traditional learning methods based on peer-to-peer knowledge transfer or learn-by-doing tend to be time-consuming and often lack the subtleties required to successfully identify and perform the need for less common releases. In contrast, ML, as part of the artificial intelligence domain, presents an opportunity to address such difficulties, particularly as it has proven to be effective in solving classification problems.[12,13]

Therefore, this paper will focus on building and validating ML algorithms classifying surgical corrections based on the encountered intra-operative sensor readings when balancing a knee in combination with the readings from a surgical navigation system, thereby aiming to demonstrate the quantified decision process underlaying valuable and hard-earned expert opinion when using these technologies.

Methods

Clinical dataset. All primary total knee surgeries performed by a single surgeon using a mid-vastus arthrotomy and posterior stabilized TKA design with a single radii femoral component between January 2017 and August 2018 were included in this IRB approved study. This resulted in 479 cases, during which surgical navigation (OrthoMap, Stryker, Kalamazoo, Michigan, USA) and smart tibial trial components (VERASENSE, OrthoSensor, Fort Lauderdale, Florida, USA) were used. Surgical navigation was used to evaluate the initial pathological deformity and its correctability in the coronal and sagittal planes. Subsequently, the instantaneous alignment during trialing was assessed also using surgical navigation as the surgeon attempted to correct the pathological deformity. The smart tibial components were used to measure the intra-articular tibiofemoral loads during trialing, with the patient in the supine position. During these load measurements, the surgeon assessed the knee's neutral position, carefully avoiding adding any additional varus or valgus stress as well as compression/distraction force to the knee joint. The load data was captured at both 10° and 90° of flexion.

During the trialing phase, the loads in each compartment and alignment in both planes were documented preceding every surgical correction leading ultimately to a balanced knee (as discussed in the next section).

Of the 479 consecutive knees, 21 knees had missing data. With an average of approximately three iterations to achieve a balanced knee per TKA, this led to 1,391 recorded steps. Some steps had ill-documented load and alignment data feeding into the chosen surgical correction and were eliminated from the dataset. For 116 steps, more than one surgical correction was chosen. This resulted in 1,437 documented decisions representing a complex, real-world dataset that was used as input for the ML algorithms as detailed inbelow.

For the purpose of this paper, this clinical dataset was interpreted as a classification problem, during which a specific surgical decision was to be binarily indicated or not based on the load and alignment readings. As described in the previous paragraph, these decisions are not necessarily mutually exclusive since there were instances where the same load and alignment readings led to multiple, simultaneous decisions. Therefore, every surgical correction was so-called "one-hot encoded", and a separate ML model was built for each surgical decision. In practice, this implies that a separate input dataset was built for every decision. This input dataset documented, for each set of load and alignment readings encountered, whether the surgical decision of interest was selected (1) or not selected (0), regardless of other decisions potentially indicated by these readings. For each decision, this resulted in 1,305 unique cases in the input dataset. This is graphically represented by Figure 1.

Surgical corrections. During total knee surgery, a strict mechanical alignment philosophy was adopted with a targeted neutral final coronal alignment. Analogous, the sagittal deformity as measured with the navigation system was targeted to be zero degrees of overall limb flexion (no flexion contracture or hyper extension). When balancing a total knee, the surgeon aimed for a medial and lateral tibiofemoral load in the range of 10 lbf to 40 lbf per compartment at both 10° and 90°. In addition, the mediolateral load difference target was to not exceed 15 lbf.[3,11] During the iterative process leading to these

