Received: 19 October 2020 | Revised: 29 December 2020 | Accepted: 15 January 2021

DOI: 10.1111/ijag.15881

INTERNATIONAL JOURNAL OF Applied Glass SCIENCE

ORIGINAL ARTICLE

Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century

Ravinder[1] | **Vineeth Venugopal**[1] | **Suresh Bishnoi**[1] | **Sourabh Singh**[1] | **Mohd Zaki**[1] | **Hargun Singh Grover**[1] | **Mathieu Bauchy**[2] | **Manish Agarwal**[3] | **N. M. Anoop Krishnan**[1,4]

[1]Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

[2]Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, CA, USA

[3]Computer Services Center, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

[4]Department of Materials Science and Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

Correspondence
N. M. Anoop Krishnan, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India.
Email: krishnan@iitd.ac.in

Funding information
Department of Science and Technology, Grant/Award Number: DST/INSPIRE/04/2016/002774; Science and Engineering Research Board, Grant/Award Number: ECR/2018/002228; National Science Foundation, Grant/Award Number: DMREF-1922167 and DMR-1944510

Abstract

Glasses have been an integral part of human life for more than 2000 years. Despite several years of research and analysis, some fundamental and practical questions on glasses still remain unanswered. While most of the earlier approaches were based on (i) expert knowledge and intuition, (ii) Edisonian trial and error, or (iii) physics-driven modeling and analysis, recent studies suggest that data-driven techniques, such as artificial intelligence (AI) and machine learning (ML), can provide fresh perspectives to tackle some of these questions. In this article, we identify 21 grand challenges in glass science, the solutions of which are either enabling AI and ML or enabled by AI and ML to accelerate the field of glass science. The challenges presented here range from fundamental questions related to glass formation and composition–processing–property relationships to industrial problems such as automated flaw detection in glass manufacturing. We believe that the present article will instill enthusiasm among the readers to explore some of the grand challenges outlined here and to discover many more challenges that can advance the field of glass science, engineering, and technology.

1 | INTRODUCTION

The progress of human civilization has always been closely associated with the discovery of new materials. This is probably why the tripartite classification of historical periods is also based on materials—stone, bronze, and iron age. Beyond these materials, there are several others which have significantly improved the quality of human life, namely, steel, aluminum, glass, plastics, the latest in the list being nanomaterials. Among these materials, glasses hold a unique place in human lives, considering their applications ranging from everyday glass utensils and kitchen-wares to

more exotic ones such as bioactive implants and energy materials.[1-5]

Although being extensively used for more than two millennia, very few glasses among the total possible glass compositions have been discovered, and even fewer have been well-studied.[6,7] This could be attributed to the compositional flexibility associated with glasses, wherein any liquid melt—a concoction of any number of elements and compounds in any proportional—when cooled fast enough, can form a glass. As such, the traditional design of glasses, thus far, has primarily been based on an Edisonian trial-and-error methodology. Although this approach has yielded some success in the past, it is highly time-consuming, inefficient, and risky (for example, it may not produce the desired result) for industrial applications. Accelerating the knowledge in glass science and the discovery of glasses in an economical fashion with a reduced to design-to-deploy period are two key areas where much work is needed.

Recent advances in computer hardware and algorithms have created a new surge of interest toward artificial intelligence (AI) and machine learning (ML) approaches for materials discovery, design, and synthesis. This has resulted in an integrated computational materials engineering (ICME) framework, which employs a combination of physics- and data-based modeling approaches to accelerate materials science.

Glasses are ideal candidates for data-driven modeling as (i) virtually any element from the periodic table, or the combination thereof, can form a glass when cooled fast enough, (ii) the properties of glasses are mainly driven by composition due to their disordered structure, and (iii) unlike crystals, the compositions of glasses can be continuously tuned, (iv) large experimental database of glass properties is available.[8-11] Due to these reasons, the glass community has also started adopting the AI/ML approaches to tackle a variety of problems such as property-prediction, tailored design, understanding the physics, accelerating the modeling, to name a few. Among these, some of the specific applications include the prediction of Young's modulus,[12-14] liquidus temperature,[15] solubility,[16] glass-transition temperature,[17,18] dissolution kinetics,[19,20] viscosity.[8,21] Other recent works have developed composition–property models for several important thermal, optical, and mechanical properties of glasses based on the available experimental dataset.[22-25] A recent work has bundled these models,[22,23] along with a database and an optimization module, in a first-of-its-kind software package for accelerating glass discovery, namely Python for Glass Genomics (PyGGi, see http://pyggi.iitd.ac.in).[23] PyGGi is a copyrighted software-as-a-service (SaaS) package, which is available both online and offline for use. At present, while the online version is free to use, the offline package has both free and paid versions.

It may be observed that most of the ML-related work in the glass literature focus mainly on property prediction and a few on compositional optimization using ML models as surrogates.[25] However, there are multitudes of opportunities in the area of AI and ML that can accelerate the field of glass science. In this article, we try to identify the main challenges that need to be addressed to accelerate innovation and development in the field of glass science and technology. The challenges are placed in such a manner that they address one of the following outstanding problems in the field of glasses, namely, (i) development of novel glasses for targeted applications, (ii) accelerating the design-to-deploy period of glasses, (iii) improving the synthesis process of glasses, (iv) fundamental understanding on the nature and response of glasses, (iv) knowledge dissemination in the area of glass science. In what follows, we briefly give an introduction to AI and ML and some of the commonly used methods. Then, we present 21 grand challenges in the field of glass science that can be addressed with the technological developments exploiting AI and ML in the 21st century.

2 | ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AI refers to a broad field that focuses on enabling machines to perform actions as humans would, based on situations or stimuli.[26,27] In this process, the machines learn the implementation of a function that maps sequences to actions in different ways possible. This approach represents a paradigm shift from the traditional physics-driven modeling, wherein, rather than instructing the computers what to do, we are allowing the system to "learn" to perform actions based on the available data. ML is thus a subset of AI, which focuses on developing algorithms or machines that can be used to detect and understand patterns in the data and extrapolate it to hitherto unexplored domains and circumstances.[28,29] It should be emphasized that, in ML, machines learn to do this task without being explicitly programmed on how to achieve it. Some of the application areas of AI include natural language processing (NLP), robotics,[30] computer vision, automated reasoning, automated programming, to name a few.[27] Similarly, some of the ML applications include classification, regression, and clustering. It should be noted that the applications of AI commonly employ the ML algorithms to achieve the objectives. For example, the NLP uses a variety of ML algorithms such as neural networks, support vectors, and random forest to extract knowledge from the text data.[31,32]

Learning algorithms in ML can be broadly classified into supervised, unsupervised, and reinforcement.[33] Figure 1 shows the hierarchy and list of algorithms commonly used in the field of ML. This list by no means is complete or exhaustive—it is representative of the broad fields in the area of machine learning.

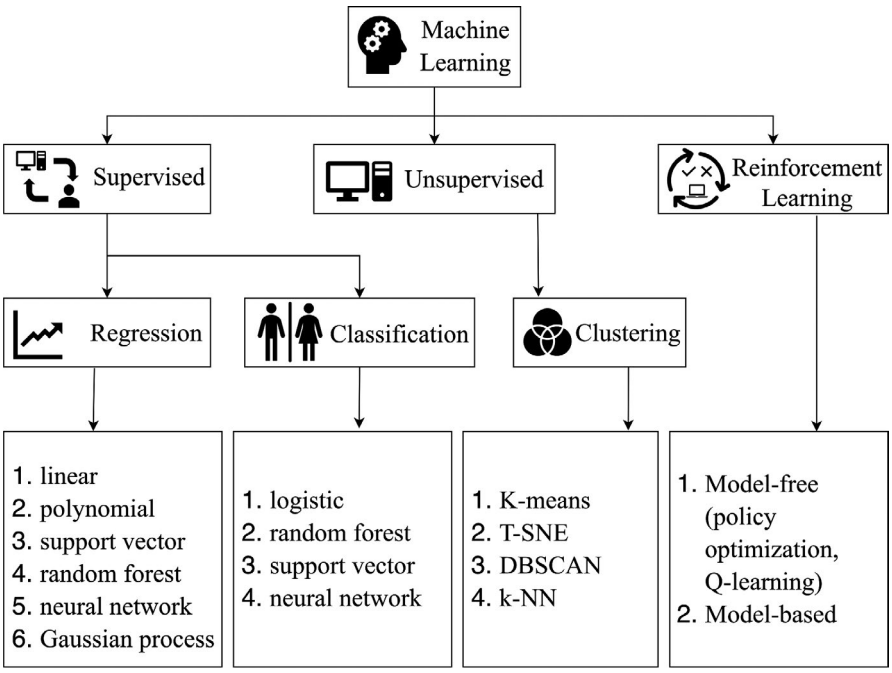

FIGURE 1 Basic machine learning paradigm

Supervised learning, which is one of the most common and useful learning approaches, involves learning a function that maps the input features to an output. To "learn" this function, available data are used to "train" the model. This model is then tested on an unseen dataset sampled from the same distribution to assess the performance. Supervised learning can be used for classification—that is, to classify the data into different classes, for example, bioactive or not bioactive, glass forming or not glass forming—or regression—that is, predicting the output as a function of features, for example, predicting Young's modulus or glass-transition temperature as a function of the glass composition. Commonly used algorithms for classification include logistic, decision trees, random forest, support vector, and neural network.

