



OPEN

Prediction of coating thickness for polyelectrolyte multilayers via machine learning

Varvara Gribova^{1,2,5}, Anastasiia Navalikhina^{4,5}, Oleksandr Lysenko⁴, Cynthia Calligaro³, Eloïse Lebaudy^{1,2}, Lucie Deiber³, Bernard Senger^{1,2}, Philippe Laval^{1,2,3} & Nihal Engin Vrana³✉

Layer-by-layer (LbL) deposition method of polyelectrolytes is a versatile way of developing functional nanoscale coatings. Even though the mechanisms of LbL film development are well-established, currently there are no predictive models that can link film components with their final properties. The current health crisis has shown the importance of accelerated development of biomedical solutions such as antiviral coatings, and the implementation of machine learning methodologies for coating development can enable achieving this. In this work, using literature data and newly generated experimental results, we first analyzed the relative impact of 23 coating parameters on the coating thickness. Next, a predictive model has been developed using aforementioned parameters and molecular descriptors of polymers from the DeepChem library. Model performance was limited because of insufficient number of data points in the training set, due to the scarce availability of data in the literature. Despite this limitation, we demonstrate, for the first time, utilization of machine learning for prediction of LbL coating properties. It can decrease the time necessary to obtain functional coating with desired properties, as well as decrease experimental costs and enable the fast first response to crisis situations (such as pandemics) where coatings can positively contribute. Besides coating thickness, which was selected as an output value in this study, machine learning approach can be potentially used to predict functional properties of multilayer coatings, e.g. biocompatibility, cell adhesive, antibacterial, antiviral or anti-inflammatory properties.

Layer-by-layer (LbL) coating is a method for surface modification based on the electrostatic interactions between two polyelectrolytes^{1,2}. Such coating is developed thanks to successive deposition of polycations and polyanions onto the surface of a material, and by performing a rinsing step after each deposition. This method is very versatile as a large number of polyelectrolytes can be used, making it possible to adapt the coating for a particular application. Different methods can be used for the build-up of LbL coatings, such as dip-coating, spin-coating, and spraying^{3,4}. The most used method and perhaps the easiest one is dip-coating, but it is also more time-consuming compared to spin-coating for instance⁵.

LbL coatings are used for multiple biomedical applications, in particular, because natural polyelectrolytes presenting good biocompatibility can be used for LbL film build-up. It is possible to develop antibacterial surfaces, smart healing materials, and coatings for tissue engineering. Moreover, LbL coatings can be used for loading drugs or other bioactive molecules, which allows their local delivery⁶⁻⁹. Non-biomedical LbL applications include construction of gas barrier films¹⁰, optical fiber sensing¹¹, and many electrochemical systems¹².

However, the empirical manner of polycation/polyanion selection is an impediment on rapid new coating development. First, the formation of the coatings can be very long, if many layers are required, and for thick films, the method can become fastidious. Secondly, the thickness of the different coatings is difficult to control, as it depends on different parameters such as temperature, pH, ionic strength, and others⁵.

Moreover, there remain difficulties in understanding how interactions between polymers occur, as they are mostly multifactorial. Thus, LbL coatings growth can be different (in most cases linear or exponential, at least up to a given number of layers deposited) depending on polymers' properties and on diffusion between layers¹³. Experimentally, different methods are used to evaluate LbL film thickness: quartz crystal microbalance with

¹Inserm UMR_S 1121, Biomaterials and Bioengineering, Centre de Recherche en Biomédecine de Strasbourg, 67000 Strasbourg, France. ²Université de Strasbourg, Faculté de Chirurgie Dentaire, 67000 Strasbourg, France. ³SPARTHA Medical, 67100 Strasbourg, France. ⁴PRESTE, 75011 Paris, France. ⁵These authors contributed equally: Varvara Gribova and Anastasiia Navalikhina. ✉email: evrana@sparthamedical.eu

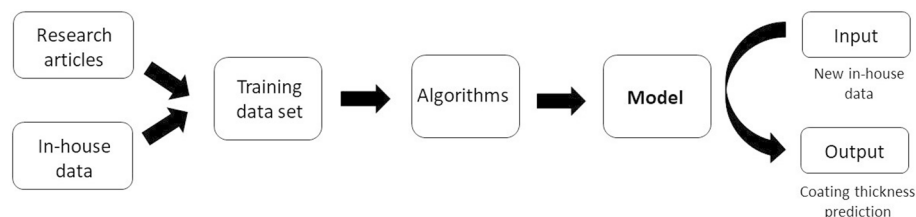


Figure 1. Workflow for coating thickness prediction using supervised machine learning approach.

dissipation monitoring (QCM-D) can be used to follow step-by-step polyelectrolyte deposition with high accuracy. Other methods such as atomic force microscopy (AFM), confocal microscopy or ellipsometry can also be used^{14,15}. However, these different methods do not use the same approach of thickness determination, and for the same LbL film, the results obtained by different techniques can differ.

As a result, and despite the progress made in the field, the data accumulated over the years do not provide predictive capacities on how a given couple of polymers will form an LbL film, which also decreases the rate of advance in the field. In this work, we hypothesize that using the current state-of-the-art data science techniques, we can determine how different parameters affect coating thickness and predict the thickness of the new coatings. To do so, we used historical and generated data for predictive model development using machine learning.

Machine learning is an approach which uses algorithms that improve upon training on large datasets and is able to find complex patterns, make predictions and decisions. In this work, we used two training datasets: one comprising the data extracted from literature (several thickness determination methods) and another containing experimental data produced in our laboratory (thickness determination by QCM-D).

In the first part of the work, the most important parameters influencing coating buildup were determined. In the second part, a thickness predictive model was built using the training set, and its performance was evaluated. Finally, we validated the model on LbLs which were not in the training set, and were able to predict the coating thickness (Fig. 1). To our knowledge, this is the first time that machine learning approach has been used for LbL coating thickness prediction.

Results and discussion

Data collection. The first step of the work consisted in data collection from the literature, which represents the first dataset. It should be stated that currently in the literature no established database is available related to LbL coating thickness, and the available experimental data is relatively scarce compared to the number of LbL related articles. The second dataset was based on the QCM-D experiments done in the laboratory. For the data extracted from the literature, different ways of thickness calculation/estimation were used, such as AFM, ellipsometry, confocal microscopy. Coating thickness determined by these methods may differ, but due to the limited amount of the available data, we selected to include all the data regardless of the thickness measurement method. All the results extracted from the literature, as well as obtained in the laboratory, were entered in the tables describing different parameters (Table S1). Of note, all the multilayers used in the study were prepared by dip-coating or a similar technique (simple polymer solution deposition on the substrate followed by adsorption time and rinsing). Thus, the coating preparation method was not among the parameters influencing thickness.

