



Comparative Analysis of Phase Prediction in High-Entropy TiZrNb Alloys Using Machine Learning Approaches

Masoud Yousefi¹, Ahmad Razaghian Arani²

¹ Postdoctoral Researcher in Materials Engineering, Department of Materials Science and Engineering, Faculty of Technical and Engineering, Imam Khomeini International University (IKIU), Qazvin, Iran.

² Professor, Department of Materials Science and Engineering, Faculty of Technical and Engineering, Imam Khomeini International University (IKIU), Qazvin, Iran.

*Corresponding Author's Email: mrzaghian@eng.ikiu.ac.ir (Ahmad Razaghian Arani)

Paper History:

Received: 2026-01-03

Revised: 2026-05-17

Accepted: 2026-08-02

Keywords:

Machine Learning,
TiZrNb,
High-Entropy Alloy,
Phase prediction

ABSTRACT A primary challenge in the development of high-entropy alloys (HEAs), particularly biocompatible variants, is the substantial cost associated with experimental trial-and-error procedures. In this study, machine learning (ML) models were employed to predict phase formation—specifically, solid solution (SS), intermetallic (IM), and composite (SS+IM) phases—to minimize experimental overhead and expedite the alloy design process. A reference database comprising approximately 400 HEAs was utilized, with key thermodynamic parameters—namely ΔH_{mix} , ΔS_{mix} , δ , VEC, and $\Delta\chi$ —serving as input features. Two primary learning frameworks were evaluated: the Random Forest (RF) algorithm and ensemble learning techniques, specifically Bagging and Boosting. The results indicate that the three-phase classification accuracy reached over 75% on average using the Random Forest and Bagging methods. This study demonstrates that these ensemble algorithms constitute effective, cost-efficient computational tools for phase prediction in HEAs, significantly accelerating the design of novel alloys tailored for biomedical applications and orthopedic implants.



<https://doi.org/10.30501/jamt.2026.569313.1353>

URL: https://www.jamt.ir/article_251345.html

1. INTRODUCTION

One of the main challenges in researching biocompatible high-entropy alloys (HEAs) is the high cost of experimental procedures (Singh, 2021). To address this issue, various simulation methods are employed to predict mechanical behavior and phase formation, thereby reducing the need for extensive experimental trials. One of the key approaches is the use of CALPHAD software.

In these simulation techniques, the dominant phase, whether a solid solution, intermetallic compound, or a combination of both, is predicted based on key parameters that influence the final phase. Parameters such as mixing enthalpy and mixing entropy are considered essential factors in these predictions.

The fundamental challenge in the classification process lies in identifying optimal ways to combine the capabilities of different classifiers to achieve higher accuracy. The primary motivation for researchers in integrating existing technologies and exploring novel approaches has been to enhance predictive performance and enable more accurate phase identification.

2. Materials and Methods

2.1. Feature Importance

As observed in Figure 1, a statistical analysis of the database was performed. The scatter plots presented in Figure 1 illustrate that the phases of high-entropy alloys are determined by more than one input feature. This finding demonstrates that relying on a single parameter is insufficient to establish a clear decision boundary between the phases.

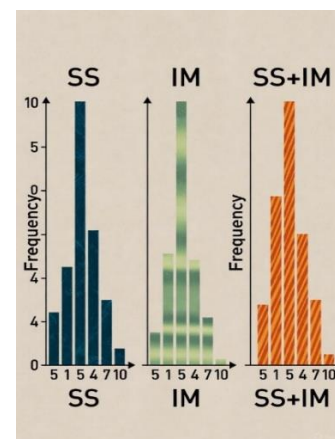


Figure 1. Distribution Histograms of SS, IM, and SS+IM Phases

The scatter plots outside the diagonal in Figure 1 display the distribution of the three phases as a function of pairs selected from the five features. All five features are utilized in the classification process by the ensemble learning algorithms. The diagonal elements of Figure 1 present histogram showing the distribution of the SS, IM, and SS+IM phases across the values of each of the five input features. The decision boundary between the SS and SS+IM phases, as defined by any pair of the five features, appears ambiguous when visualized in a two-dimensional plane, with some regions exhibiting overlap between the two classes.

Figure 2 displays the highest, average, and lowest accuracy values for the three classification tasks using the RF method.



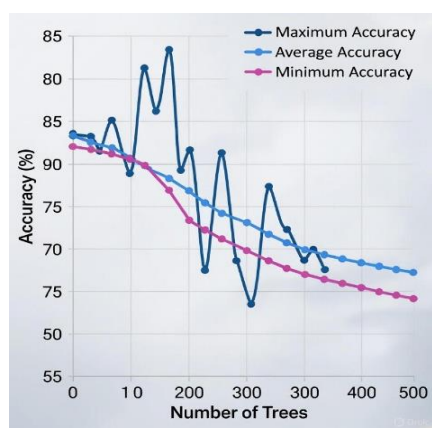


Figure 2. Display of the chart showing the maximum, average, and minimum experimental accuracy obtained from the Random Forest (RF) classifier on the high-entropy alloy dataset.