Similarly, regression is performed using algorithms such as linear, polynomial, support vector, decision trees, random forest, neural network, and Gaussian process. Unsupervised learning is used to understand the hidden patterns in data without any labels. Thus, in unsupervised learning, there is no output to "supervise" or train the model, and the model learns from the patterns in the features of the data. Some examples of unsupervised learning involve clustering, that is, to group data having similar trends or behavior, for example, grouping window glass, sealing glass, bioactive glasses, or optical glasses based on the input glass composition. Another practical application of unsupervised learning includes detecting outliers in the dataset or identifying glass samples with defects from a production pipeline. Some of the commonly used algorithms for clustering include T-SNE, DBSCAN, k-means, and k-NN. There is a third class of algorithms, namely reinforcement learning. In this approach, an agent interacts with the environment and learns the actions that maximize the reward. This approach is commonly used in many areas, including game theory, multi-agent systems, and optimization, to develop optimal solutions and algorithms. Reinforcement learning has found little application in glass science, in particular, and materials science in general. Some of the areas where reinforcement learning can have applications include the design of smart robots (robotics), which allow automated high-throughput synthesis, characterization, selection, and design of novel materials.[30] A review of some of the algorithms used in materials science and glass science can be found elsewhere.[15,34-36]

3 | GRAND CHALLENGES IN GLASS SCIENCE, ENGINEERING, AND TECHNOLOGY

Here, we aim to present some of the open challenges that are impeding the progress in the area of glass science, engineering, and technology. These challenges broadly belong to two categories: (i) ones that can be addressed using AI and ML techniques and (ii) ones that enable the application of AI and ML techniques for accelerated glass design, discovery, and manufacturing. Thus, each challenge in itself presents an opportunity to advance the area of glass science.

3.1 | Challenge 1: development of high-fidelity experimental datasets

The performance of ML algorithms is highly contingent upon the availability of high-quality, reliable, and consistent

datasets. Specifically, the dataset should consist of glasses prepared with the same experimental protocols, including the usage of the same cooling rates, heating rates, charging, stirring, crucibles, furnace, annealing time, to name a few. Similarly, testing protocols used for obtaining properties such as density, elastic moduli, hardness, glass-transition temperature, liquidus temperature, to name a few, should also be consistent. This is especially important in some of the properties such as (i) hardness—which is not a material property and depends on the testing protocol such as loading rate, indenter tip geometry, and environmental conditions; (ii) liquidus temperature—the exact determination of the lower point of which is quite challenging; (iii) glass-transition temperature—which is dependent on the cooling rate, measurement technique, and the fitting range used; (iv) viscosity—which depends on the precision of the measurement technique, furnace, and other environmental conditions.

The available glass databases include Interglad and SciGlass, both of which are compiled based on data from the literature.[11,37,38] These datasets have a large number of outliers and are highly inconsistent for many properties. Furthermore, while Interglad data are not freely available, SciGlass data are not being updated and are not available in an easily accessible and human-/machine-readable format.[10] A recent effort has made many of the properties from this database publicly available through a growing glass database, namely, PyGGi bank, a part of the Python for Glass Genomics (PyGGi) initiative.[39] There are other similar efforts such as the GlassPy, which is a free and open-source package,[40] thus following the FAIR data principle.[41] On the other hand, specialized glass companies with a track record of experimental research, such as Corning Inc., houses high-quality data, which can be used for developing high-quality data-driven models.[8,14] However, these models and data are restricted to internal use as the models give them a unique position for developing novel glass compositions.[8,14] An international collaborative initiative for the development of a publicly available experimental glass property database following a universal protocol agreed upon by the glass community—similar to the international simple glass (ISG) for studying the dissolution of nuclear waste glass[42]—will go along way in obtaining high-quality data that can be used for developing high-fidelity composition–property ML models that are publicly available. An effort toward this direction has been attempted by Citrine Informatics[43] for metals community, where users can deposit computational and experimental data. Similarly, PyGGi bank also has an option to deposit data that allow researchers to contribute and share data available with them along with the appropriate details or references, which are then made accessible to the users.

3.2 | Challenge 2: automated extraction of datasets from the literature

Most glass datasets, such as SciGlass and Interglad, relies on manual extraction of data from the literature. However, the number of unique data points in each repository is far lower than the total number of glasses that have been studied or could be studied in the future. Manual curation of online databases through the extraction of annotated composition–property pairs from literature is inherently inefficient and unsuitable for the 21st century. On the other hand, recent developments in AI can be exploited to automate the data extraction process from journal publications and patents. For example, ChemDataExtractor[44,45] is a tool that can identify chemical species through their symbols in literature. With specific data parsing algorithms such as Snowball, Chemdataextractor has been used to automatically extract large datasets of magnetic materials and their Neel's temperatures[46] and as well as a database of battery material with five leading properties.[45] Several such approaches have already been used in zeolites[47] and inorganic materials.[48] In glass science, this presents a unique challenge as glass compositions have complex representations without any uniformity. For example, a binary sodium silicate glass with a 50 mol% soda may be written as $Na_2O.SiO_2$ or $50(Na_2O).50(SiO_2)$ or $(Na_2O)_{50}.(SiO_2)_{50}$ or $(Na_2O)_{0.5}.(SiO_2)_{0.5}$ or $50(Na_2O)$–$50(SiO_2)$, all of them referencing the same glass composition. The recent development of a Named Entity Recognition (NER) system for the subject of inorganic literature suggests that NLP algorithms can be made "material science aware," allowing nuanced extraction through microlevel parsing. Figure 2 shows how NER can be used to label different parts of a text into relevant categories such as material, application, property, synthesis methods, characterization techniques, to name a few. However, a glass-specific NER system is currently unavailable and presents itself as a major challenge that needs to be addressed to enable to automated extraction of datasets from the literature.

3.3 | Challenge 3: outlier detection

Despite the importance of the quality of data for developing robust models, sufficient attention has not been given to data cleansing and outlier detection in machine learning in glass science. It could be argued that the availability of high-quality data is presently the bottleneck to developing reliable ML models for property prediction much more than the reliability and efficiency of learning algorithms. Note that outliers may arise in the data due to various factors such as instrument errors, human error, measurement conditions, system behavior, or even natural variation in the data. Detection of outliers is important not only to develop

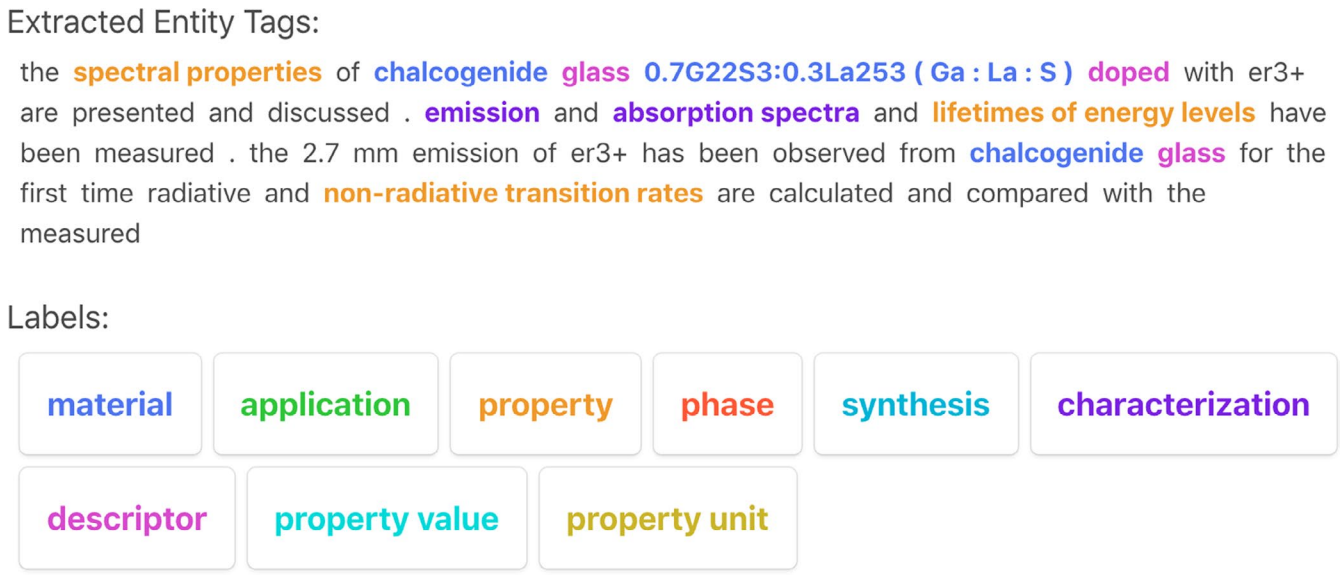

FIGURE 2 Named Entity Recognition systems, as given by Matscholar[49], can be used to label parts of a text into relevant categories such as "material," "application," "property," "synthesis," etc. Composition-property-processing tuples can be then be extracted from this annotated text block using standard routines

high-fidelity models but also to identify several anomalous behaviors—for example, some selected glass compositions exhibiting a sudden increase/decrease in the property—or system faults—for example, while doing measurement or glass manufacturing, an outlier may be due to instrument failure or system error. Several outlier detection methodologies are available in the literature depending on the nature of data and training algorithms used,[50-53] including several open-source packages such Pyod[54] and XGBOD.[55] Glass datasets in itself present unique challenges due to their inherent nonlinearities and anomalies. Therefore, the development of specialized and appropriate outlier detection algorithms for glass datasets can enable improved data cleansing and go a long way in reliable model development.

