EPiC Series in Computing

Volume 83, 2022, Pages 146–154

Proceedings of 14th International Conference
on Bioinformatics and Computational Biology

# Machine Learning Techniques in Structure-Property Optimization of Polymeric Scaffolds for Tissue Engineering

Zigeng Wang[1], Xia Xiao[1], Syam P. Nukavarapu[1,2], Sangamesh G. Kumbar[1,2], and Sanguthevar Rajasekaran[1*]

[1]University of Connecticut, Storrs, CT, USA
[2]University of Connecticut Health, Farmington, CT, USA
zigeng.wang@uconn.edu, xia.xiao@uconn.edu, syam.nukavarapu@uconn.edu, kumbar@uchc.edu, and sanguthevar.rajasekaran@uconn.edu

**Abstract**

Biomaterials and biomedical implants have revolutionized the way medicine is practiced. Technologies, such as 3D printing and electrospinning, are currently employed to create novel biomaterials. Most of the synthesis techniques are ad-hoc, time taking, and expensive. These shortcomings can be overcome greatly with the employment of computational techniques. In this paper we consider the problem of bone tissue engineering as an example and show the potentials of machine learning approaches in biomaterial construction, in which different models was built to predict the elastic modulus of the scaffold at given an arbitrary material composition. Likewise, the methodology was extended to cell-material interaction and prediction at an arbitrary process parameter.

## 1 Introduction

The advances of biomaterial discovery are changing almost every aspect of human lives, from drug design and thermal insulation to noise absorption and fuel cell. Several polymeric biomedical implants including surgical sutures, tissue substitutes (scaffolds), and drug-eluting devices have been currently in use to improve the quality of life for millions of people in the US and worldwide.

* Corresponding author

H. Al-Mubaid, T. Aldwairi and O. Eulenstein (eds.), BICOB2022 (EPiC Series in Computing, vol. 83), pp. 146–154

Biomaterials and their application in biomedical devices/implants have resulted in more than $300 billion business worldwide and the field is rapidly growing.

The major focus of tissue engineering is to create tissue substitutes to enable the repair and regeneration of damaged body parts. Scaffold-based tissue engineering creates three-dimensional porous structures to fill the tissue void, allowing cells to attach and promoting tissue ingrowth. A wide range of biodegradable polymers and their scaffolds have been developed and are used clinically today. However, several new polymers and modifications to existing polymers, are constantly being developed and applied to meet ongoing and evolving challenges in biomedical applications.

Biomaterial discovery, synthesis, and characterization to ensure biocompatibility is considered a grand challenge. Biomaterial processing techniques into implants and 3D scaffolds are dependent on the physicochemical properties of the polymer and intended end application. Scaffold fabrication techniques include salt-leaching, particle sintering, electrospinning, and 3D printing to name a few. Additionally, efforts are also made to employ micro and nanofabrication techniques to incorporate nanofeatures into the scaffold to positively influence cell-material interaction and tissue regeneration. Each varied component of the scaffold in terms of choice of polymer and fabrication methodology significantly influences the structure-property of the scaffold or device. Therefore, extensive characterization is essential to ensure functional performance. However, the current biomaterial discovery, synthesis, and characterization techniques are ad-hoc, time taking, and expensive. Ongoing efforts are looking at alternative ways to expedite biomaterial discovery by employing better analytical and computational techniques.

Computational techniques with efficient machine learning-based prediction algorithms may accurately predict material properties and have the potential to dramatically shorten the biomaterials discovery process, reduce expensive and repetitive experimentation.

This manuscript employs machine learning approaches to the bone tissue engineering scaffold to predict two parameters such as mechanical strength and cell attachment as examples [2]. The elastic modulus of the scaffold varied with varied material composition and results were generated by mechanical testing in controlled experiments. The model was built to predict the elastic modulus of the scaffold at given an arbitrary material composition. Likewise, the methodology was extended to cell-material interaction and prediction at an arbitrary process parameter.

# 2 Materials and Methods

## 2.1 Osteogenic cell survival

Biomaterials play a crucial role in tissue engineering. For example, tissue engineering constructs have been developed and studied for bone tissue engineering. The success of such engineered constructs depend on their ability to allow homogeneous cell in growth and bone regeneration throughout the construct [2]. One of the challenges in ensuring homogeneity lies in the fact that cells tend to migrate towards areas of higher nutrient levels, resulting in significantly higher cell densities at the construct's periphery compared to the interior. This will result in decreased oxygen tension and buildup of waste-products within the construct's interior regions [6].