In Figure 2, the upper and lower curves represent the maximum and minimum classification accuracies, respectively. The average accuracy curve is positioned between the maximum and minimum accuracy curves. In Random Forest, a bagging-based ensemble algorithm, increasing the number of trees generally improves the model's accuracy. As shown in Figure 2, beyond a certain number of trees, further increases have no significant effect on accuracy and only reduce the method's efficiency because of increased computational overhead.

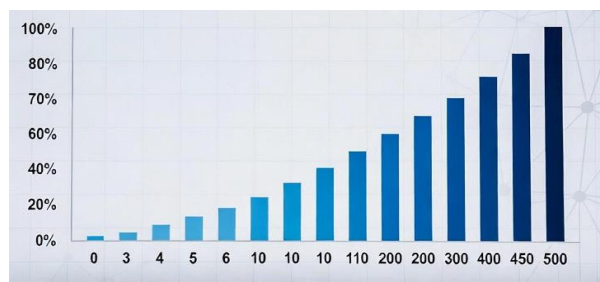


Figure 3. The effect of the number of grown trees on accuracy in the Random Forest method.

The relative importance of the five input features was further analyzed, as shown in Figure 3. It was observed that, in the classification of the SS+IM and IM phases, the order of feature relevance followed the same pattern as that obtained when classifying all three phases simultaneously. The features V (VEC) and X ($\Delta\chi$) were found to play fundamental roles in influencing classification accuracy.

In addition, the relative importance of the five input features with respect to classification accuracy was investigated by classifying the dataset using only four features at a time and calculating the resulting reduction in accuracy. As indicated in Figure 13, the features V and X exhibited significantly greater importance than the other features in terms of their contribution to classification accuracy.

Results Obtained from Ensemble Learning Algorithm Predictions.

In this study, ensemble learning algorithms were employed to predict the phases of high-entropy alloys. The fundamental principle of these algorithms is the combination of multiple diverse base models, which leads to increased flexibility (reduced bias) and

decreased sensitivity to data fluctuations (reduced variance). Two common approaches in ensemble learning are bagging and boosting.

These algorithms were implemented on a comprehensive experimental dataset comprising 401 distinct high-entropy alloy samples, including 174 samples with the SS phase, 54 samples with the IM phase, and 173 samples with the SS+IM phase. The average accuracy for simultaneous three-phase classification using the boosting and bagging methods was 70.4% and 75.3%, respectively.

Subsequently, the focus shifted to binary classification. The accuracies achieved with the bagging algorithm for distinguishing SS from IM, SS from SS+IM, and IM from SS+IM were 90.0%, 73.1%, and 85.2%, respectively. The corresponding values for the boosting algorithm were 78.1%, 71.2%, and 79.1%. Based on these results, bagging-based algorithms demonstrated superior performance in classifying the phases of high-entropy alloys.

The accuracy of the proposed method is approximately 10% higher than that of the SVM and KNN algorithms for the classification of high-entropy alloys. The decision boundary between the SS and SS+IM phases, when plotted using different pairs of the five features, is often unclear, with overlap observed in some cases. The proposed method offers two main advantages in this context: first, a reduction in overfitting in decision trees, which contributes to improved accuracy; and second, the automatic handling of missing data in the dataset.

In the Random Forest algorithm, increasing the number of trees leads to improved accuracy; however, beyond a certain threshold, adding more trees has a negligible impact on accuracy and only reduces computational efficiency because of increased computational overhead.

Conclusions

The results of this comparison revealed that the bagging method, with an average accuracy exceeding 75% in three-phase classification, demonstrated superior performance compared with boosting and baseline techniques such as SVM and KNN. In addition, Random Forest exhibited better resistance to overfitting and lower computational requirements, without the need for GPU-based computation.

This comparative evaluation not only complements and validates the findings of previous studies (Yousefi et al., 2024 #43; Yousefi et al., 2025 #45) but also addresses the lack of direct algorithmic comparisons, providing deeper insights into the strengths and limitations of each approach.

Acknowledgments

We are grateful for the data analysis assistance of Mr. Majid Oghabneshin.

REFERENCES

1. Singh, Shailesh Kumar, and Vivek K. Singh. "Design and Development of High Entropy Alloys Using Artificial Intelligence." *Advances in High-Entropy Alloys-Materials Research, Exotic Properties and Applications*. IntechOpen, 2021.



