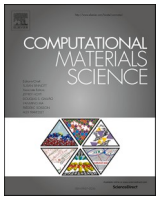
Computational Materials Science 248 (2025) 113588

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci



Full Length Article

Consolidated database of high entropy materials (COD'HEM): An open online database of high entropy materials

Mohit Singh, Eric Barr, Dilpuneet Aidhy *

Department of Materials Science and Engineering, Clemson University, Clemson, SC 29634, United States

ARTICLE INFO

Keywords:
High entropy materials
Mechanical properties
Database
Open-source website

ABSTRACT

The high entropy approach to materials design has provided access to an exponentially large compositional phase space. There are a multitude of ongoing research efforts that are continuously producing large amounts of data in the scientific journals. The community is actively and regularly consolidating the data in review articles. However, there is a pressing need to curate and digitize data in a format that is readily accessible for scientific analysis. This is particularly important to foster machine learning models that are rapidly being integrated in various materials design approaches. We present an open, web-accessible centralized database, namely Consolidated Database of High Entropy Materials (COD'HEM), that currently consists of experimentally-measured mechanical properties of high entropy alloys published over the past many years. The unique aspect of the database is that it can be queried online via a variety of filters including composition, properties and phases, thereby allowing immediate comparison and analysis of all alloys in the database. The paper provides brief overview of current features and functionalities, data collection and consolidation process, and website architecture and design.

1. Introduction

The field of materials science is experiencing an unprecedented surge in data generation, driven by advances in experimental techniques and computational methods. To leverage this wealth of information, various repositories and databases have been established, enabling researchers to share and access critical data on material properties, structures, and performance [1–8]. These platforms are essential for promoting collaborations, accelerating discoveries, and driving innovations across the disciplines.

High entropy materials (HEMs) present unique challenges due to their vast compositional space. These materials are composed of multiple principal elements in large concentrations. They require extensive exploration across a wide range of compositions and processing conditions, resulting in a significant volume of diverse data. While the data is being consolidated in timely review papers [9–14], much of this data remains underutilized, as it is often not digitally curated in a way that facilitates comprehensive analysis and comparison. To advance the discovery and optimization of HEMs, there is an increasing need to systematically curate and digitize the data, making it more accessible and actionable for the scientific community.

2. Consolidated database of high entropy materials (COD'HEM)

We have developed an open, web-accessible centralized database of high entropy alloys, namely Consolidated Database of High Entropy Materials (COD'HEM), that currently consists of experimentally-measured mechanical properties of high entropy alloys (HEAs). The data has been obtained from the published papers in the past two decades. Some important features of the database are:

- the data has been extracted from the tables published in prior papers,
- current database consists of > 4000 compositions obtained from > 400 papers,
- each datapoint is tagged to the digital object identifier (DOI) of the paper,
- each alloy has a unique_id tagged to the DOI,
- the data can be queried live on the website among various chemistries and compositions by selecting from the periodic table launchpad,
- the database can be queried based on different mechanical properties and phases,
- the data can be compared against the rule-of-mixtures (ROM),

* Corresponding author.
E-mail address: daidhy@clemson.edu (D. Aidhy).

https://doi.org/10.1016/j.commatsci.2024.113588
Received 24 October 2024; Received in revised form 29 November 2024; Accepted 1 December 2024
Available online 6 December 2024

- all or part of the filtered data can be downloaded for further analysis
- the database is constantly updated.

3. Current features and functionalities

Fig. 1 shows the periodic table of elements as the landing page of the website. On the right, the database can be queried based on different compositions, where the window of the elemental concentration can be set. On the left, the database can be queried based on different properties. Based on these queries, the filtered data is displayed. Each composition has a unique_id, and multiple data of a composition are documented with a version number (e.g., composition_1, composition_2, etc.).