Influence of different construction parameters on coating thickness. As a first step, distribution of the thickness values in the coatings made of different polymers was studied (Fig. 2). The results show that for polycations, the coatings made of poly(L-lysine) (PLL) have the greatest thickness median values with large interquartile range (IQR), which overlaps the thickness distribution of chitosan (CHI)-made coatings (Fig. 2A). In Fig. 2B, coatings having poly(L-glutamic acid) (PGA) have the greatest thickness median value with IQR overlapping with hyaluronic acid (HA)-containing coatings thickness distribution. The large thickness distribution of the LbL films containing the aforementioned polymers is probably due to the high frequency of their utilization and therefore to the wide variety of molecular weights (MW) that have been used.

Next, linear relationships between the coating thickness and different parameters were evaluated using the Pearson correlation, and non-linear relationships were analyzed with the predictive power score (PPS) method (Fig. 3). The PPS allows seeing non-symmetrical relationships between variables: features located on the x-axis are independent variables (influencers), and features on the y-axis are dependent variables (influenced by x-features).

The results show that concentrations of polymers and the number of bilayers in the coating have a strong positive linear relationship with the resulting thickness (Fig. 3A), while polyanion molecular weight and buffer properties have some non-linear relationships with this property (Fig. 3B). The effect of concentration and the number of bilayers are intuitive, however, their relative importance in an overall tendency sense cannot be extracted from a single or few types of LbL films, which is the currently common practice. Similarly, the discrepancy between the relative importance of polyanion (PA) MW and polycation (PC) MW is also not evident.

One of the most important features which have strong linear relationships with the coating thickness is the number of bilayers in the coating (Fig. 3).

As stated above, this dependency is largely obvious, and therefore, the data had to be unified by this number. For this, we decided to calculate the thickness at eight bilayers for each coating in the literature-generated data using the growth function (Eq. 1).

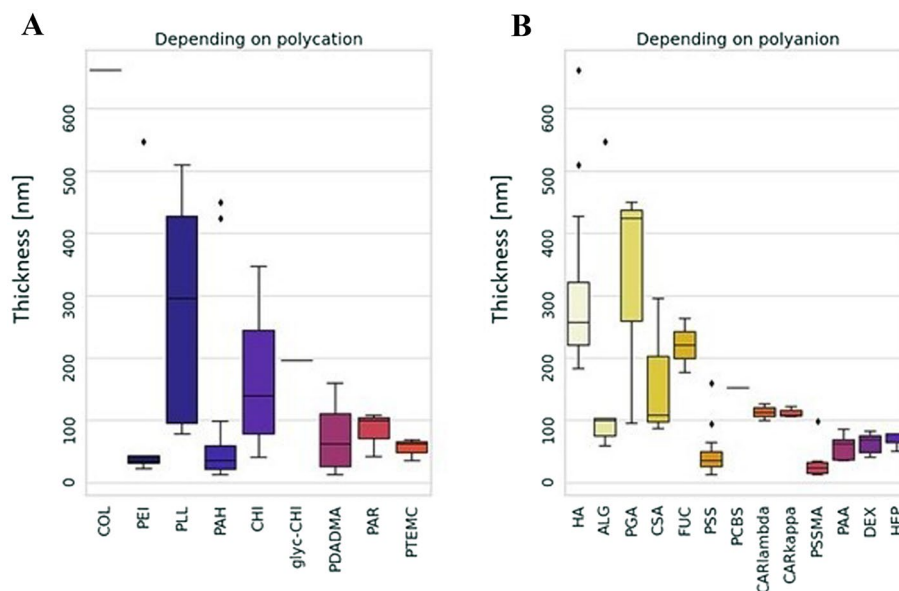


Figure 2. Boxplots showing the distribution of the thickness values in the coatings made of different polymers. **(A)** Thickness depending on polycations. **(B)** Thickness depending on polyanions. COL: collagen, PEI: polyethylenimine, PLL: poly(L-lysine), PAH: poly(allylamine hydrochloride), CHI: chitosan, glyc-CHI: glycol-chitosan, PDADMA: poly(diallyldimethylammonium chloride), PAR: poly(L-arginine), PTEMC: poly(trimethylammonium ethyl methacrylate chloride), HA: hyaluronic acid, ALG: alginate, PGA: poly(L-glutamic acid), CSA: chondroitin sulfate, FUC: fucoidan, PSS: poly(styrene sulfonate), PCBS: poly[1-[4-(3-carboxy-4-hydroxyphenylazo)benzenesulfonamido]-1,2-ethanediy], sodium salt, CARlambda: λ-carrageenan, CARKappa: κ-carrageenan, PSSMA: poly(4-styrenesulfonic acid-co-maleic acid), PAA: poly(acrylic acid), DEX: dextran, HEP: heparin.