مقاله ی کامل پژوهشی

بررسی مقایسه ای پیش بینی فاز در آلیاژهای آنتروپی بالای TiZrNb با رویکردهای یادگیری

ماشین

مسعود یوسفی^۱، احمد رزاقیان آرانی^{۲*}

^۱ دانشجوی پسادکتری مهندسی مواد، گروه مهندسی و علم مواد، دانشکده ی فنی و مهندسی، دانشگاه بین المللی امام خمینی (ره)، قزوین، ایران

^۲ استاد، گروه مهندسی و علم مواد، دانشکده ی فنی و مهندسی، دانشگاه بین المللی امام خمینی (ره)، قزوین، ایران

تاریخچه ی مقاله:

ثبت اولیه: ۱۴۰۴/۱۰/۱۳

بازنگری: ۱۴۰۵/۰۲/۲۷

پذیرش قطعی: ۱۴۰۵/۰۵/۱۱

کلیدواژه ها:

یادگیری ماشین،

TiZrNb،

آلیاژ آنتروپی بالا،

پیش بینی فاز

چکیده یکی از چالش های اصلی در توسعه ی آلیاژهای آنتروپی بالا، به ویژه انواع زیست سازگار، هزینه ی بالای آزمایش های تجربی است. در این پژوهش، به منظور کاهش تعداد آزمایش ها و تسریع در طراحی آلیاژهای جدید، از روش های یادگیری ماشین برای پیش بینی فازهای تشکیل شده (محلول جامد SS، فاز بین فلزی IM و ترکیب SS+IM) استفاده شد. پایگاه داده ای شامل حدود ۴۰۰ آلیاژ آنتروپی بالا به عنوان مرجع انتخاب شد و پارامترهای ترمودینامیکی کلیدی نظیر ΔH_{mix} ، ΔS_{mix} ، δ ، VEC و $\Delta \chi$ ویژگی های ورودی در نظر گرفته شدند. دو رویکرد اصلی بررسی شدند: الگوریتم جنگل تصادفی (Random Forest) و الگوریتم های یادگیری گروهی بگینگ (Bagging) و بوستینگ (Boosting). نتایج نشان داد که دقت طبقه بندی سه فازی با روش جنگل تصادفی و روش بگینگ به طور متوسط به بیش از ۷۵ درصد رسید. این مطالعه نشان می دهد که الگوریتم های جنگل تصادفی و یادگیری گروهی ابزارهای مؤثر و کم هزینه ای برای پیش بینی فاز در آلیاژهای آنتروپی بالا هستند و می توانند به طور چشمگیری فرایند طراحی آلیاژهای جدید به ویژه برای کاربردهای زیست پزشکی و ایمپلنت های ارتوپدی را تسریع کنند.



<https://doi.org/10.30501/jamt.2026.569313.1353>

URL: https://www.jamt.ir/article_251345.html

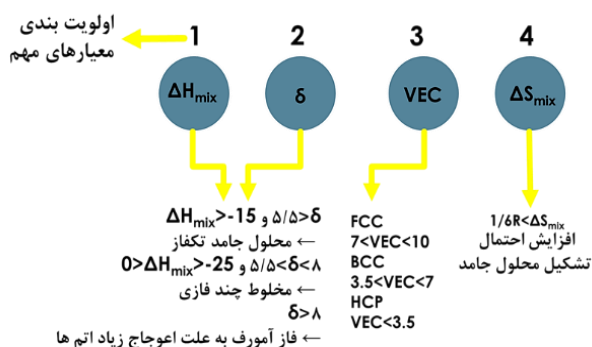
۱- مقدمه

استفاده از این روش مدل سازی نه تنها صرفه جویی چشمگیری در منابع ایجاد می کند، بلکه بهره برداری حداکثری از داده های موجود را نیز ممکن می سازد و در نهایت هزینه های آزمایش های عملی را از طریق پیش بینی های دقیق ترمودینامیکی کاهش می دهد. با توجه به اینکه تحلیل آلیاژهای آنتروپی بالا با ساختار تک فاز نسبتاً ساده تر است، نرم افزار CALPHAD نقش مهمی در پیش بینی چنین ساختارهایی ایفا می کند (Cantor, 2014: 3; Yeh, 2016: 2). بنابراین، این روش مدل سازی به عنوان ابزاری مؤثر و ساده در طراحی آلیاژهای آنتروپی بالای زیست سازگار، به ویژه در ساخت ایمپلنت های ارتوپدی، کاربرد گسترده ای دارد (Cantor, 2014: 3; Yeh, 2016: 2; Sun et al., 2014).

