Machine learning as a tool to design glasses with controlled dissolution for healthcare applications

Article *in* Acta Biomaterialia · February 2020

CITATIONS
0

READS
177

5 authors, including:

Nicholas Stone-Weiss
Rutgers, The State University of New Jersey
4 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Jie Huang
Missouri University of Science and Technology
104 PUBLICATIONS **881** CITATIONS
SEE PROFILE

Ashutosh Goel
Rutgers, The State University of New Jersey
100 PUBLICATIONS **1,538** CITATIONS
SEE PROFILE

Accepted for publication in Acta Biomaterialia

Machine learning as a tool to design glasses with controlled dissolution for healthcare applications

Taihao Han,[1] Nicholas Stone-Weiss,[2] Jie Huang,[3] Ashutosh Goel,[2,†] Aditya Kumar[1,†]

1. Department of Materials Science and Engineering, Missouri University of Science and Technology, Rolla, MO, USA 65409
2. Department of Materials Science and Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA 08854
3. Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO, USA 65409

[†] Corresponding author
A. Goel, Email: ag1179@soe.rutgers.edu; Ph: +1-848-445-4512
A. Kumar, Email: kumarad@mst.edu; Ph: +1-573-341-6994

Abstract

The advancement of glass science has played a pivotal role in enhancing the quality and length of human life. However, with an ever-increasing demand for glasses in a variety of healthcare applications – especially with controlled degradation rates – it is becoming difficult to design new glass compositions using conventional approaches. For example, it is difficult, if not impossible, to design new gene-activation bioactive glasses, with controlled release of functional ions tailored for specific patient states, using trial-and-error based approaches. Notwithstanding, it is possible to design new glasses with controlled release of functional ions by using artificial intelligence-based methods, for example, supervised machine learning (ML). In this paper, we present an ensemble ML model for reliable prediction of time- and composition-dependent dissolution behavior of a wide variety of oxide glasses relevant for various biomedical applications. A comprehensive database, comprising of over 1300 data-records consolidated from original glass dissolution experiments, has been used for training and subsequent testing of prediction performance of the ML model. Results demonstrate that the ensemble ML model can predict chemical degradation behavior of glasses in aqueous solutions over a wide range of pH relevant for their usage in a human body where the environment can be highly acidic (for example, pH = 3), for example, due to secretion of citric acid by osteoclasts, or highly alkaline (pH ≈10) due to the release of alkali cations from bioactive glasses. Outcomes of this study can be leveraged to design glasses with controlled dissolution behavior in various biological environments.

Keywords: biomedical; glass dissolution; ensemble machine learning; random forest; and additive regression

1. Introduction

Biomaterials are typically divided into the following three categories – (i) first-generation; (ii) second-generation; and (iii) third-generation biomaterials – principally based on their dissolution behavior in aqueous biological media (e.g., human blood plasma and saliva). Such interaction critically dictates a given biomaterial's overall behavior, and thus its function, in the human body [1]. The first-generation biomaterials were supposed to be bio-inert; as such, these materials were designed to exhibit high chemical durability (i.e., "practically insoluble," with little-to-no dissolution) when in contact with biological or non-biological fluids (e.g., when used in dental applications). The most common examples of first-generation biomaterials are metallic or ceramic (e.g., Al_2O_3 and ZrO_2) implants, Y_2O_3-Al_2O_3-SiO_2 based radioactive glass microspheres used for the treatment of hepatocarcinoma [2], and lithium disilicate based glass-ceramics, generally designed in the Li_2O-K_2O-Al_2O_3-SiO_2 glass-forming system, for application in dentistry [3]. The second-generation biomaterials were designed with moderate chemical durability with respect to their first-generation analogues. These materials were designed to interact – in a gradual manner – with the biological fluids and progressively degrade, thus resulting in mineralization (e.g., formation of hydroxyapatite layer on the glass surface) and bonding with tissue or bone. The first bioactive material ever produced was a glass designed in the Na_2O-CaO-SiO_2 ternary system, known as 45S5 Bioglass® [4]. Since its invention, several bioactive glass compositions with varying dissolution kinetics (when in contact with biological fluids) have been designed and reported in the scientific literature [5–13]. In the case of third-generation biomaterials, the synthetic material is expected to be bioactive and resorbable, such that, once implanted in the human body, the material would aid in the healing process. When in contact with physiological fluids, the bioactive glass is expected to deliver chemical conditions (i.e., ionic species and complexes) that commence a synchronized sequence of cellular level responses, ultimately resulting in expression of genes required for living tissue regeneration [14]. The most common examples of third-generation biomaterials are borate and borosilicate based glasses that are designed to undergo controlled chemical degradation (followed by resorption), resulting in a controlled release of functional ions [7,15,16].

It is evident from the abovementioned description that although the three generations of biomaterials may differ in their chemical make-ups and final applications, one common property that is crucial for their design is their dissolution behavior (chemical durability) upon contact with biological or non-biological fluids. Therefore, accurate prediction of aqueous dissolution of glasses – in relation to their pristine compositions – can facilitate the design of glasses for application in a broad spectrum of healthcare. However, predicting dissolution behavior of glass in an aqueous solution (solvent) is an arduous task owing to its dependence on a myriad of factors, thus giving rise to prodigiously large degrees of freedom. The aforesaid factors include, but are not limited to, bulk and surface chemistry of glass, pH and chemistry of the solvent, temperature, and the methodology used to study glass corrosion [17,18]. The two major challenges that impede the design and development of biomedical glasses with controlled dissolution behavior are: (i) lack of knowledge about underlying compositional and structural drivers that control the mechanism and kinetics of glass dissolution in aqueous solutions; and (ii) lack of reliable computational models to decipher the dominant compositional and structural descriptors which upon convergence with the experimental data on glass dissolution can be used to develop predictive *Quantitative Structure-Property Relationship (QSPR)* models. These knowledge-gaps have, in effect, necessitated the prevailing trend to design biomedical glasses — especially those that exhibit desired dissolution behavior — using semi-empirical, expensive, and time-consuming Edisonian approaches involving rigorous and iterative synthesis-testing cycles [19]. A notable example is 13-93B3 glass, which has primarily been derived from a well-known silicate-based bioactive glass 13-93 through *trial-and-error* based iterative cycles of glass syntheses (through partial replacements of SiO_2 with B_2O_3) and property-assessments, with a little rationale or revelation of the underlying glass-formation mechanisms [7,20].

A promising approach towards the design of original glasses for application in the healthcare industry is through the analysis of a large set of experimental data using artificial intelligence-based methods. The idea of using data-driven methods – such as supervised machine learning (ML), a tributary of artificial intelligence – for prediction and optimization of materials' properties (e.g., dissolution behavior) forms the premise of the United States Materials Genome Initiative [21,22]. In pursuit of the idea,

researchers from various scientific domains have compiled extensive datasets of materials and subsequently employed ML models to better understand and reveal the underlying Hume-Rothery-like QSPR models [23–26]. Knowledge of such relationships for any given material can be leveraged to promptly and reliably predict a material's properties or to design a new material that meets the desired set of performance criteria [27–32]. Accordingly, the work presented here is an attempt to unify the powers of experimental materials science and artificial intelligence to design glasses with tailored dissolution behavior for their potential applications in healthcare.

2. Approach to predict the dissolution of glasses using machine learning

In the context of glasses, ML is highly promising owing to its unique abilities: (i) to reveal – in an explicit manner – the influential factors that dictate properties (e.g., dissolution behavior) of the material even when the underlying structure-property correlations are highly nonlinear and non-monotonic; (ii) to leverage the knowledge of structure-property relationships to make predictions in previously untrained compositional domains; and (iii) to determine optimum composition of the material that is expected to deliver desired properties. Indeed, past studies – albeit a few – have shown that ML models, when trained using a sufficiently large and diverse database, can reliably predict various properties of glasses, including dissolution behavior [27,33,34], mechanical properties [35], glass-forming ability [36], and crystalline-to-amorphous phase transition temperatures [37,38]. However, when using ML as a predictive tool to design biomedical glasses with disparate dissolution behavior, the problem is two-fold. The first major challenge is the unavailability of comprehensive – yet consistent – experimental datasets of biomedical glass dissolution in scientific literature. This may be attributed to several reasons, majority of which originate from inconsistencies in synthesis and processing conditions of glasses further aggravated by variations in methods of characterization of their properties as has been discussed in our previous paper [15]. The second challenge is the choice of an appropriate ML model that is capable of handling highly nonlinear and non-monotonic *cause-effect* relationships. The majority of prior studies have used artificial neural networks (ANN), support vector machine (SVM), and decision trees (DT) based ML models to predict glass'

properties [33,34]. It is pertinent to point out that ANN and SVM – while often considered reliable – could falter at making accurate predictions in data-domains that feature highly nonlinear (e.g., periodic) and/or highly non-monotonic functional relationship between the output and inputs [39–42] – as is expected to be the case with structure-property relations in glasses. This is because ANN and SVM are premised on local optimization-and-search algorithms (e.g., the back-propagation mechanism that is used in neural network-based ML models for optimization of activation functions' parameters), which significantly augment the prospect of convergence at (or around) a local minimum as opposed to the global minimum. Because of this drawback, ANN and SVM models could produce disparate – even inferior – prediction performances as they are progressively trained with new datasets [39–42]. The other commonly-used family of ML models, that is, the DT models (e.g., the M5Prime model [43] and the classification-and-regression-trees model [44]), split the overall training database into multiple smaller ones in a manner that the input-output relationship within every small subset is linear (i.e., described by multivariate linear functions) [45–47]. Since this 'flowchart-like' structure of the DT model makes it efficient (in terms of computational complexity) and simple-to-use, its prediction accuracy is substantially relegated when the actual input-output correlation within the training database is highly nonlinear [42]. In recent studies [33,35,42], it has been shown that the random forests (RF) model, based on modification of the bootstrap aggregation decision tree algorithm, outperforms several standalone ML models in terms of prediction of materials' properties. These studies have attributed the RF model's superior prediction performance to its unique ability to process variables – regardless of whether they are discrete or non-discrete (continuous) – over data-domains featuring various permutations and combinations of complexities [e.g., (non)monotonicity and (non)linearity] [48]. In spite of these merits, developing an RF model for a progressively growing database – for example, when new data-records are periodically amended to an existing database – can be challenging [49,50]. This is because the RF model comprises of two hyper-parameters (i.e., number of trees in the forest, and number of leaves per tree) that need to be optimized (by the user) in relation to nature as well as volume of the database. Typically, such optimizations are carried out using trial-and-error based approaches or using a multi-fold cross-validation method [51], which can be time-consuming and result in

issues like high bias (underfitting) or high variance (overfitting) [24,49,50]. In recent studies [42,52], it has been shown that this deficiency of the RF model can be overcome – and its prediction performance can be boosted – by hybridizing it with an optimization algorithm (e.g., the firefly algorithm). The optimized algorithm is able to vary-and-optimize the aforementioned hyper-parameters of the RF model autonomously (i.e., without any user intervention) in relation to characteristics of the training database. This not only improves the prediction performance of the RF model but also facilitates its adaptation (through adjustments in values of the RF model's hyper-parameters) when it is re-trained using newer or larger training databases. The standalone RF model has previously been used to predict the dissolution behavior of silicate glasses [27,33,34]. However, to the best of the authors' knowledge, the RF model – or any other RF-based ML model, for that matter – have never been used for biomedical glasses. Furthermore, in prior studies [27,33,34], ML models have been used for prediction of dissolution kinetics exclusively during stage I of glass corrosion.[‡] This is good in the cases where the controlled release of functional ions is important for specific biological applications, for example, release of Ag^{+} from the glass to promote antibacterial activity, or Cu^{2+} to promote angiogenic behavior [16]. However, in the cases where long-term performance of glasses is important, characterization and understanding of stage II (residual rate regime) of glass dissolution along with stage I is of great practical significance. As an example, during stage II of dissolution of bioactive glasses, the precipitation of amorphous calcium phosphate on the silica–rich gel layer and its slow transformation of crystalline hydroxyapatite occurs, which further interacts with collagen fibrils of damaged bone to form a bond [53]. Nevertheless, experimental measurements of stage II dissolution of biomedical glasses are scarce in literature. Furthermore, prediction of long-term glass dissolution behavior is difficult because of the complexities that arise from large changes in surface area of glass particulates and solvent chemistry (i.e., pH, ionic strength, and under-saturation) with respect to their

