Chapter 4

Artificial Intelligence in Material Genomics

Joy Hoskeri H¹,*, Nivedita Pujari S¹, Badrinath Kulkarni², Arun K. Shettar³
¹Department of Bioinformatics, Karnataka State Akkamahadevi Women’s University, Vijayapur-586109, Karnataka, India
²P.C. Jabin Science College, Vidya Nagar, Hubballi -580031, Karnataka, India
³Cytxon Biosolutions Pvt. Ltd., Lingaraj Nagar, Hubballi -580031, Karnataka, India
joybioinformatics@gmail.com

Abstract
The invention of new materials with desired properties is always a matter of interest. The material genome projects and material genome initiative have bought new insights into material genomics. Artificial intelligence (AI) is the decision-making ability of a computable machine. AI in material genomics has stimulated the field of material science. AI tools like Atom2vec, MATLAB, ICSD, MPIinterfaces, PyCDT, and AFLOWLIB have opened ways for the discovery of various materials. These AI tools and databases are also efficient in the property prediction of new materials, improvement in characterization protocols, experimental parameter standardization, fastening simulation scale, development of high throughput methods, and data analysis. The current chapter is focused on AI-based developments in the material genomics.

Keywords
Material Genomics, Artificial Intelligence, Machine Learning, AI Tools, and Databases

Contents

Artificial Intelligence in Material Genomics ...................................................87
1. Introduction ...................................................................................................88
2. Material genomics .......................................................................................89
3. Strength of artificial intelligence ...............................................................92
4. Artificial intelligence in material genomics

Conclusion

References

1. Introduction

Materials in general are an important part of science and technology. For a material to be desirable for use in different fields, it has to possess certain unique attributes such as durability, economic friendly, sustainability, etc. [1]. These properties are important for the commercialization of materials. Material science is the domain of material research that mainly deals with the design, manufacturing, innovation, and development of novel materials with a wide range of applications. Material genomics approach is an important contribution to the material genome initiative (MGI) introduced by Barak Obama, the former President of the United States [2]. The initiative mainly focused on bringing scientists and researchers involved in various disciplines together for the design and discovery of new materials. It mainly aimed at making commercialization easier for such novel materials, thus bringing the market place closer to research. Material genomics is the combination of computational tools, experimental tools, and data analysis. It makes use of virtual screening and could be used for designing and standardization \textit{in-silico} [3]. Artificial intelligence is an emerging field that has applications in various fields. AI because of its ability to mimic human intelligence, has made several processes easier, time efficient, and also cost-efficient [4]. AI has several applications including health care, agriculture, industries, and material science [5, 6]. Two important domains, of material genomics, which include deep learning and machine learning have made numerous changes in the field of material science [7]. With the help of ML and DL, several tools, databases, and high throughput methods to solve material science problems are being discovered. These developments have made a remarkable change in material genomics. Artificial intelligence material genomics has developed tools for framing the periodic table of elements [8], predicting material properties with the aid of tools based on machine learning [9], pattern reorganization [10], and parameters optimization of materials [11]. The flow chart (Fig. 1) summarizes the functioning of material genomics with the help of artificial intelligence.
Figure 1. Flow chart depicting the contribution of AI in the acceleration of developments in materials genomics.

The current chapter is focused on development in material genomics caused due to material genome project and material genome imitative, artificial intelligence and its role in improving material genomics, various tools, datasets, and methods reported so far that help in the efficient screening of novel and potential materials.