یکی از چالش های عمده در پژوهش بر روی آلیاژهای آنتروپی بالای زیست سازگار هزینه های بالای فرایندهای تحقیقاتی مرتبط با آنها است (Singh & Singh, 2021: 1). به همین سبب، رویکردهای متنوعی برای شبیه سازی رفتار مکانیکی و پیش بینی فازهای این آلیاژها به کار گرفته می شود تا حجم آزمایش های تجربی به حداقل برسد. یکی از این رویکردها بهره گیری از نرم افزار CALPHAD است. مدل سازی مبتنی بر CALPHAD با محاسبه ی دیگرام های فازی و ارزیابی خواص ترمودینامیکی سیستم های آلیاژی امکان پیش بینی تشکیل فاز محلول جامد تک فاز را در ترکیبات شیمیایی گوناگون فراهم می آورد (Yeh, 2016: 2).

فناوری های موجود و کاوش ایده های نوین ارتقای قابلیت های پیش بینی و تشخیص دقیق تر فازها بوده است.

برای اجرای محاسبات و فرایند داده کاوی لازم است ابتدا یک پایگاه داده ای جامع از آلیاژهای آنتروپی بالایی که تا کنون تولید شده اند گردآوری شود و سپس داده های موجود در نرم افزار MATLAB تعریف و بارگذاری شوند. پارامترهای کلیدی و مؤثر در طراحی آلیاژهای جدید در شکل ۱ نشان داده شده اند (Fontana & Greene, 2018: 14).



شکل ۱. مهم ترین معیارها برای طراحی و پیشگویی وضعیت ترکیبی ساخته نشده در آلیاژهای آنتروپی بالا

در رابطه ی ۱، ΔH_{mix} اختلاف آنتالپی اختلاط کل است که C_i غلظت عنصر i ام برحسب درصد اتمی است و H_{ij} آنتالپی اختلاط دوتایی بین عناصر i ام و j ام است:

$$\Delta H_{mix} = \sum_{i=1, i < j}^n 4H_{ij}C_iC_j \quad (1)$$

در رابطه ی ۲، δ نسبت اختلاف شعاع عناصر با شعاع متوسط عناصر است که از رابطه ی زیر به دست می آید:

$$\delta = 100 \times \sqrt{\sum_{i=1}^n c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2} \quad (2)$$

در رابطه ی ۳، VEC غلظت الکترون ظرفیت کل و VEC_i الکترون ظرفیت عنصر i ام و C_i غلظت عنصر i ام برحسب درصد اتمی است و n تعداد عناصر شرکت کننده در ترکیب آلیاژ است.

$$VEC = \sum_{i=1}^n c_i VEC_i \quad (3)$$

2017: 4; Haase et al., 2017: 5; Harrysson & Cormier, 2005: 6.

ویژگی های مکانیکی و عملکردی آلیاژهای آنتروپی بالا عمدتاً به فازهای تشکیل شده در ساختار آنها بستگی دارد. این فازها عموماً شامل سه دسته هستند: فاز محلول جامد (SS)، فازهای بین فلزی (IM) و ترکیب هر دو (SS+IM). در بیشتر موارد، فاز محلول جامد تک فاز ساختار ایدئال در نظر گرفته می شود، زیرا خواص مکانیکی برتر و عملکرد بهتری ارائه می دهد. در مقابل، حضور فازهای بین فلزی، که معمولاً شکننده هستند، شکل پذیری آلیاژ را به طور قابل توجهی کاهش می دهد. بنابراین، پیش بینی دقیق نوع فاز تشکیل شده برای یک ترکیب شیمیایی خاص نقش کلیدی در توسعه و کاربرد موفق آلیاژهای آنتروپی بالای نسل جدید ایفا می کند.

برای تولید آلیاژهای آنتروپی بالا عمدتاً دو مسیر اصلی وجود دارد: نخست، رویکرد کاملاً تجربی که بر آزمایش های آزمایشگاهی و تحلیل نتایج آنها استوار است و هزینه های بالایی دارد و، دوم، بهره گیری از تکنیک های داده کاوی برای پیش بینی اولیه ی ساختار و سپس تأیید تجربی آن. استفاده از داده کاوی به طور قابل توجهی سرعت توسعه و تولید آلیاژهای جدید آنتروپی بالا را افزایش می دهد.