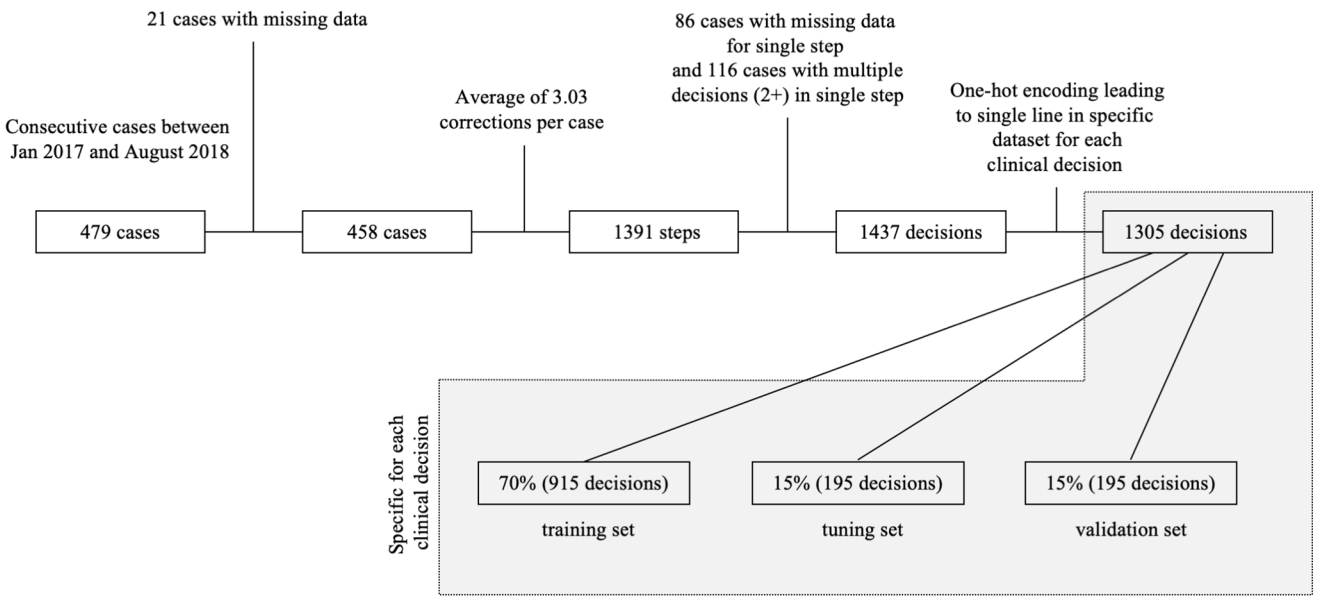


Fig. 1

Flowchart of surgical data collected leading to decision specific clinical dataset.

(ideal) target numbers for a balanced knee, the surgeon documented either of eight possible surgical corrections after evaluating the medial and lateral tibiofemoral loads at 10° and 90° of flexion, as well as the coronal and sagittal alignments at that time. These surgical corrections involve:

- **Bone recuts:**
  - Tibia recut: adding varus to the tibia. The exact amount of varus was beyond the scope of the current exercise as noted in the limitations of this paper.
  - Femur recut: proximalizing the femoral component by iterating on the distal femoral cut.
- **Soft tissue adjustments:**
  - Pie-crusting of the medial collateral ligament (MCL): through the use of a 14 gauge needle, tight bands of the MCL were poked.[14-16]
  - Arcuate release: With the knee in extension the arcuate ligament, which is a thickening of the lateral capsule posterior the iliotibial band, is cut with curved Mayo scissors along the joint line in an anterior to posterior direction.
  - Popliteus release: With the knee in 90° of flexion the popliteus, a lateral stabilizing structure situated in the posterolateral corner, is tensioned with a lamina spreader positioned between the tibial plateau and posterior femur and the mid-substance of the tendon in cut with a scalpel
  - ITB release: With the knee in extension, the iliotibial band (ITB), which lies anterior to the arcuate ligament, is cut with curved Mayo scissors along the joint line in an anterior to posterior direction.
  - Posterior capsule release: With the knee in 90° of flexion and tensioned with a lamina spreader between the femur and tibia, the posterior capsule is subperiosteally dissected off the posterior femur using electrocautery and subsequently further dissected proximally off the femur using a Cobb elevator.
- **Increase in poly insert thickness.**

The above corrections were seen as the potential decisions of the classification problem at hand. In addition, the decision to accept the current loads and alignments was also seen as a potential outcome (balanced, well-aligned knee). This led to a finite set of nine surgical decisions that could be selected based on readings from the navigation and sensor systems.

Machine learning. Deploying ML for classification problems belongs to the domain of "supervised ML"; models are built with the primary aim to reproduce what humans (i.e. expert surgeons) have decided before. In reality, this often leads to a set of model parameters that are fit on a dataset with known model inputs and outputs, such that new model inputs can then be fed to the model and the model can make an accurate prediction as to which output is most appropriate. Therefore, the clinical dataset was split in three groups such that – in relative terms – a similar prevalence of each surgical decision was achieved in each group. A first group, consisting of 70% of the entire dataset, was used as a training set. The second group consisting of an additional 15% of the data was used to tune the model's hyperparameters, and the remaining 15% of data was used for validation (Figure 1). By doing so, overfitting of the model (hyper-)parameters was avoided.[17] The results shown in this paper represent the values obtained when deploying the developed, tuned models in combination with the validation set.

Table I. Overview of different feature sets considered for ML algorithms.

Feature Set	Medial load @ 10°	Lateral load @ 10°	Medial load @ 90°	Lateral load @ 90°	Varus/Valgus deformity pre-op	Max. Extension deformity pre-op	Varus/Valgus during trialing	Max. extension during trialing
FS_1	√	√	√	√				
FS_2	√	√	√	√			√	√
FS_3	√	√	√	√	√	√	√	√

The model's input parameters represent the case-specific load and alignment data that lead to a surgical decision; these are commonly referred to as the feature set. In this paper, three different feature sets are considered for the ML models (Table I). The first feature set (FS_1) is limited to four numbers representing the medial and lateral intra-articular loads at 10° and 90° of flexion, respectively. The second feature set (FS_2) contains the previous numbers in addition to the coronal deformity in extension (varus/valgus) as well as the maximum (hyper) extension during the trialing phase. The third feature set (FS_3) uses all features in the second set as well as the pre-resection, pathological deformity in the coronal and sagittal plane, as recorded by the surgical navigation system before making any bone cuts.