To overcome the above problem, the authors of [2] have developed novel oxygen tension-controlled matrices that support more homogeneous oxygen levels throughout the constructs. These experimental studies resulted in the development of polylactic co-glycolic acid (PLGA) scaffolds with optimized pore distribution and the percent pore volumes.

Amini, et al.'s in vitro experimental results show that their approach results in significantly decreased oxygen and pH gradient from the exterior of the construct to the interior [2]. The study also tested the ability of the constructs to support the maintenance of two clinically relevant progenitor cell

populations for bone tissue engineering and vascularization, namely mesenchymal stem cells (MSCs) and endothelial progenitor cells (EPCs). The expression of key bone and vascular markers have been confirmed via immunofluorescence.

In addition to homogeneous tissue formation, the scaffolds developed for bone also need to show bone compatible mechanical characteristics. Therefore, the study has identified an optimal pore size range that will ensure the constructs to have compressive modulus and strength in the range of human cancellous bone. Once developed, the scaffolds were investigated by seeding with bone and blood-vessel forming cell populations that are clinically relevant for bone tissue engineering and vascularization, namely mesenchymal stem cells (MSCs) and endothelial progenitor cells (EPCs). This study was done over a period of around three years at the cost of more than $300K. In this paper we show that Machine Learning techniques can be employed fruitfully to speedup the discovery of scaffolds at a much reduced cost.

## 2.2 The complexity of the problem

In this study [2], the role of pore size and volume in improving oxygen tension and pH gradients was studied experimentally. The variables here are, pore size, pore volume and cell proliferation. For each combination of parameter values, an experiment can be performed to see how good this combination is. However, even a single experiment could take days to complete and can be expensive. The number of combinations for the parameter values could be very huge. For instance, if there are n parameters, and if each parameter is binary, then the total number of possible combinations is 2n. Even if n = 20, this number is more than a million! Various techniques are currently in use to evaluate scaffold architectures, including scanning electron microscopy (SEM) analysis, flow and mercury porosimetry, gas pycnometry and adsorption. Unfortunately, these techniques have crucial drawbacks such as being destructive or toxic, time consuming and not resulting in highly accurate measurements.

For example, for every single set of values for the control variables, an experiment was conducted by Amini, et al. [2]. A summary of an experiment follows. 1) PLGA microspheres were prepared by an oil-in-water method that took several hours; 2) Scaffold porosity measurements (n=3/scaffold group) were carried out using cone-beam micro-focus X-ray computed tomography (CT) analysis, Specifically, two-dimensional density profile images were obtained at every 6 micron depth of the scaffold and these images were analyzed using a software; 3) Bone marrow derived mesenchymal stem cells (MSCs) were isolated from New Zealand White rabbits; 4) Needle-type fiber optic oxygen microsensors and pH microsensors were utilized to analyze oxygen levels and pH levels in the interior of MSC-seeded control and macro-porous scaffolds; 5) Live-dead cell viability assay was used to analyze cell survival in the interior of cell-seed constructs; and 6) To assess the ability of the oxygen tension controlled matrices to support osteogenic and vasculogenic stem cell growth, MSCs and EPCs were seeded on the scaffolds and cultured for 2 days in vitro. Clearly, this entire process spans several days and is expensive.

## 2.3 Machine learning backgrounds

In simple terms, machine learning can be defined as follows: Consider a case where we are interested in guessing a function f(x1,x2,...,xn) where x1,x2,...,xn are variables (i.e., parameters) that we can control. For instance, in the case of a degradable polymer, the variables could be the chemistry, hydrolytic group chemistry, monomer composition and the molecular weight. The function f could be a specific property of the polymer. In practice, when a materials scientist wants to synthesize a material with a specific desired value for f, (s)he will identify values for the control variables (using intuition or otherwise), fabricate a material using these values, and experimentally measure the property f. If the fabricated material indeed has the desired value for f, the scientist stops. Otherwise, (s)he will change the values of the control variables and repeat this process. Clearly, this is

a very time consuming and costly exercise. Machine learning can come to the rescue. A machine learning algorithm has to be supplied with training data in order for it to come up with a guess for f. Training data, typically, will be a series of examples. An example takes the form: (a1,a2,...,an,f(a1,a2,...,an)), where a1,a2,...,an are specific values for the control variables.