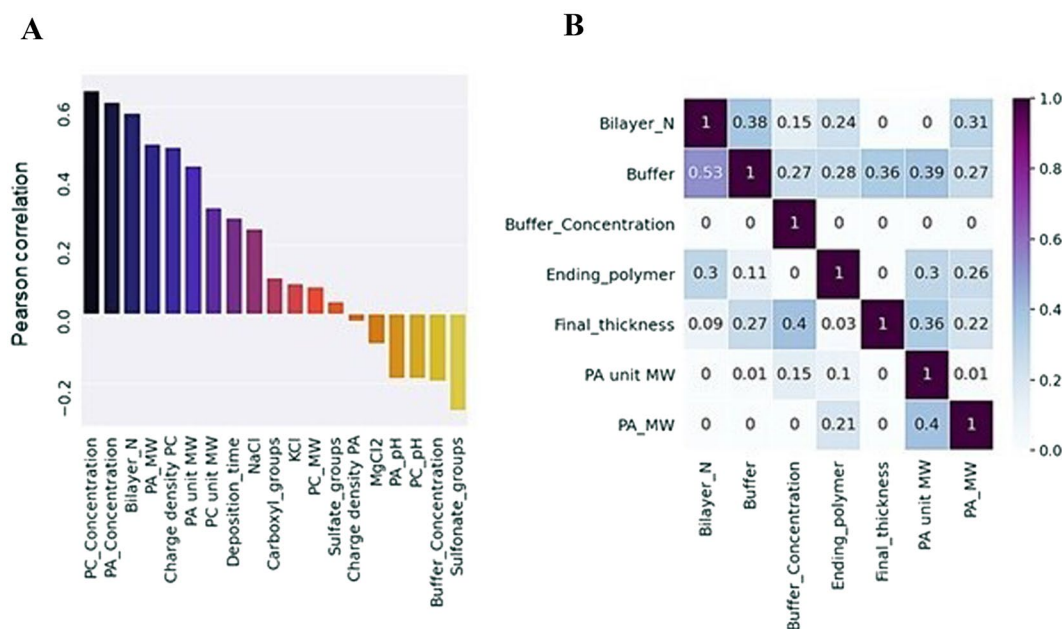


Figure 3. Pearson correlation **(A)** and Predictive Power Scores (PPS) **(B)** calculated for the final thickness of the coating and coating features. The first seven Pearson correlations are statistically significant ($p \leq 0.05$). Only features having PPS > 0.001 with the target value are included. Full names of polymer features are provided in SI Table S3.

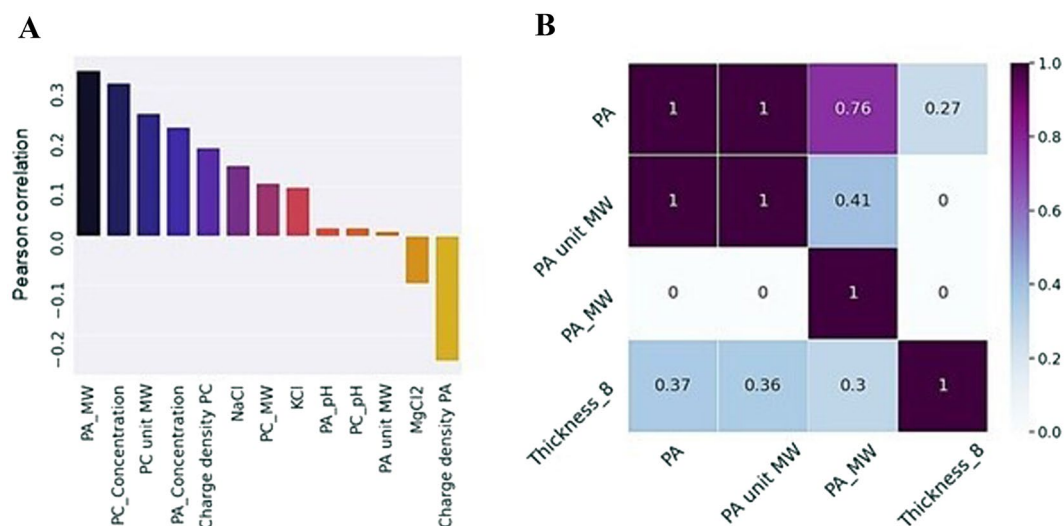


Figure 4. Pearson correlation (A) and Predictive Power Scores (PPS) (B) calculated for the thickness of the 8-bilayered coating and coating features. The first three and the last Pearson correlations are statistically significant ($p \leq 0.05$). Only features having PPS > 0.001 with the target value are included. Full names of polymer features are provided in SI Table S3.

RMSE (nm)	QSPR Bagging regressor	QSPR SVR	Bagging regressor
Training set	46.7	76	50.9
Test set	68.7	113.8	73.8
Validation set	74.7	123.1	226.6

Table 1. Performance of the constructed models measured as root-mean-square error (RMSE, nm) for three data sets. QSPR: quantitative structure–property relationship, SVR: support vector regression.

The growth function, describing changes in coating thickness d with the number of layers, N , has three coefficients, a_0 , a , and b , which vary depending on the coatings^{16,17}. In this function parameter, b defines function curvature: for $b \geq 0.05$, the growth is exponential, for $b < 0.05$, it is nearly linear.

$$d = a_0 + a * e^{bN} \quad (1)$$

We extracted data on the dynamics of each coating growth from the original research papers. Then we used these data to calculate coefficients of growth function. Having these coefficients, we defined the thickness of coatings having eight bilayers ($N = 16$). With this, we created a new target value, the thickness of the 8-bilayer coating, which was not dependent on the number of bilayers.

In this configuration, the type of polyanion, its concentration and its molecular weight were found to positively influence the thickness of the coating, while the charge density of polyanion negatively correlated with the coating thickness (Fig. 4A). Three features (on the x-axis) contributed to the coating thickness (on the y-axis) (Fig. 4B): type of polyanion, polyanion unit molecular weight, and resulting molecular weight. The reason why polyanion characteristics appear more important than those of polycation can be explained by pH values at which polyelectrolyte multilayers were built (pH values are close to the pKa of the acid groups).

Coating thickness predictions. After the determination of the most influential parameters, the next step was the build-up of a predictive model. We constructed a Bagging Regressor model¹⁸ to make predictions about coating thickness using ten features from the original data set: presence of HA, presence of poly(styrene sulfonate) (PSS), resulting MW of the polycation and of the polyanion, unit MW of the polycation and of the polyanion, concentration of polycation and concentration of polyanion, the concentration of NaCl, and charge density of polyanion. Two quantitative structure–property relationship (QSPR) regression models were constructed using the selected features, Bagging Regressor and support vector regression (SVR) (Table 1).

These models have smaller root-mean-square error (RMSE) values for the validation set than the Bagging Regressor constructed using the original dataset features. This fact indicates their greater potential for generalization. The model performance was evaluated with RMSE values calculated for the training set, test set, and validation set. All the models were fivefold cross-validated. All metrics are mean values for scores from different folds from cross-validation.