2. Material genomics

Materials are strong grounds for technology and innovation, discovery and commercialization of new, novel, and advanced materials which play a vital role in solving various problems associated with economic growth, innovation, and the environment [1]. However, the discovery of new material includes the hit and trial method and also consumes time to standardize a material for technological application and commercialization. Therefore, an organized and scalable approach that supports both researchers and the material community is required. The concept of material genomics started in 2011 and was announced as a material genome initiative (MGI) [2]. MGI confronted the societal science and engineering platforms to speed up innovations of novel materials which help in innovative designing, manufacturing, eco-friendly machines, and many more [12]. MGI encouraged scientists to integrate it with other fields to develop genuine products at an increased rate. Material genomes were brought up and studied to innovate, research and build novel materials that can address problems and help in improving social well-being [13]. Material genome initiative or MGI was initiated by US
president Barak Obama to enhance global competitiveness by reducing the time and cost to half for bringing novel materials from research labs to market place in the nation [2].

MGI is an extensive association formed with scientists of both experimental and computational backgrounds to develop scientific computational methods to screen, interpret and standardize at an unbiased scale and rate. The material genomics approach (Fig.2) consists of three important requirements such as experimental tools, computational tools, and digital data (Figure 2). Several research groups have employed material genomics approaches, some of which are discussed here: one of the studies done by Khaira et al. combined small-angle X-ray scattering, physics-based molecular modeling, and progressive standardization to frame films for experiments [14]. Another study reported the use of high-throughput virtual screening that combines machine learning, quantum chemistry, and experimental optimization to explore around 1.6 million OLED molecules [15]. One of the studies applied a combination of quantum mechanical stimulations to develop in-silico polar metal showing variable stability. This theory-guided material genomics approach revealed an infamous class of materials that can be important concerning innovations and technologies [16].

Research of high-throughput materials with string physiochemical properties is quite challenging, however, the material genomics approach makes use of lesser mechanical sensitivity and high-energy materials for the innovation of materials with higher energy density. Such an approach was used in a study to accelerate the process of innovation of novel insensitive high-energy explosives by distinguishing rapid genetic features, screening, and molecular design, through which they synthesized two target molecules 2,4,6-triamino-5-nitropyrimidine-1,3-dioxide [17].

Figure 2. Material genomics approach
The material genomics approach utilizes image screening of materials that can be used for standardizing materials \emph{in-silico} [3]. Many studies have already used this highly efficient approach to analyze thousands of compounds as efficient technological materials. Some of the putative examples are listed below in Table 1.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Sl.No. & Examples & Reference \\
\hline
1 & Thermoelectrics & [3] \\
2 & Hydrogen storage material & [17] \\
3 & Topological insulators & [1] \\
4 & Solar photovoltaics & [18] \\
5 & Li-ion batteries & [16] \\
\hline
\end{tabular}
\caption{Different approaches in material genomics.}
\end{table}

Another recent development in material genome includes the integration of computational tools and information technology, this methodology has allowed access to different computed material datasets to various research communities and has allowed new collaborations for material discovery [19]. The material project, an important part of the material genome initiative aims to speed up material design by generating open collaborative systems that target each step in the design process (data acquisition, validation, analysis, and design) [20]. Polymer nanocomposites are a diverse class of materials, whose nano-range particles, chemistry, and polymer resin combination have made them potential compounds [21]. With the help of the material genome initiative, a data-guided web-based platform called nano-mine has been developed. Nano-mine can analyze and construct polymer nanocomposite systems under the material genome initiative concept [22].

Machine learning is a fast-growing field that has shown its way in several fields. A detailed review [23] is available on materials databases, analytical tools, and material data used in machine learning algorithms. Such an approach is designed to achieve extraordinary results in material discovery and design.

Material genomics, therefore, aims at connecting researchers of different communities with various technologies that help in bringing novel materials to the marketplace. Therefore material genomics approach can be summarized as a combination of innovation, discovery, development, and collaboration. Material genomics makes use of researchers that work on computational tools and then connect them with experimental approaches for further validation. A combined output of such an approach is analyzed using databases. Since the approach is interdisciplinary and benefits people of different backgrounds, material
genomics has emerged as a potential field. An interrelationship of the material genomics approach is presented in Figure 3.