A number of machine learning models and algorithms have been proposed in the literature. Examples include regression, support vector machines, random forests, neural networks, and probably approximately correct (PAC) learning. Deep learning refers to large neural networks (i.e., networks with a large number of layers). In the past decade neural networks have been extensively used in many domains.

Several papers have been published that focus on applying machine learning to solve problems in biomedicine and biomaterials (to a small extent). For instance, Mamoshina, et al. [8] review different applications of deep learning in biomedicine. Applications included the identification of biomarkers from biological data, metagenomics classification, transcriptomics, drug discovery and repurposing, and multiomics. In [7], Li, et al. study the link between chemical structures of peptides and their corresponding hydrogel properties using ML. They have generated a structurally diverse hydrogel library with more than 2,000 peptides and evaluated their corresponding properties. The ML approach they have developed was able to link the chemical structures with their self-assembly behavior and hence can accelerate the design of novel peptide structures for biomedical use.

In [3], Chen, et al. have developed a shape phenotyping framework to associate cell morphology with cell-material interactions. This framework is based on support vector machines (SVMs). They have first employed a feature selection algorithm to select the most significant combination of cell shape metrics before applying the SVM model. Tuning cell shape by altering the biophysical properties of biomaterial substrates on which cells operate is the subject of [13]. The authors train a machine learning model to classify cell shape phenotypes. This work has potential applications in fundamental mechanobiology studies and 3D bioprinting of tissue constructs.

Accumulation of biofilms on biomaterial surfaces is a major health problem. Vyas, et al. offer an ML based approach for removing biofilms [14]. ML is used here to segment biofilm from scanning electron microscopy images. Ren, et al. focus on the problem of identifying a new system of metallic glasses in the Co-V-Zr ternary [12]. With the help of ML and high throughput experiments they have discovered three new glass-forming systems.

## 2.4 Different machine learning models

Computational techniques could indeed help to cope with the above problems. Specifically, machine learning can be fruitfully employed. A simple approach could be to perform experiments for a small set of possible combinations for the parameter values, use the resulting data to train a model, and then employ the model to predict the performance for any possible combination of parameter values.

Gaussian process regression is a popular machine learning model which has been frequently used in material related machine learning tasks [5,15]. Gaussian process is a nonparametric, Bayesian-type of approach, mainly designed for regression. Gaussian process can work well with comparatively small datasets. Due to its Bayesian nature, Gaussian process can provide meaningful predictions together with uncertainty measurements [10].

Neural network models or artificial neural network models are receiving growing research and industrial application interests in recent decades due to their ability of estimation complex non-linear distributions. Neural network models are inspired by the functioning of real neural network in brain which processes information through layers of neurons. Each artificial neuron is combined with an activation function which introduces nonlinearity into the estimation [4].

Kernel ridge regression employs kernel trick in ridge regression, in which a linear function is learned by minimizing the linear least squares with l2norm regularization in a non-linear transformed

space induced by the kernel. The kernel trick enables the model to learn corresponding nonlinear relations between the attributes and the target and the least square objective makes the model fitting efficient.

Support vector regression also naturally supports kernel trick to learn non-linear distributions by mapping inputs into high-dimensional feature space. Different from kernel ridge regression, the support vector regression uses epsilon-insensitive loss as the main loss function. This loss function enables the model to construct a hyperplane through an acceptable error margin which provides flexibility of defining the degree of error acceptance.

# 3 Results

We employ the four machine learning models above on bone tissue engineering [2] to show the potentials of machine learning approaches in biomaterial construction. Scaffolds for bone tissue engineering require higher mechanical strength to withstand the load and ambulatory forces. The compressive modulus of the cellulose acetate (CA) scaffold was measured at varied compositions with cellulose acetate phthalate (CAP). Elastic modulus of the CA:CAP scaffold changed as a function of CA ratio. We collect four groups of controlled experiment results based on four different CA proportion values. We build our model to predict the elastic modulus given an arbitrary CA proportion.

As shown in Fig. 1, we employ four different machine learning models to predict elastic modulus, including Gaussian process regression (GPR), neural network (NN), kernel ridge regression (KRR) and support vector regression (SVR). For each individual machine learning model, we divide the data into four folds according to the four groups of experiments of distinct CA proportion values. For each model, we leave one fold out for testing and the rest three folds for training.

