Predicting the polyethylene wear rate in pin-on-disc experiments in the context of prosthetic hip implants: Deriving a data-driven model using machine learning methods

A. Borjali, K. Monson, B. Raeymaekers

Department of Mechanical Engineering, University of Utah, Salt Lake City, UT, 84112, USA

ARTICLE INFO

Keywords:
- Prosthetic hip implant
- Polyethylene wear
- Pin-on-disc
- Machine learning

ABSTRACT

Pin-on-disc (PoD) experiments are widely used to quantify and rank wear of different material couples for prosthetic hip implant bearings. However, polyethylene wear results obtained from different PoD experiments are sometimes difficult to compare, which potentially leaves information inaccessible. We use machine learning methods to implement several data-driven models, and subsequently validate them by quantifying the prediction error with respect to published experimental data. A data-driven model can supplement results from PoD wear experiments, and enables predicting polyethylene wear of new PoD experiments based on its operating parameters. It also reveals the relative contribution of individual PoD operating parameters to the resulting polyethylene wear, thus informing design of experiments, and potentially reducing the need for time consuming PoD wear measurements.

1. Introduction

A prosthetic hip implant typically comprises a femoral component that articulates with an acetabular component to replace the natural hip function and alleviate pain and disability from degenerative joint diseases such as (osteo)arthritis [1]. Metal-on-polyethylene (MoP) is the most commonly used bearing material couple in state-of-the-art prosthetic hip implants used in the United States [2], typically pairing a CoCrMo femoral head with a polyethylene acetabular liner. Many studies document the effect of polyethylene wear on the longevity of MoP prosthetic hip implants (see e.g. Refs. [3–5]). Polyethylene wear debris may cause osteolysis (“weakening of the bone”) [4], which could potentially lead to implant loosening and mechanical instability [5]. Research to reduce polyethylene wear and increase longevity of MoP prosthetic hip implants involves changing the implant design and improving the mechanical properties of the polyethylene liner. For instance, highly cross-linked polyethylene (HXPE) and vitamin-E infused/blended cross-linked polyethylene (VEXPE) show significantly reduced wear compared to conventional ultra-high molecular weight polyethylene (UHMWPE) both in-vitro [6] and in-vivo [7]. On the other hand, using new materials for the femoral component, such as titanium [8], zirconia [9–11], silicon nitride [12], and tungsten [13], and manufacturing ultra-smooth ceramic bearing surfaces [14] or microtexturing the metal bearing surface [15–19] also reduces polyethylene wear.

Pin-on-disc (PoD) wear experiments are widely used as a screening method to quantify, compare, and rank wear of different implant bearing material couples as a function of operating parameters and environmental conditions. A PoD wear measurement in the context of MoP prosthetic hip implants typically consists of a polyethylene pin that is loaded against a metallic disc, while relative motion between the pin and the disc causes polyethylene wear. Many researchers have performed PoD wear experiments attempting to obtain clinically relevant polyethylene wear, using a variety of configurations. Fig. 1 shows eight different PoD wear measurement configurations documented in the literature and used in the context of prosthetic hip implants. $u_x$ and $u_y$ are the velocity magnitude in the $x$- and $y$-directions, respectively, and $\omega_z$ is the angular velocity about the $z$-direction.

Fig. 1 (a) shows a configuration in which the pin is stationary and loaded onto a disc that performs a reciprocating motion along the $x$-direction with velocity $u_x$ [20–30]. Conversely, Fig. 1 (b) depicts a pin that performs a reciprocating motion along the $x$-direction with velocity $u_x$ while loaded onto a stationary disc [31]. Fig. 1 (c) displays a stationary pin loaded onto a disc that rotates around its center axis with angular velocity $\omega_z$ [32–35], whereas Fig. 1 (d) depicts a pin loaded onto a stationary disc while it rotates around its center axis with angular velocity $\omega_z$ [16,36]. The PoD wear measurement configurations of Fig. 1 (a)-(d) all create unidirectional relative motion between the pin and the disc. While these configurations allow ranking wear of different...
bearing material couples, the resulting polyethylene wear rate is typically one to two orders of magnitude lower than the polyethylene wear measured in retrieved prosthetic hip implants of the same material, because long polyethylene molecules align in the sliding direction [37–39]. In contrast, multidirectional motion creates cross-shear, i.e., the relative motion between pin and disc changes direction with respect to the surface of the pin, thereby avoiding polymer molecule alignment and typically resulting in polyethylene wear that is similar in magnitude to what is observed in-vivo.