The performance of the ML models described in the next paragraph is evaluated using the area under the receiver operator curve. This performance metric considers the specificity and sensitivity of the models and is often preferred over the model accuracy, also evaluated in this paper.

Within the scope of this paper, three different ML models have been implemented and trained for each surgical decision: a random forest (RF), a linear support vector machine (SVM) and an artificial neural network (ANN). A description of these models is provided in Supplementary Figure A1 (online supplementary figure 1).

These ML models have been implemented using Python v.3.2 with the Tensorflow 2.0 and Keras package for the ANN and SciKitLearn for the random forest and linear support vector machines. When comparing different models and feature sets, non-parametric Kruskal-Wallis tests were used to compare model performance characteristics such as the area under the receiver-operator curve or the model accuracy.

Results

Target alignment and loads. When reviewing the pre-resection pathological deformity, a wide range of varus/valgus deformities were observed ranging between 11° valgus to 19.5° varus. With the implants cemented in their final position, a narrower coronal deformity was seen from the surgical navigation readings, largely limited to ± 3° varus/valgus deformity (Figure 2a). Similarly, the pre-resection sagittal deformity ranged between 11° of hyperextension to 22° of flexion contracture. With the final implants cemented in place, a narrow range of ± 2.5° was achieved (Figure 2b).

When looking at the final compartmental loads (Figure 2c&d), the average medial and lateral loads of 32.5 lbf and 25.5 lbf, respectively, were achieved at 10° of flexion. At 90° of flexion, the average medial load was 23.7 lbf and the average lateral load was 22.9 lbf.

Surgical decisions. Not every surgical decision was equally prevalent in the clinical dataset, thus potentially limiting the ability of the ML algorithms to learn from these past events and thus accurately predict future incidences. For all considered surgical decisions, this is schematically shown in Figure 3a. As expected, every case eventually ended in a decision to stop the balancing process (458 observations). In contrast, popliteus and ITB releases as well as femur recuts were among the least prevalent decisions taken during surgery to balance a knee with respectively 19, 13, and 23 documented observations in the database.

Effect of ML model. After tuning the algorithms' hyper-parameters, the algorithm performance on the validation set was evaluated using both the area under the receiver-operator curve (AUC) and the prediction accuracy. Looking only at the model performance considering the full feature set, FS_3, the resulting AUCs are summarized in Table II and Figure 3b. Overall, the random forest model performs superior to the support vector machine and artificial neural network for this set of features. The median AUC for the RF model was 0.89, while the SVM and ANN scored 0.82 and 0.83 respectively. This difference was strongly significant based on a Kruskal-Wallis test ($p = 0.001$ for ANN vs. RF and $p < 0.001$ for SVM vs. RF). Focusing on the individual surgical decisions, model performance for decisions that are less prevalent in the clinical dataset (e.g. popliteus release, femoral recuts) is clearly more scattered and often inferior to the decisions that are more prevalent (e.g. arcuate release, MCL pie-crusting). In contrast, looking at the accuracy of the predictions, no significant differences were observed between the models. The accuracy was more closely linked to the number of events in the dataset; accuracies of 98% to 99% were observed for the clinical decisions that are less prevalent in the dataset such as the ITB release or popliteus release. Focusing on the ITB releases for instance, this is understood since a naive algorithm that avoids recommending