For the Gaussian process regression model, we use default Constant Kernels. For neural network, we design the network with one hidden layer of size 100. The activation function for the hidden layer is sigmoid and the activation functions for the output layer is linear. We use Adam optimizer with learning rate 5e−2 and we conduct training for 1000 epochs. For kernel ridge regression, we use sigmoid kernel with no weight penalization. For support vector regression, we use RBF kernels with default regularization parameter as 1.0. We specify the epsilon-tube with no penalty as 0.2.

| Models | GPR | NN | KRR | SVR |
|---|---|---|---|---|
| Mean Absolute Error | 37.92 | 24.46 | 6.26 | 43.73 |

**Table 1**: Elastic Modulus Prediction Accuracy Comparison

We see that the KRR model can predict the elastic modulus with a high accuracy. The mean absolute error (MAE) we achieved is 6.26 and the mean percentage absolute error is 4.53%. Machine learning models have the potential to accelerate the discovery of ideal experimental settings and to speed up the whole biomaterial discovery process.

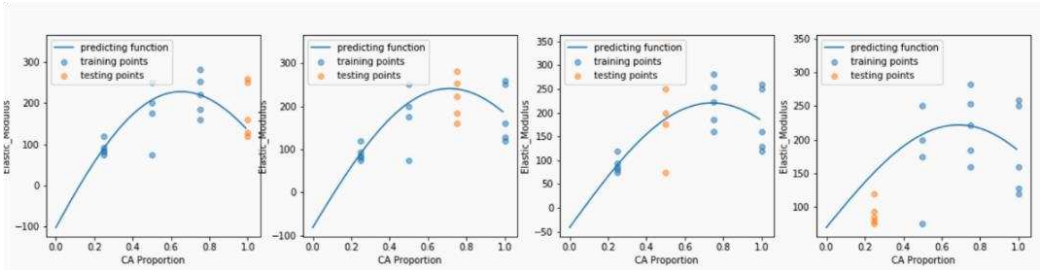
| Models | GPR | NN | KRR | SVR |
|---|---|---|---|---|
| Mean Absolute Error | 2.82 | 0.96 | 2.90 | 0.99 |

**Table 2**: MSC Prediction Accuracy Comparison

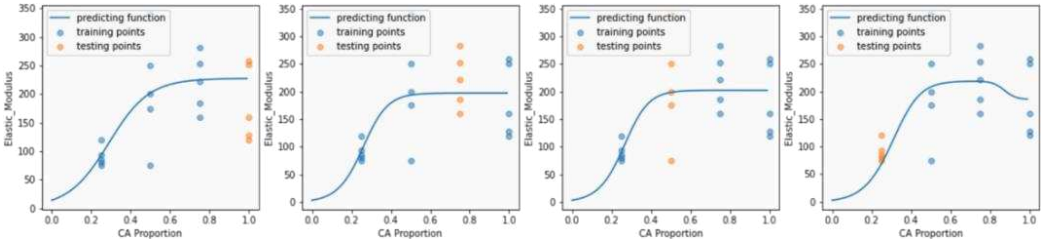
Besides elastic modulus, MSC is another key property in bone tissue engineering [2]. Concentration is an important feature that influences the MSC. We also employ the four machine learning models to predict the MSC with different given concentration values. Since the experiments

are conducted in groups with five concentration values, we train our models and evaluate their prediction performance with 5-fold cross validation. We show the results in Fig. 2. Among the four models, neural network can predict the MSC with the highest accuracy as shown in Fig. 2(b), with the MAE 0.96 and support vector regressor can also achieve a relatively low MAE as 0.99. It is meaningful to predict that MSC and concentration follow a complex non-linear relation that neural network and support vector regressor can nicely estimate.

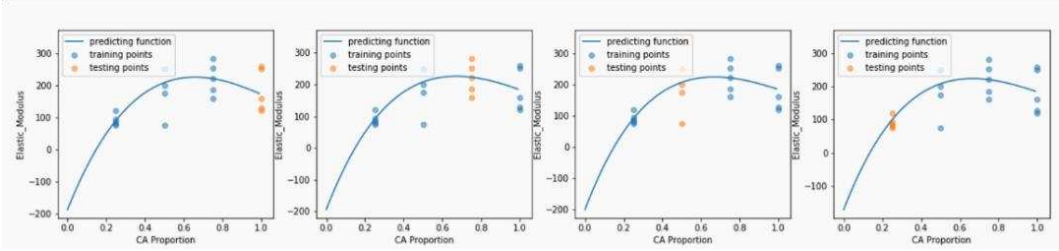
Besides the prediction accuracy comparison, we also summarize the computational cost of the machine learning models with respect to time usage during model training/prediction and hardware expense. The GPR, KRR and SVR are implemented in scikit-learn toolbox [9] and the NN model is implemented with Tensorflow 2.3.0 [1]. We run our experiments with Python 3.7.4 on Ubuntu 18.04 with AMD Threadripper 2950X CPU and single NVIDIA Titan RTX with 24 GB GPU memory.