[‡] When a glass is exposed to dilute solution, the process begins by corrosion of the glass surface with no solution feedback at an initial rate of r_0 (sometimes called the forward rate). This period, termed Stage I, produces a rate that is maximum for the given temperature and pH; and is controlled by the hydrolysis of network forming species. As the concentration of glass formers in solution increases, the driving force for dissolution decreases, and alteration layers begin to form. The change in mechanistic drivers leads to a nearly constant and slow residual rate (r_r) during a period termed Stage II.

initial values. Given the superior prediction performance of the hybridized RF model, in terms of predicting physiochemical behavior of complex material systems [42,52], it is deemed important to examine if the same (or a similar) model would be able to produce high-fidelity predictions of composition-dependent dissolution behavior of biomedical glasses, not just during stage I but also through stage II.

To overcome the aforesaid challenges, in the past five years, the authors have collected a comprehensive experimental dataset on the dissolution of oxide glasses over a broad compositional space encompassing the Na_2O-B_2O_3-SiO_2, Na_2O-Al_2O_3-B_2O_3-SiO_2, Na_2O-B_2O_3-P_2O_5-SiO_2, and Na_2O-Al_2O_3-B_2O_3 systems. The choice of glass composition systems has been made based on their chemical and structural diversity along with their proposed potential applications in healthcare. As examples: aluminosilicate, aluminoborate, and aluminoborosilicate glasses are used as glass-ionomer cements [54–57] ; borate, aluminoborate, and borosilicate based bioactive glasses are used in fields of hard and soft tissue engineering [7,15,58–60]; and borate and aluminoborate glasses are used for radiation synovectomy treatment of rheumatoid arthritis [61,62]. Emphasis has been given to exhaustively, and in a consistent manner, account for the influential factors – for example, time, initial pH of solvent, and temperature – that affect the dissolution behavior of glasses. The aim is to use the experimental data along with an ensemble ML model – developed by combining the random forests (RF) model with an optimization technique, that is, additive regression (AR) – to predict the time-dependent dissolution (measured in terms of surface area normalized critical ion mass loss) of the glasses in relation to their compositions and influential physiochemical properties (e.g., temperature) of the system. The performance of the ensemble ML model is benchmarked – on the basis of five statistical parameters as well as a singular composite performance index (CPI) – against six commonly-used standalone ML models [i.e., linear regression (LR); elastic net regression (ENR); Gaussian Process Regression (GPR); multilayer perceptron ANN (MLP-ANN); SVM; and RF]. On the basis of such comparisons, it is shown that the hybrid ML model is able to produce high-fidelity predictions of glass dissolution while outperforming all standalone ML models in terms of prediction performance.

3. Machine learning

Section 3.1, below, presents a concise overview of six standalone models (MLP-ANN; SVM; LR; ENR; GPR; and RF) that have been used in this study. **Section 3.2** briefly describes the ensemble ML model (AR-RF), developed by hybridizing the RF model with an optimization scheme (i.e., AR: additive regression). Further details (e.g., underlying algorithms) pertaining to the ML models are provided in **Supplementary Information**.

3.1. Overview of standalone machine learning models

Artificial neural network – abbreviated as ANN – consists of multiple neurons arranged in hierarchical layers. Each of the neurons serves as a computational element and is responsible for processing information relayed from the previous layer of neurons (using sigmoidal or logistic-transfer activation functions) and transmitting the processed information to the next layer of neurons [63]. The structure of ANN resembles the network of interconnected neurons within the human brain – wherein information is processed (and simplified) and transmitted from one layer to another in a hierarchical manner. Multilayer perceptron artificial neural network (MLP-ANN) is a subclass of ANN with multiple layers of neurons, and, therefore, strong self-learning capabilities [64]. Support vector machine (SVM), a commonly used ML model for both classification and regression purposes, approximates the correlation – either in the form of multivariate linear or nonlinear functions – between inputs and output of a dataset. This is accomplished by employing an optimization scheme – as opposed to a regression approach – geared towards minimizing an objective cost function (i.e., ε-insensitive loss function), or simply put, to transform input data into a higher-dimensional structure such that data with similar characteristics are sequestered from dissimilar ones [65]. Linear regression (LR) is a simple ML technique that uses piecewise linear functions – driven by independent predictors – to predict a numerical target based on a set of independent inputs [66]. Elastic net regression (ENR) is a modified form of linear regression. The ENR model was originally developed [67] to improve the prediction performance of the LR model, especially when applied to datasets pertaining to heterogeneous systems with large number of variables. The improvement involves identifying significant

variables and then assigning larger coefficients to them as compared to less significant variables. If a given variable is deemed inconsequential, its coefficient is reduced to 0. Gaussian process regression (GPR) is a regression algorithm based on the Bayes' theorem, which identifies the most probable outcome (or hypothesis) on the basis of prior knowledge acquired from the training database [68]. The GPR model employs a stochastic process to collect random variables, any finite number of which have a joint Gaussian distribution [69]. The variables – which represent prior knowledge – are used to estimate the probability of a given outcome, and compare it against the probabilities of all possible outcomes. Through such comparisons, the prior knowledge is updated iteratively throughout the training process; ultimately, the outcome with the highest probability is selected as the final prediction. Random forests (RF) is a modification of [decision tree algorithm + bootstrap aggregation], premised on amalgamation of *bagging* and *adaptive nearest neighbors* to achieve logic-based inference of input-output correlations in a dataset. RF employs two-stage randomization to grow a large number of uncorrelated "deep" trees, all without any pruning or smoothening (unlike conventional decision trees-based models, which do require pruning during the training process) [70,71].

The authors would like to point out that all standalone ML models described above comprise of hyper-parameters that needs to be adjusted by the user to improve their prediction performance. In some models, user intervention is also required to select optimal functions (e.g., type of kernel function used for transformation of dimensionality of data in SVM; and type of activation function used for activation of neurons in MLP-ANN). In this study, for the selection of optimal functions and hyper-parameters for each ML model, 10-fold cross-validation (CV) method [47,51,72] was used as the primary performance-assessment method. In short, the 10-fold CV method randomly splits the training database into 10 equisized folds. The ML model – the hyper-parameters/functions of which need to be optimized – is trained using data-records from 9 folds, and subsequently blind-tested against data-records in the 10^{th} fold. This process is iteratively repeated 9 times – each time using a unique combination of folds for training of the ML model and its blind-testing. During each iteration, the relevant parameters and functions of the ML models are fine-tuned such that prediction errors [measured in terms of root mean squared error (RMSE)] are

progressively minimized. Functions and parameters' values after the last iteration are selected as "optimum," and used – without further changes – for subsequent testing of the ML model against a blind test dataset.

3.2. The ensemble machine learning model

In this study, an ensemble ML model (AR-RF) was developed by hybridizing the RF model with the additive regression (AR) technique. AR is an optimization algorithm based on the gradient boosting approach [73]. With the ensemble model, AR was paired with the RF model to semi-autonomously determine optimum values of the two hyper-parameters of the RF model (i.e., n_t and n_{LV}, representing number of trees and number of leaves per tree, respectively), thereby improving its prediction performance. The ensemble ML model's structure constitutes three separate stages of processing of the database.

- In *stage A*, the RF model is employed, wherein the values of n_t and n_{LV} are set at 450 and 5 (values determined as optimum by the 10-fold CV method), respectively. An objective function, F, is defined. The objective function is a numerical value equivalent to the total RMSE of RF model's predictions, as estimated at the end of *stage A*. The "deep" unpruned trees constructed by the RF model are retained.
- In *stage B*, the AR technique is implemented. The residuals of the predictions (i.e., RMSE) are used to construct a second set of trees, along with the original "deep" trees. The objective here is to train the second set of trees – by varying-and-optimizing the values of two hyper-parameters of the RF model (n_t and n_{LV}) – to fit the residuals such as that the overall training error is reduced. This tandem between the aforementioned pair of steps – of performing predictions using the RF model, and subsequently refining the prediction performance by fitting the residuals – is repeated over several iterations until convergence is reached, that is, reduction in training error is $< 10^{-6}$ units for 3 successive iterations.
- Finally, in *stage C*, the RF model is employed to make predictions against new test dataset(s). For such predictions, the AR-determined optimum values of n_t and n_{LV} are used without any further changes.

4. Experimental section

4.1. Glass compositional design

Alkali borosilicate glasses: Figure 1a presents the batched and analyzed compositions of the sodium borosilicate glasses highlighted in the $Na_2O–B_2O_3–SiO_2$ ternary diagram. This composition-space was chosen so as to provide a large range of homogenous glasses in the Na_2O-B_2O_3-SiO_2 system, encompassing both sodium silicate and sodium borate endpoints as well as glasses of intermediate compositions. The details about the composition, processing, and dissolution behavior of the glasses can be found elsewhere [15].

Alkali aluminoborate glasses: Figure 1b presents the ternary diagram of the investigated glasses, overlaying target, and measured compositions, as well as glass-forming range. The glasses were synthesized using the melt-quench technique. Overall, the glasses encompass a broad compositional domain including peralkaline (Na/Al>1), metaluminous (Na/Al=1), and per-aluminous (Na/Al<1) regions of the Na_2O-Al_2O_3-B_2O_3 ternary system. The details about the synthesis, processing and dissolution behavior of glasses have been published elsewhere [74].

Alkali aluminoborosilicate glasses: The glasses in the system $Na_2O–Al_2O_3–B_2O_3–SiO_2$ were synthesized using the melt-quench technique. These glasses cover a broad composition space in the Na_2O-Al_2O_3-B_2O_3-SiO_2 system, extending from well-studied sodium aluminosilicate compositions to less-studied B_2O_3-rich compositions. The details about the composition, processing, and dissolution behavior of the glasses can be found elsewhere [75].