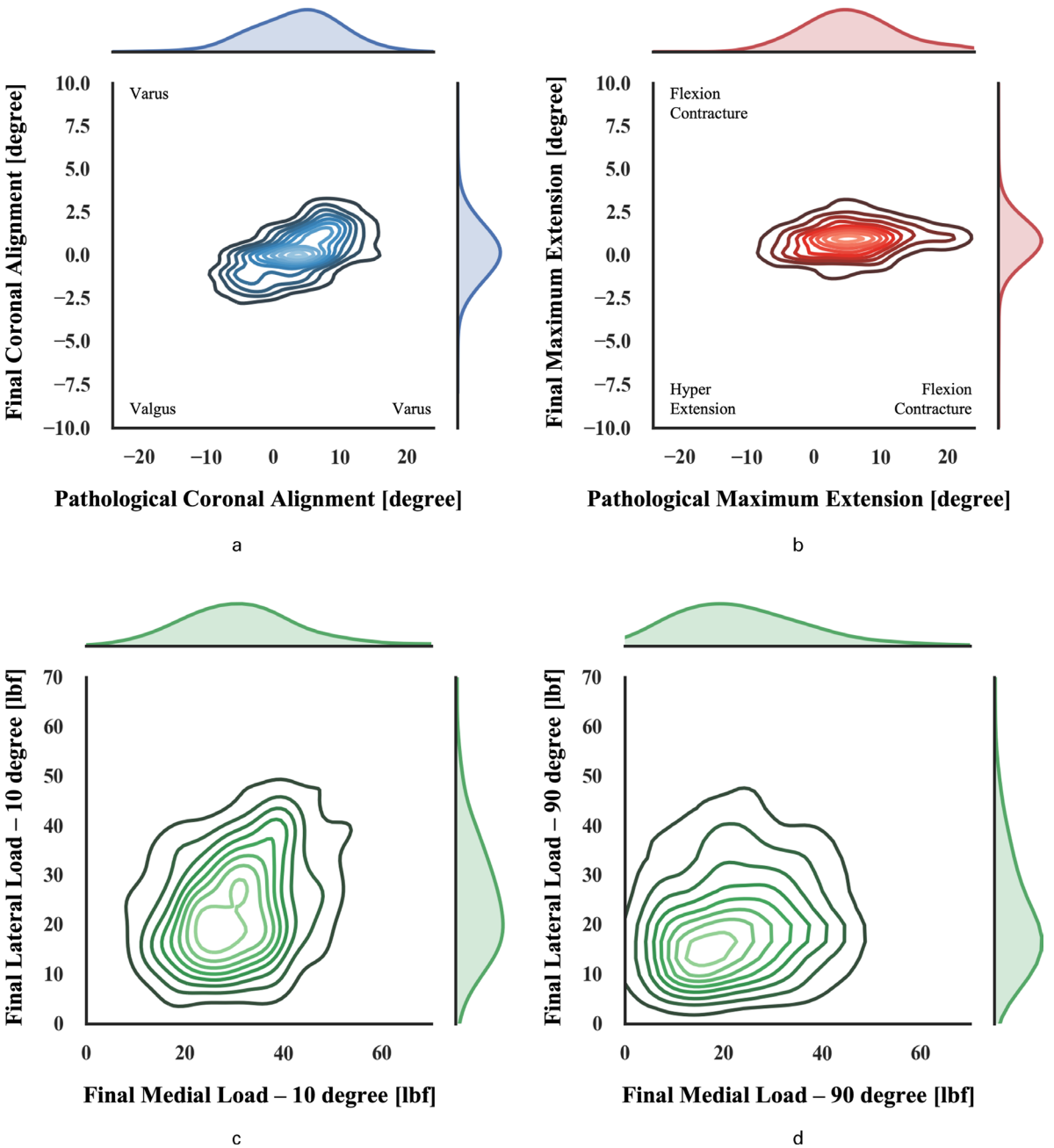


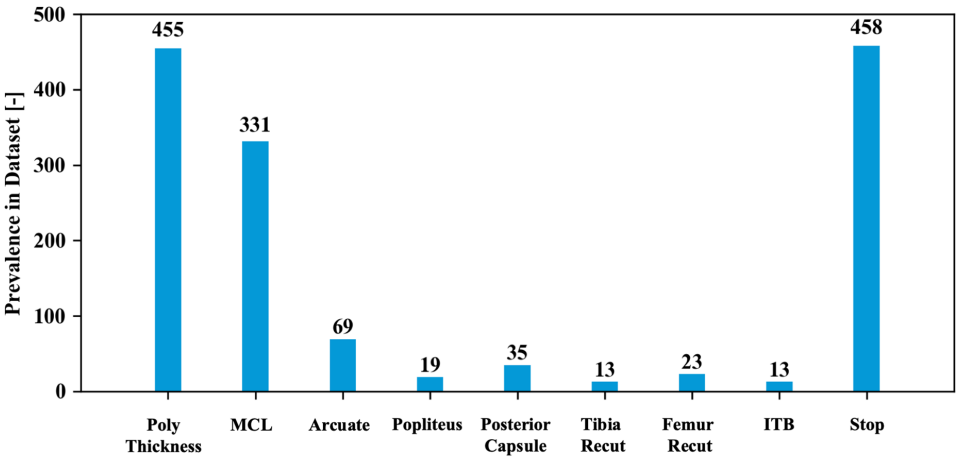
Fig. 2

Kernel density approximation showing distribution of pre-resection alignment and alignment with the final implants cemented in place in the coronal (a) and sagittal (b) plane as well as the final medial and lateral loads at 10 (c) and 90 (d) degree of flexion.