Fig. 1 (a)-(h) show PoD wear measurement configurations that allow creating multidirectional relative motion between the pin and the disc. Fig. 1 (e) depicts a stationary pin loaded onto a disc that reciprocates with velocities $u_x$ and $u_y$ in the $x$- and $y$-directions, respectively [6,37,40–62]. Furthermore, Fig. 1 (f) shows a pin reciprocating in the $x$-direction with velocity $u_x$ and loaded onto a disc reciprocating in the $x$-direction with velocity $u_e$ [63]. Fig. 1 (g) displays a pin that rotates around its center axis with angular velocity $\omega_x$ and is loaded onto a disc that reciprocates in the $x$-direction with velocity $u_x$ [64–67]. Finally, Fig. 1 (h) shows a pin that is loaded onto the disc and rotates around an axis parallel to the disc axis with eccentricity $e$ and with angular velocity $\omega_{z,1}$, while the disc rotates around its center axis with angular velocity $\omega_z$ [15,68–75].

PoD wear experiments typically require defining several operating parameters. For instance, different multidirectional wear paths have been documented in the literature, such as rectangular [15,40,53,57–60,68,71], elliptical [42,44,48,49,61], circular [6,43,45,47,51,52], square [56,69,70,72–75], and random [54,55]. The circular and elliptical wear paths create cross-shear on the surface of the pin throughout the entire wear path, whereas the rectangular and square wear paths create cross-shear when the pin changes direction along the wear path. Furthermore, several researchers report a strong correlation between polyethylene wear and both contact area and contact pressure between the pin and the disc (or the normal load applied to the pin) [51,55,69]. Polyethylene wear is also dependent on the surface roughness of the disc surface [26,43,76,77] and on the lubricant used during the PoD wear experiment. Although bovine serum is typically used as lubricant for PoD wear experiments in the context of prosthetic hip implants [78], the optimal composition of bovine serum remains subject to debate. Studies have reported that polyethylene wear is a function of bovine serum protein concentration [52,79], protein type [80], lipid concentration [45], dilution method [75], and anti-bacterial and fungal additives [73,81]. Another important factor reported in the literature is the radiation dose of HXPE [82]; increasing the radiation dose increases polyethylene cross-linking, which in turn increases its wear resistance. However, some reports also document decreasing fracture resistance with increasing radiation dose [83]. Radiation may also leave residual free radicals in the polyethylene that can cause oxidation over time [84]. Re-melting [85] or adding free radical scavenger agents such as vitamin-E to the HXPE [41,86,87] can reduce the risk of oxidation.

A large number of polyethylene wear datasets obtained using PoD wear experiments, conducted in the context of prosthetic hip implants, exists in the literature. These experiments are performed by different research groups, using different devices, configurations, and operating conditions. Thus, results of different PoD wear experiments are sometimes difficult to compare, which potentially leaves valuable inaccessible. Also, several limitations exist to conducting PoD wear experiments. The viscoelastic nature of polyethylene necessitates performing PoD wear experiments at a strain rate that is identical to what occurs in-vivo [88], resulting in a kinematic cycle of 1–2 Hz to mimic the human gait cycle frequency [89]. Many million kinematic cycles are needed to obtain measurable wear of the polyethylene bearing surface, which is time consuming. In addition, manufacturing pin and disc specimens to specific standards [90–92], and performing gravimetric polyethylene wear measurements also requires trained personnel [78,93].