3.4 | Challenge 4: development of consistent synthetic datasets

While the development of an experimental dataset is ideal for the development of realistic composition–property dataset, experiments can be expensive, time-consuming, and mostly inconsistent due to different synthesis, sample preparation, and measurement protocol. Besides, many properties at the atomic scale may not be easily obtainable from experiments. An alternate approach is to use high-throughput atomistic and first principle simulations to generate simulation data, referred to as synthetic data. Note that such a system has been successfully implemented by The Materials Project[56] for inorganic materials, wherein a "self-healing" workflow has been implemented for the development of consistent high-throughput simulation data. The approach in glasses would be slightly more challenging as glasses being a non-equilibrium system, the simulation protocol such as the number of atoms, ensembles used, cooling rate used, cooling protocol implemented (step-wise vs. continuous), timestep, and equilibration steps can all affect the final glass structure. Several efforts have already independently used large-scale glass simulation data to develop ML models.[13,57] An international collaborative effort to establish a consistent simulation dataset, including the interatomic potentials used for glass simulation, glass structures developed through simulation, properties computed, will accelerate the development of a consistent simulation dataset that can be used for ML applications.

3.5 | Challenge 5: feature discovery and selection—atomic fingerprinting

Once a dataset on glass with the relevant composition and properties is obtained, selecting the appropriate input feature is a major challenge. Feature selection is an important aspect as the predictive capability of the model will be contingent upon the ability of input features to cover the domain appropriately. Thus, the selection of the input features should be made in such a way that all the parameters controlling a certain property are included, whereas all irrelevant features are removed. So far, there have been two broad classes of features that have been used for property prediction: (i) features based on the glass composition and (ii) features based on physics or other fundamental

properties. In the first approach, the percentage of the components or elements present in the glass composition itself is used as an input. For example, the mole percentage of Na_2O and SiO_2 in a binary $(Na_2O)_x.(SiO_2)_{1-x}$ or directly the amount of Na, Si, and O present in the glass may be used as the input for the ML algorithm, and the output would be the glass property. While this approach may yield satisfactory results for a select glass composition, the models trained on one glass composition will not be transferable to another set of compositions with different components. A more generic approach would be to develop novel physics-based features—some examples of this include topology—,[20] interatomic potential parameter—,[57] and periodic table-based descriptors.[8,21] The advantage of such descriptors is that the models developed have the potential to be universal and transferable. Furthermore, these models might be able to provide deeper insights into the factors controlling the respective properties. However, physics-based descriptors developed so far have been limited to a few select glass compositions or properties. Furthermore, most existing ML models for glasses presently are not informed by the glass structure, which may also play a crucial role in addition to the glass composition. This is an important problem if the models need further extended to include materials such as glass-ceramics. As such, it still remains as an open challenge to discover universal physics-based features, also known as the atomic fingerprints of glass, that can predict glass properties.

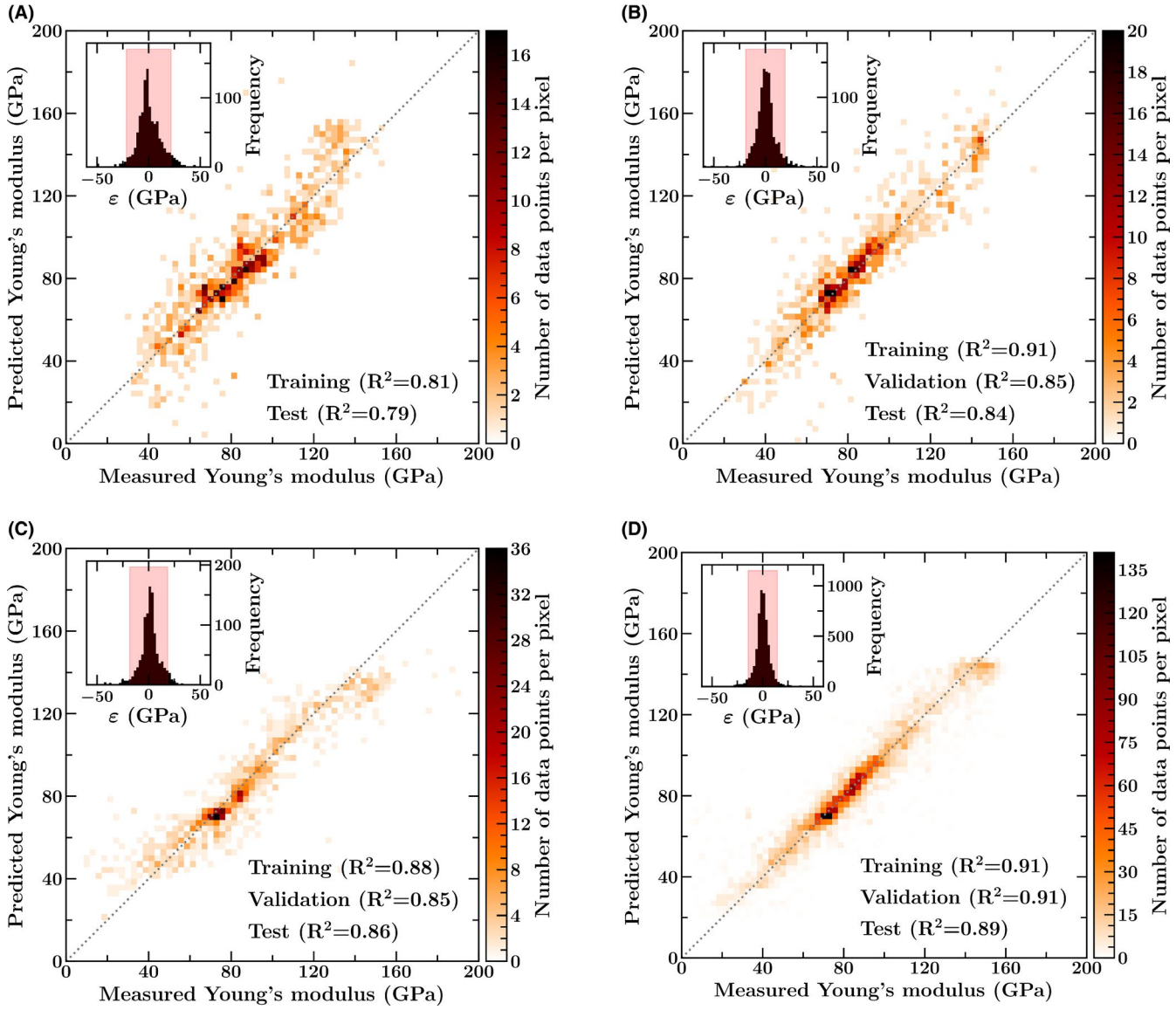

FIGURE 3 Predicted Young's modulus with respect to measured value for (A) linear regression, (B) XGBoost, (C) deep neural network, (D) KISS-GP. Inset shows the distribution of error (ε), that is, measured value minus predicted value for the test dataset

3.6 | Challenge 6: composition–property models—selection of training algorithms

As mentioned earlier (see Figure 1), the choice of an appropriate algorithm (or the development of a new one) is a crucial issue while developing composition–property models. Algorithms may range from simple linear regression to complex deep neural networks, ensemble-based methods, or Gaussian process regressions. These algorithms try to learn the input–output relationship from the available data. The preferred algorithm may be chosen depending on the size and nature of data and the desired output. Figure 3 shows the prediction of Young's modulus using four different methods, namely, linear regression, XGBoost, deep neural networks, and Gaussian process regression (GPR). We observe that the predictions for each of these models are different, with the GPR exhibiting superior performance. Although certain thumb rules exist for the choice of algorithm and approach, thus far, there are no "intelligent" systems that suggest the best approach for a particular dataset. These models also present some deficiencies or superiorities when they are used for glass discovery using surrogate model-based optimization. As a surrogate model, neural networks may perform better in contrast to decision tree-based methods such as random forest or XGBoost. It should be noted that choosing the best model is a general problem in the area of ML. However, this problem will have additional constraints in glass science in terms of the underlying physics. For example, glasses are unique in the sense that a given property changes continuously as a function of composition. Thus, the composition–property relationship should be continuous and differentiable. Similarly, the complexity and interpretability of ML models are inversely proportional. It is well understood based on the glass physics that some properties are "fairly linear," whereas some others are highly nonlinear. This prior knowledge may go into the ML model instead of a brute force approach. As such, the development of an automated "intelligent" system that can suggest the best learning algorithm for a given dataset, respecting the physics of glasses, can significantly reduce the efforts associated with a trial-and-error approach in choosing a particular ML model. In addition, ensuring the reproducibility of ML models is also a major issue that has received lesser attention. Making the final ML models available (free and open-source), along with the complete code and data could be the solution to address this challenge. Some recent works[21,25] can be taken as examples in this regard.