روش های متنوعی برای پیش بینی فازهای تشکیل شده در این آلیاژها به کار گرفته می شود که از جمله ی مهم ترین آنها می توان به الگوریتم های KNN، SVM و ANN و سایر تکنیک های مرتبط اشاره کرد. هدف اصلی این روش های داده کاوی تسهیل طراحی و ساخت آلیاژهای آنتروپی بالای نوین است (Singh & Singh, 2021: 1; Huang et al., 2019: 7; King et al., 2016: 8; Krishna et al., 2021: 9; Roy & Balasubramanian, 2021: 10; Qiao et al., 2021: 11; Machaka, 2021: 12; Rickman et al., 2020: 13). در این تکنیک های شبیه سازی، با تکیه بر پارامترهای کلیدی مؤثر بر فاز نهایی، که می تواند محلول جامد، فازهای بین فلزی یا ترکیبی از هر دو باشد، نوع فاز غالب پیش بینی می شود. پارامترهایی مانند آنتالپی و آنتروپی اختلاط از جمله عوامل ضروری در این پیش بینی ها به شمار می روند. چالش اساسی در فرایند طبقه بندی یافتن راهکارهایی برای ترکیب بهینه ی توانایی های طبقه بندی کننده ها به منظور دستیابی به دقت بالاتر است. انگیزه ی اصلی پژوهشگران در ادغام

7. سپس، تحلیل داده‌کاوی برای پیش‌بینی ساختار نهایی شامل فاز محلول جامد، فازهای بین‌فلزی یا ترکیبی از هر دو صورت می‌پذیرد و در نهایت اعتبار و دقت این پیش‌بینی‌ها از طریق آزمایش‌های ساختاری و تجربی تأیید می‌شود.

استفاده از داده‌کاوی ذکر شده می‌تواند در شناسایی درصد ترکیب مناسب عناصر مختلف قبل از ساخت کمک جدی کند. هرچه بانک اطلاعاتی مورد استفاده کامل‌تر باشد، نتایج با دقت بیشتری خواهند بود (Huang et al., 2019: 7).

سه رویکرد اصلی و پذیرفته‌شده برای پیش‌بینی فازها در آلیاژهای آنتروپی بالا به شرح زیر است:

۱- روش‌های پارامتری ژانگ و همکاران (Zhang et al., 2008: 16) پارامترهایی را معرفی کردند که بر تعیین نوع فاز در آلیاژهای آنتروپی بالا تأثیرگذار هستند، از جمله تفاوت اندازه‌ی اتمی (δ)، آنتروپی اختلاط (ΔS_{mix}) و آنتالپی اختلاط (ΔH_{mix}).

۲- نظریه‌ی تابعی چگالی (DFT) هان و ویدوم (Huhn & Hu et al., 2020: 17) با استفاده از محاسبات مبتنی بر نظریه‌ی تابعی چگالی، انتقال فاز را در آلیاژهای آنتروپی بالا پیش‌بینی کردند. باین‌حال، یکی از محدودیت‌های اصلی این روش هزینه‌ی محاسباتی بسیار بالای آن است (Feng et al., 2017: 19).

۳- یادگیری ماشینی (ML) به‌عنوان تکنیکی پیشرفته در پیش‌بینی فازهای آلیاژهای آنتروپی بالا کاربرد گسترده‌ای دارد. این رویکرد با دو مزیت اصلی یعنی دقت بالا و سرعت زیاد همراه است. روش‌های یادگیری ماشینی امکان استخراج دانش درباره‌ی کیفیت آلیاژها از داده‌های موجود را بدون نیاز به برنامه‌نویسی پیچیده فراهم می‌کند (Ward et al., 2016: 20). لی و همکاران (Lee et al., 2021) از تکنیک‌های مبتنی بر یادگیری عمیق برای بهینه‌سازی، تولید و تفسیر داده‌ها استفاده کردند تا خروجی‌های مدل را بهبود بخشند و عوامل کلیدی طراحی را در پیش‌بینی فاز آلیاژهای آنتروپی بالا شناسایی کنند. همچنین، هوانگ و همکاران (Huang et al., 2019: 22) با بهره‌گیری از الگوریتم‌های-K نزدیک‌ترین همسایه (KNN) ماشین بردار پشتیبان (SVM) و شبکه‌ی عصبی مصنوعی (ANN) فازهای این آلیاژها را پیش‌بینی کردند. یکی از چالش‌های اصلی در به‌کارگیری یادگیری ماشینی انتخاب مناسب ویژگی‌های ورودی مرتبط است.

ΔS_{mix} اختلاف آنتروپی اختلاط کل است که C_i غلظت عنصر آی ام برحسب درصد اتمی است و R همان ثابت جهانی گازها است. $\bar{r}$ شعاع اتمی متوسط است.

$$\bar{r} = \sum_{i=1}^n c_i r_i \quad (4)$$

$$\Delta S_{mix} = -R \sum_{i=1}^n c_i \ln c_i \quad (5)$$

علاوه بر موارد ذکر شده، پارامترهای دیگری نظیر اختلاف الکترونگاتیویته مهم هستند. Δx اختلاف الکترونگاتیویته اجزا با الکترونگاتیویته متوسط است و $\bar{x}$ الکترونگاتیویته متوسط است که از روابط زیر به دست می‌آید و X_i الکترونگاتیویته پائولینگ (متداول‌ترین مقیاس که مبتنی بر مقادیر تجربی انرژی‌های پیوندی است) عنصر i ام است.