However, in recent years, materials researchers (among others) have used machine learning methods in combination with existing datasets, to facilitate modeling complex relationships between material constituents, structure, and the corresponding mechanical properties [94]. Such data-driven models enable comparing existing datasets and predicting new results based on the existing knowledge embedded in the model, which are otherwise difficult or time consuming to obtain using traditional experimental methods [95].
Hence, the objective of this work is to aggregate published PoD polyethylene wear datasets specifically performed in the context of prosthetic hip implants, and use machine learning methods to implement a data-driven model that allows predicting the polyethylene wear rate of PoD wear experiments based on its operating parameters. Such a model potentially supplements PoD wear experiments and may reveal hidden relationships between polyethylene wear and PoD operating parameters. Furthermore, the model assists researchers with design of experiments (DoE) by identifying and ranking the operating parameters that most significantly affect polyethylene wear in PoD wear experiments. This allows prioritizing operating parameters considered in future PoD wear experiments, and ultimately reducing the number of experiments one must conduct. Finally, a data-driven model based on the published literature also facilitates validating new experiments and detecting outlier results.

2. Methods

2.1. Data acquisition

We perform a literature survey to collect published polyethylene wear data from PoD wear experiments, conducted by others in the context of prosthetic hip implants. The entire dataset is available in the Supplementary Material. We search the Google Scholar and PubMed databases, using keywords including “UHMWPE”, “wear”, “hip” and “pin-on-disc/disk”. We restrict our search to these keywords because polyethylene wear is dependent on the operating conditions of the PoD wear experiments, which may differ significantly depending on the application for which they are intended; e.g. operating conditions for knee and hip PoD wear experiments could be significantly different [96]. We only consider studies that imposed multidirectional motion between a polyethylene pin and a CoCr disc, with flat-on-flat geometry to control for the effect of specimen geometry, and with bovine serum as lubricant. Furthermore, we only retain studies with clearly defined and reported operating parameters (which we refer to as input attributes) and polyethylene wear rate results (which we refer to as the target attribute), and we eliminate studies where this information is either ambiguous, such as, a random wear path, or is not fully reported (more than two attributes with missing values). All data in this work is based on gravimetric polyethylene wear measurements only, which avoids inaccuracies due to plastic deformation, creep, and fluid absorption (when a soak control specimen is used) [97]. We quantify polyethylene wear using the “wear rate [mg/MC]”, which is defined as the material loss per million cycles (MC), and prescribed in the ASTM F732 standard. Some studies report the wear factor instead, which is the wear volume per unit of normal load and sliding distance [39]. Since the wear factor implicitly assumes that wear is independent of contact area [98], which may contradict in-vitro and in-vivo polyethylene wear observations in prosthetic hip implants [39], we convert the wear factor to wear rate by multiplying it by the sliding distance, normal load, and polyethylene density. We use the polyethylene density reported in each study; in cases of missing polyethylene density, we use the UHMWPE density reported in the literature (0.93 mg/mm³ [43]). For dynamic normal loading used in some PoD wear experiments, we report the maximum value. We average a parameter's value if it is reported as a range in any of the studies we consider.

Because their effect on the polyethylene wear rate is well-documented in the literature, we include the following PoD wear experiment operating parameters (input attributes) in the data-driven model that describes and predicts the polyethylene wear rate (target attribute) in PoD wear experiments: normal load [N], contact area [mm²], frequency [Hz], sliding distance per cycle [mm/C], wear path aspect ratio, lubricant temperature [°C], lubricant protein concentration [mg/ml], disc average surface roughness (Rₐ) [µm], polyethylene radiation dose [kGy], and test duration [MC].

2.2. Descriptive statistics

We quantify the linear (Pearson’s) correlation coefficient between each input attribute and the target attribute, normalized by the maximum correlation coefficient computed between any of the input attributes and the target attribute, to determine the relative contribution of each input attribute to the target attribute. We also determine the minimum, maximum, average, standard deviation, and stability (S) of each input attribute to characterize the dataset. The stability S is the ratio of the number of occurrences of the most frequent value of a dataset and the total number of values in that dataset, which indicates how constant an input attribute is. An input attribute with high S is almost constant and likely does not capture the entire range of that attribute’s possible values. Thus, the data-driven model may underestimate the effect of that input attribute on the target attribute.