The filtered data can be plotted between different properties to compare different alloys. For example, Fig. 2a shows a plot between ductility and yield strength of the queried compositions. Data visualization is the unique feature of the website that enables comparison of properties among the selected alloys, allowing opportunities to capture patterns, trends, and relationships. Such visual analysis is crucial for forming hypotheses and driving deeper investigations into material properties. Beyond merely presenting data, these graphs also prompt new questions, encouraging researchers to explore mechanistic behaviors, which can lead to further scientific insights. For example, we observed that addition of a softer element (Ti) to a stronger alloy Mo alloy, shown in green in Fig. 2b, increases the yield strength, which is unintuitive. The conventional trend, i.e., strength increase by adding a stronger element (e.g. Mo) is observed in various other alloys, such as alloys shown in blue and purple in Fig. 2b. Our follow-up DFT investigation revealed that the softer element induces lower bond stiffnesses which increases the local lattice distortion, that ultimately leads to solid solution strengthening thereby raising the yield strength. This is one example of identification of patterns due to data visualization. Looking forward, we aim to expand our visualization tool capabilities, including developing methods to extract data directly from plots and converting into a tabular format that can be further analyzed. This advancement will improve the ability to fully utilize graphical data and integrate it into our research database, enhancing both the scope and depth of analytical capabilities.

An additional feature is the interactive interface in which hovering a cursor over the data-point reveals its corresponding DOI and the unique_id, as shown in Fig. 2a. The data can be further filtered by phases and properties of alloys. Finally, experimentally-obtained property value can be compared to ROMs to develop a perspective on the deviation that may be expected in an alloy compared to its individual elements. Fig. 3a shows the comparison of modulus of elasticity (E) between ROM and experimental data. The plot shows that single phase BCC and FCC agree very well, whereas multi-phase alloys don't, as expected. Fig. 3b shows the comparison of yield strength between ROM and experimental data. The agreement is poor because the yield strength is largely guided by the microstructure (i.e., solid solution strengthening), which is absent in pure metals, and not captured by ROM.