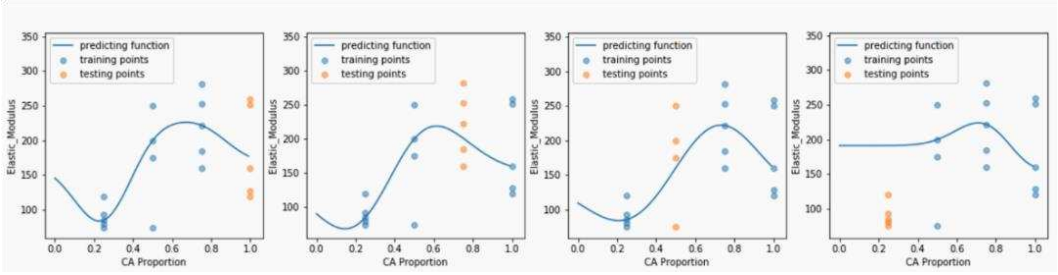

(a) Gaussian Process Regression




(b) Neural Network

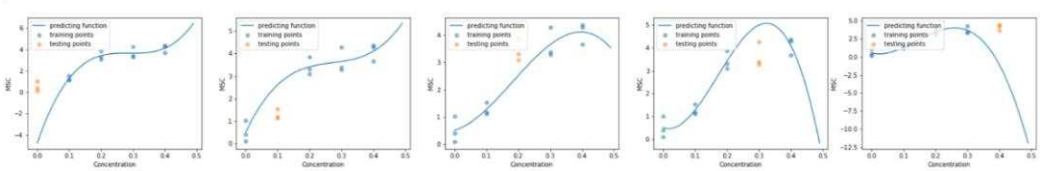



(c) Kernel Ridge Regression

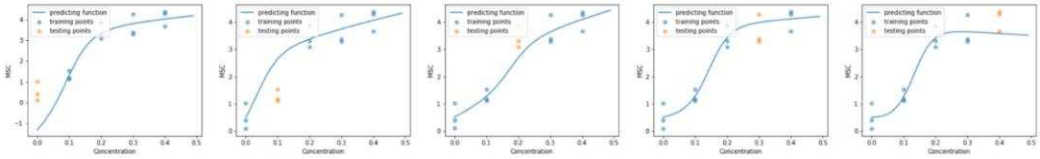



(d) Support Vector Regression

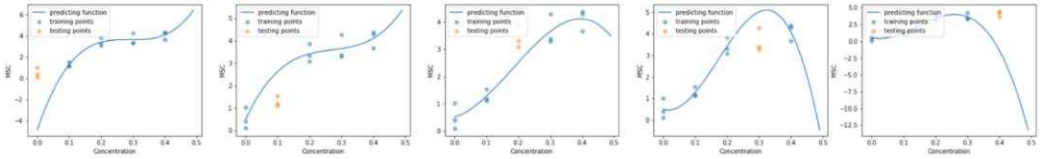
**Figure 1**: Elastic Modulus Prediction Using Different Machine Learning Models



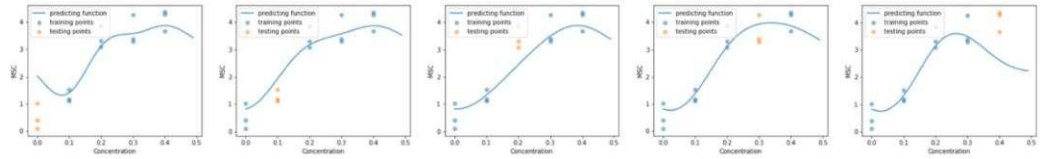
(a) Gaussian Process Regression



(b) Neural Network



(c) Kernel Ridge Regression



(d) Support Vector Regression

**Figure 2**: MSC Prediction Using Different Machine Learning Models

| Models | GPR | NN | KRR | SVR |
|---|---|---|---|---|
| Average Training Time (ms) | 1.34 | 2415 | 0.59 | 0.38 |
| Average Prediction Time (ms) | 0.15 | 28.2 | 0.16 | 0.07 |