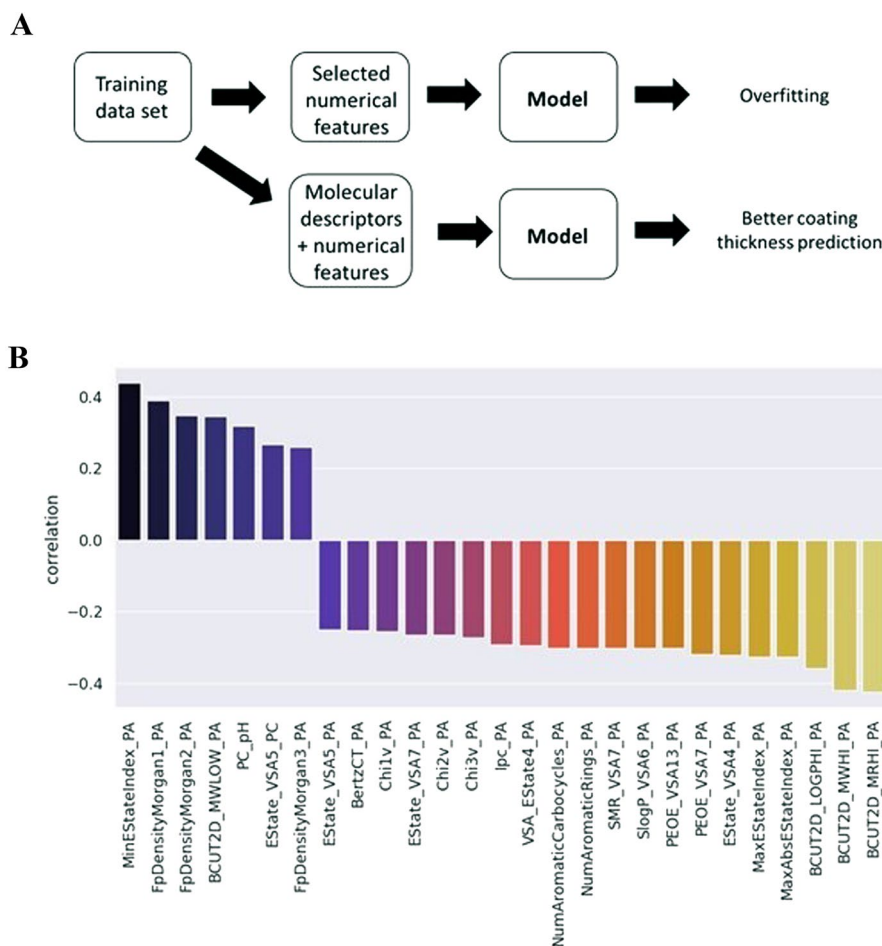


Figure 5. (A) Schematic presentation of the model building process. (B) Correlation between RDKit descriptors and thickness of the coating. Only descriptors with $r \geq 0.25$ are shown. All coefficients are statistically significant ($p \leq 0.05$).

However, we encountered a classical Machine Learning challenge: overfitting, when the model makes good predictions for the instances it was built on (training set), but fails to “generalize”, i.e. make good predictions for the unseen items (validation and test sets). Therefore, the large gap between training and test/validation error values is the major sign of overfitting. This is the case for the Bagging Regressor constructed on the original data set features. As we can see, this model generates a large RMSE value for the validation set, which is 4 times larger than the error for the training set. From here, we conclude that polymers as specific chemical entities are not good features by themselves, and generating numerical features that describe the chemical properties of polymers can improve the model performance.

To get features of a molecule, firstly we had to get information about its structure, which is commonly represented in simplified molecular-input line-entry system (SMILES) format, available in the PubChem database. This information is further used to predict molecular features by deep learning pre-trained models available in the DeepChem library.

For each polycation and polyanion, 123 molecular descriptors were generated, therefore each coating in the dataset was characterized by 246 molecular descriptors. Many molecular features have a moderate correlation with the thickness of the coating, so we assume that they can be used to predict this target value (Fig. 5). Mostly, molecular descriptors with high correlations with the coating thickness are related to polyanions, not polycations; this is in line with the observations in the previous section.

In the next step, we combined features generated by DeepChem and numerical features of the polymers from the original data set. After that, we performed feature elimination with the recursive feature elimination (RFE) algorithm leaving ten features that will be used by the models. These features are MW of the polycation, MW of the polyanion, NaCl concentration, six polyanion RDKit descriptors (MinAbsEStateIndex, FpDensityMorgan3, BalabanJ, PEOE_VSA8, VSA_EState2, VSA_EState6), and one polycation descriptor Kappa1. The most significant molecular descriptors demonstrate the importance of the topological features of the polyanions (such as Balaban distance connectivity index (BalabanJ) and also polycations (Kappa1) in the formation of the supramolecular LbL assemblies in addition to the electrostatic interactions which are described by molecular operation environment

Film	Ground truth, thickness (nm)	Predicted thickness, QSPR RFE/Bagging	Predicted thickness, QSPR RFE/SVR	Predicted thickness, RFE/Bagging
PAR30/DEX40	51.30	68.55	43.86	308.61
CHI50/DEX5	42.70	62.02	91.45	265.33
CHI50/FUC	123.00	125.84	197.79	274.23
CHI50/CARiota	109.30	161.07	232.98	278.50
PAR30/CARiota_1	65.80	146.22	279.71	322.28
PAR30/PSS0.2	7.80	97.00	126.70	285.69
PAR30/PSS4	31.50	97.00	126.89	288.20
PAR30/DEX20	46.00	62.74	42.77	308.61
CHI30/HA108	162.80	308.23	285.14	335.13
CHI20/HA108	218.20	306.78	287.21	335.13
PAR30/CARiota_2	45.90	146.22	279.71	322.28

Table 2. The thickness of the coatings predicted with three models and thickness determined experimentally (ground truth).