Fig. 4 is a snapshot of the website Dashboard summarizing the total

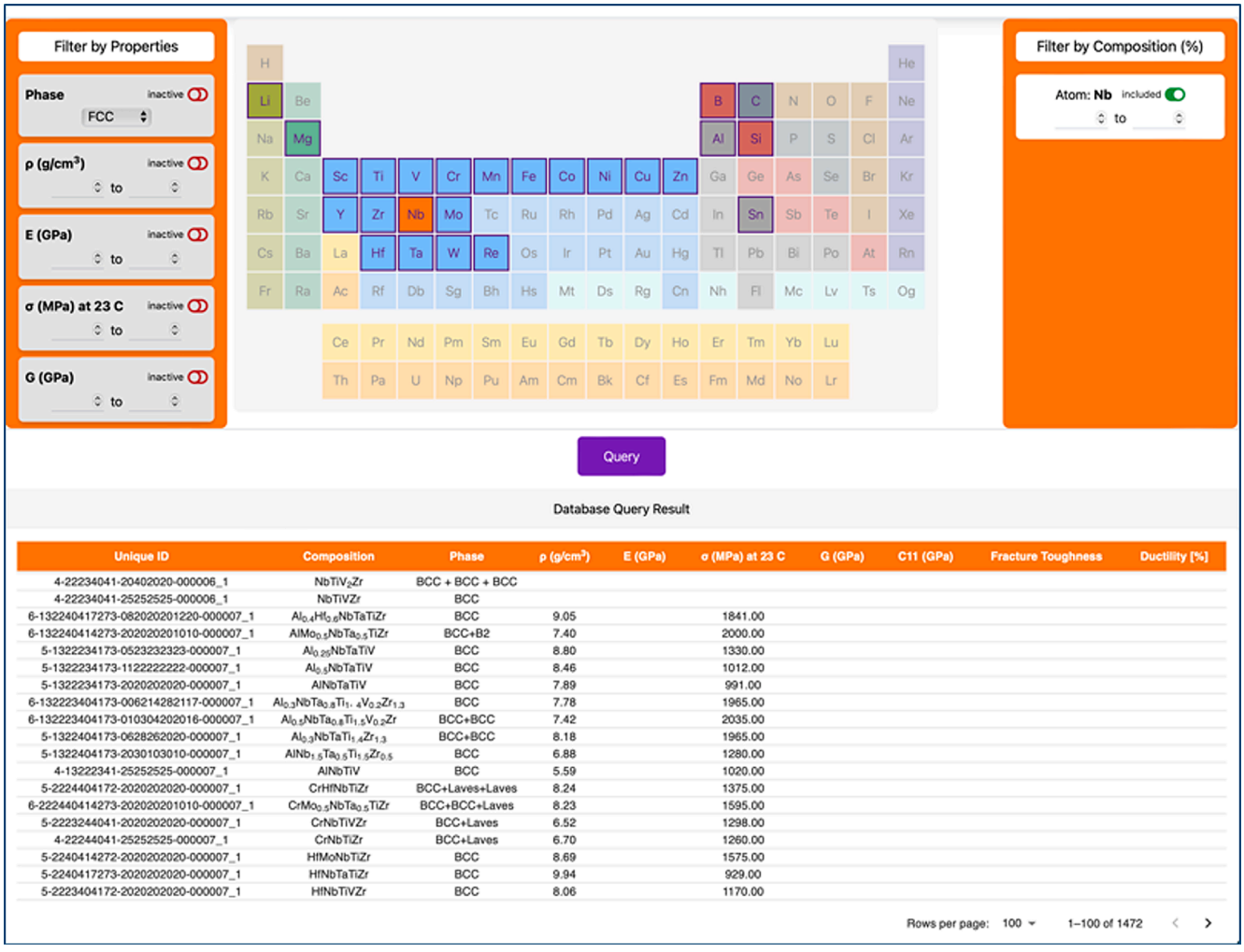


Fig. 1. Snapshot of the periodic-table based materials property search. The materials can be searched based on property-filter on the left of the panel, and/or composition-filter on the right of the panel.

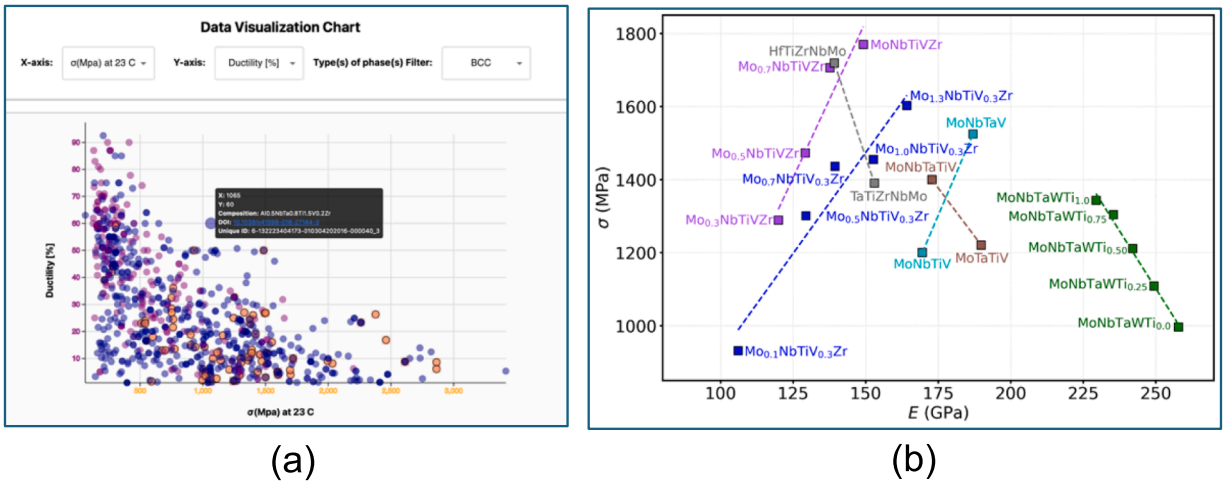


Fig. 2. (a) Snapshot of the plotted data queried between ductility and yield strength. (b) Data reveals unintuitive strength vs elasticity trends. The Ti-alloys show unintuitive strength increase by adding softer Ti, whereas Mo-alloy show conventional behavior.