![Figure 3. Interrelationship of material genomics with various fields and domains.](image)

3. **Strength of artificial intelligence**

Artificial intelligence (AI) is a newly emerging branch of science and technology that is based on computer programs to perform various tasks of human needs. The field is based on the concept of – “Can a machine think?” [4]. AI refers to mimicking human intelligence by a computational system or a machine. Such decision-making technology is used in various fields including health science, machine learning, material science, etc. The first breakthrough of AI was back in 1996 when Geoffrey Hinton and his co-workers proposed an approach to building deeper neural networks [24]. This approach redefined AI research and also different algorithms like deep learning (DL) and machine learning (ML). DL is a part of ML that forms multiple layers of neural networks which represent learning [7]. However, ML is part of AI which a computer system or program can use to learn and acquire intelligence without human intervention. A few implications of artificial intelligence are given below in Table 2. (Fig.4)
Table 2. Implications of artificial intelligence.

<table>
<thead>
<tr>
<th>SL. No</th>
<th>Area/field</th>
<th>Purpose</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intensive care unit</td>
<td>AI tools function as intelligent assistants to clinicians for constant monitoring and thus help in reducing costs and improving the efficiency of machines.</td>
<td>[5]</td>
</tr>
<tr>
<td>2</td>
<td>Higher education</td>
<td>AI tools in education are used for assessment, evaluation, profiling, and intelligent tutoring and help in the dynamic development of the education system.</td>
<td>[25]</td>
</tr>
<tr>
<td>3</td>
<td>Covid-19</td>
<td>AI tools and databases are used to address challenges posed by covid19 at molecular, clinical, and diagnostic levels.</td>
<td>[6]</td>
</tr>
<tr>
<td>4</td>
<td>Urology</td>
<td>AI tools are used for accurate prediction and analysis of large data charts to facilitate personalized and evidence-based care of patients.</td>
<td>[26]</td>
</tr>
<tr>
<td>5</td>
<td>Solid waste management</td>
<td>Software of AI and machine learning are used for waste characteristic prediction, prose output prediction, bin level detection, and process parameter prediction.</td>
<td>[27]</td>
</tr>
<tr>
<td>6</td>
<td>Cancer genomics</td>
<td>AI applications are used in testing cancer genetics and for variant calling in diagnostics.</td>
<td>[28]</td>
</tr>
<tr>
<td>7</td>
<td>Health care</td>
<td>AI application in health care helps in data acquisition, algorithmic transparency, and real-time assessment and also helps in preparing patients and physicians for modern and digitalized healthcare</td>
<td>[29]</td>
</tr>
<tr>
<td>8</td>
<td>Agriculture</td>
<td>The application of AI in agriculture aims at accuracy, high performance, flexibility, and cost-effectiveness of soil, disease, and weeds management.</td>
<td>[30]</td>
</tr>
</tbody>
</table>
4. Artificial intelligence in material genomics

The collaboration of material genomics and artificial intelligence has brought in a new dimension of studies in material science. Material genomics which mainly dealt with the design, manufacture, and discovery of new materials was brought to light by the material genome initiative project as discussed earlier. To modernize, and initiate new approaches to the material genome, artificial intelligence was merged in. Various advances and research associated with material genomics and artificial intelligence are discussed here.

Artificial intelligence emerging with material science has brought in a new area of development in material science[31]. After a research period of over 60 years development has taken place from basic perceptron to complex neural networks[32][33]. AI has evolved as an important algorithm framework and has proven its efficiency in various fields. Such an extraordinary data-storing ability of artificial intelligence has gained a broad range of attention from researchers of material science. AI is a data-driven science that compresses large data information into undefined theories that guide scientific innovations [34]. Such an approach can address many problems in material research, where large-scale composite space and nonlinear processes are used. The combination of artificial intelligence with material science is defined as material informatics, it is an interdisciplinary branch that helps scientists to efficiently obtain unknown relationships between different aspects, determines specific properties of materials, optimizes different methods of chemical
synthesis and process parameters and also improves the quality of current characterization methods of materials. Machine learning (ML) is an important part of AI. Material learning research in the field of material science is growing rapidly, where it has a major role in the synthesis of new materials and the prediction of various routes of synthesis [35]. The relationship between AI and material science can be explained in three generations (Figure 5). The first generation mainly deals with structure performance which is predicted by the local optimization algorithm and the performance of materials present in the structure. The second generation follows a global optimization algorithm that can carry out crystal structure elemental composition predictions. The third generation refers to statically oriented designs that make use of machine learning algorithms to find out the structure, configuration, and efficiency of chemical and physical data [36, 37]. The current research process involving the combination of material science and artificial intelligence refers to the third-generation relationship.