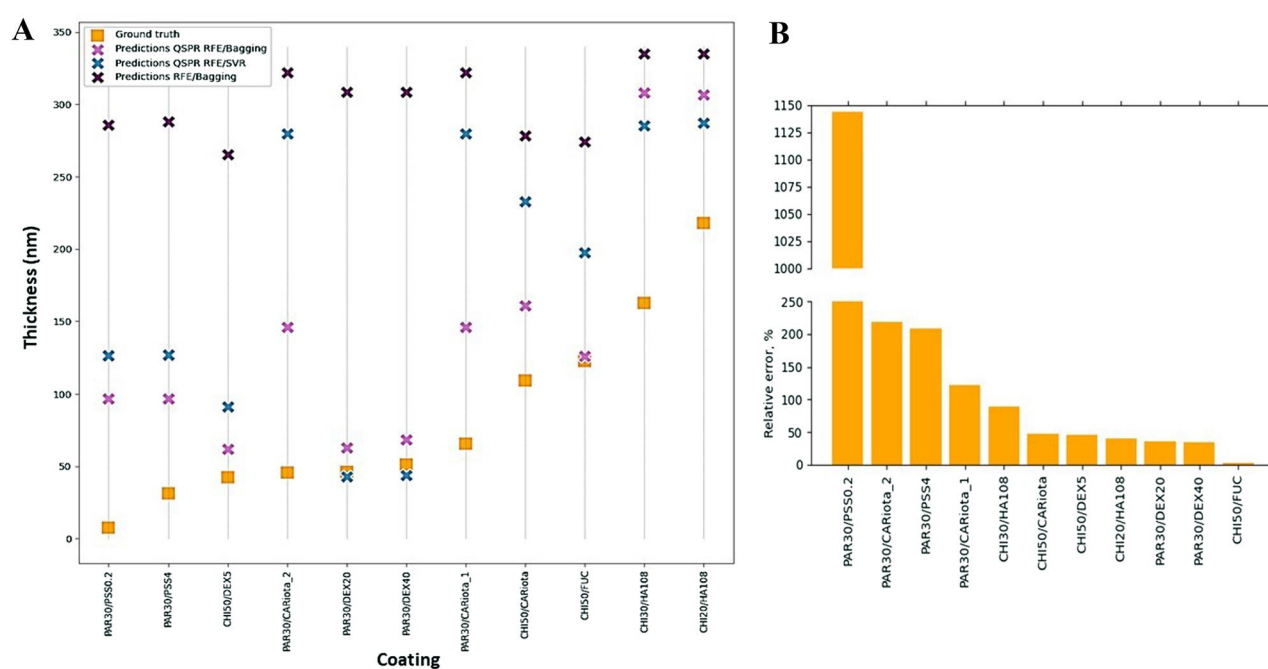


Figure 6. (A) Real (ground truth) and predicted thickness of coatings in the validation data set. (B) Relative errors for predictions made with the best of constructed models (QSPR RFE/Bagging).

(MOE) type electrotopological descriptors. Descriptors related to van der Waals forces and also partial charges underline the highly intricate nature of the LbL film formation at molecular level.

As the last step, in order to further test the generalization capacity of the final model, we have tested polymers which were not in the training set that are described solely by molecular descriptors.

In this configuration, we observed more inaccurate predictions (Table 2, Fig. 6). However, this could be expected, given the size of the training dataset. As we can see, the best results are obtained with the use of the QSPR RFE/Bagging regressor. It can predict the thickness of the coating better than two other models. We can also see that predictions of QSPR models are better than for the model RFE/Bagging constructed with original features, which predicts almost constant thickness values for all coatings (302 ± 25 nm). Predictions of QSPR models are correlating, and there are four coatings for which both regressors failed to predict thickness correctly (relative error > 100%): PAR30/PSS0.2, PAR30/CARiota_2, PAR30/PSS4, PAR30/CARiota_1 (Fig. 6B).

Two other coatings, that have inaccurate predictions, contain CARiota polymer which was not in the dataset during training. The fact that the model fails to make predictions for the coatings which are made of unseen polymers confirms the assumption the generalization is not complete. It is interesting that despite this, there is one coating with CARiota for which relative error is small (47%), and the model was able to make an accurate prediction in this case.

Below, we discuss some of the possible reasons that may have caused large errors in the predictions made by the best of the constructed models, QSPR RFE/Bagging. For the PAR30/PSS0.2 and PAR30/PSS4, a large error

may be caused by the too low molecular weight of the PSS polyanion compared to the one that the model has seen in the training set. The smallest polyanion in the dataset used for training was DEX with MW = 7.2 kDa, whereas in the validation set we have coatings where PSS has lower MW values (0.2 and 4 kDa). Moreover, in the training set, PSS MW was greater than in the validation set (60–70 kDa), so the combination of polyanion/MW of polyanion is new to the model. Because the model has made inaccurate predictions for the kind of polymers that it did not see during training, we assume that it can not generalize well and this is the reason for the prediction failures.

In the end, we observe that the model makes acceptable predictions for the coatings made of combinations of polymers that were present in the data set (like CHI/FUC, CHI/HA, and PAR/DEX). However, it fails to do so for the unseen polymers tested, due to the lack of extensive training data. We believe that the model potentially can be improved by generating and using more data for training. More data points will cover more chemical parameters, and the dependencies between features and thickness will be more informative and will have more predicting power.

Conclusions and perspectives

In this work, we analyzed how different parameters such as polymer molecular weight, concentration, etc. affect LbL film thickness. After the determination of the most influential parameters, we used machine learning approach to verify if we can predict coating thickness from different parameters. We found that construction parameters alone were insufficient to build a robust model for thickness prediction because of the overfitting. To overcome this problem, we hypothesized that generating numerical features that describe the chemical properties of polymers can improve the model performance. Thus, we analysed the relationship between 123 molecular descriptors and the coating thickness, and found that molecular features had a moderate correlation with the thickness of the coating. Finally, we combined molecular descriptors and numerical features of the polymers from the original data set to build new models, and these models had better performance for the validation set than the model constructed using the original dataset features, which indicates their greater potential for generalization.

In conclusion, the generalization capacities of an algorithmic model predicting coating thickness can be improved by delving into the determining properties of the polymers in the context of LbL film formation dynamics. Harnessing the available data science techniques in biomaterial design and development such as multifunctional coatings will decrease lead time, empirical experimental load and also establish relationships between structure and function, which are otherwise hard to guess or estimate. The ultimate goal is to be able to predict coating functionalities based on polymer structure and buildup conditions, to develop innovative coatings for different applications.

Materials and methods

Materials. Alginate (ALG), λ -carrageenan (CAR λ), κ -carrageenan (CAR κ), ι -carrageenan (CAR ι), chitosan (Mw = 50 and 100 kDa; CHI50 and CHI100), chondroitin sulfate (CSA), dextran (Mw = 5, 7.2, 20, 40 and 500 kDa; denoted respectively DEX5, DEX7, DEX20, DEX40 and DEX500), fucoidan (FUC), heparin (HEP), poly(styrene sulfonate) (0.2, 4.2, 15, 29, 70, 80, 145 and 2070 kDa; denoted respectively PSS02, PSS4, PSS15, PSS29, PSS70, PSS80, PSS145, PSS2600) were purchased from Sigma Aldrich, France. Chitosan (Mw = 20, 30 and 250 kDa; CHI20, CHI30 and CHI250) were purchased from Glentham Life Sciences, United Kingdom. Hyaluronic acid (Mw = 29, 108, 823 and 2670 kDa, HA29, HA108, HA823 and HA2670) were purchased from Lifecore Biomedical, USA. Poly(L-arginine) with 30 residues of arginine (PAR30) was purchased from Alamanda Polymers, USA.