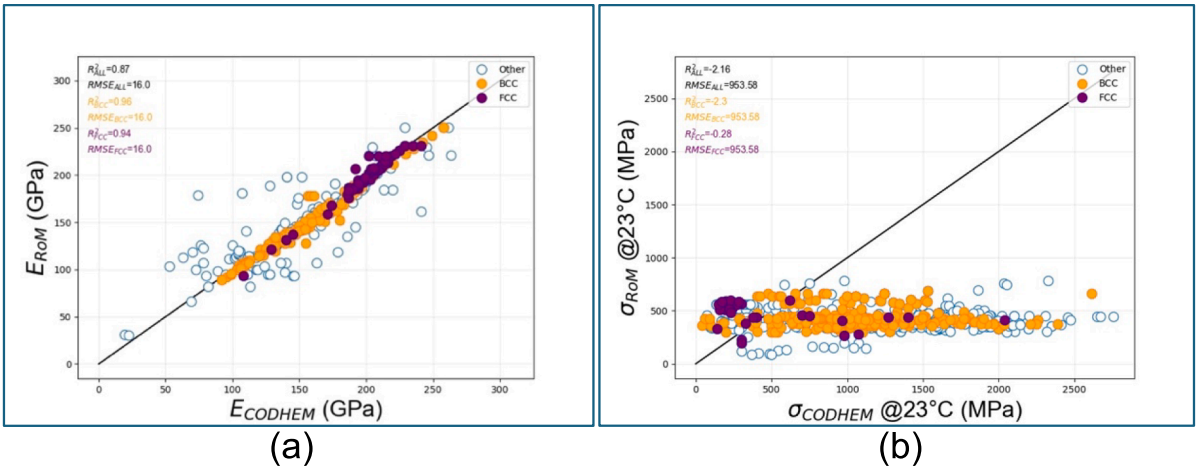


Fig. 3. (a) Comparison of modulus of elasticity between the literature data (CODHEM) and the rule of mixtures (RoM) data. (b) Comparison of yield strength (at room temperature) between the CODHEM and RoM data. Orange data points are BCC alloys, purple data points are FCC alloys, and empty circles correspond to multiphases.

number of compositions, number of DOIs, phases of alloys, distribution of alloys with different number of elements, etc. The dashboard is built in MondoDB framework and is automatically updated and provides an overview of the latest data availability.

4. Data collection and consolidation

The current version of the database focuses on the mechanical properties of high entropy alloys gathered from papers published over more than ten years. The main sources of the papers are Google Scholar and ScienceDirect. The former offers unique access to resources, including full-text downloads from ScienceDirect, which is pivotal for efficient, large-scale data collection. ScienceDirect allows bulk download of tens of papers at once, streamlining the process by facilitating the rapid acquisition of large data volumes. Additionally, we utilize large language models (LLMs) such as ChatGPT and SCISPACE for targeted searches, to refine the results to meet specific research criteria. All papers are systematically named and organized to maintain a curated dataset for subsequent analysis. This methodical approach not only optimizes efficiency but also ensures the quality and relevance of the collected data.

Data extraction is a crucial part of the process that requires a variety of tools. Primarily, we use Tabula for its user-friendly interface, allowing for a quick conversion of tables in pdfs to.csv format. Additionally, Nanonets is used to handle complex data structures. Looking ahead, we plan to develop techniques to extract data directly from graphical plots, converting visual data points into a structured tabular format for enhanced analysis. This advancement will significantly broaden the scope and depth of data utilization, improving the analytical capabilities to evaluate HEAs. With the advancements in the LLMs, there is greater ability to extract and analyze data at a faster rate. ChatGPT is also employed to extract data from tables in.csv format. All papers are tagged via ChatGPT to curate them in categories such as microstructure, processing methods, properties, etc. Finally, a primary conversational/prompt tool is created to deal with any inquiry regarding the content of the papers. This is a multi-purpose tool that converses with the user, learns papers' contents and compares data when queried. These LLM based tools has greatly increased productivity and growth of the website's content.