$$\bar{x} = \sum_{i=1}^n c_i x_i \quad (6)$$

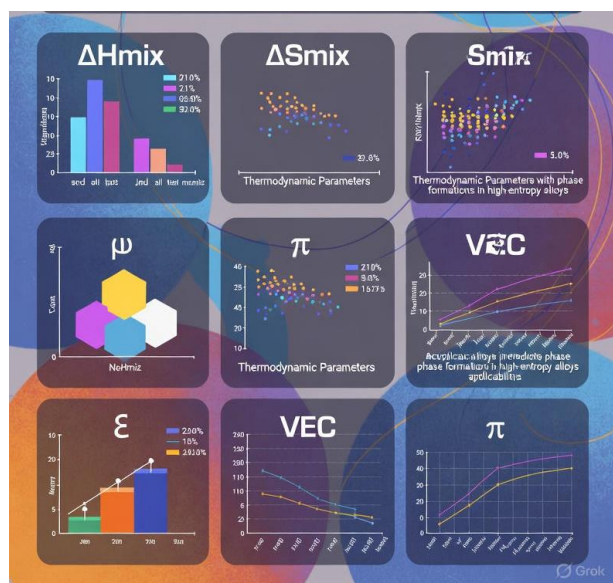
$$\Delta x = \sqrt{\sum_{i=1}^n c_i (x_i - \bar{x})^2} \quad (7)$$

بنابراین، براساس اولویت‌های ذکر شده در شکل ۱، میزان آنتالپی نشان‌دهنده‌ی تشکیل احتمالی محلول جامد تک‌فاز و آنتروپی نیز حاکی از تشکیل محلول جامد تک‌فاز است. ضمناً پارامتر VEC نیز حاکی از ایجاد ساختار BCC است. از طرف دیگر، می‌توان پیش‌بینی کرد با توجه به بالا بودن میزان پارامتر δ با ساختاری شکل مواجه خواهیم بود که مقاومت به خوردگی نزدیک به تانتالم با فلزات نجیب ایجاد می‌کند (Fontana & Greene, 2018: 14; Tsai & Yeh, 2014: 15).

با توجه به پارامترهای ذکر شده و همچنین داده‌های به‌دست‌آمده از ویژگی‌های ترمودینامیکی کلیدی که در تعیین فازهای احتمالی ساختار نهایی نقش دارند، فرایند شبیه‌سازی مبتنی بر داده‌کاوی اجرا و نتایج حاصل از آن با یافته‌های آزمایش‌های عملی مقایسه شد.

باید توجه شود که با بهره‌گیری از محاسبات ترمودینامیکی پیش‌گفته، ابتدا ارزیابی‌های اولیه بر روی داده‌های پایگاه اطلاعاتی گردآوری‌شده انجام می‌شود (Huang et al., 2019: 15).

یادگیری گروهی در حل انواع گوناگون و وسیعی از مسائل تشخیصی در حوزه ی پزشکی به کار گرفته شده است (Zhang et al., 2019: 28; Han et al., 2014: 29).



شکل ۲. نمایش توزیع پیش بینی فازهای محلول جامد فاز بین فلزی و محلول جامد به همراه فاز بین فلزی براساس فاکتورهای مهم در طراحی آلیاژهای آنروپی بالا (Huang et al., 2019: 7)

یکی از ضرورت های اصلی در پژوهش های مرتبط با آلیاژهای آنروپی بالا مقایسه ی سیستماتیک روش های پیش بینی فاز پیشین است تا نقاط قوت و محدودیت های هر رویکرد در شرایط یکسان مشخص شود. در مقالات قبلی نویسندگان یوسفی و همکاران (Yousefi et al., 2025a; Yousefi et al., 2025b) الگوریتم های جنگل تصادفی و یادگیری گروهی (Bagging و Boosting) جداگانه بررسی شدند و دقت های قابل توجهی مانند ۷۵/۳ درصد برای طبقه بندی سه فازی به دست آمد، اما فقدان مقایسه ی مستقیم این روش ها با یکدیگر و با تکنیک های دیگر مانند SVM و KNN خلأیی در درک کارایی کلی ایجاد کرده است. این مقاله، با تمرکز بر مقایسه ی این رویکردها بر پایه ی پایگاه داده ی تجربی مشترک (۴۰۱ آلیاژ)، جزئیاتی مانند تأثیر پارامترهای ترمودینامیکی (ΔH_{mix} , ΔS_{mix} , δ , VEC و $\Delta \chi$) و کاربرد عملی در آلیاژهای TiZrNbX را توجیه می کند و لزوم نگارش آن را در پر کردن این خلأ و پیشنهاد روش های