Some of the important aspects of research and developments of artificial intelligence in material genomics are discussed below:

**Atom2vec**: is a tool for the predictive output of artificial intelligence in material science; the tool is designed on the bases of a machine learning concept, which rebuilt the periodic table in a few hours. Atom2vec works by initially differentiating atoms by investigating the series of compounds available in the online database. Later with the help of natural language processing, the characteristic of each word is derived. Chemical compounds are grouped according to the environment of the chemical. This can provide an efficient way of quantitative portrayal in the coming years [38].
Property prediction of new materials: Researchers of material science generally focus on optimizing desired properties such as electrolytes conductivity, power transformation potential of organic-inorganic hybrids, and Seebeck coefficient of thermoelectric materials [39]. The process of optimizing such properties is time-consuming and based on trial and error experiments where many times the experimental results are not satisfactory. Machine learning approach of AI devolved models that can predict shapes of materials with considerable accuracy before synthesis. One such model is MATLAB; this can be used to search for a minute amount of electrolytes that are solid among more than 12000 materials [40]. Similar methods like MATLAB are applied to design monochromatic catalysts [41, 42], light-emitting diodes [9], organic–inorganic hybrid perovskite [43], and organic light-emitting diode (OLED) [44]. Therefore, materials science has emerged not just as a trial and error method it is also efficient in reducing the number of experiments.

Improvement in characterization protocols: For material preparation using high throughput methods, material genome projects and analysis of data using AI tools have become more evident in recent years [45]. Once such development is in use of neural networks in deep learning, this approach has made considerable achievement in recognition [46]. Image characterization of micromaterials can be easily transferred by pattern recognition. Such images can be easily visualized by electron microscopy to analyze the microstructures and properties of various materials. With the help of AI, automatic defect reorganization and classification can be incorporated into electron microscopy, so that a large number of images can be screened for statically significant information. Such an approach was used by Li et al. where material information such as size and kind of defect was identified using combined machine learning, computational, and image analysis approaches [47]. Similarly, data obtained by X-ray diffraction can also be analyzed using a machine-learning approach [48].

Experimental parameter standardization: In conventional material developments, several parameters have to be considered, analyzed, and also adjusted manually during synthesis. The process is less efficient and might not find optimum parameters. The machine learning approach has the ability of nonlinear regression that can be used to search suitable areas or areas in a large diameter space [44]. One of the fields where such an approach is used is the welding process, called friction stir welding. It is a new solid-state welding process that can be used in shipbuilding, automobile, and aerospace. A study involved 108 independent experimental data for training with machine learning models that included decision trees and neural networks. The study focused on parameters of original welding such as stain rate, shear stress, temperature, and strong variables on void formation [45]. The algorithm of AI used here can find 96.6% of defects. Making use of such model standardization of parameters in the welding process can be fulfilled, and the
occurrence of unfavorable parameters such as void formation in FSW from ML can be avoided. In future years, the process of material synthesis will be fully automated and combined with industries for manufacturing, for example, digital manufacturing of polymers from high-throughput programs [49].