The process of uploading data uses DOI from the original papers, which links each data piece directly to its source. This linkage provides a permanent and verifiable connection to the original research, ensuring proper attribution to the original works. The data formatted into.csv files plays a critical role during the uploading phase. These files are

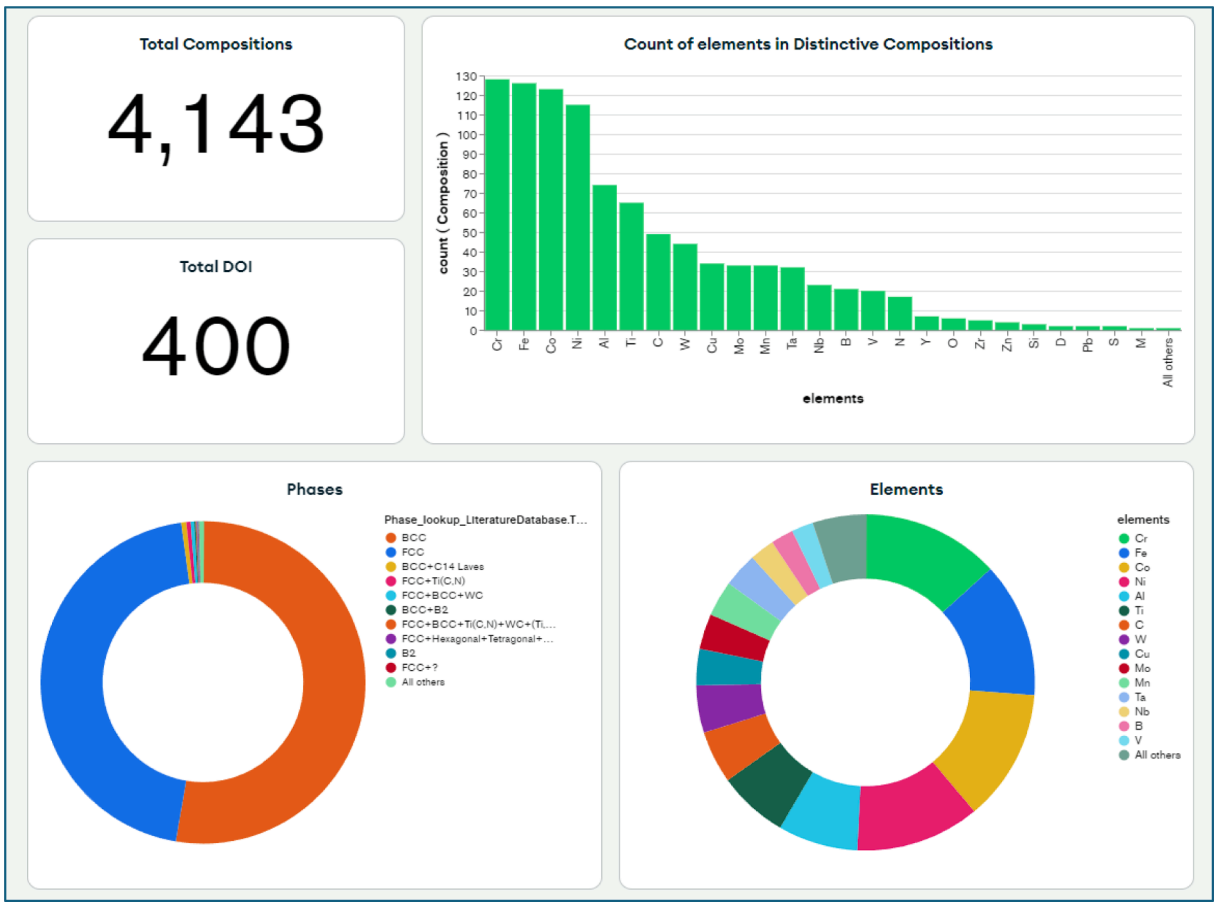


Fig. 4. Snapshot of the dashboard of the website providing overview of the available data.

meticulously prepared with data arranged in structured rows and columns, complete with appropriate headers. This organization is essential for smooth integration into the research database, promoting consistency and data reliability. The uploading steps are as follow: first, the. csv files are uploaded to ensure all relevant data is included in the database. Second, each data set is linked to its corresponding DOI, establishing a clear connection to its source. Finally, the data is interrogated for verification and validation process to ensure accuracy and address any discrepancies, thus maintaining data integrity. Data visualization capability is a direct and convenient way to verify and validate the data. Any anomaly such as an outlier data for a yield strength of an alloy deviated from rest of its composition family, is noted and corrected.