2.3. Numerical experiment

We conduct two sets of numerical experiments. First, we apply machine learning methods to the entire polyethylene wear rate dataset and the corresponding PoD wear experiment operating parameters, to determine the method that represents the entire dataset with the highest prediction accuracy. Second, we divide the polyethylene wear rate dataset into subgroups based on the polyethylene radiation dose, because it is well-known that polyethylene radiation dose affects polyethylene wear and, thus, we expect these subgroups to have similar wear rates. The three subgroups are: (1) non-irradiated, 0 kGy radiation dose, (2) conventional with radiation dose between 20 and 55 kGy, and (3) HXPE with radiation dose in excess of 70 kGy. We then apply machine learning methods to each subgroup and compare the prediction accuracy of each method to the one obtained without subgroups.

2.4. Machine learning methods

We use three types of machine learning methods, which we briefly describe in this section, and we cite references that contain details of each method, as these are not the focus of this paper. First, we employ interpretable model-based methods, in which the relationship between input and target attributes is explicitly defined, including linear regression [99], CART [100], M5 [101], random forest [102] and gradient boosting [103]. These methods train a data-driven model on the polyethylene wear rate dataset to predict the polyethylene wear rate of PoD wear experiments based on the operating parameters (input attributes, see last paragraph of Section 2.1). We use linear regression based on the least-squares method to fit weighting factors to each operating parameter and quantify its contribution to the resulting polyethylene wear rate. This allows understanding whether the relationship between the operating parameters and the polyethylene wear rate can be captured by a single linear model. CART builds a decision tree model, where nodes represent decision points, and where each branch of the tree is a separate linear regression model, i.e., different parts of the data are modeled by distinct regression models. Thus, we use this method to investigate the effect of modeling different segments of the polyethylene wear rate dataset with various linear regression models, optimizing the CART tree by tuning the depth of the tree and the number of leaves per node, and pruning. M5 is similar to CART but minimizes the sum rather than the mean of the error of all linear regression models that constitute the CART decision tree. We use the random forest method to investigate whether combining multiple CART trees reduces the prediction error of the data-driven model compared to using a single CART tree. The random forest method predicts the polyethylene wear rate by averaging the predicted polyethylene wear rate of multiple CART trees. In addition to the CART parameters, we also tune the number of trees in the random forest to minimize the prediction error. Gradient boosting also allows combining several CART trees into one data-driven model. While in the random forest method each tree is
added to the forest independent of the other trees, gradient boosting adds each new tree specifically to improve the performance of weak trees.

Second, we use non-interpretable model-based methods including artificial neural network (ANN) [104] and support vector machine (SVM) [105]. These methods train a model on the dataset, without explicitly defining the relationship between input and target attributes, but creating a black-box model instead. ANN relates the operating parameters to the polyethylene wear rate by means of neurons that communicate with each other in a non-linear fashion, trained by the polyethylene wear rate dataset. We implement ANN by tuning the number of hidden layers, number of nodes per layer, and the learning rate of the neural network as commonly implemented in the machine learning literature [106]. SVM fits a hyperplane through the polyethylene wear rate dataset by minimizing the error between the predicted and actual wear rate.

Third, we implement instance-based methods such as the k-nearest neighbor (KNN) method [107], which predicts the polyethylene wear rate based on the most similar instances in the dataset. Such methods handle complicated datasets that cannot be captured by a single model. Specifically, the KNN method compares unseen data against all other instances in the dataset to find its k nearest neighbors, i.e., its k most similar data points. Then, the unseen data is assigned a value based on the weighted average value of its neighbors. We tune the number of nearest neighbors k and the weighting function in our model.

We employ tenfold cross-validation to evaluate the prediction error of the data-driven models of the polyethylene wear rate that we implement using different machine learning methods. Cross-validation involves randomly dividing the dataset into m equal subsets (so-called folds). Then, we use m-1 subsets to train the data-driven model and we use one remaining subset to validate the model. We repeat this process m times such that we validate the model on each subset exactly once. Finally, we obtain a single prediction error for each model by averaging the results of the m iterations. It is important to note that the model validation process is always performed on the one subset that has not been used to train the data-driven model, i.e., it is validated using data not used to train the model. Thus, the advantage of cross-validation compared to other validation methods, such as e.g. the hold out method, is that every data point is in the validation dataset exactly once, and is in the training dataset m-1 times [108]. In contrast, using the hold-out method for validation requires partitioning the data in a training and validation set and, thus, the validation may be significantly different depending on how the partitioning is performed.