Alkali borophosphosilicate glasses: The glasses in the system $Na_2O–B_2O_3–P_2O_5–SiO_2$ were designed in the per-alkaline (Na/B >1), sub-boric (Na/B = 1), and meta-boric (Na/B <1) regimes. Accordingly, three P_2O_5–free sodium borosilicate glasses, as plotted in **Figure 1a**, were chosen as the baseline compositions. Further, P_2O_5 was incorporated in the glasses in a manner that the Na/B molar ratios of the baseline compositions remain constant across the composition walk. The details of the investigated glass compositions (mol.%) are presented below.

(i) Glasses with Na/B >1: $x\ P_2O_5 – (100\text{-}x)\ (25Na_2O\text{-}20B_2O_3\text{-}45SiO_2)$

(ii) Glasses with Na/B =1: $x\ P_2O_5 - (100\text{-}x)\ (25Na_2O\text{-}25\ B_2O_3\text{-}50\ SiO_2)$

(iii) Glasses with Na/B <1: $x\ P_2O_5 - (100\text{-}x)\ (25Na_2O\text{-}30B_2O_3\text{-}45SiO_2)$

where, x varies between 0 – 5 mol.%. The glasses were synthesized using the melt-quench technique as per the methodology described in our previous article [15]. The details about the dissolution behavior of glasses will be published in our forthcoming article.

4.2. Experimental dataset of glass dissolution

The dissolution tests were performed on glass particles with their particle size varying between 300 – 425 μm. A detailed methodology of sample preparation, and surface area analysis (using ImageJ software) has been published in our previous article [15]. The dissolution behavior of glasses was studied by submerging 20 – 60 mg of glass particles in 30 – 50 mL of the selected aqueous solutions (corresponding to SA/V [i.e., the ratio of surface area (SA) of glass to volume (V) of solution] of 2 – 7.5 m^{-1}; SA/V was held constant in each set of experiments as can be seen in **Table 1**). Acidic pH experiments (pH 0-to-4), were performed in HCl (1 M, 0.01 M, and 0.0001 M HCl for pH 0, 2, and 4, respectively). Near-neutral pH experiments (pH 5-9) were performed in either DI water or 0.1 M Tris-HCl/Tris-HNO_3 buffer solutions. Tris-HCl and Tris-HNO_3 solutions were prepared by adding Tris(hydroxymethyl)aminomethane to DI water and adjusting the pH to the desired value by adding appropriate amounts of 1 M HCl or 1 M HNO_3. Basic solutions (pH 10-to-13) were prepared by diluting concentrated tetramethylammonium hydroxide to 0.0001 M and 0.1 M solutions with DI water for pH 10 and pH 13, respectively. All powder–solution mixtures were immediately sealed into sterilized polypropylene flasks (static media; aqueous solution was not replenished with time) and placed in an oven at temperatures ranging from 35 – 65 ºC. The dissolution experiments ranged from as little as 2 minutes to 120 days, depending on the glass composition and experimental conditions used. A summary of the dissolution protocols for each system studied can be found in **Table 1**. In addition to analyses of neat (unused) and blank (glass-free) control solutions, all the experiments were performed in triplicates to evaluate uncertainty of final results. The pH evolution of

solution during glass corrosion was measured at room temperature using a pH meter (Mettler Toledo InLab® Pro-ISM). The liquid aliquots of the corrosion solutions were filtered and acidified (with HNO_3) before being chemically analyzed by either inductively coupled plasma – optical emission spectroscopy (ICP-OES, PerkinElmer 7300 or 8300) or inductively coupled plasma – mass spectroscopy (ICP-MS, Agilent 7700s), depending upon expected elemental concentration levels (where ICP-MS has detection limits near <0.02-0.05 and ICP-OES has detection limits near <0.1-0.2 ppm; each of the detection limits listed are element-specific). The normalized loss (NL) of each element (Na, Al, B, P, and Si) released from glasses into the surrounding solution was calculated using **Eq. 1**,

$$NL_i = \frac{C_i - C_o}{\left(\frac{SA}{V}\right) f_i} \quad \textbf{(1)}$$

where C_i is the concentration of element i in the solution as detected by ICP-OES or ICP-MS; f_i is the mass fraction of the element i in the glass; and C_o is the background concentration (as determined from blank solutions). When the normalized loss data is plotted against time and linearly fit, the approximate forward dissolution rate can be estimated.

5. Database Development and Assessment of Prediction Accuracy of ML Models

5.1. Development of database

Datasets collected from the glass dissolution studies (**section 4.2**) were consolidated into a singular database, and then used for training the ML models (**section 3.0**; and **Supplementary Information)** and subsequently for evaluation of the models' abilities to predict time-dependent glass dissolution – assessed in terms of critical ion (CI) normalized mass loss – in previously untrained compositional domains. The database consisted of 1364 unique data-records, wherein each data-record had 10 inputs and 1 output. The inputs included pertinent physiochemical properties of the [glass + solution] system: initial pH of the solution; elemental composition of the glasses (in mass fraction); temperature (in °C) at which the dissolution studies were carried out; specific surface area (in $m^2.g^{-1}$) of the glass; volume of the solution (in cm^3); and time (in min). It is acknowledged that there are additional parameters relevant to solution

chemistry (e.g., ionic strength, and type of ionic species) that could potentially affect dissolution behavior of the glasses (although not as appreciably as solution pH does [76]). However, as these parameters were not consistently available for all experiments, they were not included in the database. The output parameter included the extent of glass dissolution, expressed in the form of CI normalized mass loss (g. m^{-2}) of the glass. For glasses consisting of boron (B), B was considered as a critical ion; and for selected glasses (i.e., only 98 out of 1364) that were devoid of any boron, silicon (Si) was chosen as the critical ion. Here it should be noted that for the 98 glasses devoid of boron, the alteration rate calculation from silicon release during Stage II probably underestimates the rate values because Si is not a tracer (unlike boron). Statistical parameters pertaining to the database are shown in **Table 2**.

5.2. Assessment of prediction accuracy of ML models

For the purposes of training, and to assess the prediction performance, of ML models, the experimental database (described in **section 5.1**) was randomly apportioned into two subsets: one for training and the other for testing. 75% of data-records of the parent database (i.e., training set) were used for training of the ML models (i.e., optimization of functions and values of hyper-parameters); and the remaining 25% of the data-records (i.e., testing set) were utilized for assessment of prediction performance of the models. Various past studies [42,47,77,78] have also used such split of 75%-to-25% in the parent database for training and testing of ML models as such splitting permits not only adequate optimization of the ML models functions and parameters but also the ability to test prediction performances of the ML models over previously untrained data-records. It is clarified that while splitting of the parent database was done in a randomized manner, care was taken to ensure that the training dataset was archetypal – albeit a shortened version – of the parent database. To this end, the training set was formulated in manner that each input variable spanned over a wide range between (and excluding) its minimum and maximum values in the parent database. Such formulation of the training dataset is important to ensure that, during training, the ML models are able to learn composition-dissolution behavior links in the glasses not just during stage I (i.e., within minutes of dissolution) but also during the later stages (i.e., after several hours of dissolution).

To quantitatively measure prediction performance of the ML models (against the testing set), 5 distinct statistical parameters were extracted through comparisons of the models' predictions against actual measurements. These parameters include: coefficient of determination (R^2); root mean squared error (RMSE); Person correlation coefficient (R); mean absolute percentage error (MAPE); and mean absolute error (MAE). Equations used to calculate the aforesaid parameters can be found in [42].

$$CPI = \frac{1}{N}\sum_{j=1}^{j=N}\frac{P_j - P_{min,j}}{P_{max,j} - P_{min,j}} \quad \textbf{(2)}$$

The five statistical parameters, described above, were integrated into the composite performance index (CPI, see **Eq. 2**) [42,47,79], to obtain a singular, unified measure of prediction performance of each ML model. In **Eq. 2**: N (=5) is the total number of statistical parameters used to measure performance of the models; P_j is magnitude of the j^{th} statistical parameter; and $P_{j,\,min}$ and $P_{j,max}$ are the minimum (i.e., worst) and maximum (i.e., best) values of the j^{th} statistical parameter. With **Eq. 2** formulated in the manner described here, CPI of any given ML model can vary between 0-and-1. The best ML model would acquire a CPI value of 0 (or the lowest value) and the worst ML model would obtain a value of 1 (or the highest value). Therefore, on the sole the basis of CPI values – which accounts for all five statistical parameters (i.e., performance measures) – the ML models can be ranked (from best to worst) in terms of their prediction performances.

6. Results and Discussion

6.1. Prediction of Glass Dissolution Behavior

In typical regression-oriented ML models, the overall quality of the model is assessed in terms of two aspects: (i) capability of the model to uncover underlying *cause-effect* correlations from the training set; and (ii) to apply the learned knowledge to perform predictions in previously unseen data-domains (i.e., new sets of inputs). To put it simply, prediction performance of a ML model distills down to its capability of identifying one or more trends in the dataset linking the input variables with the output, and subsequently using such trends for interpolation – and even extrapolation, in some cases – in blank data-domains. To

assess the accuracy of predictions produced by ML models – and to order them on the basis of their prediction performances – an experimental database, comprised of glasses' compositions, experimental process variables (i.e., SA/V, initial pH of the solvent, and temperature), and time-dependent dissolution behavior of the glasses (described in **section 5.1**) was used. Readers are reminded that, in both the training and testing datasets, each data-record featured 10 inputs and 1 output (i.e., normalized CI mass loss of the glass). Predictions of dissolution of glasses belonging to the testing set, as yielded by the ML models, are compared against experimental measurements in **Figures 2-5**; for reference, predictions of dissolution of glasses from the training set are also included. Statistical parameters pertaining to prediction of glasses from the testing set are itemized in **Table 3**; here, parameters pertaining to glasses from the training set are not included as their magnitudes are similar to their counterparts from the testing set.

As can be seen in **Table 3**, predictions of glass dissolution, produced by the ML models, range from poor to excellent, with the Pearson correlation coefficient (R) ranging from 0.39 to 0.99 and the root mean squared error ($RMSE$) ranging from 11.2 to 75.2 g. m^{-2}. If we consider the maximum absolute percentage error *(MAPE)* of the best performing ML model, the margin of error in predictions of glass dissolution is ± 9 g m^{-2}. This is excellent considering that even in experimental measurements of glass dissolution, typical standard deviations are within the same order of magnitude as 10 g m^{-2} [80]. On the basis of the values of CPI – the integrated measure of prediction performance – the ML models can be ranked as ensemble ML > standalone RF > SVM > MLP-ANN > LR > GPR > ENR.