Data collection. To collect the data from the literature, we searched articles describing LbL film buildup using PubMed website and Google Scholar (keywords: LbL, polyelectrolyte, film, thickness). In total, 31 articles were found. Among them, only articles specifying the parameters of the film buildup, as well as film thickness, in the text and/or in the figures, were selected. Thus, 19 articles were used in this study^{16,19–36}.

The second dataset consisted of experimental data produced in the laboratory using QCM (Q-Sense, Sweden). For this, a SiO₂ coated-crystal was excited at different resonance frequencies (fundamental frequency, third, fifth, and seventh overtones), and changes in frequency and dissipation were measured during the successive deposition of polymers and rinsing steps (Fig. 7A).

Before each experiment, the crystal was cleaned for 30 min with 2% Hellmanex, then rinsed with water. The final cleaning was done with 1 M HCl for 10 min, then rinsed with water. Poly(L-arginine) with chains composed of 30 arginine units (PAR30) and chitosan with different molecule weights (20, 30, 50, 100 and 250 kDa) (respectively CHI20, CHI30, CHI50, CHI100 and CHI250) were used as polycations. For the films constructed with PAR30 as polycation, a Tris 10 mM/NaCl 0.15 M at pH 7.4 buffer was used for the preparation of solutions and for rinsing steps. For the films constructed using chitosan as polycation, a buffer consisting of sodium acetate 70 mM/NaCl 80 mM at pH5 was used. Buffers were filtered with 0.2- μ m filters. Twenty-eight different polymers were used as polyanions (see [Materials](#) part). QCM-D experiments were performed as described previously³⁷. Briefly, polyanions and polycations used for the experiment were prepared at 0.5 mg mL⁻¹ in acetate or Tris/NaCl buffers, as explained above. Polycations were first adsorbed to the surface for 5 min. A rinsing step was performed using the buffer for 5 min after each polyelectrolyte deposition, and 8 bilayers were constructed on the SiO₂-coated crystal with a flow rate of 250 μ L min⁻¹.

Sauerbrey's equation gives the relation between the mass deposited on the vibrating crystal per unit area and the change of resonance frequency when the deposit behaves like a stiff coating of the crystal, i.e. when the deposit changes only the mass of the crystal. This is, however, generally not the case when a polyelectrolyte film is built up by successive depositions of polycations and polyanions from solutions. Indeed, on the one hand, the film is a viscoelastic body and on the other hand, the film is in contact with the solution. For both reasons,

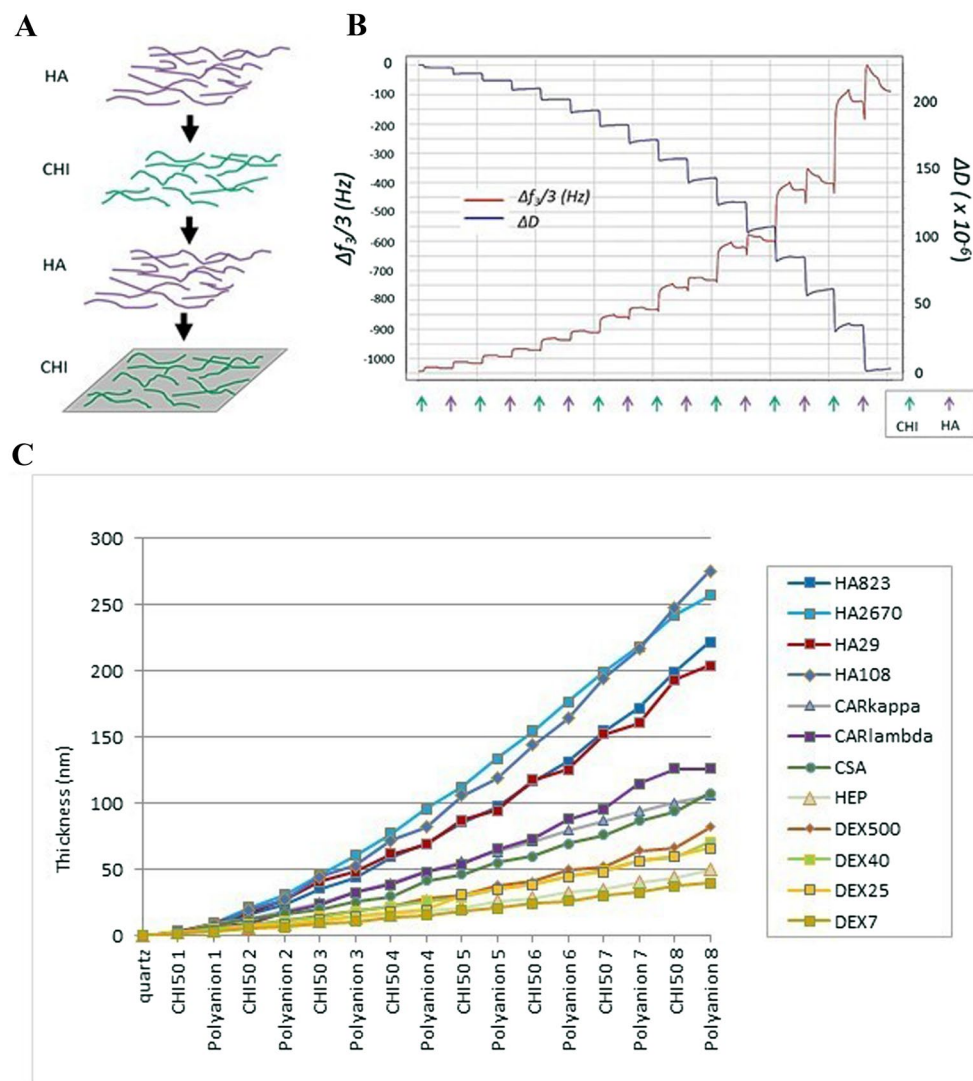


Figure 7. Film thickness determination by QCM-D. (A) Polyelectrolytes (chitosan, CHI, and hyaluronic acid, HA) are deposited on the crystal, where they form a multilayer film. (B) Growth of the film is followed by the measurements of the frequency and the dissipation variations with respect to the crystal in contact with the buffer solution only. As an example, the figure shows the measurements corresponding to the third overtone. (C) Film growth depending on polyanion.