الگوریتم های طبقه بندی مبتنی بر یادگیری ماشینی در حوزه های گوناگون به طور گسترده ای استفاده می شوند. به همین دلیل، پژوهشگران به طور مداوم در حال ابداع روش های نوین برای ارتقای دقت طبقه بندی هستند. یادگیری گروهی،^۱ چه همگن و چه ناهمگن، یکی از این روش های مؤثر به شمار می رود. نمونه های اولیه ی این رویکرد شامل تجمع بوت استرپ^۲ یا بگینگ^۳ (Ho, 1998: 23)، تقویت یا بوستینگ^۴ و جنگل های تصادفی (Schapire & Singer, 1998: 24) هستند. استقرار این مجموعه ها معمولاً به بهبود کلی عملکرد طبقه بندی منجر می شود. علاوه بر این، محققان دیگری از راهبردهای متنوعی برای دستیابی به یادگیری گروهی بهره برده اند، از جمله تکنیک هایی که بر پایه ی رأی اکثریت یا روش های دیگر برای ترکیب چندین طبقه بندی کننده یا تقسیم داده ها عمل می کنند.

لئون و همکاران (Leon et al., 2017: 25) اخیراً مطالعه ای منتشر کرده اند که به بررسی تأثیر سیستم های رأی گیری در مسائل طبقه بندی اختصاص دارد. در این پژوهش، نویسندگان به تحلیل چگونگی اثرگذاری روش های مختلف رأی گیری در عملکرد دو الگوریتم طبقه بندی در مجموعه داده هایی با سطوح دشواری متفاوت پرداخته اند. با وجود اینکه به کارگیری رأی اکثریت در منابع علمی رایج است، نتایج تجربی حاصل نشان می دهد که روش رأی تک گانه می تواند گزینه ی جایگزین مناسبی باشد. همچنین، بانفیلد و همکاران (Banfield et al., 2007: 26) ارزیابی ای مقایسه ای از روش های مختلف ساخت مجموعه ی درخت های تصمیم انجام دادند که شامل هفت تکنیک مبتنی بر تصادفی سازی بود.

در حالی که برخی رویکردهای یادگیری گروهی بر ترکیب طبقه بندی کننده های پایه ی متفاوت تمرکز دارند، مجموعه ها همچنین می توانند از طریق تقسیم مجموعه داده به زیرمجموعه های کوچک تر و سپس ادغام نتایج بخش های جداگانه تشکیل شوند. علاوه بر این، بدون نیاز به تکنیک های بسته بندی، تقویت یا جنگل های تصادفی، امکان تولید مجموعه داده های متنوع وجود دارد. روتا و همکاران (Ruta et al., 2010: 27)، با بهره گیری از جایگشت تصادفی و تقسیم مجموعه داده ی اصلی، چندین مجموعه داده ی جدید ایجاد کردند. در نهایت،

محلول جامد اغلب مسئول آلیاژهای آنتروپی بالایی است که ویژگی های مطلوب را نشان می دهند. علاوه بر این، وجود یک فاز بین فلزی شکننده، از سوی دیگر، شکل پذیری را کاهش می دهد (King et al., 2016: 8).

۲-۱-۲- طرح طبقه بندی

۲-۱-۲-۱- ویژگی های ورودی

به طور کلی، پذیرفته شده است که تعیین نوع فاز در آلیاژها با ویژگی های عنصری مانند غلظت الکترون لایه ی ظرفیت و آنتالپی ایجاد، ارتباط نزدیکی دارد و این امر به تدوین مجموعه ای از معیارهای پارامتری برای انتخاب فاز منجر شده است (Ye et al., 2016: 35). در پیش بینی فاز آلیاژهای آنتروپی بالا، از رویکردهای پارامتریک بهره گرفته شده است. در این زمینه، ژانگ و همکاران (Zhang, 2010: 36) پیشنهاد کردند که پارامترهای اختلاف شعاع اتمی نسبی (δ)، آنتروپی اختلاط (ΔS_{mix}) و آنتالپی اختلاط (ΔH_{mix}) نقش تعیین کننده ای در شناسایی فاز نهایی آلیاژهای آنتروپی بالا ایفا می کنند. در مقابل رویکردهای پارامتریک، یادگیری ماشین ابزاری پیشرفته ارائه می دهد که قادر است اطلاعات مرتبط با خواص کلیدی آلیاژها را از داده های موجود، بدون نیاز به برنامه نویسی پیچیده، به سرعت استخراج و ارائه کند.