3.7 | Challenge 7: composition–property models—hyperparametric optimization

The performance of a model upon training is highly contingent on the use of appropriate hyperparameters during the training process. Note that hyperparameters are different from the parameters of an equation. For example, for a straight line $y = mx + c$, m, and c correspond to the parameters. On the other hand, the parameters that are employed while fitting the straight line such as the order of equation (e.g., linear or quadratic), step-size for error minimization, the algorithm used for error minimization (e.g., gradient descent), error measure (e.g., mean squared error) are the hyperparameters. Thus, parameters are learned and updated during training, whereas hyperparameters are fixed before the training. Poor or no hyperparametric optimization will lead to poor training of the model leading to overfitting or underfitting. There are several approaches that can be employed for hyperparametric optimization ranging from the traditional grid search or random search to more sophisticated approaches such as Bayesian optimization. While there are several packages to perform hyperparametric optimization such as hyperopt[58] or optuna,[59] an intelligent system that can suggest the optimal hyperparameters associated with a given training algorithm for a glass specific dataset, such as Young's modulus, can significantly reduce the training time associated with developing composition–property models while avoiding underfitting or overfitting. Furthermore, making these models publicly accessible will enable direct access to the high-fidelity composition–property models for scientists and nonscientists alike. Already such an effort is undertaken by the project PyGGi through PyGGi Seer, wherein neural network for nine glass properties with up to 34 components have been developed (see: https://pyggi.iitd.ac.in).

3.8 | Challenge 8: physics-informed machine learning

While most machine learning approaches simply use data as an input and train the model, this might result in absurd predictions, especially for regions where data are sparse. To infuse "common-sense" into the model, which allows reasonable extrapolation using the basic physical laws available. This alternate approach, known as the gray-box neural network, learns the parameters associated with a functional relationship between the input-output,[8,21] instead of directly learning the input and output. For example, a recent work used the neural network to learn the parameters associated with the MYEGA equation, which was then used to predict the viscosity.[8] Another approach in this direction is to redefine the loss function in terms of some physics-based differential equations. For example, Hamiltonian and lagrangian neural networks[60] is an attempt in which the neural learns the lagrangian of the function, thus exhibiting superior performance in terms of physical laws such as energy conservation. The development of such physics-based neural networks may aid the development of models that are interpretable and robust against extrapolation.

3.9 | Challenge 9: uncertainty quantification in predictions

Most of the ML algorithms discussed thus far are deterministic in nature—for a given output, we always obtain the same output. While these algorithms exhibit excellent predictive capabilities for interpolation, the ability of the models to extrapolate beyond the training domain is questionable. In such cases, it is important to quantify the uncertainty associated with the predictions. To this extent, GPR presents as an ideal candidate as the prediction in GPR is done in a probabilistic framework—associated with every predicted value, the standard deviation in the prediction can also be obtained from GPR, thereby giving insights into the reliability of the model. However, GPR is extremely computationally intensive, and training is limited to a small dataset (due to the matrix inversion operation associated with the covariance matrix of GP). Another recent approach to address this challenge is the Monte Carlo dropout[61] to estimate model uncertainty. The advantage of this approach is that it can be applied to already trained models as well. Alternate approaches under the structured kernel interpolation (SKI) framework, such as kernel interpolation for scalable structured GP or KISS-GP[62] have been developed to address this challenge. Figure 4 shows the predicted values of density using the KISS-GP approach.[22] We observe that the prediction provides reasonable agreement with the experimental values along with the uncertainty in predictions. Development and applications of algorithms that provide tighter bounds on uncertainty while having reduced computational cost is an open challenge that needs to be addressed for novel glass discovery. Furthermore, developing such models a large number of glass properties and making this publicly available will accelerate the design of glasses for targeted applications. Already such an effort is undertaken by the project PyGGi[39] through PyGGi Seer, wherein KISS-GP models for nine glass properties with up to 34 components have been developed (see: https://pyggi.iitd.ac.in).

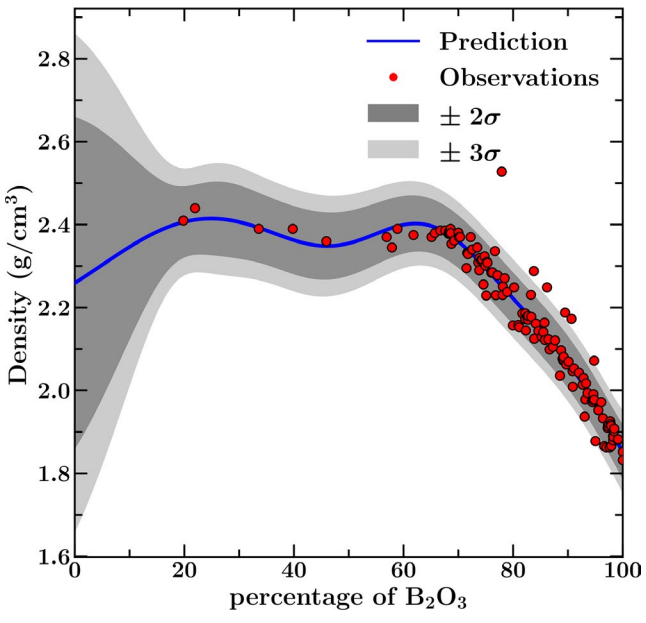

FIGURE 4 The predicted density of B_2O_3 using KISS-GP for a binary $(Na_2O)_x.(B_2O_3)$ glass

3.10 | Challenge 10: optimized predictions of glass compositions

While the ML model allows us to predict the properties of given glass composition, surrogate model-based optimization can be used to solve the reverse problem—that is, predict the composition of glass with desired properties. To this extent, various optimization algorithms such as gradient descent, ant colony, particle swarm, and genetic algorithm can be used to predict the possible compositions for a given target property. It should be noted obtaining the global minima is challenging in many cases, and thus, the optimization algorithms provide a family of glass compositions. A similar approach has been used in recent work to design optical glasses using genetic algorithm with ML models for refractive index (n_d) and glass-transition temperature (T_g) as surrogates[25]. In this work, two criteria, namely $n_d > 1.7$ and $T_g < 500$ºC, was applied to discover glasses. Additional criteria can also be applied in terms of compositional constraints. Such an approach can significantly accelerate the design and discovery of glasses. However, at present, this approach is limited to those who are already having ML models for the properties of interest. In order for this approach to be accessible to a larger public, the surrogate models for a large number of properties and components need to be developed and shared along with various optimization algorithms publicly. Already such an effort is undertaken by the project PyGGi through PyGGi Zen, wherein optimization models with four different algorithms are shared along with the ML models for nine glass properties with up to 34 components for optimized design of glasses (see: http://pyggi.iitd.ac.in).

3.11 | Challenge 11: interpreting the machine learning models for composition–property relationships

ML methods are notoriously known as black-box methods, as it is extremely challenging to interpret the nature of input–output relationships for them. However, to develop reliable models and to interpret the input relationship meaningfully, the interpretability of ML models is an

important aspect that needs to be explored. Recent developments in the ML field have resulted in a variety of "interpreting" tools for ML models. Some of these algorithms are Shapley additive explanations[63] (SHAP), Locally Interpretable Models and Effects (LIME), partial dependency plots (PDP), to name a few.[64] Applications of these approaches to the ML models will allow one to infer the role of each of the components in controlling the respective property. For example, interpreting the ML model for hardness can provide insights into the role of network former and modifier in controlling the overall response to indentation. As such, this area of interpretable ML can have a huge impact on glass science and technology in understanding the composition–property relationships.

3.12 | Challenge 12: development of machine-learned interatomic potentials

Glass simulation is an area where there is a lot of interest as the simulations can provide insights into the structure and properties of glasses. Most of the glass simulations are conducted at the atomic scale as it provides a balance between accuracy (better than mesoscale simulations) and computational cost (lower than first principle simulations). Since the atomistic simulations are only as accurate as the interatomic potentials used, the development of realistic interatomic potentials for glass is an active field of research. In contrast to crystalline systems, potential development and refinement for glasses are more challenging due to their disordered structure. The development of high accuracy (similar to the first principle) potentials for glasses remains an open challenge, which can be addressed through ML. Potential parametrization using machine learning has mainly been of two broad categories: (i) use machine learning to learn and optimize the parameters of a known functional form,[65,66] for example, BKS, (ii) develop a machine-learned potential,[67] where the functional itself is approximated using a machine learning model such as neural network, random forest, or gaussian approximation potential.[68] While the former has been applied to some simple glasses, the latter has barely been attempted for glasses. The challenge can be addressed by developing a collaborative cloud-based service, which provides an interatomic potential for any system given enough training data. Thus, the service should be able to take user-given input trajectories of first principle simulation, use this training data to develop a machine-learned potential, and share it with the public. These potentials can keep improving with more and more training data with an increasing user base. Such a trainable machine learning interatomic potential can accelerate the glass simulations allowing deeper insights into the structure and response of glassy materials.