The ENR model had the poorest prediction performance of all ML models implemented in this study. As can be seen in **Figure 2a**, across the 342 cases used for evaluation of the model's prediction performance, majority of predictions exhibited significant departures from the actual values. LR and GPR models (**Figure 2b-c**) produced superior prediction performances compared to the ENR model, albeit significantly lower than the SVM, MLP-ANN, and RF-based models. Such poor prediction performances of the ENR, LR, GPR models are not surprising because the former two models use multivariate linear functions and the latter uses Gaussian distribution functions over datasets that – in all likelihood – feature far more complex correlations between the input variables and the output. It is worth pointing out that non-

linear and non-monotonic composition-properties correlations are not unique to biomedical glasses; glasses of other families (e.g., silicates) also feature such complex correlations. Indeed, in a recent study [33], it was shown that ML models based on linear (LR) and Gaussian distribution (ENR) functions – even when trained and optimized in a comprehensive manner (using a multi-fold CV method, as used in this study) – failed to capture the inherently nonlinear relationship between dissolution rate of silicate glasses and the pH of the contacting solution. The models' prediction performances deteriorated even further when, in addition to pH, other influential variables of the system (e.g., composition of silicate glasses) were varied [33]. Therefore, based on the above discussion, it can be said that for prediction of properties of any given material, it is crucial to select ML models founded on mathematical formulations that are in alignment with the material's elementary composition-properties relationship. In the specific case of glass, it is important to select ML models premised on nonlinear mathematical formulations; otherwise, the models' predictions are rendered inaccurate, thereby making them unsuitable for blind predictions as well as for optimizations.

As shown in **Figure 3**, prediction performances (assessed in terms of R^2) of MLP-ANN and SVM models were similar, and superior compared to LR, GPR, and ENR models (**Table 3**). As the MLP-ANN model employs a series of nonlinear logistic-transfer functions – as activation functions for the neurons – to develop input-output correlations, it is not surprising that its prediction performance supersedes those of models constructed using linear functions. Similarly, SVM, on account of its classification-and-optimization-based structure and use of nonlinear kernel function (to develop hyperplanes for data classification), is proficient at handling nonlinear input-output correlations, thus making it suitable for prediction of composition-dependent dissolution behavior of glasses. As previously stated in **section 2.0**, both MLP-ANN and SVM models utilize local search-and-optimization mechanism during their training [39–42]. While this mechanism generally leads to faster convergence and reliable detection of extrema in the database, it endures a characteristic disadvantage of converging to a (one of several) local minimum rather than the global minimum. This drawback is often inconsequential – with little-to-no effect on prediction performance – in datasets wherein the functional relationship between the input variables and output is broadly linear and/or monotonic [41,42]. However, in the case of glass dissolution, the input-

output correlations are expected to be complex (presumably highly nonlinear). This could, potentially, render the predictions of MLP-ANN and SVM models inaccurate – especially in comparison to ML models that are better at handing highly nonlinear input-output correlations within the database (e.g., the RF model; see **Table 3**). In this study, the hyper-parameters (i.e., MLP-ANN: number of hidden layers and number of neurons per hidden layer; and SVM: kernel function and parameters) of the models were rigorously optimized through the 10-fold CV method (see **section 3.1**). On account of this optimization, it is expected that the inherent drawbacks of the MLP-ANN and SVM models were overcome – at least partially – thereby allowing them to produce predictions with superior accuracy compared to other ML models that were based on linear or Gaussian distribution functions (see **Table 3**). In the context of MLP-ANN and SVM models, it is important to point out that several past studies have posited that the models' prediction performances can be bolstered by using: Genetic programing [81,82]; or bootstrap aggregation based methods (e.g., boosting) [45,47]; or by using theoretically-guided constraints (e.g., to eliminate negative, thus unrealistic, values of CI normalized mass loss that were produced as outputs of ML models in **Figure 3**). While, in this study, these aspects – of improving prediction performance of MLP-ANN and SVM models – have not been examined, the performance of an ensemble ML model, wherein the RF model is hybridized with a gradient boosting-based AR technique, has been examined and benchmarked against the standalone RF model (described below).

On the basis of statistical parameters listed in **Tables 3,** and results shown in **Figure 4**, it is clear that in terms of prediction accuracy, the RF model is superior compared to all standalone ML models that were used in this study. This result is not surprising – and, in fact, in very good agreement with prior studies [33,34,42,78] that have also reported that the prediction performance of RF model often supersedes those of several standalone ML models. Superior prediction performance of the RF model is attributed to its structure – which comprises of a large number of "deep" trees that are grown without any smoothening or pruning. The unpruned "deep" trees allow data in the training set to be split in a logical manner, which, in turn, not only reduces generalization errors but also serves to mitigate overfitting (high bias) of the training data. Furthermore, the two-stage randomization employed in the RF model (see **Supplementary**

Information for more details) diminishes correlation among the unpruned trees, thus reducing variation (underfitting) and ensuring homogeneity among the data represented in each of the tree-nodes. It is interesting to note that prediction performance of the RF model was bolstered when it was hybridized with the AR technique. This is better shown in **Figure 4b** and **Table 3**. As can be seen, the ensemble (AR-RF) model reliably captured the functional relationship between the input variables and output, and was able to leverage the learned relationship to reliably interpolate in previously untrained data-domains. Such improvement in prediction accuracy of the ensemble model – vis-à-vis the RF model – is elicited by the inclusion of AR, which iteratively varies-and-optimizes the two hyper-parameters of the RF model (i.e., number of trees and leaves per tree) until the deviation between actual values and model's predictions reaches its ultimate minimum. It is also interesting to note in **Table 3** that compared to the standalone RF model, prediction performance of the ensemble model is better, albeit only slightly. The implication of the aforesaid slight improvement – brought about by incorporation of AR – is that the application of AR is not particularly beneficial when the training and testing datasets comprise of only tens or hundreds of data-records. Conversely, when the parent database is large – for example, the database (with > 50000 data-records) that was used in a prior study [78] – the application of AR is expressly advantageous as it could elicit substantial improvements in prediction performance of the RF model.

Finally, based on our extensive literature review, we found that prediction errors of the ensemble ML model (measured in terms of R^2 or RMSE) presented in this study were consistently and substantially lower than those reported in prior studies that were focused on prediction of glass dissolution [33,34]. It is, however, acknowledged that RMSE and R^2, by themselves, do not conclusively prove that the union of RF and AR will dependably produce more accurate predictions compared to other ML models. This is because the overall prediction performance of any given ML model is affected by various factors, and, therefore, it is difficult – if not impossible – to compare (or rank) different ML models, especially when they are devised and operated by different users. The aforementioned factors include, but are not limited to: (i) nature and volume of the parent database, and its splitting into training and testing sets; (ii) pre-processing (or lack thereof) of the parent database, or datasets derived from it (i.e., training and testing sets); (iii) type and

number of statistical parameters (e.g., R^2 and RMSE) used for assessment of prediction performance; and (iv) techniques used to optimize the hyper-parameters of the ML models. In spite of the abovementioned factors, the high values of R and R^2 – combined with the low values of RMSE, MAE, and MAPE (see **Table 3**) – strongly suggest that the ensemble AR-RF model is a reliable tool for predictions of glass dissolution, especially glasses of the biomedical family, in relation to physical and chemical characteristics of the [glass + solvent] system. Given the model's excellent prediction performance, it could be said that with further improvements – for example, by using a larger and more diverse database for training – the model can potentially be used for performing prediction in reverse, that is, optimization: determination of optimal glass compositions and/or experimental process parameters that lead to desired dissolution behavior.

6.2. Predicting the dissolution behavior of unknown glass compositions

Results shown in **section 6.1**, and the accompanying discussions, show that the ensemble ML model – comprised of AR and RF models – is able to reliably predict the dissolution behavior of glasses in relation to their composition and relevant physiochemical characteristics of the system (e.g., SA/V and pH). In this section, the training of the ensemble ML model (on the basis of which the model learned the functional correlation(s) between input variables and the output of the database), and its ability to reliably predict dissolution of glasses were leveraged to sketch generic composition-reactivity trends in two distinct families of biomedical glasses. The first family of glasses selected for the analyses was $(Na_2O)_{0.25}(Al_2O_3)_{0.25}(B_2O_3)_x(SiO_2)_{0.5-x}$ glasses. For these glasses, the SA/V and temperature were fixed at 2.5 or 7.5 m^{-1} and 35 ^oC, respectively. Next, the ensemble ML model was employed to estimate the extent of dissolution (i.e., CI normalized mass loss at different times) of the glasses, at different reaction times, in relation to the glasses' original composition (i.e., B_2O_3 content, or the value of x) and pH of the solvent (i.e., pH = 0 and pH = 2). The second family of glasses selected for the analyses was $(Na_2O)_{0.8-x}(Al_2O_3)_{0.20}(B_2O_3)_x$ glasses. For these glasses, the SA/V, temperature, and solvent pH were fixed at 5 m^{-1}, 65 ^oC, and 7, respectively. The ensemble ML model was then used to estimate the extent of dissolution of

the glasses – at different times – in relation to their composition (i.e., B_2O_3 content, or the value of *x*). Composition-reactivity trends generated by the ensemble ML model are shown in **Figure 5**. In addition to the ML-predicted trends, experimentally-obtained values are added to the plots for the purposes of benchmarking.

In the case of $(Na_2O)_{0.25}(Al_2O_3)_{0.25}(B_2O_3)_x(SiO_2)_{0.5-x}$ glasses, at pH = 0, the ML model predicts non-monotonic variation in the extent of glass dissolution with respect to increasing B_2O_3 content in the glass (**Figure 5a**). Conversely, at pH = 2, the model predicts monotonic increase in extent of glass dissolution with increasing B_2O_3 content (**Figure 5b**). In $(Na_2O)_{0.8-x}Al_2O_3)_{0.20}(B_2O_3)_x$ glasses, at shorter reaction times (i.e., at t = 1 h), the extent of glass dissolution decreases with increasing B_2O_3 content. At later ages (at $t \geq$ 12 h), the extent of glass dissolution is broadly unchanged with respect to B_2O_3 content. As can be seen in **Figure 5**, the experimentally-obtained data points are in good agreement with the ML-predicted trends, thus validating the ensemble ML model's ability to predict glass dissolution behavior in new, previously untrained, compositional domains. It is worth pointing out that the trends shown in **Figure 5** – which are essentially obtained via interpolation – are not smooth; they feature sharp changes in curvature (indicating non-differentiable functional relationship between dissolution and composition) between successive experimentally-measured data points. The abrupt curvature changes are attributed partly to the intrinsic limitation of the RF model (i.e., inability to produce smooth inputs-output correlations), and primarily to a specific limitation in the database: its diversity. More specifically, the database used for training of the ensemble ML – albeit reasonably large (> 1300 distinct data-records) – features limited variations in the inputs (e.g., pH of the solvent, temperature, and SA/V and composition of the glass). For instance, in the case of $(Na_2O)_{0.25}(Al_2O_3)_{0.25}(B_2O_3)_x(SiO_2)_{0.5-x}$ glasses, the database includes only 11 different compositions, 2 different SA/V, and a singular temperature. Given the ensemble ML model's excellent prediction performance – as demonstrated in **Figures 4** and **5** – it could be said that with further diversification of the database, the model can potentially be used for high-fidelity predictions (i.e., interpolations as well as some extrapolations) of glass dissolution behavior. A larger and more diverse database will also enable the ability to perform prediction in reverse, that is, optimization: determination of optimal biomedical glass

compositions, and the accompanying experimental process parameters (e.g., SA/V), that lead to desired dissolution behavior (or controlled release of ions).