In addition to the technical steps, the final phase of the uploading process involves structuring the website to present the data effectively. The website is designed to offer a comprehensive overview of high entropy alloys' compositions, properties, and origins in the field. Future enhancements planned for the website include adding detailed sections on material properties, alloy compositions, and their industrial applications, as well as summaries of key research insights and directions. These improvements aim to enhance the website's navigation and usefulness, making it a valuable resource for researchers and industry stakeholders alike.

5. Website architecture and design workflow

The website's architecture is designed for performance, scalability, and security, hosted on Clemson Computing and Information Technology (CCIT) servers. It consists of several key components working together seamlessly:

- Nginx**:** acts as the gateway, managing HTTP requests between users and the platform.
- Backend: The core server is built with Go using the Echo framework, handling routing, business logic, and database interactions, ensuring efficient request processing and data management.
- Databases: PostgreSQL is used for structured data, while MongoDB handles unstructured data, offering flexibility and robust data handling.
- Frontend: Developed with ReactJS and TypeScript, the frontend delivers a dynamic, responsive user experience, with TypeScript enhancing code reliability through strict typing.
- Version Control: GitHub is used for version control, supporting collaborative development across teams.

When a user interacts with the platform, Nginx routes their request to the Go server, which processes the request, communicates with the databases, and sends back the necessary data. This ensures a fast and seamless experience for the user. Fig. 5 provides overview of the website architecture.

The platform is designed as a Single-Page Application (SPA), which enhances user experience by significantly improving load times. By dynamically updating content without reloading the entire page, users can interact with the application more fluidly and efficiently. This architecture allows for seamless transitions between different views and features, reducing wait times and providing a more engaging interface. As users navigate through the platform, only the necessary data is fetched and rendered, minimizing server requests and optimizing performance. Overall, this approach not only speeds up interactions but also creates a more responsive and enjoyable user experience.

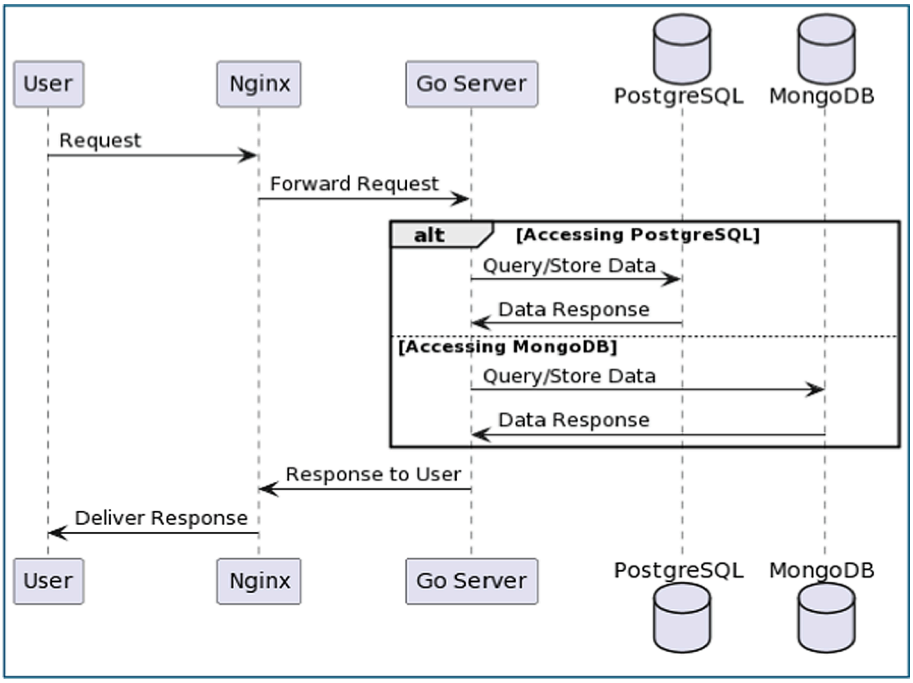


Fig. 5. Overall website architecture.