3.13 | Challenge 13: deciphering the relationship between structure and dynamics of glasses

A comment by Anderson[1] emphasizes the "theory of the nature of glass and glass transition to be the deepest and most interesting unsolved problem in solid-state theory". Recent studies suggest that ML may provide a new angle to attack this unsolved problem.[69-72] These works suggested a machine-learned parameter, namely, softness,[69,72] which connects the local structure to the dynamics during the glass transition. The softness parameter addresses the long-standing question regarding whether there is a structural signature associated with the dynamics during the glass transition. Although the physical meaning of "softness" itself and its direct role in solving the mystery of the glass transition remains unclear, the work suggests how ML can be used to decipher the hidden relationship between the structure and dynamics of the glass transition. A fresh data-driven approach using ML may thus provide insights into one of the greatest unsolved problems of the 21st century—the glass transition.

3.14 | Challenge 14: transfer learning—applying knowledge learned in one area to another

In materials science, obtaining large-scale data associated with a system are always a challenge. However, it is well-understood many of the physical, chemical, mechanical, thermodynamic, and electronic properties are interrelated. Similarly, the nature of the relationship between the input features and output may also be interrelated in many different materials. Thus, the knowledge learned from one system or property can be used to develop models for a similar system or property. This approach, known as transfer learning, has recently found some applications in the area of materials science.[73,74] The advantage in this approach is that a model pretrained on a large database can then be used to learn the function on a much smaller database. Thus, for glasses or properties in which only sparse data are available, transfer learning can be used to develop reliable models. However, confounding effects that different input components may have could be different for each of the properties. Such an effect may, thus, make the transfer learning approach inferior as well. To this extent, the development of a reliable transfer learning module for the development of models for glasses with a sparse dataset remains an open challenge.

3.15 | Challenge 15: semantic search and analysis from literature

The number of publications relevant to the keyword "glass" is over 1.8 million in Science Direct and 1.6 million in

Semantic Scholar. Yet, at the same time, our ability to query and navigate these databases are rudimentary. For example, we are unable to rapidly answer questions such as: (i) What glass compositions have been studied for a property P? (ii) What experimental results are available in the literature for a given glass composition G? (iii) Which synthesis steps S are available for a composition G to obtain a property $P_1 > e_1$ & $P_2 < e_2$?, etc. Generically, this comes down to the identification and extraction of composition-property-processing tuples from textual databases. NLP is increasingly being recognized as a means to meaningfully engage with scientific literature and to extract user-specific information from a large corpus of texts.[31] Select demonstrations of this approach range from regular expression and syntax-based identification of structure-property-values pairs from literature (battery materials,[45] phase diagrams,[75] inorganic oxides[48]), variational autoencoder-driven prediction of synthesis parameters (inorganic oxides),[76] the discovery of new thermoelectrics using word embeddings,[31] and the identification of broad synthesis recipes using neural networks.[77-80]

The caption-cluster plot[81,82] is a graphical summary of an entire field of knowledge, providing insights into the availability and distribution of research interests within the community and their interrelationships. The example in Figure 5 is built using standard NLP techniques such as vectorization and clustering on over 10 000 figure captions selected from 3000 papers selected randomly from glass literature. The pixelated figure captions are found to cluster based on the information contained in the images, such as different types of spectra, polarization measurements, X-ray diffraction images, thermal measurements, etc. Combined with composition and processing markers, this changes into a tool that helps answer some of the questions raised at the beginning. Each point in the image represents a figure caption that has been vectorized and clustered. The color of the point represents the label assigned to the caption based on the type of information it represents. Thus, in Figure 5, blue point corresponds to IR spectra, whereas green corresponds to DSC images.

However, for the most part, the overwhelming textual, graphic, and ontological information on glass literature is grossly underutilized and poorly assimilated, even by subject experts—a condition that is endemic in the physical sciences. Moreover, the absence of a standard subject ontology for the physical sciences, variations in terminology, syntax, and representation, along with poor writing practices, have resulted in disjoint and often indecipherable dependency relations in text, which complicate routine NLP methods. The automation of information extraction in the physical sciences has, therefore, lagged behind fields such as bioinformatics, where standard representation and presentation systems have existed for years. The development of automated routines to

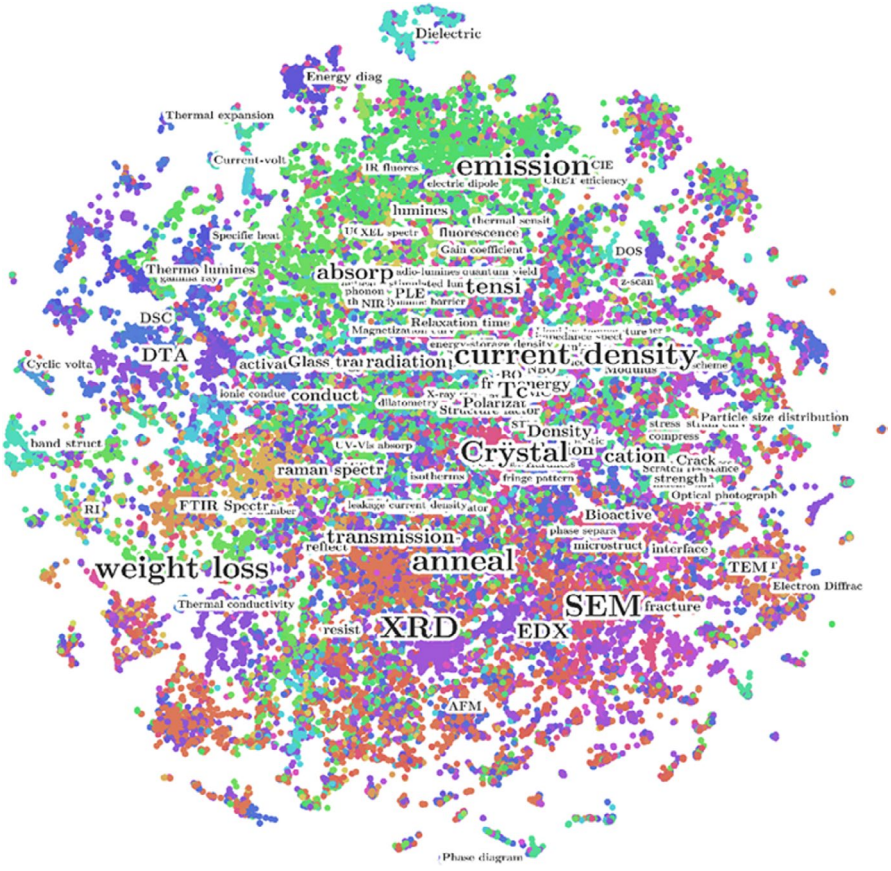

FIGURE 5 The caption cluster plot of 10 000 figure captions selected from 3000 papers on the keyword "glass." The captions are vectorized and clustered using standard NLP techniques

extract meaningful data from glass literature is one of the biggest challenges in the field for the 21st century.

3.16 | Challenge 16: extracting the synthesis protocols from the literature

It is well known that processing conditions and synthesis recipes influence the final microstructure and properties of glasses. Therefore, any predictive ML/AI model—as outlined in Challenges 5 and 6, for example—would benefit enormously from the inclusion of synthesis parameters as a feature. In addition, the extracting of synthesis protocols[47,83,84] can provide deeper insights into the literature by answering some questions such as: (i) what are the methods commonly used to prepare a glass?, (ii) what method should be used for a specific glass, (iii) what is the effect of a method on a property, (iv) what method should be adopted to obtain a target property, to name a few. One way to address this issue is to develop a consistent dataset with the same protocol followed throughout, as mentioned in Challenge 1. An alternate approach would be to extract the synthesis protocol[48,80] directly from the published scientific literature. This dataset will identify a glass composition and relate the synthesis and processing parameters associated with it. These parameters, in turn, can range from high-level synthesis routines (sol–gel, hydrothermal, solid-state), middle-level synthesis steps (annealing, quenching, rolling) to low-level synthesis parameters (annealing temperature, quenching rate, indentation stress, etc.). Similar approaches have been demonstrated for inorganic oxides where specific synthesis parameters for a given reaction have been calculated using variational auto encoders.[76] Figure 6 shows the synthesis steps for hydrothermal, sol–gel, and solid-state reactions that have been automatically generated from published literature.[77] A random forest model trained on synthesis routines of oxide inorganics, in combination with a Markov chain model, has been shown to "learn" the synthesis steps directly from text without any human input.[77] This routine was able to identify differences between necessary and optional subroutines and to correctly order the steps of a reaction chronologically. It has also been shown that Long Short Term Neural Network (LSTM) trained on synthesis literature could currently identify the most appropriate chemical precursor for a synthesis routine.[80] The development of an automated pipeline to extract synthesis protocols for glasses will be extremely useful for scientists and industry alike, and thus, remains an open challenge.