7. Implications on the design of new glasses for healthcare

The kinetics and mechanism of aqueous corrosion are two aspects that are critical to the design of any glass/ceramic based biomaterial, irrespective of its application. This is because glass, and its derivatives, are used in myriad applications in the human body – starting from mouth where they interact with alcoholic and non-alcoholic (non-dairy) drinks with their pH varying between 2 – 7 [83] to their application as bioactive materials for bone regeneration and tissue engineering where the local pH can reach as high as ~10 due to the release alkali cations from the glass into the biological fluids [83,84]. A significant, and ever-increasing, body of evidence in the literature indicates that the ionic dissolution products from inorganic materials are key to understanding their biological performance in the context of tissue engineering applications. The controlled release of trace elements, for example, strontium, cobalt copper or zinc, from the biomaterials (e.g., glass-based scaffolds) into the physiological fluids is believed to favorably affect the behavior of human cells and enhance the bioactivity of the scaffolds related to both osteogenesis and angiogenesis [16]. Therefore, the possibility of predicting the dissolution dynamics of glasses, and the ability to design glasses (of desired compositions) with controlled release of functional ions using machine learning brings a paradigm shift in the field of biomaterials research. To the best of authors' knowledge, there are very few studies – most of which are cited in a recent review [28] – that describe the application of machine learning in the field of glasses for healthcare applications. An example of one such study is by Echezarreta – Lopez and Landin [29], wherein the authors placed emphasis on describing the relationship between critical bioactive glass characteristics and their antibacterial behaviors. Thus, the present article is an important effort in the direction of predictive modeling of dissolution behavior of glasses with focus on designing new biomaterials with controlled degradation rates. It is worth mentioning that the ML formalisms applied in this study can be readily extended to predict and optimize other important properties of biomaterials, such as glass forming ability (i.e., probability of glass formation in relation to its

composition, and ambient pressure and temperature), crystalline-to-amorphous phase transition temperature, liquidous temperature, and mechanical properties (e.g., elastic modulus and fracture toughness). For such extension of ML models, comprehensive databases linking composition of the materials with their properties would be needed, and using such databases the models would have been to trained and tested in the same manner as that described in this study.

8. Conclusion

The design and application of biomedical glasses – such as Theraspehere®, 45S5 Bioglass®, and DermaFuse™ – have charted new, efficient pathways for treatment of various human ailments, which has revolutionized the healthcare industry. As the demand for biomedical glasses – in particular with controlled degradation rates – continues to increase, it has become apparent that the design and discovery of new biomedical glasses can only be expedited if their dissolution behavior in biological fluids can be reliably predicted *in vitro* and, ultimately, controlled (by optimizing glass composition). Conversely, due to the prodigiously large physiochemical degrees of freedom in [glass + solvent] systems, accurate prediction of aqueous dissolution behavior of glass is an insurmountable task if the tools used for such prediction are based on conventional, trial-and-error based (i.e., Edisonian) approaches. Furthermore, such Edisonian approaches – which are time-consuming, cumbersome, and expensive – are largely unable to reveal the underlying *Quantitative Structure-Property Relationships (QSPRs)*, thus making it infeasible to discover or design new biomedical glasses with desired performance (e.g., dissolution behvaior).

This study, focused on prompt and reliable prediction of aqueous dissolution behavior of potential glassy biomaterials, transcends beyond the conventional approach into an entirely new paradigm of artifical intelliegence. Speficially, this study presents an original ensemble machine learning (ML) model – developed by uniting the random forests (RF) model with the additive regression (AR) technique – to predict dissolution behavior of four different types of biomedical glasses (sodium borosilicates; sodium aluminoborates; sodium aluminoborosilicates; and sodium boro-phosphosilicates) in aqueous solutions maintained at different temperatures, pH environments (ranging from 0-to-13), and surface area-to-volume

ratios (SA/V). The ensemble ML model was trained using 75% of a parent database, comprising of >1300 experimentally-obtained, distinct data-records. Each data-record featured 10 inputs (i.e., glass composition; initial pH of the solvent; temperature; surface area-to-volume ratio of the system; and time) and 1 output (extent of glass dissolution, expressed as normalized mass loss of critical ion). After the model's training, input-output correlations learned during the training process were leveraged by the model to make predictions in previously untrained data-domains (i.e., remaining 25% of the parent database).

Results show that the ensemble ML model is able to promptly and reliably predict composition- and time-dependent aqueous dissolution behavior of all four families of glasses. In terms of prediction performance, the ensemble model consistently, and comprehensively, outperformed other standalone ML models. Excellent prediction performance of the ensemble model was attributed to its primary component, RF, which comprises of a large number of "deep" trees that are grown without any smoothening or pruning. The prediction performance of the RF model was further bolstered when it was combined with AR. This enhancement was attributed to the AR technique's ability to iteratively vary-and-optimize the two hyper-parameters of the RF model (i.e., number of trees and leaves per tree) until the deviation between actual values and the RF model's predictions was reduced to its ultimate minimum.

High-fidelity predictions produced by the ensemble ML model suggests that the model could be as a tool for optimization of glass dissolution behavior. In this study, ≈1300 data-records were used for training of the model. Inclusion of more data-records into the database (e.g., dissolution behavior of P_2O_5-containing 45S5 Bioglasses and other biomedical glasses) will enhance its volume and diversity, which will, in turn, further reduce the model's prediction errors, thereby making the model more amenable for optimization-based tasks. More specifically, a larger and more diverse database will enable the ability to perform prediction in reverse, that is, optimization: determination of optimal, new biomedical glass compositions, and the accompanying experimental process parameters, that lead to desired dissolution behavior. Lastly, the coupling of ML models with physics-based computational tools – for example, density functional theory, molecular dynamics, and phase-field based models – could elicit substantial improvement in the general performance of ML models [28,85,86]. Through such coupling, the models could mutually inform

each other (e.g., physics-based model could provide theoretical guidance to the ML model), thus reducing the overall dimensionality of the problem and enabling improved predictions and optimizations.

Acknowledgements

Computational tasks were conducted in the Materials Research Center (MRC) of Missouri S&T. This material is based upon the work supported by the National Science Foundation under Grant No. DMR: 1507131, CMMI: 1661609, and CMMI: 1932690, and the Leonard Wood Institute (LWI).

References

[1] M. Navarro, A. Michiardi, O. Castaño, J.A. Planell, Biomaterials in orthopaedics, Journal of the Royal Society Interface. 5 (2008) 1137–1158. https://doi.org/10.1098/rsif.2008.0151.

[2] J.K. Christie, A. Tilocca, Molecular Dynamics Simulations and Structural Descriptors of Radioisotope Glass Vectors for In Situ Radiotherapy, The Journal of Physical Chemistry B. 116 (2012) 12614–12620. https://doi.org/10.1021/jp304200f.

[3] W. Hland, V. Rheinberger, M. Schweiger, Control of nucleation in glass ceramics, Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences. (2003). https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2002.1152 (accessed May 21, 2019).

[4] L.L. Hench, The story of Bioglass®, Journal of Materials Science: Materials in Medicine. 17 (2006) 967–978. https://doi.org/10.1007/s10856-006-0432-z.

[5] A. Goel, S. Kapoor, R.R. Rajagopal, M.J. Pascual, H.-W. Kim, J.M.F. Ferreira, Alkali-free bioactive glasses for bone tissue engineering: A preliminary investigation, Acta Biomaterialia. 8 (2012) 361–372. https://doi.org/10.1016/j.actbio.2011.08.026.

[6] A. Goel, S. Kapoor, A. Tilocca, R.R. Rajagopal, J.M.F. Ferreira, Structural role of zinc in biodegradation of alkali-free bioactive glasses, Journal of Materials Chemistry B. 1 (2013) 3073–3082. https://doi.org/10.1039/C3TB20163E.

[7] M.N. Rahaman, D.E. Day, B. Sonny Bal, Q. Fu, S.B. Jung, L.F. Bonewald, A.P. Tomsia, Bioactive glass in tissue engineering, Acta Biomaterialia. 7 (2011) 2355–2373. https://doi.org/10.1016/j.actbio.2011.03.016.

[8] F.A. Shah, D.S. Brauer, R.M. Wilson, R.G. Hill, K.A. Hing, Influence of cell culture medium composition on in vitro dissolution behavior of a fluoride-containing bioactive glass, Journal of Biomedical Materials Research Part A. 102 (2014) 647–654. https://doi.org/10.1002/jbm.a.34724.

[9] D.S. Brauer, N. Karpukhina, M.D. O'Donnell, R.V. Law, R.G. Hill, Fluoride-containing bioactive glasses: Effect of glass design and structure on degradation, pH and apatite formation in simulated body fluid, Acta Biomaterialia. 6 (2010) 3275–3282. https://doi.org/10.1016/j.actbio.2010.01.043.

[10] M. Mneimne, R.G. Hill, A.J. Bushby, D.S. Brauer, High phosphate content significantly increases apatite formation of fluoride-containing bioactive glasses, Acta Biomaterialia. 7 (2011) 1827–1834. https://doi.org/10.1016/j.actbio.2010.11.037.

[11] D.S. Brauer, M. Mneimne, R.G. Hill, Fluoride-containing bioactive glasses: Fluoride loss during melting and ion release in tris buffer solution, Journal of Non-Crystalline Solids. 357 (2011) 3328–3333. https://doi.org/10.1016/j.jnoncrysol.2011.05.031.

[12] S. Kapoor, Â. Semitela, A. Goel, Y. Xiang, J. Du, A.H. Lourenço, D.M. Sousa, P.L. Granja, J.M.F. Ferreira, Understanding the composition–structure–bioactivity relationships in diopside (CaO·MgO·2SiO2)–tricalcium phosphate (3CaO·P2O5) glass system, Acta Biomaterialia. 15 (2015) 210–226. https://doi.org/10.1016/j.actbio.2015.01.001.

[13] S. Kapoor, A. Goel, A. Tilocca, V. Dhuna, G. Bhatia, K. Dhuna, J.M.F. Ferreira, Role of glass structure in defining the chemical dissolution behavior, bioactivity and antioxidant properties of zinc and strontium co-doped alkali-free phosphosilicate glasses, Acta Biomaterialia. 10 (2014) 3264–3278. https://doi.org/10.1016/j.actbio.2014.03.033.

[14] L.L. Hench, Genetic design of bioactive glass, Journal of the European Ceramic Society. 29 (2009) 1257–1265. https://doi.org/10.1016/j.jeurceramsoc.2008.08.002.

[15] N. Stone-Weiss, E.M. Pierce, R.E. Youngman, O. Gulbiten, N.J. Smith, J. Du, A. Goel, Understanding the structural drivers governing glass–water interactions in borosilicate based model bioactive glasses, Acta Biomaterialia. 65 (2018) 436–449. https://doi.org/10.1016/j.actbio.2017.11.006.