Commonly used metrics to evaluate the prediction error of a data-driven model include the mean absolute error (MAE), root mean square error (RMSE), and the square of the correlation coefficient (R^2) [94]. A combination of these three metrics yields a good indication of the accuracy of the data-driven model [109]. The MAE is given as

\[
\text{MAE} = \frac{\sum_{i=1}^{n} |a_i - p_i|}{n},
\]

where \(a_i\) and \(p_i\) are the actual and predicted values of the ith data point, respectively, and \(n\) is the total number of data points in the dataset. RMSE is computed as

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (a_i - p_i)^2}{n}},
\]

and, \(R^2\) is calculated as

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (a_i - p_i)^2}{\sum_{i=1}^{n} (a_i - \bar{a})^2},
\]

3. Results and discussion

Table 1 shows the descriptive statistics of the entire polyethylene wear rate dataset consisting of 129 data points from 29 different studies that we retain from published literature based on the criteria we define in Section 2.1, and that we use to implement the different machine learning methods of Section 2.4. The literature shows that the size of a dataset to predict material properties is typically small compared to other research fields [110]. Other studies use datasets ranging from just a few data points (14 data points [111]) to tens and hundreds of data points (82 data points [112], 121 data points [113], 157 data points [114], and 218 data points [115]).

We report minimum and maximum values in Table 1 with the same number of significant digits as in their respective publications. Note that the wear path shape reports the minimum and maximum number of occurrences, i.e., one experiment used a 10 × 20 mm rectangular wear path, whereas 39 experiments used a \(d = 10\) mm circular wear path. We exclude the input attributes “lubricant protein concentration” and “lubricant temperature” from our analysis due to a high number of missing values, i.e., they are often not reported in their respective publications. We replace the four missing average disc surface roughness \(R_a\) value with the average value of the dataset (0.05 μm), which is a common practice in the machine learning literature [116].

Fig. 2 shows the normalized linear correlation coefficient (see Section 2.2) between each operating parameter (input attribute) and the polyethylene wear rate (target attribute), quantifying the relative contribution of each input attribute to the target attribute. Different published studies evaluate the relative contribution of one or two operating parameters to the polyethylene wear rate. In contrast, the results of

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive statistics of the polyethylene wear rate dataset, showing the same number of significant digits as in their respective publications.</strong></td>
</tr>
<tr>
<td><strong>Minimum</strong></td>
</tr>
<tr>
<td>Publication year</td>
</tr>
<tr>
<td>Normal load [N]</td>
</tr>
<tr>
<td>Contact area [mm²]</td>
</tr>
<tr>
<td>Frequency [Hz]</td>
</tr>
<tr>
<td>Sliding distance per cycle [mm/C]</td>
</tr>
<tr>
<td>Wear path shape</td>
</tr>
<tr>
<td>Wear path aspect ratio</td>
</tr>
<tr>
<td>Lubricant temp. [°C]</td>
</tr>
<tr>
<td>Lubricant protein concentration [mg/ml]</td>
</tr>
<tr>
<td>Average disc surface roughness (R_a) [μm]</td>
</tr>
<tr>
<td>Polyethylene radiation dose [kGy]</td>
</tr>
<tr>
<td>Test duration [MC]</td>
</tr>
<tr>
<td>Polyethylene wear rate [mg/MC]</td>
</tr>
</tbody>
</table>
Fig. 2 leverage all PoD polyethylene wear rate data published in the literature (and retained for this work) and quantify the relative contribution of all operating parameters described in this work (see Section 2.1) to the resulting polyethylene wear rate. We observe that the contact area between the pin and the disc is the most important factor that affects the polyethylene wear rate, as expected, and in agreement with clinical results. For instance, MoP prosthetic hip implants show increasing wear rate with increasing femoral head size (i.e., increasing contact area between femoral head and polyethylene liner) [39]. The polyethylene radiation dose, which indicates the level of cross-linking, is the second most important input attribute, followed by the normal load, average surface roughness $R_a$, wear path aspect ratio (i.e., cross-shear), test duration, and sliding distance per cycle. Ranking of the relative contribution of operating parameters to the polyethylene wear rate provides guidance to designing and conducting future PoD wear experiments. Indeed, operating parameters should be included in PoD wear experiments prioritized according to this ranking, as the effect of lower ranked operating parameters on polyethylene wear could be within the noise level of the higher ranked ones. We note that although the normalized linear correlation between the frequency and the polyethylene wear rate is almost zero, its high stability value ($S = 44.9\%$) indicates that the dataset only spans a small range (1.8 Hz) and, thus, the effect of frequency on the polyethylene wear rate might not be evident from the aggregate dataset, because most researchers recognize that polyethylene is viscoelastic, and therefore performed the PoD wear experiments at a frequency that is similar to the human gait frequency.