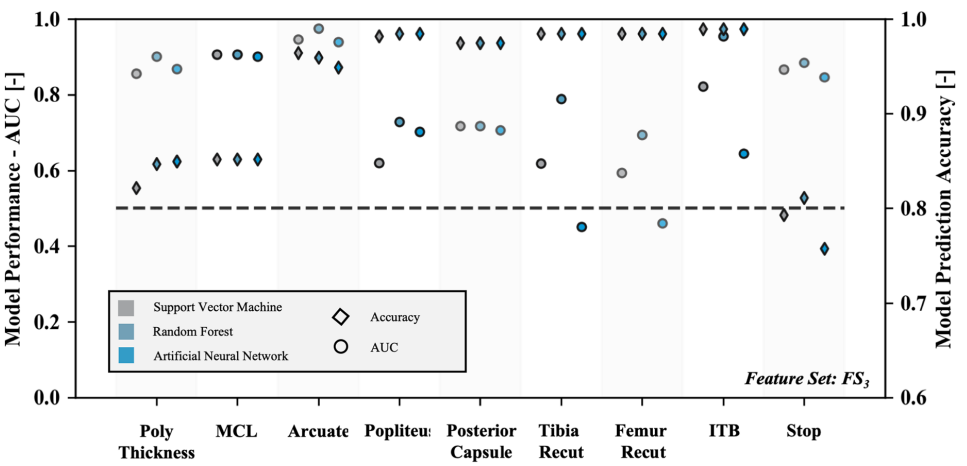
this surgical decision (always false) predicts the decision process correctly for a large relative number of cases (13 occurrences in 1,305 observations represents a success ratio of (1 - 13/1,305) x 100 = 99.00%) .

Effect of feature set

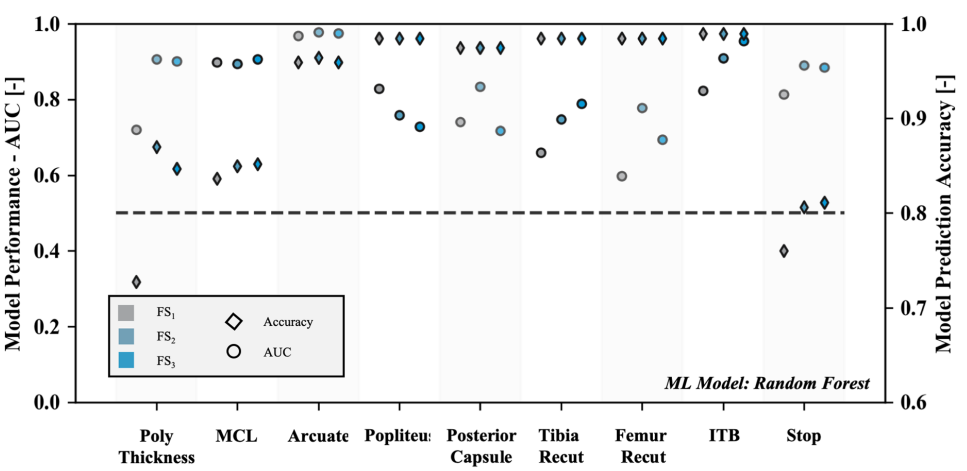
Using the random forest algorithm, the various feature sets are used to train the algorithm for each of the clinical decisions. In general, the algorithm performance increases with adding the alignment parameters during trialing and adding pre-resection, pathological alignment conditions. More specifically, adding the intra-operative alignment parameters to the feature set (FS_1 versus FS_2) increased model performance in predicting the surgical corrections; the median AUC was 0.80 for FS_1 and 0.89 for FS_2. This difference was statistically significant based on a Kruskal-Wallis test (p = 0.004). In contrast, adding the pre-resection alignment information (FS_3) did not affect the model performance relative to the condition with the intra-operative alignment information (FS_2). No



a



b



c

Fig. 3

Prevalence of surgical decisions for which ML model is built in clinical dataset (a) with performance of models expressed by area under the receiver-operator curve for each considered model using the full feature set (FS3) (b) and evaluation of selected feature set on performance of random forest algorithm (c) with dotted line representing an area under the curve of 0.5.

Table II. Performance of machine learning algorithms developed using feature set FS_3 expressed by area under the receiver-operator curve (above the line) and prediction accuracy (below the line).

	Poly thickness	MCL	Arcuate	Popliteus	Posterior capsule	Tibia recut	Femur recut	ITB	Stop
Support Vector Machine	0.86	0.91	0.95	0.62	0.72	0.62	0.59	0.82	0.87
	---	---	---	---	---	---	---	---	---
	0.82	0.85	0.96	0.98	0.97	0.98	0.98	0.99	0.79
Random Forest	0.90	0.91	0.98	0.73	0.72	0.79	0.69	0.95	0.89
	---	---	---	---	---	---	---	---	---
	0.85	0.85	0.96	0.98	0.97	0.98	0.98	0.99	0.81
Artificial Neural Network	0.87	0.90	0.94	0.70	0.71	0.45	0.46	0.64	0.85
	---	---	---	---	---	---	---	---	---
	0.85	0.85	0.95	0.98	0.97	0.98	0.98	0.99	0.76

Table III. Performance of Random Forest algorithms developed for various feature sets expressed by area under the receiver-operator curve (above the line) and prediction accuracy (below the line).