**Fastening simulation scale:** machine learning approach can find out irregular repetition present in the calculation of atomic force field theoretically, and the corresponding energy can be easily calculated. Another important feature is that movement of several atoms can be enlarged into millions of atoms within a period of a few Pico-seconds. This strategy greatly increases the time range and length of simulation calculations to give better results. Apart from these complex material structures like crystallinity and amorphous, chemical reactions like interfacial reactions and corrosion can be stimulated. Such interatomic potential based on artificial neural networks has provided an unbiased way for building surface systems that are difficult to elaborate by conventional potential. A study carried out by Artrith et al. made use of zinc oxide and copper as a reference system to validate the efficiency of the interatomic strategy of artificial neural networks and explored the CuZnO ternary combination system of oxide-supported copper clusters [46].

**Databases:** Databases play important role in any research domain of science and technology. In the field of material science, some of the important parameters such as defects in patterns, physical properties of the compound of interest, ionization potentials, electron facilities, and numerical quantity of material properties associated with research studies are maintained in well-equipped and well-established databases. One such database is ICSD [50] database, which contains data extracted from experimental outcomes. It mainly consists of original data that are reported in various scientific articles. Later the information repositories are further screened for potential material with desired applications, with the aid of high throughput artificial intelligence methods [51]. Newly framed databases consist of sophisticated algorithms for semi-automated and automated representation and interpretation of information such as scripting language retrieval. Some of the leading databases include material project [20] and the open quantum materials database [51], which contain around 105 density functional theory (DFT) calculated crystal structures.

**High throughput methods:** this method of analysis is one of the prominent procedures used by researchers of several disciplines. It has been a potential tool for finding out material innovation requirements over recent years [52]. Such type of screening is commonly seen in pharmaceutical industries for screening potential compounds for drug development. However, investigation in material sciences is associated with computational tools that are employed for the synthesis of compounds experimentally, leading to the
discovery of novel magnetic and dielectric materials [53]. Figure 6 explains the workflow of high throughput methods.

![Diagram of high-throughput material synthesis workflow](image)

**Figure 6. The flow of high-throughput material synthesis**

Materials synthesized by high-throughput methods have targeted properties. AFLOWLIB is one of the huge initiatives in computed data [95]. It consists of an elaborated database of compounds, metastable structures, and binary phase diagrams computed in the AFLOWLIB workflow [54].

**Data analysis:** Data analysis appears as one of the limiting factors in material science in the discovery of new materials. Even though novel techniques are discovered for manufacturing and identifying numerous novel compounds and computational tools are required to analyze such large data. Various studies have proposed different software infrastructures, the majority of those developed by research groups, these support storage, collection, retrieval, and analysis of data generated by various material stimulators. Some of the examples include MPI interfaces [55]. This tool automates high-throughput computational screening and investigation of interfacial systems, which enables computational tools-based screening of nanoparticle/ligand pairs by analyzing various degrees of freedom within a system. Matminer is software that facilitates data-oriented methods for predicting and analyzing material properties [55]. PyCDT is user-friendly software and provides a direct interface with the material project database [20].
Conclusion

Material genomics is an upcoming branch of material science mainly supported by the material genome initiative, it mainly aims to bring researchers from laboratories of different material science research and market place of materials closer. However, with an increase in material science research outcomes, novel materials with varied properties and parameters are being discovered. Therefore, manual screening, analysis, and interpretation are time-consuming and also need huge financing. The intervention of artificial intelligence in combination with material learning, machine learning, and deep learning algorithms has accelerated the field of material science. AI has bought in new tools, high throughput methods, and databases for analyzing thousands of compounds to find and develop potential and functional materials and further synthesis of the same. Therefore artificial intelligence in material genomics has significantly stimulated material science research and bought in new materials of potential interest.

Declaration: All the figures and tables mentioned in the manuscript are photographed and tabulated by authors and hence no copyright permissions are required.

References


[25] O.Z. Richter, V.I. Marin, M. Bond, F. Gouverneur, Systematic review of research on artificial intelligence applications in higher education—where are the educators?, Int. J.