Table 2 shows the prediction error ($MAE$, $RMSE$, and $R^2$) of the different machine learning methods we implement in this work, based on the entire polyethylene wear rate dataset. From Table 2 we observe that KNN yields the smallest prediction error ($MAE = 1.38$ mg/MC) of the polyethylene wear rate compared to all other methods, after tenfold cross-validation. Thus, the KNN model predicts the polyethylene wear rate within 1.38 mg/MC for any new PoD experiment with input attributes that fall within the range of those of the dataset used to develop the model. Furthermore, these results show that an instance-based method (KNN) outperforms both interpretable and non-interpretable model-based methods, which indicates that the relationship between the operating parameters and the polyethylene wear rate is not easily captured by one single model.

Fig. 3 (a) shows the experimental polyethylene PoD wear rate (red square markers) of all published studies included in our dataset, ranked by descending wear rate, and the corresponding wear rate predicted using the cross-validated data-driven model based on the KNN method (blue circle markers). In addition, Fig. 3 (b) shows the prediction error between the KNN data-driven model and the published wear data, defined for each individual study as the absolute value of the difference between the experimental and the predicted result, divided by the experimental result (black triangle markers). From Fig. 3 (a) we observe that when the wear rate exceeds 15 mg/MC (indicated as region (a) in Fig. 3 (a)), the predicted polyethylene wear rate deviates from the corresponding experimental results by 5–47%. This is due to the lack of data to train the model in this region, as only eight out of 129 experiments report a polyethylene wear rate in excess of 15 mg/MC. Since KNN is an instance-based method, it requires more “instances” to train itself and lower the prediction error for polyethylene wear rate higher than 15 mg/MC.

Furthermore, Fig. 3 (a) highlights several studies using labels (b) to (l), for which the data-driven model results in a prediction error that exceeds 4%. The high prediction error for these specific studies is because they display a unique feature that differs significantly from the rest of the dataset, which cannot be fully captured by the data-driven model. Studies (b) and (d) are performed by the same research group [69] and are the only experiments in the dataset that change the composition of the lubricant while the experiment was ongoing; specifically, the lubricant composition changes after 0.5 MC and continues with a different composition for 1 MC. Study (c) intentionally uses a significantly higher lubricant protein concentration (the maximum lubricant protein concentration in the dataset of 64.8 mg/ml) than what...
other wear experiments typically use (approximately 20–30 mg/ml [52]) to generate a higher polyethylene wear rate [15], which causes the data-driven model to underestimate the polyethylene wear rate for this specific study. On the other hand, study (g) uses a significantly lower lubricant protein concentration (the minimum protein concentration in the dataset of 0.69 mg/ml) than what others commonly use [56], which results in the data-driven model overestimating the polyethylene wear rate for that study. Studies (e) [40] and (k) [41] used polyethylene infused with vitamin-E antioxidant, which is not an input attribute to the data-driven model because few published studies document the PoD polyethylene wear rate of VEXPE. Thus, the model is not trained to account for these parameters. Study (f) [45] is the only study that uses a frequency (0.2 Hz) outside of the ASTM F732 standard recommended frequency range (0.5–2.0 Hz [78]). All the other studies employ frequencies between 1 and 2 Hz. Study (h) [57] performs heat treatment on the polyethylene after cross-linking, which we do not specifically consider in the data-driven model as an input attribute. Furthermore, the polyethylene of study (h) is the only polyethylene that is manufactured by the same research group who evaluates polyethylene wear of several commercially available cross-linked polyethylene materials. Hence, the prediction error of the data-driven model also potentially identifies manufacturing defects. Study (i) and (j) are performed by the same research group [68] who use a rectangular wear path for PoD wear experiments, with the highest (1 × 9 mm) and the second highest (2 × 8 mm) aspect ratio in the dataset, indicating one direction is dominant. A unidirectional wear path is well-known to generate a significantly lower polyethylene wear rate compared to a multidirectional wear path because of cross-shear. Since the data-driven model in this work is trained on the polyethylene wear rate dataset obtained with multidirectional relative motion between the pin and the disc, it significantly overestimates the polyethylene wear rate in these two studies. Finally, study (l) uses the smallest contact area and normal load in the dataset [47], which are well below the ASTM F732 recommended values (contact area 63.6 mm² and normal load corresponding to 2–10 MPa contact pressure [78]).