	Poly thickness	MCL	Arcuate	Popliteus	Posterior Capsule	Tibia Recut	Femur Recut	ITB	Stop
FS_1	0.72	0.90	0.97	0.83	0.74	0.66	0.60	0.82	0.81
	---	---	---	---	---	---	---	---	---
	0.73	0.84	0.96	0.98	0.97	0.98	0.98	0.99	0.76
FS_2	0.91	0.89	0.98	0.76	0.83	0.75	0.78	0.91	0.89
	---	---	---	---	---	---	---	---	---
	0.87	0.85	0.96	0.98	0.97	0.98	0.98	0.99	0.81
FS_3	0.90	0.91	0.98	0.73	0.72	0.79	0.69	0.95	0.89
	---	---	---	---	---	---	---	---	---
	0.85	0.85	0.96	0.98	0.97	0.98	0.98	0.99	0.81

significant difference was seen between model performance when considering FS_2 versus FS_3 ($p = 0.758$). This data is summarized in Table III and Figure 3c. With respect to the prediction accuracy, no significant differences were observed between the different feature sets.

Discussion

With the introduction of intra-operative sensor technology, the balance state of the soft tissues can now be quantitatively assessed during TKA. However, achieving a balanced knee is still challenging as it is a classic over-determined problem; various strategies can be followed to achieve balance when guided by sensor technology. Based on an extensive clinical database from a single, expert user, this paper provides insight in the potential of using MLL algorithms to address various imbalance scenarios encountered during total knee surgery.

To evaluate various ML models, a number of performance parameters can be considered. In this paper, the area under the receiver-operator curve and the predication accuracy were both evaluated. Whereas the former showed significant differences between the various models and feature sets, these differences were not withheld when looking at the prediction accuracy. More specifically, the latter tends to be unrepresentative of model performance and easily gives a false impression of the model's ability to guide decisions when the occurrence of a given decision in the dataset is low. For a popliteus release, for instance, this is particularly relevant as the prevalence is limited to approx. 1% in the current clinical dataset. As such, a trivial model that never indicates a popliteus release achieves a misleading prediction accuracy of approximately 99%, not considering the model sensitivity and specificity.

Predictive models can be driven by a number of different mathematical models. In this paper, the random forest models performed superior compared to the implemented artificial neural networks and support vector machines. This is in line with a recent comparative study, comparing various supervised models in the medical field.[18] While describing the fundamental reasons of these differences for the given clinical dataset goes beyond the scope of this paper and often remains a black box, it is worth noting that exploring a variety of models should be considered when searching forML solution strategies.[18] When selecting the optimum feature set in combination with this random forest model, the overall model performance – as expressed by the area under the receiver operator curve – ranges between 0.75 and 0.98 for the various surgical corrections. Compared to other medical applications, it is concluded that these values represent acceptable to outstanding model predictions.[19-21] In fact, a 100% prediction accuracy or area under the receiver operator curve might not even be achievable using supervised ML models, as this would suggest that surgical decisions are 100% consistent, something that can be challenged given the general variability seen in the surgeon's decisions when templating

arthroplasty surgery.[22] The outstanding predictions of the model presented in this paper indicate that these well-tuned and validated ML models can make the experience of an expert surgeon/panel accessible to the broader community. Whereas the introduction of new technology has inevitably been associated with a learning curve,[5,6] deploying ML models can overcome the risk of data overload and the challenge of data interpretation for new users by providing case-specific guidance directly in the operating theatre. An idea introduced by the psychologist Ericsson and popularized by Malcolm Gladwell is that it takes an average of 10,000 hours of deliberate practice to master an activity. This has not been validated in knee arthroplasty and an argument can be made that there isn't an achievable "mastery" when it comes to surgery on humans, but the introduction of ML models in a quantified environment, driven by data sources including, but not limited to sensors, robotics, navigation and imaging, clearly has the potential to reduce the time required to become proficient at the complex task of knee arthroplasty. Even more interesting and likely is that "automation" in the form of robotics and ML instead allows the rapid advancement of the complexity of decision making and judgement made by the user with the hopeful goal of creating treatments that are safer, more personalized and more effective. Another useful aspect of ML models is the potential ability to codify the complex decision making process gained over years of experience and making it accessible in the learning environment for both training and assessment of orthopaedic trainees.[23-25]

Examining the various surgical discussions, it is clear that model performance is linked to the number of cases available in the training set. This underlines the importance of sound clinical data collection as the fundamental basis for developing ML models. Within the current study, acceptable model performance was already observed for the less prevalent surgical decisions, though it is clear that excellent to outstanding model performance was primarily achieved for the decisions that occurred at least 50 times in the training and tuning phase.