not only the change in frequency is measured but also the dissipation related to the characteristic damping time of the crystal vibration (Fig. 7B). If the normalized frequency shifts, $\Delta f_n/\nu$, corresponding to different overtone numbers are equal to $-m/C$, i.e. are independent of ν , Sauerbrey's equation stays valid. If this is not the case, a more sophisticated formalism ought to be used³⁸. In this approach, the film is characterized by its elasticity, μ , its viscosity, η , its density, ρ , and its thickness, d . In the present study, the frequency shift and the dissipation corresponding to the overtones $\nu = 3, 5$ and 7 , i.e. for excitation frequencies of about 15, 25 and 35 MHz, have been processed to extract the thickness of the film at the end of the build-up (examples of film growth are shown in Fig. 7C). Hereafter, all thickness results are given for $\rho = 1 \text{ g cm}^{-3}$.

Data analysis. The data set (Table S1), used for analysis and model construction, had 76 data points which were obtained by literature search (43 points) and experiments (33 points). The validation set used to evaluate model performance had 11 points, all generated by experiments (Table S2).

In the original data set, each data point represents one coating with 23 features and one target value (resulting coating thickness). The features are the type of polycation (PC) and type of polyanion (PA), type of ending polymer, PC unit molecular weight (MW) and PA unit MW, total PC and PA MWs, the concentration of PC and the concentration of PA, polycation pH and polyanion pH, the charge density of PC and charge density of PA, presence of carboxyl groups, presence of sulfonate groups, presence of sulfate groups, crosslinking method,

bilayer deposition time, type of buffer, buffer concentration, the concentration of KCl, the concentration of MgCl₂, and the concentration of NaCl (Table S3).

Relationships between features and target value were evaluated with the Pearson correlation coefficient and Predictive Power Score³⁹. Statistical significance of the Pearson correlation coefficient was checked using confidence intervals for $p = 0.05$ after Fisher's Z-transformation.

Three regression models were built to predict coating thickness: one based on features from the original data set, and two Quantitative Structure–Property Relationships (QSPR-type) models based on the RDKit molecular descriptors. To build QSPR models, firstly the SMILES (simplified molecular-input line-entry system) representations of molecules were obtained from the PubChem database. Then, for each polymer SMILES, 123 molecular descriptors were generated using RDKit Descriptors from DeepChem library⁴⁰.

Features, that are further used to build models, were selected from original feature space and from molecular descriptors space with the Recursive Feature Elimination method having Random Forest as the basis. Finally, two methods were used to construct three coating thickness prediction models: Bagging Regression, QSPR Bagging Regression, and QSPR Support Vector Regression (SVR). All the described methods are implemented in the scikit-learn library²².

