**Editorial**

# Informatics is fueling new materials discovery

**Tong-Yi Zhang[1,2], Xing-Jun Liu[1]**

[1]School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, Guang Dong, China.
[2]Materials Genome Institute, Shanghai University, Shanghai 200444, China.

**Correspondence to:** Prof. Tong-Yi Zhang, School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, Guang Dong, China. E-mail: zhangty@hit.edu.cn; Prof. Xingjun Liu, School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, Guang Dong, China. E-mail: xjliu@hit.edu.cn

As we enter the third decade of the 21st century, we have witnessed artificial intelligence (AI) and machine learning (ML) playing an increasing and wide spreading role in our society. Materials science and engineering (MS&E) is quickly and tightly grasping AI and ML as another powerful tool, as evidenced by a worldwide push towards big data and data mining in materials to accelerate discovery.

Surrogate models are trained on available data (either from legacy in-house data, high throughput combinatorial characterization, high throughput multi-scale computations, or a combination of these three) for instant property predictions, thereby allowing for screening of promising material candidates with specific designed target properties. Questions regarding composition, processing routes to create a target material, are being explored using statistical means. Data-driven strategies are being studied to tackle unique challenges resulting from the extraordinary physical, chemical, as well as structural diversity of materials and myriad interactions spanning across multiple scales. Methods to solve inverse problems, wherein materials recommendations are made using advanced AI algorithms that meet application targets, are also being investigated.

Given the rapid changes in this field, it is challenging to understand both the breadth of opportunities and the best practices for their use. In this context, we launched the *Journal of Materials Informatics* (*JMI*), providing a platform for sharing the growing roles of ML/informatics in an array of aspects of MS&E, and

for encouraging more detailed discussion of some of the challenges and opportunities regarding using ML for predicting material properties and accelerating the design, discovery, and innovation of new materials.

The first issue collected five articles, all of which are invited contributions from leading experts in their particular areas. The cover article, "Generative models for inverse design of inorganic solid materials", authored by Chen *et al.*[1], set the stage for this discussion by introducing the general aspects of inverse materials design and providing a brief overview of two generative models, namely variational autoencoder and generative adversarial network, which can be utilized for generating and optimizing inorganic solid materials according to their properties. The authors compared reversible representation schemes for generative models between molecular and crystalline structures and discuss the challenges to the latter. They also summarized the recent applications of generative models in the exploration of chemical space with compositional and configurational degrees of freedom and pointed out potential future directions.

In "Generative deep learning as a tool for inverse design of high entropy refractory alloys", Debnath *et al.*[2] continued to address the issue of inverse design and presented novel computational infrastructure and workflow for the inverse design of new alloys powered by the generative deep learning method. In a case study regarding the design of high entropy alloys (HEAs), the authors employed the generative models as an inexpensive, low-fidelity approach for generating new and interesting HEA samples, which were then paired with more expensive, high fidelity experimental validations. The preliminary results well demonstrate the great potentials of generative models in learning complex relationships for inverse design of materials on demand, making them a valuable tool for material informatics.

It is also important to point out that the data-driven mode will in no way replace the other two modes. Instead, it is a powerful extension and upgrade of conventional cognitive paradigms. As presented in the third paper "Modelling of phase stability: integrating computational materials science and materials informatics", Song *et al.*[3] first gave an overview about the current status of the computational studies on phase stabilities in alloys in general and in Sm-Co permanent magnetic material in particular. The authors also compared various computational models for studying phase stabilities and data-driven approaches, through which the advantages of the latter are identified. They further proposed a step forward for developing interpretable ML algorithms through dynamic interaction between ML and multi-scale simulations.

The article titled "Prediction of the atomic structure and thermoelectric performance for semiconducting $Ge_1Sb_6Te_{10}$ (GST) from DFT calculations[4]" leads us away from materials informatics by discovering atomic structure of GST through *ab initio* calculations. Gan *et al.*[4], using first-principle calculations in conjunction with the Boltzmann transport theory, carried out a systematic study on the atomic stacking configurations, electronic structures, transport properties, as well as thermoelectric performance of layered GST, known as the most popular phase change materials for data-storage applications. This study led to the discovery of two novel atomic arrangements, namely GST-II and GST-IV, with the former and the latter being a semiconductor and a semimetal, respectively. The unique but distinctive properties associated with these two atomic stacking, discovered by *ab initio* calculations, highlight the great potential of GST for thermoelectric applications and inspire further experimental study. The authors further envisaged from the perspective of materials informatics that crystals with multilayered structures may open a viable route for the discovery of new functional materials.

Finally, Lou *et al.*[5] gave an update of the development of robust surfaces for harsh service environments from the perspective of phase transformations, extending methods, tools (both experimental and

computational), and databases to embrace metastable the formation of metastable phase(s). The dawn of such experimental and computational databases will make the integration of analysis, prediction, and discovery using data-driven approach the key theme in the accelerated research for coating materials in the near future.

## DECLARATIONS

**Authors' contributions**
Preparation of the manuscript: Zhang TY, Liu XJ

**Availability of data and materials**
Not applicable.

**Financial support and sponsorship**
None.

**Conflicts of interest**
Both authors declared that there are no conflicts of interest.

**Ethical approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Copyright**
© The Author(s) 2021.

## REFERENCES

1. Chen L, Zhang W, Nie Z, Li S, Pan F. Generative models for inverse design of inorganic solid materials. *J Mater Inf* 2021;1:4. DOI
2. Debnath A, Krajewski AM, Sun H, et al. Generative deep learning as a tool for inverse design of high entropy refractory alloys. *J Mater Inf* 2021;1:3. DOI
3. Song X, Guo K, Lu H, Liu D, Tang F. Modelling of phase stability: integrating computational materials science and materials informatics. *J Mater Inf* 2021;1:[Accept]. DOI
4. Gan Y, Zhou J, Sun Z. Prediction of the atomic structure and thermoelectric performance for semiconducting $Ge_1Sb_6Te_{10}$ from DFT calculations. *J Mater Inf* 2021;1:2. DOI
5. Lou M, Xu K, Chen L, et al. Development of robust surfaces for harsh service environments from the perspective of phase formation and transformation. *J Mater Inf* 2021;1:5. DOI