Another critically important aspect of model performance is the considered feature set. In our study, the relevance of adding intraoperative alignment information (the limb alignment in both planes at each step resulting from previous surgical decisions) to the feature set in addition to the intra-articular load measurements cannot be underestimated. The area under the curve increased significantly when adding information on the sagittal and coronal alignment. This mirrors clinical practice, where the intra-articular loads cannot be seen as the sole driver of surgical decisions when balancing a total knee. This is particularly pronounced for predicting the increase in poly thickness and stopping condition in our study. The former is explained as an increase in poly thickness is often considered when the leg hyper-extends, something not necessarily captured when looking at the loads at 10 (or 90) degrees of flexion. The latter relates to the fact that a perfectly balanced knee at 10° and 90° will still not be clinically acceptable if the knee hyper-extends or suffers from a residual flexion contracture. Meanwhile, it is interesting that the relevance of the pathological, pre-resection alignment condition on the surgical decisions taken is limited. The surgical corrections are primarily driven by the assessments during the trialing phase. This might be different when looking at the initial implant planning, which is likely more directly driven by the pre-resection alignment condition. It is therefore clear that alignment information – both prior to and following the initial bone resections – needs to be taken into account when striving for a balanced knee. Overall, these observations support the idea of quantifying various aspects of surgery to improve model performance and reliability of the predictions. This will thus present increasing opportunities as robotics and detailed 3D imaging become more widely adopted within the orthopaedic practice.[26]

It is worth noting that this study has a number of limitations. First, it is important to note that these models have been developed based on the expert opinion of a single surgeon, using a single implant design and alignment philosophy while also recognizing that soft tissue corrections can inherently be subjective. Deploying such models in the operating theatre of the novice user might require more diversified surgeon input to train the models in order to avoid bias, meanwhile providing the potential for peers to select the surgeon or surgeon group that most closely reflects their preferences, surgical technique or types of surgical corrections. Second, the use of even larger datasets could be recommended to improve model performance for those, more rarely observed clinical decisions while potentially also including other variables not considered in this study (e.g. component sizing). It is for that reason that, for instance, exact quantification of the additional varus alignment considered during a tibial recut or the type and magnitude for a femoral recut could not be discussed and analyzed in this paper. Meanwhile, the current work has proven that even for a relatively limited number of (sometimes generalized) observations, and with the challenges that come with using a real-world data set (such as missing values and double decisions for a given set of sensor readings), an acceptable model can already be built if the correct features are chosen and the models are selected and tuned appropriately. A third limitation is that these sensors only assess the tibiofemoral joint during the trialing phase. As such they overlook the relevance of the patellofemoral joint to achieve a balanced knee and shall be seen as a tool to correct for tibiofemoral imbalance encountered following the initial bone resections. It is for that reason that mid-flexion readings have not been discussed in this paper since these are little actionable

(e.g. when pointing to mid-flexion instability as a result of an excessive joint line shift).[27]

Conclusion

In conclusion, this paper presents a validated ML algorithm to guide the complex multi-dimensional classification problem encountered when balancing a total knee and selecting a surgical correction for a given imbalance scenario. The presented model has the potential to make experience available to the (new) adopters of technology, bringing expert opinion in their operating theatre, but also provides insight in the surgical decision process. As such, the present study demonstrates the relevance of including alignment information when making surgical decisions and balancing a total knee. Furthermore, this paper demonstrates the relevance of using the area under the receiver operator curve as a sensitive and reliable characteristic when evaluating model performance, particularly when the prevalence of a predicted event in the clinical dataset is limited.

Supplementary material

Description of machine learning models implemented and trained for each surgical decision: a random forest (RF), a linear support vector machine (SVM) and an artificial neural network (ANN).