**Table 3**: Elastic Modulus Computational Efficiency Comparison

| Models | GPR | NN | KRR | SVR |
|---|---|---|---|---|
| Average Training Time (ms) | 1.67 | 2774 | 1.01 | 0.29 |
| Average Prediction Time (ms) | 0.25 | 29.8 | 0.13 | 0.06 |

**Table 4**: MSC Computational Efficiency Comparison

As shown in the tables above, all of the machine learning models can learn and predict the data distribution very efficiently. For GPR, KRR and SVR model, the processing is in millisecond level. For NN model, the processing time is within several seconds. Compared with the human experiment time which is usually months or even years, the machine learning models are 7-9 orders more efficient. The computational resources such as the computer desktop are within 5K US dollars and highly reusable. Compared with real experiments like 300K US dollars, the machine learning based prediction is much more cost-effective. Moreover, the machine learning models can provide highly accurate predictions with arbitrary attribute values which can nicely guide the high-potential attribute settings for real experiments. We believe machine learning approaches can greatly improve the efficiency and effectiveness of real experiments and bring a more environmentally friendly bio-material study ecosystem.

# 4 Conclusions

In this paper, we have developed machine learning algorithms to predict the elastic modulus of the scaffold at an orbitrary material composition. Likewise we have also predicted cell-material interaction and cell permeation within the scaffold at the chosen arbitrary cell culture parameters. These studies present the utility of this tool as an alternative to significantly reduce the cost and time to expedite the biomaterial and process optimization. Specifically, we have illustrated the use of ML in optimizing bone tissue engineering scaffold parameters. The experimental data Amini, et al. [2] was used to train four different ML models and predicted the performance values accurately.

# References

1. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.
2. A. Amini and S.Nukavarapu. Oxygen-tension controlled matrices for enhanced osteogenic cell survival and performance. Annals of Biomedical Engineering, 42:1261–1270, 2014.
3. D. Chen, S. Sarkar, J. Candia, S.J. Florczyk, S. Bodhak, M.K. Driscoll, C.G. Simon Jr., J.P. Dunkers, and W. Losert. Machine learning based methodology to identify cell shape phenotypes associated with microenvironmental cues. Biomaterials, 104:104–118, 2016.
4. I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. Deep learning, volume 1. MIT press Cambridge, 2016.
5. C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, and R. Ramprasad. Polymer genome: a data-powered polymer informatics platform for property predictions. The Journal of Physical Chemistry C, 122(31):17575–17585, 2018.

6. M.C. Lewis, B.D. Macarthur, J. Malda, G. Pettet, and C.P. Please. Heterogeneous proliferation within engineered cartilaginous tissue: the role of oxygen tension. Biotechnol Bioeng., 91(5):607–615, 2005.
7. F. Lia, J. Hana, T. Caob, W. Lamc, B. Fand, W. Tangd, S. Chena, K.L. Fokc, and L. Lia. Design of self-assembly dipeptide hydrogels and machine learning via their chemical features. PNAS, 116(23):11259– 11264, 2019.
8. P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov. Applications of deep learning in biomedicine. Molecular Pharmaceutics, 13:1445–1454, 2016.
9. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.
10. C. Rasmussen. Gaussian processes in machine learning. In Summer school on machine learning, pages 63–71. Springer, 2003.
11. B.D. Ratner. Biomaterials: Been there, done that, and evolving into the future. Annual Review of Biomedical Engineering, pages 171–191, 2019.
12. F. Ren, L. Ward, T. Williams, K.J. Laws, C. Wolverton, J. HattrickSimpers, and A. Mehta. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. Science Advances, 4, 2018.
13. F. Tourlomousis, C. Jia, T. Karydis1, A. Mershin, H. Wang, D.M. Kalyon, and R.C. Chang. Machine learning metrology of cell confinement in melt electrowritten threedimensional biomaterial substrates. Microsystems & Nanoengineering, 5(15), 2019.
14. N. Vyas, R.L. Sammons, O. Addison, H. Dehghani, and A.D. Walmsley. A quantitative method to measure biofilm removal efficiency from complex biomaterial surfaces using sem and image analysis. Scientific Reports, 2016.
15. Z. Wang, X. Xiao, and S. Rajasekaran. Novel and efficient randomized algorithms for feature selection. Big Data Mining and Analytics, 3(3):208–224, 2020.