5.1. Server and Client

The backend is built using Go and the Echo framework, which allows for efficient routing, middleware integration, and robust performance. Go was chosen for its strong concurrency model, enabling the application to handle multiple requests simultaneously. Echo's built-in features, such as secure CORS handling and logging, enhance the overall security and reliability of the API.

For the frontend, ReactJS and TypeScript are used to create a dynamic and responsive user interface. ReactJS facilitates the development of reusable components, while TypeScript helps reduce errors and improve code quality through strict type checking. This combination ensures a smooth user experience and facilitates easier maintenance and scalability as the platform grows.

To streamline deployment and enhance the development workflow, Docker is employed to containerize both the backend and frontend applications. Docker simplifies the packaging of applications along with their dependencies, ensuring consistency across different environments. This containerization allows for rapid development and deployment, making it easy to test new features and roll out updates without disrupting the existing platform. As a result, rich features and improvements can be quickly implemented, keeping the platform responsive to user needs and market demands.

By integrating these technologies and practices, the application is made efficient, secure, and user-friendly, ultimately providing a better experience for users.

5.2. Database management

In developing the web application using Go and React, both PostgreSQL and MongoDB were chosen as database solutions, each serving distinct purposes to enhance functionality. PostgreSQL, a robust relational database management system (RDBMS), excels in managing structured data. Its ACID compliance ensures reliable transactions and data integrity, which are vital for handling sensitive information. Additionally, PostgreSQL's ability to perform complex SQL queries supports intricate data analysis and reporting.

In contrast, MongoDB was selected as a NoSQL database for managing unstructured and semi-structured data. Its document-oriented storage model provides flexibility in data representation, allowing for rapid development without extensive schema changes. MongoDB's horizontal scalability is crucial for handling large volumes of data and maintaining performance during high-traffic scenarios.

The combination of PostgreSQL and MongoDB effectively addresses the platform's diverse data management needs. PostgreSQL manages structured data with complex queries and data integrity, while MongoDB facilitates agile development for unstructured data. This dual-database approach optimizes performance and prepares the application for future enhancements, ensuring adaptability to evolving user requirements.

In summary, choosing both PostgreSQL and MongoDB allows the application to take advantage of what each database does best. This combination not only boosts the overall performance and reliability of the web app but also sets it up for future growth and ensures a great user experience.

6. Summary

We have developed an interactive high entropy alloy database consisting of mechanical properties of HEAs. The website enables analysis and comparison among different alloys. Currently, the database consists of $>$ 4000 compositions from $>$ 400 papers. In future, we plan to expand to computational data, include other properties, and diversify to high entropy ceramics. We plan to integrate machine learning models for predictive capabilities.

Author contributions

MS: Website architecture development (lead), Writing (equal), **EB**: Database collection and consolidation (lead), Writing (equal), **DSA**: Conceptualization (lead); Project administration (lead); Supervision (lead); Writing – review & editing.

CRediT authorship contribution statement

Mohit Singh: Writing – original draft, Software, Methodology. **Eric**

Barr: Writing – original draft, Formal analysis, Data curation. **Dilpuneet Aidhy:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work is supported by the National Science Foundation project # 2302763 titled CDSE: Charge-density based ML framework for efficient exploration and property predictions in the large phase space of concentrated materials. The computational resources are provided by Clemson Computing and Information Technology (CCIT).

Data availability

Data will be made available on request.

References

[1] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013).

[2] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD), JOM 65 (11) (2013) 1501–1509.

[3] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, AFLOW: an automatic framework for high-throughput materials discovery, Comput. Mater. Sci 58 (2012) 218–226.