Table 2 also shows that CART is the interpretable model-based method with the lowest mean absolute error (1.95 mg/MC). Although the CART method shows a higher prediction error than the KNN method for the dataset of this work, it allows creating an interpretable model of the dataset. Fig. 4 illustrates the first seven nodes of the CART model of the polyethylene wear rate dataset. Each node shows an operating parameter and reflects a decision whether that operating

![Fig. 3. Polyethylene wear rate for all experimental studies considered in our dataset ranked in descending order, (a) showing the experimental results documented in the literature (red square markers) and the corresponding predicted results using the data-driven model based on the KNN method (blue circle markers), and (b) showing the prediction error for the corresponding experimental study (black triangle markers). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)](image-url)
parameter is smaller than or equal to \( \leq \) (i.e., left branch) or larger than \( > \) (i.e., right branch) the specified value at the node. From Fig. 4 we observe that “contact area” is the highest-level attribute in the CART decision tree for predicting the polyethylene wear rate, i.e., CART considers it to have the most significant effect on the polyethylene wear rate of all operating parameters included in the model. Furthermore, we observe that for a contact area larger than 113.05 mm\(^2\) the polyethylene wear rate is dominated by the normal load acting on the polyethylene pin, whereas for a contact area smaller than 113.05 mm\(^2\) the polyethylene radiation dose is the deciding attribute. Similarly, one can interpret each level of the CART decision tree to obtain an understanding of how the data-driven model interprets the data and predicts results.

We also create subgroups of polyethylene wear rate data based on the polyethylene radiation dose, and implement the machine learning methods for each of these subgroups. Table 3 shows the prediction error (MAE, RMSE, \(R^2\)) of the different machine learning methods we implement in this study, based on each polyethylene wear rate subgroup, as defined in Section 2.3. From Table 3 we observe that clustering the data into subgroups based on the polyethylene radiation dose reduces the prediction MAE for the wear rate of conventional and HXPE by 10% and 64% respectively, whereas it increases the prediction MAE for non-irradiated polyethylene by 57%, compared to the data-driven model without clustering the data into subgroups (Table 2). The inter-quartile range of the polyethylene wear rate (the difference between the third and first quartiles of each subgroups, \(Q_3-Q_1\)) is 5.70 for non-irradiated polyethylene, whereas it is 4.64 for conventional polyethylene and 1.48 for HXPE. This indicates that the machine learning methods perform

Table 3
Machine learning methods prediction error based on the polyethylene wear rate data for three subgroups.