3.17 | Challenge 17: image repository of glasses

In addition to text, the majority of the knowledge in the literature is presented in the form of images. These images may include graphs, microscopic images, XRDs, spectroscopies, to name a few. The total number of images in literature databases on glasses number in the millions. Despite the abundance of data, the glass community, as well as the broader material science community, suffer from the absence of a curated image library. This has prevented the development of robust computer vision applications to accelerate the development and discovery of new glass compositions. For example, convolutional neural networks have been shown to perform as well as human agents in correctly identify the composition and phase from an XRD image.[85] They have similarly been used to automate the identification of vacancies and point defects in TEM images.[86] Such approaches can vastly improve the quality of glass research by automating tedious and labor-intensive characterization methods. However, both deep learning and machine learning

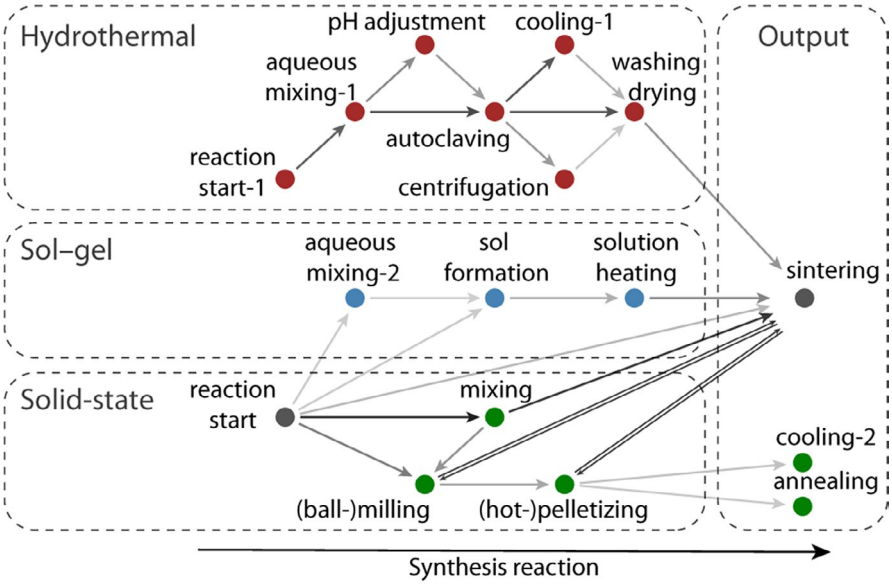

FIGURE 6 Synthesis steps for hydrothermal, sol–gel, and solid-state reactions that have been automatically generated from published literature. Relevant experimental parameters were trained using random forest, and the features of this model were ordered using a Markov chain to generate these protocols. The figure is reproduced with permission from Ref [77]

algorithms require a sufficiently large training dataset of annotated images with positive and negative labels. Not only as these image sets required for training the models, but they also set the benchmark for future models—which is necessary for the standardization of the field. However, we currently lack a large image repository similar to the MNIST handwritten images dataset or the IMAGENET—which has played a pivotal role in the development of CV models. Tools such as the caption-cluster plot mentioned in Challenge 15 is very useful in extracting and identifying the labels of an image. We need packages such as the ImageDataExtractor[87] to extract and quantify images from the literature. However, the currently available tools are limited to optical microscopy or electron microscopy images and do not extend to other characterizations and spectra. Building a searchable repository of labeled images will allow easy access to the knowledge hidden in images and buried deep in the literature.

Furthermore, images tell a detailed story of material for those who know how to read it, and for those who do not know how to read it, there is AI. Recent studies have shown that the material properties such as ionic conductivity can be directly predicted from the microstructure.[88] Similar approaches have been used in other materials such as composites and polycrystals[89,90] to extract the structure-property correlations. Interestingly, a recent study has employed graph neural networks to study the structure and dynamics of glass[71] from the 3D images obtained from atomistic simulations. These studies suggest that learning the information hidden in images using computer vision and AI can provide deeper insights into the structure–property relationship in materials.

3.18 | Challenge 18: automated high-throughput glass synthesis

High-throughput glass synthesis is time-consuming and involves a lot of manual labor. In addition, there might be human errors (or rather, habits) associated with the synthesis that makes it less consistent—that is, two different experimentalists performing the same protocol might do it slightly differently. Automating the glass synthesis process using robotics can significantly reduce the design to discovery period for a new product through high-throughput experiments at a much faster rate and in a more economical fashion. These approaches are extremely complex involving robots, which can sense the physical environment based on partial visual, auditory and other sensory cues, and make decisions instantaneously. It involves the implementation of several concepts in tandem such as state estimation, perception, unsupervised and reinforcement learning, optimization, and scheduling.[27,30] Such approaches have already been implemented in a modular robotic platform to discover thin-film materials with optimized optical and electronic properties[91] and to synthesize and sinter ceramics in seconds[92]. Such "self-driving laboratories" developed for glass synthesis and characterization can be a disruptive change allowing for the accelerated discovery of optimized glasses.

3.19 | Challenge 19: scheduling problems and optimization during glass manufacturing

In addition to being able to tackle the scientific questions, AI can also be used to simplify the problems faced in glass industries. Glass factories always face the issue of scheduling the tasks and optimizing the workflow of interdependent tasks composed in a job. This scheduling is subject to a large number of constraints in terms of time, money, resources, and environmental impact. A minor misjudgment in the scheduling can lead to major financial losses for the company. To this extent, AI can be used for automated scheduling of tasks in the industry, which can take into account the dynamic changes in the situations and respond intelligently using an on-the-fly optimization approach.[26,27] Furthermore, AI could be further used to monitor and tune the manufacturing process parameters such as furnace temperature, charging rate, etc. Dynamically optimizing the temperature of furnaces in real-time could lead to an enormous reduction in the cost and energy associated with glass manufacturing. Such an approach can significantly reduce the human efforts associated with the scheduling and process optimization while ensuring a reliable and faster solution.

3.20 | Challenge 20: automated detection of flaws in a large-scale glass synthesis

It is important for glass industries, producing glasses on a large scale, to detect flaws in their products. These flaws may range from inhomogeneities and localized crystallization to microcracks and surface scratches. Visual identification of such flaws are extremely challenging and relies on the ability of an expert. To address this challenge, smart computer vision-based AI systems can be developed, allowing for automated detection of flaws during the glass manufacturing process. Such AI-based flaw detections have been used widely for engineering materials.[93] Once the AI system is sufficiently trained to detect the flaws, they can potentially be used to optimize the processing parameters so as to minimize the flaws. Overall, AI may be used to optimize and improve the glass manufacturing process.

3.21 | Challenge 21: automated warning and safety systems for glass industries

Maintenance and replacement of machinery and equipment form one of the major expenses in industrial applications.

While most of the industries focus on the diagnosis of damage through nondestructive methods, damage prognosis is disruptive as it can avoid disasters due to material or machinery failure. Furthermore, the prognosis allows the authority to plan the maintenance or replacement schedule, thereby leading to significant savings in terms of energy, cost, and human resources. It is well known that workplace accidents can have a significant impact on productivity when measured in terms of the number of days lost. While these accidents seem to be erratic with no structure or pattern, it has been shown that ML approaches can indeed predict these accidents reasonably. As part of Industry 4.0, great emphasis has been placed on the Industrial Internet of Things (IIoT) to provide intelligent solutions based on machine monitoring to improve workplace safety.[94] Thus, the development of automated warning and safety systems for glass industries as part of IIoT can thus enhance workplace safety while ensuring reduced expenses associated with the maintenance of the machinery.

4 | CONCLUSIONS

Altogether, in this article, we discuss various avenues in glass science, where AI and ML can bring about a disruptive change. Specifically, we discuss 21 grand challenges related to glass science, technology, and engineering that can bring about drastic changes in the status quo. This list by no means is complete or exhaustive—there are many more avenues for the application of AI and ML in the field of glasses. We hope that this article will instill enough enthusiasm in the readers to explore those avenues. Finally, we believe that this article will provide the impetus for beginners to explore the exciting field of AI and ML for glass science and experts to provide novel and innovative ideas to accelerate the field of glass science through AI and ML.

ACKNOWLEDGMENTS

N.M.A.K. acknowledges the financial support for this research provided by the Department of Science and Technology, India under the INSPIRE faculty scheme (DST/INSPIRE/04/2016/002774), and DST Science and Engineering Research Board Early Career Award (ECR/2018/002228). Part of this work was supported by the National Science Foundation under Grants No. DMREF-1922167 and DMR-1944510. The authors thank the IIT Delhi HPC facility for providing the computational and storage resources.