[4] K. Choudhary, K.F. Garrity, A.C.E. Reid, B. DeCost, A.J. Biacchi, A.R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A.G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B.G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe, F. Tavazza, The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design, npj Comput. Mater. 6 (1) (2020).

[5] A. Zakutayev, N. Wunder, M. Schwarting, J.D. Perkins, R. White, K. Munch, W. Tumas, C. Phillips, An open experimental database for exploring inorganic materials, Sci. Data 5 (2018) 180053.

[6] B. Puchala, G. Tarcea, E.A. Marquis, M. Hedstrom, H.V. Jagadish, J.E. Allison, The materials commons: a collaboration platform and information repository for the global materials community, JOM 68 (8) (2016) 2035–2044.

[7] T.J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A.G. Ricciardulli, A. Vijayan, A. Kulkarni, A. Y. Anderson, B.P. Darwich, B. Yang, B.L. Coles, C.A.R. Perini, C. Rehermann, D. Ramirez, D. Fairen-Jimenez, D. Di Girolamo, D. Jia, E. Avila, E.J. Juarez-Perez, F. Baumann, F. Mathies, G.S.A. González, G. Boschloo, G. Nasti, G. Paramasivam, G. Martínez-Denegri, H. Näsström, H. Michaels, H. Köbler, H. Wu, I. Benesperi, M. I. Dar, I. Bayrak Pehlivan, I.E. Gould, J.N. Vagott, J. Dagar, J. Kettle, J. Yang, J. Li, J.A. Smith, J. Pascual, J.J. Jerónimo-Rendón, J.F. Montoya, J.-P. Correa-Baena, J. Qiu, J. Wang, K. Sveinbjörnsson, K. Hirselandt, K. Dey, K. Frohna, L. Mathies, L. A. Castriotta, M.H. Aldamasy, M. Vasquez-Montoya, M.A. Ruiz-Preciado, M. A. Flatken, M.V. Khenkin, M. Grischek, M. Kedia, M. Saliba, M. Anaya, M. Veldhoen, N. Arora, O. Shargaieva, O. Maus, O.S. Game, O. Yudilevich, P. Fassl, Q. Zhou, R. Betancur, R. Munir, R. Patidar, S.D. Stranks, S. Alam, S. Kar, T. Unold, T. Abzieher, T. Edvinsson, T.W. David, U.W. Paetzold, W. Zia, W. Fu, W. Zuo, V.R. F. Schröder, W. Tress, X. Zhang, Y.-H. Chiang, Z. Iqbal, Z. Xie, E. Unger, An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles, Nat. Energy 7 (1) (2021) 107–115.

[8] W. Li, L. Raman, A. Debnath, M. Ahn, S. Lin, A.M. Krajewski, S. Shang, S. Priya, W. F. Reinhart, Z.-K. Liu, A.M. Beese, Design and validation of refractory alloys using machine learning, CALPHAD, and experiments, Int. J. Refract Metal Hard Mater. 121 (2024).

[9] O.N. Senkov, D.B. Miracle, S.I. Rao, Correlations to improve room temperature ductility of refractory complex concentrated alloys, Mater. Sci. Eng. A 820 (2021).

[10] C.K.H. Borg, C. Frey, J. Moh, T.M. Pollock, S. Gorsse, D.B. Miracle, O.N. Senkov, B. Meredig, J.E. Saal, Expanded dataset of mechanical properties and observed phases of multi-principal element alloys, Sci. Data 7 (1) (2020) 430.

[11] S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, Data Brief 21 (2018) 2664–2678.

[12] J.P. Couzinie, O.N. Senkov, D.B. Miracle, G. Dirras, Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys, Data Brief 21 (2018) 1622–1641.

[13] S. Gorsse, D.B. Miracle, O.N. Senkov, Mapping the world of complex concentrated alloys, Acta Mater. 135 (2017) 177–187.

[14] E.P. George, W.A. Curtin, C.C. Tasan, High entropy alloys: a focused review of mechanical properties and deformation mechanisms, Acta Mater. 188 (2020) 435–474.