Received: 29 March 2021; Accepted: 6 September 2021

Published online: 21 September 2021

References

- Decher, G. Fuzzy nanoassemblies: Toward layered polymeric multicomposites. *Science* **277**, 5 (1997).
- Decher, G. & Hong, J. D. Buildup of ultrathin multilayer films by a self-assembly process: II. Consecutive adsorption of anionic and cationic bipolar amphiphiles and polyelectrolytes on charged surfaces. *Ber. Bunsenges. Phys. Chem.* **95**, 4 (1991).
- Richardson, J. J. *et al.* Innovation in layer-by-layer assembly. *Chem. Rev.* **116**, 14828–14867 (2016).
- Zhao, S. *et al.* The future of layer-by-layer assembly: A tribute to ACS nano associate editor Helmuth Möhwald. *ACS Nano* **13**, 6151–6169 (2019).
- Campbell, J. & Vikulina, A. S. Layer-by-layer assemblies of biopolymers: Build-up, mechanical stability and molecular dynamics. *Polymers (Basel)* **12**, 1949 (2020).
- Boudou, T., Crouzier, T., Ren, K., Blin, G. & Picart, C. Multiple functionalities of polyelectrolyte multilayer films: New biomedical applications. *Adv. Mater.* **22**, 441–467 (2010).
- de Avila, E. D. *et al.* Anti-bacterial efficacy via drug-delivery system from layer-by-layer coating for percutaneous dental implant components. *Appl. Surf. Sci.* **488**, 194–204 (2019).
- Gribova, V., Auzely-Velty, R. & Picart, C. Polyelectrolyte multilayer assemblies on materials surfaces: From cell adhesion to tissue engineering. *Chem. Mater.* **24**, 854–869 (2012).
- Pahal, S., Gakhar, R., Raichur, A. M. & Varma, M. M. Polyelectrolyte multilayers for bio-applications: Recent advancements. *IET Nanobiotechnol.* **11**, 903–908 (2017).
- Priolo, M. A., Holder, K. M., Guin, T. & Grunlan, J. C. Recent advances in gas barrier thin films via layer-by-layer assembly of polymers and platelets. *Macromol. Rapid Commun.* **36**, 866–879 (2015).
- Rivero, P. J., Goicoechea, J. & Arregui, F. J. Layer-by-layer nano-assembly: A powerful tool for optical fiber sensing applications. *Sensors* **19**, 683 (2019).
- Lutkenhaus, J. L. & Hammond, P. T. Electrochemically enabled polyelectrolyte multilayer devices: From fuel cells to sensors. *Soft Matter* **3**, 804–816 (2007).
- Picart, C. *et al.* Molecular basis for the explanation of the exponential growth of polyelectrolyte multilayers. *PNAS* **99**, 4 (2002).
- Abtahi, S. M. *et al.* Micropollutant rejection of annealed polyelectrolyte multilayer based nanofiltration membranes for treatment of conventionally-treated municipal wastewater. *Sep. Purif. Technol.* **209**, 470–481 (2019).
- Richert, L. *et al.* Improvement of stability and cell adhesion properties of polyelectrolyte multilayer films by chemical cross-linking. *Biomacromol* **5**, 10 (2004).
- Boulmedais, F. *et al.* Buildup of exponentially growing multilayer polypeptide films with internal secondary structure. *Langmuir* **19**, 440–445 (2003).
- Müller, M. The anomalous influence of polyelectrolyte concentration on the deposition and nanostructure of poly(ethyleneimine)/poly(acrylic acid) multilayers. *Molecules* **24**, 2141 (2019).
- Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Boudou, T., Crouzier, T., Auzély-Velty, R., Glinel, K. & Picart, C. Internal composition versus the mechanical properties of polyelectrolyte multilayer films: The influence of chemical cross-linking. *Langmuir* **25**, 13809–13819 (2009).
- Abtahi, S. M., Ilyas, S., Joannis Cassan, C., Albasi, C. & de Vos, W. M. Micropollutants removal from secondary-treated municipal wastewater using weak polyelectrolyte multilayer based nanofiltration membranes. *J. Membr. Sci.* **548**, 654–666 (2018).
- Alves, N. M., Picart, C. & Mano, J. F. Self assembling and crosslinking of polyelectrolyte multilayer films of chitosan and alginate studied by QCM and IR spectroscopy. *Macromol. Biosci.* **9**, 776–785 (2009).
- Benbow, N. L. *et al.* Odd-even effects on hydration of natural polyelectrolyte multilayers: An in situ synchrotron FTIR microspectroscopy study. *J. Colloid Interface Sci.* **553**, 720–733 (2019).
- Buron, C. C. *et al.* Surface morphology and thickness of a multilayer film composed of strong and weak polyelectrolytes: Effect of the number of adsorbed layers, concentration and type of salts. *Thin Solid Films* **517**, 2611–2617 (2009).
- Crouzier, T. & Picart, C. Ion Pairing and hydration in polyelectrolyte multilayer films containing polysaccharides. *Biomacromol* **10**, 433–442 (2009).
- Czerwieńec, B., Strawski, M., Granicka, L. H. & Szklarczyk, M. AFM study of adhesion and interactions between polyelectrolyte bilayers assembly. *Colloids Surf. A* **555**, 465–472 (2018).
- Elzbięciak, M. *et al.* Influence of pH on the structure of multilayer films composed of strong and weak polyelectrolytes. *Langmuir* **25**, 3255–3259 (2009).
- Elzbięciak-Wodka, M., Kolasinińska-Sojka, M. & Warszyński, P. Effect of mono- and divalent ions on the formation and permeability of polyelectrolyte multilayer films. *J. Electroanal. Chem.* **789**, 123–132 (2017).
- Fernández, R., Ocando, C., Fernandes, S. C. M., Eceiza, A. & Tercjak, A. Optically active multilayer films based on chitosan and an azopolymer. *Biomacromol* **15**, 1399–1407 (2014).
- Hillberg, A. L., Holmes, C. A. & Tabrizian, M. Effect of genipin cross-linking on the cellular adhesion properties of layer-by-layer assembled polyelectrolyte films. *Biomaterials* **30**, 4463–4470 (2009).
- Holmes, C. A. & Tabrizian, M. Enhanced MC3T3 preosteoblast viability and adhesion on polyelectrolyte multilayer films composed of glycol-modified chitosan and hyaluronic acid. *J. Biomed. Mater. Res. Part A* **100A**, 518–526 (2012).

31. Ngankam, A. P., Mao, G. & Van Tassel, P. R. Fibronectin adsorption onto polyelectrolyte multilayer films. *Langmuir* **20**, 3362–3370 (2004).
32. Picart, C. *et al.* Primary cell adhesion on RGD-functionalized and covalently crosslinked thin polyelectrolyte multilayer films. *Adv. Func. Mater.* **15**, 83–94 (2005).
33. Strawski, M., Granicka, L. H. & Szklarczyk, M. Redox properties of polyelectrolyte multilayer modified electrodes: A significant effect of the interactions between the polyelectrolyte layers in the films. *Electrochim. Acta* **226**, 121–131 (2017).
34. Tjipto, E., Quinn, J. F. & Caruso, F. Assembly of multilayer films from polyelectrolytes containing weak and strong acid moieties. *Langmuir* **21**, 8785–8792 (2005).
35. Trybała, A., Szyk-Warszyńska, L. & Warszyński, P. The effect of anchoring PEI layer on the build-up of polyelectrolyte multilayer films at homogeneous and heterogeneous surfaces. *Colloids Surf. A* **343**, 127–132 (2009).
36. Zhang, J. *et al.* Natural polyelectrolyte films based on layer-by-layer deposition of collagen and hyaluronic acid. *Biomaterials* **26**, 3353–3361 (2005).
37. Mutschler, A. *et al.* Unexpected bactericidal activity of poly(arginine)/hyaluronan nanolayered coatings. *Chem. Mater.* **28**, 8700–8709 (2016).
38. Voinova, M. V., Rodahl, M., Jonson, M. & Kasemo, B. Viscoelastic acoustic response of layered polymer films at fluid-solid interfaces: continuum mechanics approach. *Phys. Scr.* **59**, 391–396 (1999).
39. Wetschorek, F., Krael, T. & Krishnamurthy, S. *8080labs/ppscore: Zenodo Release (Version 1.1.2)* (Zenodo, 2020).
40. Ramsundar, B. *et al.* *Deep Learning for the Life Sciences* (O'Reilly Media, 2019).

Acknowledgements

This project received funding from the European Union's Horizon 2020 PANBioRA research and innovation program under Grant Agreement No. 760921, from ANR TerminAnion and Bourse Frenchtech Emergence grant.

Author contributions

P.L. and N.E.V. conceived the experiments; C.C., E.L. and L.D. conducted the experiments; V.G., A.N., O.L. and B.S. analyzed the results. All authors reviewed the manuscript. A.N.'s new affiliation is Eagle Genomics Ltd., Cambridge, UK.

Competing interests

Nihal Engin Vrana is the majority shareholder of SPARTHA Medical which is a coating development company. Philippe Lavalley is a shareholder of SPARTHA Medical and Cynthia Calligaro is a full-time employee of SPARTHA Medical. The article does not contain any information about SPARTHA products. Anastasiia Navalikhina and Oleksandr Lysenko are employees of Preste, a Data Science company. They provided impartial data analysis. Varvara Gribova, Bernard Senger, Lucie Deiber, Eloise Lebaudy have no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98170-x>.

Correspondence and requests for materials should be addressed to N.E.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021