ORCID

Mathieu Bauchy https://orcid.org/0000-0003-4600-0631
N. M. Anoop Krishnan https://orcid.org/0000-0003-1500-4947

REFERENCES

1. Anderson DL. Through the glass lightly. Science. 1995;267(5204):1618.
2. Mauro JC, Zanotto ED. Two centuries of glass research: historical trends, current status, and grand challenges for the future. Int J Appl Glass Sci. 2014;5(3):313–27. https://doi.org/10.1111/ijag.12087
3. Mauro JC. Grand challenges in glass science. Glass Sci. 2014;1:20. https://doi.org/10.3389/fmats.2014.00020
4. Mauro JC. Through a glass, darkly: dispelling three common misconceptions in glass science. Int J Appl Glass Sci. 2011;2(4):245–61. https://doi.org/10.1111/j.2041-1294.2011.00069.x
5. Bhaskar P, Kumar R, Maurya Y, Ravinder R, Allu AR, Das S, et al. Cooling rate effects on the structure of 45S5 bioglass: Insights from experiments and simulations. J Non-Cryst Solids. 2020;534:119952. https://doi.org/10.1016/j.jnoncrysol.2020.119952
6. Varshneya AK, Mauro JC. Fundamentals of inorganic glasses. Amsterdam: Elsevier; 2019.
7. Zanotto ED, Coutinho FAB. How many non-crystalline solids can be made from all the elements of the periodic table? J Non-Cryst Solids. 2004;347(1):285–8. https://doi.org/10.1016/j.jnoncrysol.2004.07.081
8. Tandia A, Onbasli MC, Mauro JC. Machine learning for glass modeling. In: Musgraves JD, Hu J, Calvez L, editors. Springer handbook of glass. Cham: Springer International Publishing; 2019:1157–92. https://doi.org/10.1007/978-3-319-93728-1_33
9. Priven AI, Mazurin OV. Glass property databases: their history, present state, and prospects for further development. Adv Mater Res. 2008;39–40:147–52. https://doi.org/10.4028/www.scientific.net/AMR.39-40.147
10. https://github.com/epam/SciGlass
11. Ver INTERGLAD. 7. n.d. https://www.newglass.jp/interglad_n/gaiyo/info_e.html
12. Bishnoi S, Singh S, Ravinder R, Bauchy M, Gosvami NN, Kodamana H, et al. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. J Non-Cryst Solids. 2019;524:119643. https://doi.org/10.1016/j.jnoncrysol.2019.119643
13. Yang K, Xu X, Yang B, Cook B, Ramos H, Krishnan NMA, et al. Predicting the young's modulus of silicate glasses using high-throughput molecular dynamics simulations and machine learning. Sci Rep. 2019;9(1):8739. https://doi.org/10.1038/s41598-019-45344-3
14. Onbaşlı MC, Tandia A, Mauro JC. Mechanical and compositional design of high-strength corning Gorilla® glass. Handb Mater Model. 2020;1997–2019: https://doi.org/10.1007/978-3-319-44680-6_100
15. Mauro JC, Tandia A, Vargheese KD, Mauro YZ, Smedskjaer MM. Accelerating the design of functional glasses through modeling. Chem Mater. 2016;28(12):4267–77. https://doi.org/10.1021/acs.chemmater.6b01054
16. Brauer DS, Rüssel C, Kraft J. Solubility of glasses in the system P2O5–CaO–MgO–Na2O–TiO2: experimental and modeling using artificial neural networks. J Non-Cryst Solids. 2007;353(3):263–70. https://doi.org/10.1016/j.jnoncrysol.2006.12.005
17. Cassar DR, de Carvalho ACPLF, Zanotto ED. Predicting glass transition temperatures using neural networks. Acta Mater. 2018;159:249–56. https://doi.org/10.1016/j.actamat.2018.08.022

18. Alcobaça E, Mastelini SM, Botari T, Pimentel BA, Cassar DR, de Carvalho ACPdLF, et al. Explainable machine learning algorithms for predicting glass transition temperatures. Acta Mater. 2020;188:92–100. https://doi.org/10.1016/j.actamat.2020.01.047
19. Krishnan NMA, Mangalathu S, Smedskjaer MM, Tandia A, Burton H, Bauchy M. Predicting the dissolution kinetics of silicate glasses using machine learning. J Non-Cryst Solids. 2018;487:37–45. https://doi.org/10.1016/j.jnoncrysol.2018.02.023
20. Liu H, Zhang T, Anoop Krishnan NM, Smedskjaer MM, Ryan JV, Gin Stéphane, et al. Predicting the dissolution kinetics of silicate glasses by topology-informed machine learning. Npj Mater Degrad. 2019;3(1):32. https://doi.org/10.1038/s41529-019-0094-1
21. Cassar DR. ViscNet: Neural network for predicting the fragility index and the temperature-dependency of viscosity. Acta Mater. 2021;206:116602. http://doi.org/10.1016/j.actamat.2020.116602
22. Bishnoi S, Ravinder R, Grover HS, Kodamana H, Krishnan NMA. Scalable Gaussian processes for predicting the optical, physical, thermal, and mechanical properties of inorganic glasses with large datasets. Mater Adv. 2021;2(1):477–87. https://doi.org/10.1039/D0MA00764A
23. Ravinder R, Sridhara KH, Bishnoi S, Grover HS, Bauchy M, Jayadeva, et al. Deep learning aided rational design of oxide glasses. Mater Horiz. 2020;7(7):1819–27. http://doi.org/10.1039/d0mh00162g
24. Cassar DR, Mastelini SM, Botari T, Alcobaça E, de Carvalho ACPLF, Zanotto ED. Predicting thermal, mechanical, and optical properties of oxide glasses by machine learning using large datasets. 2020.
25. Cassar DR, Santos GG, Zanotto ED. Designing optical glasses by machine learning coupled with a genetic algorithm. Ceram Int. 2020. http://doi.org/10.1016/j.ceramint.2020.12.167
26. Nilsson NJ. Principles of artificial intelligence. Morgan Kaufmann; 2014.
27. Russell S, Norvig P. Artificial intelligence: a modern approach. Englewood Cliffs, NJ: Series in Artificial Intelligence; 2002.
28. Bengio Y, Goodfellow I, Courville A. Deep learning. Cambridge, MA: MIT Press; 2017.
29. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539
30. Yang G-Z, Bellingham J, Dupont PE, Fischer P, Floridi L, Full R, et al. The grand challenges of Science Robotics. Sci Robot. 2018;3(14). https://doi.org/10.1126/scirobotics.aar7650
31. Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 2019;571(7763):95–8. https://doi.org/10.1038/s41586-019-1335-8
32. Manning C, Schutze H. Foundations of statistical natural language processing. Cambridge, MA: MIT Press; 1999.
33. Bishop CM. Pattern recognition and machine learning (information science and statistics). Berlin, Heidelberg: Springer-Verlag; 2006.
34. Gubernatis JE, Lookman T. Machine learning in materials design and discovery: examples from the present and suggestions for the future. Phys Rev Mater. 2018;2(12):120301. https://doi.org/10.1103/PhysRevMaterials.2.120301
35. Liu H, Fu Z, Yang K, Xu X, Bauchy M. Machine learning for glass science and engineering: a review. J Non-Cryst Solids. 2019;119419: https://doi.org/10.1016/j.jnoncrysol.2019.04.039
36. Montazerian M, Zanotto ED, Mauro JC. Model-driven design of bioactive glasses: from molecular dynamics through machine learning. Int Mater Rev. 2019;1–25. https://doi.org/10.1080/09506608.2019.1694779
37. Jantzen CM, Crawford CL.Letter Report on SRNL Modeling Database: Accelerated Leach Testing of GLASS (ALTGLASS)-Version 2.1. SRNL-L3100-2014-00229 and FCRD-SWF-2014-000249; 2015.
38. Bansal NP, Doremus RH. Handbook of glass properties. Amsterdam: Elsevier; 2013.
39. Python for Glass Genomics (PyGGi). n.d. http://pyggi.iitd.ac.in
40. Cassar DR. drcassar/glasspy. 2020. https://github.com/drcassar/glasspy
41. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018. https://doi.org/10.1038/sdata.2016.18
42. Gin S, Abdelouas A, Criscenti LJ, Ebert WL, Ferrand K, Geisler T, et al. An international initiative on long-term behavior of high-level nuclear waste glass. Mater Today. 2013;16(6):243–8. https://doi.org/10.1016/j.mattod.2013.06.008
43. Open Citrination Platform - Citrine Informatics: The AI Platform for Materials Development. Citrine Inform. n.d.
44. Swain MC, Cole JM. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. J Chem Inf Model. 2016;56(10):1894–904. https://doi.org/10.1021/acs.jcim.6b00207
45. Huang S, Cole JM. A database of battery materials auto-generated using ChemDataExtractor. Sci Data. 2020;7(1):260. https://doi.org/10.1038/s41597-020-00602-2
46. Court CJ, Cole JM. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. Sci Data. 2018;5(1):180111. https://doi.org/10.1038/sdata.2018.111
47. Jensen Z, Kim E, Kwon S, Gani TZH, Román-Leshkov Y, Moliner M, et al. A machine learning approach to zeolite synthesis enabled by automatic literature data extraction. ACS Cent Sci. 2019;5(5):892–9. https://doi.org/10.1021/acscentsci.9b00193
48. Kononova O, Huo H, He T, Rong Z, Botari T, Sun W, et al. Text-mined dataset of inorganic materials synthesis recipes. Sci Data. 2019;6(1):203. https://doi.org/10.1038/s41597-019-0224-1
49. Weston L, Tshitoyan V, Dagdelen J, Kononova O, Trewartha A, Persson KA, et al. Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. J Chem Inf Model. 2019;59(9):3692–702. https://doi.org/10.1021/acs.jcim.9b00470
50. Domingues R, Filippone M, Michiardi P, Zouaoui J. A comparative evaluation of outlier detection algorithms: experiments and analyses. Pattern Recognit. 2018;74:406–21. https://doi.org/10.1016/j.patcog.2017.09.037
51. Hodge V, Austin J. A survey of outlier detection methodologies. Artif Intell Rev. 2004;22(2):85–126. https://doi.org/10.1023/B:AIRE.0000045502.10941.a9
52. Knorr EM, Ng RT. A unified approach for mining outliers. In Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research. 1997:11.
53. Breunig MM, Kriegel H-P, Ng RT, Sander J.LOF: identifying density-based local outliers. *ACM Sigmod Rec*, Vol. 29. ACM; 2000:93–104.
54. Zhao Y, Nasrullah Z, Li Z. PyOD: a python toolbox for scalable outlier detection. J Mach Learn Res. 2019;20(96):1–7.
55. Zhao Y, Hryniewicki MK. XGBOD: improving supervised outlier detection with unsupervised representation learning. *2018 Int. Jt.*