[16] A. Hoppe, N.S. Güldal, A.R. Boccaccini, A review of the biological response to ionic dissolution products from bioactive glasses and glass-ceramics, Biomaterials. 32 (2011) 2757–2774. https://doi.org/10.1016/j.biomaterials.2011.01.004.

[17] P.M. Dove, N. Han, J.J. De Yoreo, Mechanisms of classical crystal growth theory explain quartz and silicate dissolution behavior, Proceedings of the National Academy of Sciences. 102 (2005) 15357–15362. https://doi.org/10.1073/pnas.0507777102.
[18] A.C. Lasaga, A. Luttge, Variation of Crystal Dissolution Rate Based on a Dissolution Stepwave Model, Science. 291 (2001) 2400–2404. https://doi.org/10.1126/science.1058173.
[19] T. Mueller, A.G. Kusne, R. Ramprasad, Machine Learning in Materials Science, in: Reviews in Computational Chemistry, John Wiley & Sons, Ltd, 2016: pp. 186–273. https://doi.org/10.1002/9781119148739.ch4.
[20] W. Huang, D.E. Day, K. Kittiratanapiboon, M.N. Rahaman, Kinetics and mechanisms of the conversion of silicate (45S5), borate, and borosilicate glasses to hydroxyapatite in dilute phosphate solutions, Journal of Materials Science: Materials in Medicine. 17 (2006) 583–596. https://doi.org/10.1007/s10856-006-9220-z.
[21] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials. 1 (2013) 011002. https://doi.org/10.1063/1.4812323.
[22] J. Warren, Materials Genome Initiative, in: AIP Conference Proceedings, American Institute of Physics, Ste. 1 NO 1 Melville NY 11747-4502 United States, 2012.
[23] J. Carrasquilla, R.G. Melko, Machine learning phases of matter, Nature Physics. 13 (2017) 431–434. https://doi.org/10.1038/nphys4035.
[24] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, Scientific Reports. 3 (2013) 2810. https://doi.org/10.1038/srep02810.
[25] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, Npj Computational Materials. 2 (2016) 16028. https://doi.org/10.1038/npjcompumats.2016.28.
[26] L. Zdeborová, New tool in the box, Nature Physics. 13 (2017) 420–421. https://doi.org/10.1038/nphys4053.
[27] Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning, Journal of Materiomics. 3 (2017) 159–177. https://doi.org/10.1016/j.jmat.2017.08.002.
[28] M. Montazerian, E.D. Zanotto, J.C. Mauro, Model-driven design of bioactive glasses: from molecular dynamics through machine learning, International Materials Reviews. (2019) 1–25. https://doi.org/10.1080/09506608.2019.1694779.
[29] M.M. Echezarreta-López, M. Landin, Using machine learning for improving knowledge on antibacterial effect of bioactive glass, International Journal of Pharmaceutics. 453 (2013) 641–647. https://doi.org/10.1016/j.ijpharm.2013.06.036.
[30] A. Tandia, M.C. Onbasli, J.C. Mauro, Machine Learning for Glass Modeling, in: J.D. Musgraves, J. Hu, L. Calvez (Eds.), Springer Handbook of Glass, Springer International Publishing, Cham, 2019: pp. 1157–1192. https://doi.org/10.1007/978-3-319-93728-1_33.
[31] E.D. Guire, L. Bartolo, R. Brindle, R. Devanathan, E.C. Dickey, J. Fessler, R.H. French, U. Fotheringham, M. Harmer, E. Lara-Curzio, S. Lichtner, E. Maillet, J. Mauro, M. Mecklenborg, B. Meredig, K. Rajan, J. Rickman, S. Sinnott, C. Spahr, C. Suh, A. Tandia, L. Ward, R. Weber, Data-driven glass/ceramic science research: Insights from the glass and ceramic and data science/informatics communities, Journal of the American Ceramic Society. 102 (2019) 6385–6406. https://doi.org/10.1111/jace.16677.
[32] J.C. Mauro, A. Tandia, K.D. Vargheese, Y.Z. Mauro, M.M. Smedskjaer, Accelerating the Design of Functional Glasses through Modeling, Chemistry of Materials. 28 (2016) 4267–4277. https://doi.org/10.1021/acs.chemmater.6b01054.
[33] N.M. Anoop Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning, Journal of Non-Crystalline Solids. 487 (2018) 37–45. https://doi.org/10.1016/j.jnoncrysol.2018.02.023.

[34] H. Liu, Z. Fu, K. Yang, X. Xu, M. Bauchy, Machine learning for glass science and engineering: A review, Journal of Non-Crystalline Solids. (2019) 119419. https://doi.org/10.1016/j.jnoncrysol.2019.04.039.
[35] K. Yang, X. Xu, B. Yang, B. Cook, H. Ramos, N.M.A. Krishnan, M.M. Smedskjaer, C. Hoover, M. Bauchy, Predicting the Young's Modulus of Silicate Glasses using High-Throughput Molecular Dynamics Simulations and Machine Learning, Scientific Reports. 9 (2019) 8739. https://doi.org/10.1038/s41598-019-45344-3.
[36] Y.T. Sun, H.Y. Bai, M.Z. Li, W.H. Wang, Machine Learning Approach for Prediction and Understanding of Glass-Forming Ability, The Journal of Physical Chemistry Letters. 8 (2017) 3434–3439. https://doi.org/10.1021/acs.jpclett.7b01046.
[37] X. Yu, Support vector machine-based QSPR for the prediction of glass transition temperatures of polymers, Fibers and Polymers. 11 (2010) 757–766. https://doi.org/10.1007/s12221-010-0757-6.
[38] D.R. Cassar, A.C.P.L.F. de Carvalho, E.D. Zanotto, Predicting glass transition temperatures using neural networks, Acta Biomaterialia. 159 (2018) 249–256. https://doi.org/10.1016/j.actamat.2018.08.022.
[39] P. Cunningham, J. Carney, S. Jacob, Stability problems with artificial neural networks and the ensemble solution, Artificial Intelligence in Medicine. 20 (2000) 217–225. https://doi.org/10.1016/S0933-3657(00)00065-8.
[40] Xin Yao, Evolving artificial neural networks, Proceedings of the IEEE. 87 (1999) 1423–1447. https://doi.org/10.1109/5.784219.
[41] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks:: The state of the art, International Journal of Forecasting. 14 (1998) 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7.
[42] R. Cook, J. Lapeyre, H. Ma, A. Kumar, Prediction of Compressive Strength of Concrete: A Critical Comparison of Performance of a Hybrid Machine Learning Model with Standalone Models, ASCE Journal of Materials in Civil Engineering. 31 (2019) 04019255. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902.
[43] A. Behnood, V. Behnood, M.M. Gharehveran, K.E. Alyamac, Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm, Construction and Building Materials. 142 (2017) 199–207. https://doi.org/10.1016/j.conbuildmat.2017.03.061.
[44] W.-Y. Loh, Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 1 (2011) 14–23. https://doi.org/10.1002/widm.8.
[45] R. Polikar, Ensemble based systems in decision making, IEEE Circuits and Systems Magazine. 6 (2006) 21–45. https://doi.org/10.1109/MCAS.2006.1688199.
[46] S. Keon Lee, On generalized multivariate decision tree by using GEE, Computational Statistics & Data Analysis. 49 (2005) 1105–1119. https://doi.org/10.1016/j.csda.2004.07.003.
[47] J.-S. Chou, C.-F. Tsai, A.-D. Pham, Y.-H. Lu, Machine learning in concrete strength simulations: Multi-nation data analytics, Construction and Building Materials. 73 (2014) 771–780. https://doi.org/10.1016/j.conbuildmat.2014.09.054.
[48] L. Breiman, Random forests, Machine Learning. 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.
[49] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature. 559 (2018) 547–555. https://doi.org/10.1038/s41586-018-0337-2.
[50] D. Howard, A.E. Eiben, D.F. Kennedy, J.-B. Mouret, P. Valencia, D. Winkler, Evolving embodied intelligence from materials to machines, Nature Machine Intelligence. 1 (2019) 12–19. https://doi.org/10.1038/s42256-018-0009-9.
[51] C. Schaffer, Selecting a classification method by cross-validation, Machine Learning. 13 (1993) 135–143. https://doi.org/10.1007/BF00993106.
[52] I.A. Ibrahim, T. Khatib, A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm, Energy Conversion and Management. 138 (2017) 413–425. https://doi.org/10.1016/j.enconman.2017.02.006.

[53] J.R. Jones, Review of bioactive glass: From Hench to hybrids, Acta Biomaterialia. 9 (2013) 4457–4486. https://doi.org/10.1016/j.actbio.2012.08.023.
[54] R.A. Pires, I. Abrahams, T.G. Nunes, G.E. Hawkes, The role of alumina in aluminoborosilicate glasses for use in glass–ionomer cements, Journal of Materials Chemistry. 19 (2009) 3652–3660. https://doi.org/10.1039/B822285A.
[55] J.W. Nicholson, Chemistry of glass-ionomer cements: a review, Biomaterials. 19 (1998) 485–494. https://doi.org/10.1016/S0142-9612(97)00128-2.
[56] A.D. Neve, V. Piddock, E.C. Combe, Development of novel dental cements. I. Formulation of aluminoborate glasses, Clinical Materials. 9 (1992) 7–12. https://doi.org/10.1016/0267-6605(92)90004-D.
[57] M.S. Baig, G.J.P. Fleming, Conventional glass-ionomer materials: A review of the developments in glass powder, polyacid liquid and the strategies of reinforcement, Journal of Dentistry. 43 (2015) 897–912. https://doi.org/10.1016/j.jdent.2015.04.004.
[58] H. Deng, Y. Fang, Q. Fu, J.C. Mauro, Bioactive aluminoborate glasses, US20170342382A1, 2017. https://patents.google.com/patent/US20170342382A1/en (accessed September 11, 2019).
[59] H. Deng, Y. Fang, Q. Fu, L. Ma, J.C. Mauro, M.J. Snyder, D.L. Widmer-Quiel, Bioactive glass microspheres, US20170340666A1, 2017. https://patents.google.com/patent/US20170340666A1/en (accessed September 11, 2019).
[60] W. Jia, H. Hu, A. Li, H. Deng, C.L. Hogue, J.C. Mauro, C. Zhang, Q. Fu, Glass-activated regeneration of volumetric muscle loss, Acta Biomaterialia. (2019). https://doi.org/10.1016/j.actbio.2019.12.007.
[61] S.D. Conzone, D.E. Day, Preparation and properties of porous microspheres made from borate glass, Journal of Biomedical Materials Research Part A. 88A (2009) 531–542. https://doi.org/10.1002/jbm.a.31883.
[62] S.D. Conzone, R.F. Brown, D.E. Day, G.J. Ehrhardt, In vitro and in vivo dissolution behavior of a dysprosium lithium borate glass designed for the radiation synovectomy treatment of rheumatoid arthritis, Journal of Biomedical Materials Research. 60 (2002) 260–268. https://doi.org/10.1002/jbm.10047.
[63] R.J. Schalkoff, Artificial neural networks, McGraw-Hill New York, 1997.
[64] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmospheric Environment. 32 (1998) 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0.
[65] C.-C. Chia, I. Rubinfeld, B.M. Scirica, S. McMillan, H.S. Gurm, Z. Syed, Looking Beyond Historical Patient Outcomes to Improve Clinical Models, Science Translational Medicine. 4 (2012) 131ra49–131ra49. https://doi.org/10.1126/scitranslmed.3003561.
[66] I.H. Witten, E. Frank, M.A. Hall, Data mining: practical machine learning tools and techniques, 3rd ed, Morgan Kaufmann, Burlington, MA, 2011.
[67] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology). 67 (2005) 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.
[68] O. Bousquet, U. von Luxburg, G. Rätsch, eds., Advanced lectures on machine learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003 [and] Tübingen, Germany, August 4-16, 2003: revised lectures, Springer, Berlin ; New York, 2004.
[69] J. Bernardo, J. Berger, A. Dawid, A. Smith, others, Regression and classification using Gaussian process priors, Bayesian Statistics. 6 (1998) 475.
[70] C. Strobl, A.-L. Boulesteix, A. Zeileis, T. Hothorn, Bias in random forest variable importance measures: Illustrations, sources and a solution, . 8 (2007) 25. https://doi.org/10.1186/1471-2105-8-25.
[71] K.J. Archer, R.V. Kimes, Empirical characterization of random forest variable importance measures, Computational Statistics & Data Analysis. 52 (2008) 2249–2260. https://doi.org/10.1016/j.csda.2007.08.015.