<table>
<thead>
<tr>
<th>Polyethylene radiation dose [kGy]</th>
<th>0 (Non-irradiated)</th>
<th>20-55 (Conventional)</th>
<th>&gt; 70 (HXPE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>MAE, RMSE, (R^2)</td>
<td>MAE, RMSE, (R^2)</td>
<td>MAE, RMSE, (R^2)</td>
</tr>
<tr>
<td>Model-based (interpretable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Regression</td>
<td>3.54, 4.84, 0.81</td>
<td>4.51, 6.19, 0.62</td>
<td>0.67, 0.85, 0.78</td>
</tr>
<tr>
<td>CART</td>
<td>2.67, 3.43, 0.77</td>
<td>2.39, 3.62, 0.81</td>
<td>0.70, 0.99, 0.77</td>
</tr>
<tr>
<td>M5</td>
<td>3.29, 4.96, 0.77</td>
<td>4.32, 6.55, 0.65</td>
<td>0.68, 0.90, 0.72</td>
</tr>
<tr>
<td>Random Forest</td>
<td>2.76, 3.54, 0.09</td>
<td>2.58, 3.93, 0.77</td>
<td>0.70, 0.92, 0.82</td>
</tr>
<tr>
<td>Gradient boosting</td>
<td>3.16, 4.45, 0.68</td>
<td>2.84, 3.91, 0.68</td>
<td>0.81, 1.03, 0.79</td>
</tr>
<tr>
<td>Model-based (non-interpretable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>3.35, 4.22, 0.77</td>
<td>2.77, 4.46, 0.79</td>
<td>0.88, 1.15, 0.76</td>
</tr>
<tr>
<td>SVM</td>
<td>3.42, 4.54, 0.79</td>
<td>4.07, 6.08, 0.43</td>
<td>0.57, 0.80, 0.82</td>
</tr>
<tr>
<td>Instance-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>2.16, 3.08, 0.87</td>
<td>1.24, 1.97, 0.81</td>
<td>0.50, 0.75, 0.88</td>
</tr>
</tbody>
</table>
better on the subgroups with less polyethylene wear rate variability.

The primary limitation of this study is the size of the polyethylene wear rate dataset, which is inherently limited by what is available in the published literature. Ultimately, increasing the size of the dataset, as more studies are published, could increase the prediction accuracy and reduce the sensitivity of the model to noise and outlier data, such as unique features or operating parameters of specific experiments. More data would also allow considering additional input attributes in the data-driven polyethylene wear model, such as lubricant protein concentration, lubricant anti-bacterial and fungal additives, and lubricant temperature, which are parameters known to have an effect on polyethylene wear. The size of the dataset also directly affects the prediction error of the data-driven model. Furthermore, the KNN method, which results in the best prediction accuracy in our work, is limited to predict the wear rate for input attributes that fall within the range of the input attributes considered in the dataset. The model based on KNN predicts results based on the weighted average value of its k nearest neighbors, and it can only predict accurately when close neighbors exist in the dataset to that new unseen data. Considering additional input attributes, such as lubricant protein concentration, could change the structure of the dataset and, thus, the distance between the data points and nearest neighbors, which could change the KNN prediction accuracy. Another limitation of this study is that we do not distinguish between static and dynamic loading during the PoD experiments. Instead, we use the maximum load value in cases of dynamic loading.

4. Conclusion

We have aggregated a dataset of published PoD polyethylene wear rate data, performed in the context of prosthetic hip implants. Using several model-based and instance-based machine learning methods both with and without clustering of the data, we have implemented a data-driven model that allows predicting the PoD polyethylene wear rate based on its operating parameters.

We find that the KNN method with clustering into subgroups based on polyethylene radiation dose results in the lowest prediction error, i.e., this instance-based method outperforms interpretable and non-interpretable model-based methods, because the PoD polyethylene wear rate dataset cannot easily be captured by a single model.

The data-driven model reveals the relative contribution of PoD wear experiment operating parameters (input attributes) to the polyethylene wear rate (target attribute). This provides guidance for designing future PoD wear experiments. Operating parameters should be included in these experiments prioritized according to their relative contribution to the polyethylene wear rate, as the effect of lower ranked operating parameters on polyethylene wear could be within the noise level of the higher ranked ones.

Using cross-validation of the data-driven model we predict the polyethylene wear rate of all the experimental studies in our dataset. The data-driven model predicts results based on the subset of the dataset that is not used to train the model, at each iteration of the tenfold cross-validation process. This demonstrates that the data-driven model predicts the polyethylene wear rate for new PoD experiments with operating parameters that fall within the ranges of those of the dataset used to implement the model. This could potentially reduce the need for more experimental studies or shed light on experiment design. Furthermore, this data-driven model facilitates validating new experimental results and detecting outliers, by comparing them to results in the literature.

Acknowledgments

A.B. and B.R. acknowledge support from the National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases under grant 1R03AR066826-01A1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.triboint.2019.01.014.

References

A. Borjali et al.

Flotation Int (2013) 101–110


[56] Guenther TE, Turgeon TR, Bohm ER, Brandt J-MM. The biochemical