Conf. Neural Netw. IJCNN. 2018:1–8. https://doi.org/10.1109/IJCNN.2018.8489605
56. Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. APL Mater. 2013;1(1):011002. https://doi.org/10.1063/1.4812323
57. Hu Y-J, Zhao G, Zhang M, Bin B, Del Rose T, Zhao Q, et al. Predicting densities and elastic moduli of SiO 2 -based glasses by machine learning. Npj Comput Mater. 2020;6(1):1–13. https://doi.org/10.1038/s41524-020-0291-z
58. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures; 2013:115–23.
59. Akiba T, Sano S, Yanase T, Ohta T, Koyama M.Optuna: a next-generation hyperparameter optimization framework. *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min*. Anchorage, AK, USA: Association for Computing Machinery. 2019:2623–2631. https://doi.org/10.1145/3292500.3330701
60. Cranmer M, Greydanus S, Hoyer S, Battaglia P, Spergel D, Ho S. Lagrangian neural networks. 2020.
61. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. *Int Conf Mach Learn*. PMLR; 2016:1050–9.
62. Wilson A, Nickisch H. Kernel interpolation for scalable structured Gaussian processes (KISS-GP). Int Conf Mach Learn. 2015;1775–84.
63. Lundberg SM, Lee S-I.A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, editors. Advances in neural information processing systems 30. Curran Associates, Inc.; 2017:4765–74.
64. Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. *ArXiv160605386 Cs Stat*. 2016.
65. Liu H, Fu Z, Li Y, Sabri NFA, Bauchy M. Balance between accuracy and simplicity in empirical forcefields for glass modeling: insights from machine learning. J Non-Cryst Solids. 2019;515:133–42. https://doi.org/10.1016/j.jnoncrysol.2019.04.020
66. Liu H, Fu Z, Li Y, Sabri NFA, Bauchy M. Parameterization of empirical forcefields for glassy silica using machine learning. MRS Commun. 2019;9(2):593–9. https://doi.org/10.1557/mrc.2019.47
67. Deringer VL, Caro MA, Csányi G. Machine learning interatomic potentials as emerging tools for materials science. Adv Mater. 2019;31(46):1902765. https://doi.org/10.1002/adma.201902765
68. Bartók AP, Payne MC, Kondor R, Csányi G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. Phys Rev Lett. 2010;104(13):136403. https://doi.org/10.1103/PhysRevLett.104.136403
69. Schoenholz SS, Cubuk ED, Sussman DM, Kaxiras E, Liu AJ. A structural approach to relaxation in glassy liquids. Nat Phys. 2016;12(5):469–71. https://doi.org/10.1038/nphys3644
70. Cubuk ED, Schoenholz SS, Rieser JM, Malone BD, Rottler J, Durian DJ, et al. Identifying structural flow defects in disordered solids using machine-learning methods. Phys Rev Lett. 2015;114(10):108001. https://doi.org/10.1103/PhysRevLett.114.108001
71. Bapst V, Keck T, Grabska-Barwińska A, Donner C, Cubuk ED, Schoenholz SS, et al. Unveiling the predictive power of static structure in glassy systems. Nat Phys. 2020;16(4):448–54. https://doi.org/10.1038/s41567-020-0842-8
72. Schoenholz SS, Cubuk ED, Kaxiras E, Liu AJ. Relationship between local structure and relaxation in out-of-equilibrium glassy systems. Proc Natl Acad Sci. 2017;114(2):263–7. https://doi.org/10.1073/pnas.1610204114
73. Jha D, Choudhary K, Tavazza F, Liao W-K, Choudhary A, Campbell C, et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. Nat Commun. 2019;10(1):1–12. https://doi.org/10.1038/s41467-019-13297-w
74. Yamada H, Liu C, Wu S, Koyama Y, Ju S, Shiomi J, et al. Predicting materials properties with little data using shotgun transfer learning. ACS Cent Sci. 2019;5(10):1717–30. https://doi.org/10.1021/acscentsci.9b00804
75. Court CJ, Cole JM. Magnetic and superconducting phase diagrams and transition temperatures predicted using text mining and machine learning. Npj Comput Mater. 2020;6(1):1–9. https://doi.org/10.1038/s41524-020-0287-8
76. Kim E, Huang K, Jegelka S, Olivetti E. Virtual screening of inorganic materials synthesis parameters with deep learning. Npj Comput Mater. 2017;3(1):53. https://doi.org/10.1038/s41524-017-0055-6
77. Huo H, Rong Z, Kononova O, Sun W, Botari T, He T, et al. Semi-supervised machine-learning classification of materials synthesis procedures. Npj Comput Mater. 2019;5(1):62. https://doi.org/10.1038/s41524-019-0204-1
78. Johansson S, Thakkar A, Kogej T, Bjerrum E, Genheden S, Bastys T, et al. AI-assisted synthesis prediction. Drug Discov Today Technol. 2019;32-33:65–72. https://doi.org/10.1016/j.ddtec.2020.06.002
79. Vaucher AC, Zipoli F, Geluykens J, Nair VH, Schwaller P, Laino T. Automated extraction of chemical synthesis actions from experimental procedures. Nat Commun. 2020;11(1):3601. https://doi.org/10.1038/s41467-020-17266-6
80. Kim E, Jensen Z, van Grootel A, Huang K, Staib M, Mysore S, et al. Inorganic materials synthesis planning with literature-trained neural networks. J Chem Inf Model. 2020;60(3):1194–201. https://doi.org/10.1021/acs.jcim.9b00995
81. Venugopal V, Broderick SR, Rajan K. A picture is worth a thousand words: applying natural language processing tools for creating a quantum materials database map. MRS Commun. 2019;9(4):1134–41. https://doi.org/10.1557/mrc.2019.136
82. Venugopal V, Sahoo S, Zaki M, Agarwal M, Gosvami NN, Krishnan NMA. Looking through glass: knowledge discovery from materials science literature using natural language processing. ArXiv210101508 Phys. 2021.
83. Kim E, Huang K, Kononova O, Ceder G, Olivetti E. Distilling a materials synthesis ontology. Matter. 2019;1(1):8–12. https://doi.org/10.1016/j.matt.2019.05.011
84. Kim E, Huang K, Saunders A, McCallum A, Ceder G, Olivetti E. Materials synthesis insights from scientific literature via text extraction and machine learning. Chem Mater. 2017;29(21):9436–44. https://doi.org/10.1021/acs.chemmater.7b03500
85. Oviedo F, Ren Z, Sun S, Settens C, Liu Z, Hartono NTP, et al. Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. Npj Comput Mater. 2019;5(1):60. https://doi.org/10.1038/s41524-019-0196-x
86. Maksov A, Dyck O, Wang K, Xiao K, Geohegan DB, Sumpter BG, et al. Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2. Npj Comput Mater. 2019;5(1):12. https://doi.org/10.1038/s41524-019-0152-9

87. Mukaddem KT, Beard EJ, Yildirim B, Cole JM. ImageDataExtractor: a tool to extract and quantify data from microscopy images. J Chem Inf Model. 2020;60(5):2492–509. https://doi.org/10.1021/acs.jcim.9b00734
88. Kondo R, Yamakawa S, Masuoka Y, Tajima S, Asahi R. Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics. Acta Mater. 2017;141:29–38. https://doi.org/10.1016/j.actamat.2017.09.004
89. Ling J, Hutchinson M, Antono E, DeCost B, Holm EA, Meredig B. Building data-driven models with microstructural images: generalization and interpretability. Mater Discov. 2017;10:19–28. https://doi.org/10.1016/j.md.2018.03.002
90. Yang Z, Yabansu YC, Jha D, Liao W-K, Choudhary AN, Kalidindi SR, et al. Establishing structure-property localization linkages for elastic deformation of three-dimensional high contrast composites using deep learning approaches. Acta Mater. 2019;166:335–45. https://doi.org/10.1016/j.actamat.2018.12.045
91. MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20):eaaz8867. https://doi.org/10.1126/sciadv.aaz8867
92. Wang C, Ping W, Bai Q, Cui H, Hensleigh R, Wang R, et al. A general method to synthesize and sinter bulk ceramics in seconds. Science. 2020;368(6490):521. https://doi.org/10.1126/science.aaz7681
93. Margrave FW, Rigas K, Bradley DA, Barrowcliffe P. The use of neural networks in ultrasonic flaw detection. Measurement. 1999;25(2):143–54. https://doi.org/10.1016/S0263-2241(98)00075-X
94. McNinch M, Parks D, Jacksha R, Miller A. Leveraging IIoT to improve machine safety in the mining industry. Min Metall Explor. 2019;36(4):675–81. https://doi.org/10.1007/s42461-019-0067-5

How to cite this article: Ravinder , Venugopal V, Bishnoi S, et al. Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century. *Int J Appl Glass Sci*. 2021;12:277–292. https://doi.org/10.1111/ijag.15881