[72] T.G. Dietterich, Ensemble methods in machine learning, in: International Workshop on Multiple Classifier Systems, 2000: pp. 1–15. https://doi.org/10.1007/3-540-45014-9_1.
[73] J.H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis. 38 (2002) 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.
[74] S. Kapoor, R.E. Youngman, K. Zakharchuk, A. Yaremchenko, N.J. Smith, A. Goel, Structural and Chemical Approach toward Understanding the Aqueous Corrosion of Sodium Aluminoborate Glasses, The Journal of Physical Chemistry B. 122 (2018) 10913–10927. https://doi.org/10.1021/acs.jpcb.8b06155.
[75] N. Stone-Weiss, R. E. Youngman, R. Thorpe, N. J. Smith, E. M. Pierce, A. Goel, An insight into the corrosion of alkali aluminoborosilicate glasses in acidic environments, Physical Chemistry Chemical Physics. 22 (2020) 1881-1896. https://doi.org/10.1039/C9CP06064B.
[76] M. Cerruti, D. Greenspan, K. Powers, Effect of pH and ionic strength on the reactivity of Bioglass® 45S5, Biomaterials. 26 (2005) 1665–1674. https://doi.org/10.1016/j.biomaterials.2004.07.009.
[77] J.-S. Chou, C.-K. Chiu, M. Farfoura, I. Al-Taharwa, Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques, Journal of Computing in Civil Engineering. 25 (2010) 242–253. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000088.
[78] B.A. Young, A. Hall, L. Pilon, P. Gupta, G. Sant, Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods, Cement and Concrete Research. 115 (2019) 379–388. https://doi.org/10.1016/j.cemconres.2018.09.006.
[79] V. Chandwani, V. Agrawal, R. Nagar, Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks, Expert Systems with Applications. 42 (2015) 885–893. https://doi.org/10.1016/j.eswa.2014.08.048.
[80] M. Fournier, A. Ull, E. Nicoleau, Y. Inagaki, M. Odorico, P. Frugier, S. Gin, Glass dissolution rate measurement and calculation revisited, Journal of Nuclear Materials. 476 (2016) 140–154. https://doi.org/10.1016/j.jnucmat.2016.04.028.
[81] P. Chopra, R.K. Sharma, M. Kumar, Prediction of compressive strength of concrete using artificial neural network and genetic programming, Advances in Materials Science and Engineering. 2016 (2016) http://dx.doi.org/10.1155/2016/7648467.
[82] V. Veloso de Melo, W. Banzhaf, Improving the prediction of material properties of concrete using Kaizen Programming with Simulated Annealing, Neurocomputing. 246 (2017) 25–44. https://doi.org/10.1016/j.neucom.2016.12.077.
[83] A. Reddy, D.F. Norris, S.S. Momeni, B. Waldo, J.D. Ruby, The pH of beverages in the United States, The Journal of the American Dental Association. 147 (2016) 255–263. https://doi.org/10.1016/j.adaj.2015.10.019.
[84] L. Drago, M. Toscano, M. Bottagisio, Recent Evidence on Bioactive Glass Antimicrobial and Antibiofilm Activity: A Mini-Review, Materials. 11 (2018) 326. https://doi.org/10.3390/ma11020326.
[85] V.L. Deringer, N. Bernstein, A.P. Bartók, M.J. Cliffe, R.N. Kerber, L.E. Marbella, C.P. Grey, S.R. Elliott, G. Csányi, Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics, The Journal of Physical Chemistry Letters. 9 (2018) 2879–2885. https://doi.org/10.1021/acs.jpclett.8b00902.
[86] J.C. Mauro, Decoding the glass genome, Current Opinion in Solid State and Materials Science. 22 (2018) 58–64. https://doi.org/10.1016/j.cossms.2017.09.001.

List of Figures

Figure 1: The batched and analyzed compositions of: **(a)** sodium borosilicate glasses in the $Na_2O–B_2O_3–SiO_2$ ternary diagram; and **(b)** sodium aluminoborate glasses in the $Na_2O–Al_2O_3–B_2O_3$ ternary diagram.

Figure 2: Predictions produced by: **(a)** ENR; **(b)** LR; and **(c)** GPR models compared against actual values of dissolution of glasses (drawn from the training and testing datasets described in s**ection 5.1**). R^2 of the predictions are shown in the legends. The dashed line represents the line of ideality and the solid lines represent a ±10% bound.

Figure 3: Predictions produced by: **(a)** MLP-ANN; and **(b)** SVM models compared against actual values of dissolution of glasses (drawn from the training and testing datasets described in s**ection 5.1**). R^2 of the predictions are shown in the legends. The dashed line represents the line of ideality and the solid lines represent a ±10% bound. s

Figure 4: Predictions produced by: **(a)** standalone RF; and **(b)** ensemble (AR-RF) ML models compared against actual values of dissolution of glasses (drawn from the training and testing datasets described in s**ection 5.1**). R^2 of the predictions are shown in the legends. The dashed line represents the line of ideality and the solid lines represent a ±10% bound.

Figure 5: Prediction of time- and composition-dependent dissolution behavior of: **(a)** $(Na_2O)_{0.25}(Al_2O_3)_{0.25}(B_2O_3)_x(SiO_2)_{0.5-x}$ glasses at pH = 0 (35 ºC) ; **(b)** $(Na_2O)_{0.25}(Al_2O_3)_{0.25}(B_2O_3)_x(SiO_2)_{0.5-x}$ glasses at pH = 2 (35 ºC); and **(c)** $(Na_2O)_{0.8-x}Al_2O_3)_{0.20}(B_2O_3)_x$ glasses at pH = 7 (65 ºC). The circular symbols represent experimental measurements. The SA/V of the glasses featured in (a), (b), and (c) are 7.5 m^{-1}, 2.5 m^{-1}, and 5 m^{-1}, respectively.

Table 1: Summary of physiochemical parameters pertaining to the [glass + solution] systems.

Glass system (mol.%)	SA/V (m^{-1})	T (ºC)	Solutions studied	Time
$25Na_2O$–xB_2O_3–$(75\text{-}x)SiO_2$	5	65	DI water	1 h – 14 d
Na_2O–Al_2O_3–B_2O_3	5	65	DI water	15 min – 24 h
$25Na_2O$–$25Al_2O_3$–xB_2O_3–$(50\text{-}x)$ SiO_2	2.5 & 7.5*	35	1 M HCl (pH 0) 0.01 M HCl (pH 2) 0.0001 M HCl (pH 4)	2 min – 7 d
19.37 Na_2O–16.93B_2O_3–63.70 SiO_2	5	65	0.1 M TMAH (pH 13)	1 h – 120 d
18.62Na_2O–3.87Al_2O_3–16.28B_2O_3–61.24 SiO_2	5	65	0.1 M TMAH (pH 13)	1 h – 120 d
Na_2O-P_2O_5-B_2O_3-SiO_2	5	37	0.1 M Tris-HCl (pH 7.4) 0.1 M Tris-HNO_3 (pH 7.4)	1 h – 28 d

* 7.5 m^{-1} was used for pH 0 and 2.5 m^{-1} was used for pH 2 and 4

Table 2: Statistical parameters relevant to the 11 attributes (10 inputs and 1 output) of the database used for training and testing of the ML models. The database comprises of 1364 distinct data-records.

Attribute	Unit	Min.	Max.	Mean	Std. Dev.
Solution pH	Unitless	0.0000	13.000	5.2977	2.9828
Na Composition	mass fraction	0.0919	0.2837	0.1613	0.0275
B Composition	mass fraction	0.0000	0.2533	0.1065	0.0703
Si Composition	mass fraction	0.0000	0.6167	0.1261	0.1316
Al Composition	mass fraction	0.0000	0.1967	0.1105	0.0855
P Composition	mass fraction	0.0000	0.0463	0.0034	0.0103
Temperature	°C	35.000	65.000	43.665	12.486
Specific SA	$m^2. g^{-1}$	0.0042	0.0050	0.0046	0.0003
Solution Volume	cm^3	30.000	100.00	60.059	23.716
Time	min	1.0000	172800	2496.8	12569
Critical ion normalized loss	$g. m^{-2}$	0.0000	294.92	73.312	83.487

Table 3: Statistical parameters pertaining to ML models' predictions of dissolution behavior (normalized CI mass loss) of glasses from the testing set. The worst and the best ML models – in terms of prediction accuracy – are highlighted in bold.