References

1. **Kayani B**, **Konan S**, **Tahmassebi J**, **Pietrzak JRT**, **Haddad FS**. Robotic-arm assisted total knee arthroplasty is associated with improved early functional recovery and reduced time to hospital discharge compared with conventional jig-based total knee arthroplasty: a prospective cohort study. *Bone Joint J*. 2018;100-B(7):930–937.
2. **Chow JC**, **Breslauer L**. The use of intraoperative sensors significantly increases the patient-reported rate of improvement in primary total knee arthroplasty. *Orthopedics*. 2017;40(4):e648–e651.
3. **Golladay GJ**, **Bradbury TL**, **Gordon AC**, **et al**. Are patients more satisfied with a balanced total knee arthroplasty? *J Arthroplasty*. 2019;34(7S):S195–S200.
4. **Grau L**, **Lingamfelter M**, **Ponzio D**, **et al**. Robotic arm assisted total knee arthroplasty workflow optimization, operative times and learning curve. *Arthroplast Today*. 2019;5(4):465–470.
5. **Kayani B**, **Konan S**, **Huq SS**, **Tahmassebi J**, **Haddad FS**. Robotic-arm assisted total knee arthroplasty has a learning curve of seven cases for integration into the surgical workflow but no learning curve effect for accuracy of implant positioning. *Knee Surg Sports Traumatol Arthrosc*. 2019;27(4):1132–1141.
6. **Lakra A**, **Sarpong NO**, **Jennings EL**, **et al**. The learning curve by operative time for soft tissue balancing in total knee arthroplasty using electronic sensor technology. *J Arthroplasty*. 2019;34(3):483–487.
7. **Fontana MA**, **Lyman S**, **Sarker GK**, **Padgett DE**, **MacLean CH**. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res*. 2019;477(6):1267–1279.
8. **Babazadeh S**. The relevance of ligament balancing in total knee arthroplasty: how important is it? A systematic review of the literature. *Orthop Rev*. 2009;1:26.
9. **Walker LC**, **Clement ND**, **Ghosh KM**, **Deehan DJ**. What is a balanced knee replacement? *EFORT Open Rev*. 2018;3(12):614–619.
10. **Elmallah RK**, **Mistry JB**, **Cherian JJ**, **et al**. Can We Really "Feel" a Balanced Total Knee Arthroplasty? *J Arthroplasty*. 2016;31(9 Suppl):102–105.
11. **Gustke KA**, **Golladay GJ**, **Roche MW**, **Elson LC**, **Anderson CR**. A new method for defining balance. *J Arthroplasty*. 2014;29(5):955–960.
12. **Hasan M**, **Kotov A**, **Carcone A**, **et al**. A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories. *J Biomed Inform*. 2016;62:21–31.
13. **Sarker IH**, **Kayes ASM**, **Watters P**. Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *J Big Data*. 2019;6(1):57. https://doi.org/
14. **Bellemans J**. Multiple needle puncturing: balancing the varus knee. *Orthopedics*. 2011;34(9):e510-2.
15. **Dubois de Mont-Marin G**, **Babusiaux D**, **Brilhault J**. Medial collateral ligament lengthening by standardized pie-crusting technique: a cadaver study. *Orthop Traumatol Surg Res*. 2016;102(4 Suppl):S209–S212.
16. **Herschmiller T**, **Grosso MJ**, **Cunn GJ**, **et al**. Step-Wise medial collateral ligament needle puncturing in extension leads to a safe and predictable reduction in medial compartment pressure during TKA. *Knee Surg Sports Traumatol Arthrosc*. 2018;26(6):1759–1766.
17. **Géron A**. Hands-On machine learning with Scikit-Learn and TensorFlow: concepts, tools. *and Techniques to Build Intelligent Systems. n.d.*
18. **Uddin S**, **Khan A**, **Hossain ME**, **Moni MA**. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019;19(1):281. https://doi.org/
19. **Mandrekar JN**. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315–1316.
20. **Van Onsem S**, **Verstraete M**, **Dhont S**, **et al**. Improved walking distance and range of motion predict patient satisfaction after TKA. *Knee Surg Sports Traumatol Arthrosc*. 2018;26(11):3272–3279.
21. **Youngstrom EA**. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol*. 2014;39(2):204–221.
22. **Miura M**, **Hagiwara S**, **Nakamura J**, **et al**. Interobserver and intraobserver reliability of computed tomography-based three-dimensional preoperative planning for primary total knee arthroplasty. *J Arthroplasty*. 2018;33(5):1572–1578.
23. **Booth RE**, **Sharkey PF**, **Parvizi J**. Robotics in hip and knee arthroplasty: real innovation or marketing ruse. *J Arthroplasty*. 2019;34(10):2197–2198.
24. **Ericsson KA**, **Krampe RT**, **Tesch-Römer C**. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363–.
25. **Gladwell M**. Outliers: Little, Brown and Company, 2008.
26. **Parsley BS**. Robotics in Orthopedics: a brave new world. *J Arthroplasty*. 2018;33(8):2355–2357.
27. **Luyckx T**, **Vandenneucker H**, **Ing LS**, **et al**. Raising the joint line in TKA is associated with Mid-flexion laxity: a study in cadaver knees. *Clin Orthop Relat Res*. 2018;476(3):601–611.

Author information:

- M. A. Verstraete, PhD, Director of Clinical Research & Development, Clinical Research & Development, Dania Beach, USA; Dept of Human Structure and Repair, Ghent University, Gent, Belgium.
- R. E. Moore, MD, PhD, Orthopedic Surgeon, St Helena Hospital, Saint Helena, California, USA.
- M. Roche, PhD, Orthopaedic Surgeon, Orthopedics, Holy Cross Hospital, Fort Lauderdale, Florida, USA.
- M. A. Conditt, PhD, VP of Clinical Development, OrthoSensor Inc, Dania Beach, Florida, USA.

Author contributions:

- M. A. Verstraete: Concept and protocol development, Carried out data analysis, Drafted the manuscript.
- R. E. Moore: Carried out data collection, Protocol development, Drafted the manuscript.
- M. Roche: Carried out data collection, Protocol development.
- M. A. Conditt: Concept development, Drafted the manuscript.

Funding statement:

- The author or one or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article.

Acknowledgements:

- The authors would like to thank Andy Van Yperen-DeDeyne, PhD (ArcelorMittal, Belgium) for his valued and critical input as well as thriving enthusiasm on developing Machine Learning models. The authors would also like to thank Tyler Westfall, Chris Hunsaker, and Nate Bernstein for intra-operative data recording, and Jennifer DeBattista for the support in data management.

Ethical review statement

- This research was supported through a research grant by OrthoSensor Inc.