ML Model	**R**	**R^2**	**MAE**	**MAPE**	**RMSE**	**CPI**
	Unitless	***Unitless***	**g. m^{-2}**	***%***	**g. m^{-2}**	***Unitless***
LR	0.73	0.53	42.8	62.4	55.7	0.66
ENR	**0.39**	**0.15**	**41.5**	**89.4**	**75.2**	**0.99**
MLP-ANN	0.78	0.61	42.8	62.4	59.3	0.64
SVM	0.82	0.67	31.9	46.5	48.0	0.48
GPR	0.71	0.50	44.1	64.3	57.7	0.69
Standalone RF	0.99	0.97	7.98	11.6	13.5	0.03
Ensemble (AR-RF)	**0.99**	**0.98**	**6.15**	**8.97**	**11.2**	**0.00**

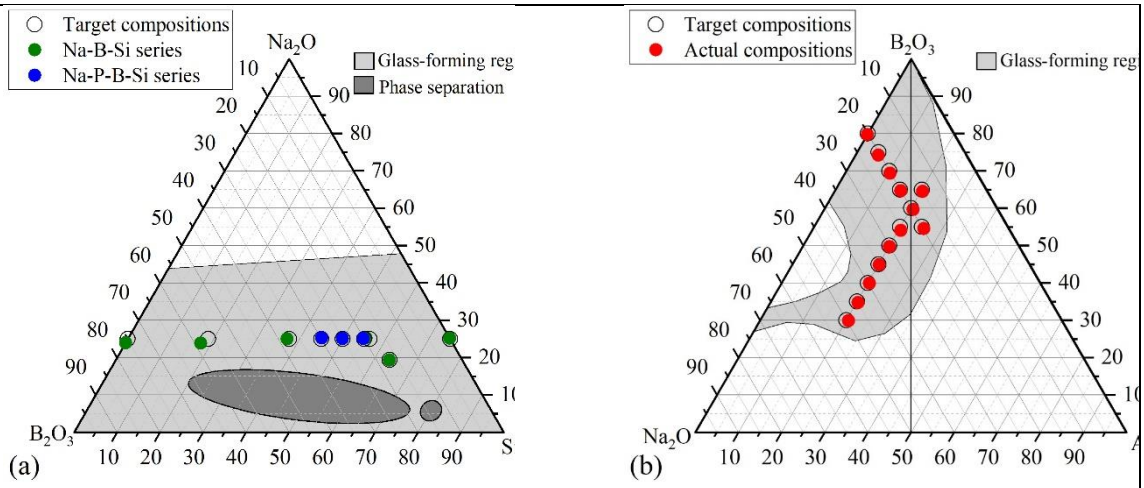
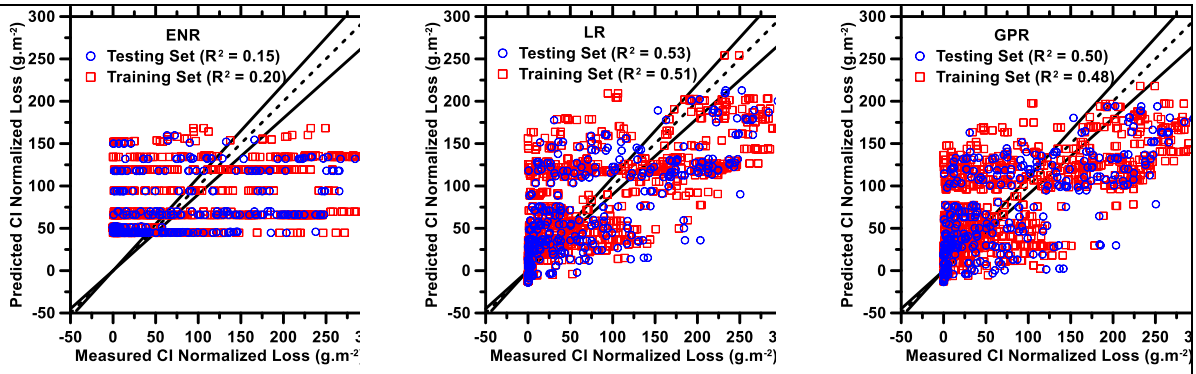

Target compositions
Na-B-Si series
Na-P-B-Si series
Glass-forming reg
Phase separation
Na_2O
B_2O_3
(a)
Target compositions
Actual compositions
Glass-forming reg
B_2O_3
Na_2O
(b)

Figure 1


ENR
Testing Set (R² = 0.15)
Training Set (R² = 0.20)
LR
Testing Set (R² = 0.53)
Training Set (R² = 0.51)
GPR
Testing Set (R² = 0.50)
Training Set (R² = 0.48)
Predicted CI Normalized Loss (g.m⁻²)
Measured CI Normalized Loss (g.m⁻²)
-50
0
50
100
150
200
250
300

(a) **(b)** **(c)**

Figure 2

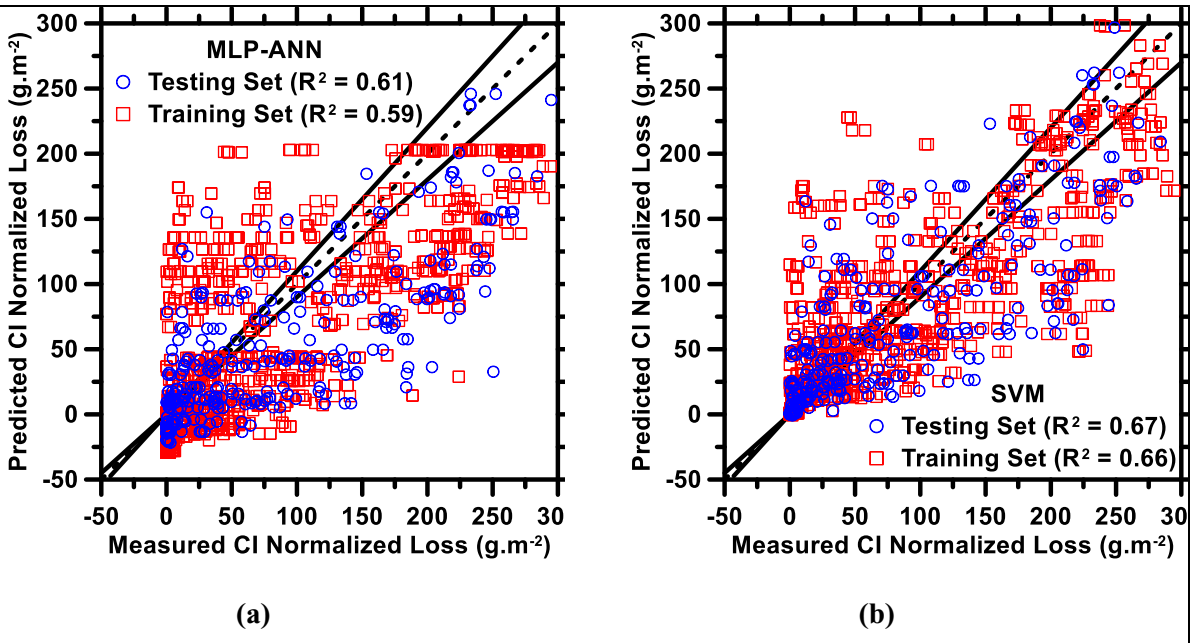
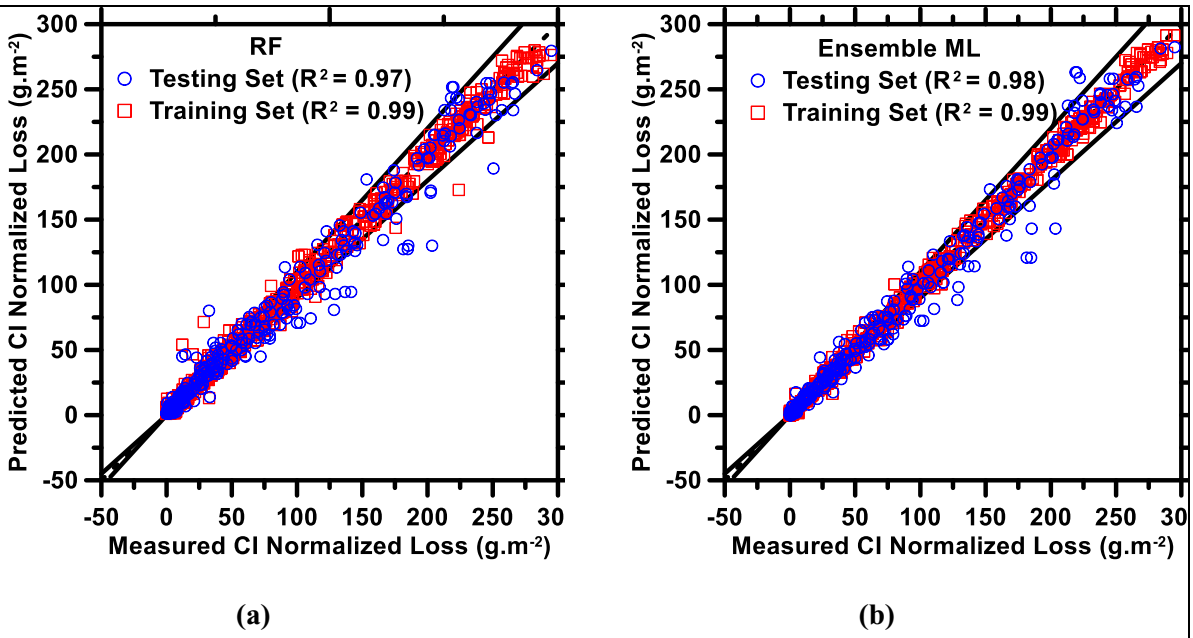
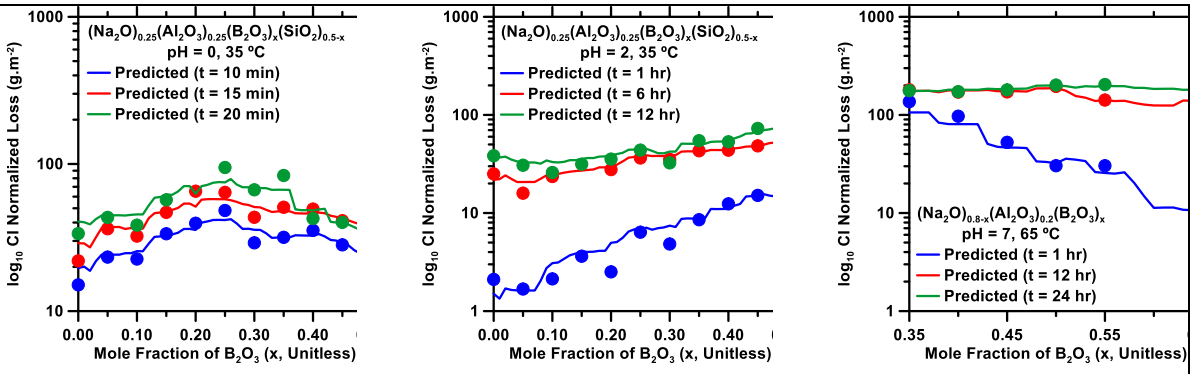

MLP-ANN
Testing Set (R² = 0.61)
Training Set (R² = 0.59)
Predicted Cl Normalized Loss (g.m-2)
Measured Cl Normalized Loss (g.m-2)
(a)
SVM
Testing Set (R² = 0.67)
Training Set (R² = 0.66)
Predicted Cl Normalized Loss (g.m-2)
Measured Cl Normalized Loss (g.m-2)
(b)

Figure 3


RF
Testing Set (R² = 0.97)
Training Set (R² = 0.99)
Predicted Cl Normalized Loss (g.m⁻²)
Measured Cl Normalized Loss (g.m⁻²)
Ensemble ML
Testing Set (R² = 0.98)
Training Set (R² = 0.99)
Predicted Cl Normalized Loss (g.m⁻²)
Measured Cl Normalized Loss (g.m⁻²)

(a) **(b)**

Figure 4



(a) **(b)** **(c)**

Figure 5