قوانین هیوم-روتیری (Smallman & Ngan, 2011: 37)

بیان می کنند که ویژگی های عنصری گوناگون، نظیر شعاع اتمی و غلظت الکترون ظرفیت (VEC)، با ساختار بلورین و پایداری آلیاژهای فلزی ارتباط مستقیم دارند و به عنوان ابزارهای عملی در توسعه ی آلیاژهای نوین به کار می روند. این قوانین تأکید دارند که پارامترهای بنیادی متعددی مانند شعاع اتمی و الکترون ظرفیت بر ساختار بلورین و پایداری آلیاژهای فلزی تأثیرگذار هستند و همین امر آنها را به ابزاری ارزشمند برای طراحی آلیاژهای جدید تبدیل کرده است. برای مثال، شباهت در شعاع اتمی و الکترونگاتیویته عناصر تشکیل دهنده ی آلیاژ ممکن است حاکی از وجود شرایط لازم برای تشکیل فاز محلول جامد باشد، اما در آلیاژهای آنتروپی بالا، این فاز لزوماً شکل نمی گیرد (Zhang et al., 2011: 38). به دلیل فقدان تحلیل جامع داده ها و نتایج تجربی، هنوز روشن نیست که این پارامترهای اضافی تا

پیشرفته تر مانند شبکه های عصبی عمیق^۱ برای تحقیقات آینده برجسته می سازد، که می تواند به کاهش هزینه های تجربی و تسریع طراحی مواد زیست سازگار کمک شایانی کند.

۲- الگوریتم های داده کاوی

۲-۱- جنگل های تصادفی RF

الگوریتم جنگل های تصادفی (RF) در زمینه های متنوعی کاربرد یافته است، از جمله کشف دارو (Lind & Anderson, 2019: 30)، تحلیل داده های جغرافیایی (Talebi et al., 2022: 31) و بررسی توزیع پوشش گیاهی روی زمین (Tokar et al., 2018: 32). همچنین، بهره گیری از RF در کاربردهای تشخیص عیوب و خرابی های ماشین آلات توجه زیادی را به خود جلب کرده است (Yang et al., 2008: 33).

مزایای اصلی الگوریتم جنگل های تصادفی (RF) عبارت هستند از دقت بالا، پایداری و استحکام قوی و همچنین قابلیت مدیریت حجم زیادی از ویژگی های ورودی.

در پژوهش حاضر، از الگوریتم جنگل تصادفی برای طبقه بندی و پیش بینی فازهای SS، IM و SS+IM در آلیاژهای آنتروپی بالا بهره گرفته شده است. این روش (RF) به طور مؤثری برای تحلیل اصول حاکم بر انتخاب فاز به کار رفت و بر پایه ی یک مجموعه داده ی تجربی گسترده شامل ۴۰۱ آلیاژ آنتروپی بالای متمایز، که متشکل از ۱۷۴ نمونه با فاز SS، ۵۴ نمونه با فاز IM و ۱۷۳ نمونه با فاز SS+IM بود، اجرا شد. یافته های تجربی حاصل، کارایی برتر و دقت بالاتر روش پیشنهادی را در مقایسه با تحقیقات پیشین دیگر تأیید می کند.

۲-۱-۱- داده های مرجع

در این تحقیق، از یک پایگاه داده ی تجربی، که میراکل و سنکوو در مقاله ی مروری آلیاژ آنتروپی بالا ارائه کردند، استفاده شده است (Miracle & Senkov, 2017: 34). پایگاه داده ی ارائه شده شامل ۶۴۸ مورد است، اما برخی از آنها حذف شده اند؛ زیرا یک ترکیب چند بار تکرار شده است. بنابراین، با اصلاحاتی که انجام شد، یک مجموعه داده ی جدید برای استفاده در الگوریتم های داده کاوی جدا شد که شامل ۴۰۱ آلیاژ آنتروپی بالا با ۱۷۴ فاز SS، ۵۴ فاز IM و ۱۷۳ فاز SS+IM بودند. وجود فاز

نمونه‌گیری تصادفی (با جایگزینی) از داده‌های آموزشی و، دوم، انتخاب تصادفی زیرمجموعه‌ای از ویژگی‌ها برای تقسیم هر گره در درخت‌ها.

۲-۱-۴- نمونه‌گیری تصادفی

یکی از اصول بنیادین الگوریتم جنگل تصادفی یادگیری هر درخت بر پایه‌ی نمونه‌گیری تصادفی (با جایگزینی) از داده‌های آموزشی است. این نمونه‌ها از طریق روش بوت‌استرپ تولید می‌شوند بدین‌معنی که برخی نقاط داده ممکن است چندین بار در مجموعه‌ی آموزشی یک درخت ظاهر شوند (و در صورت لزوم می‌توان این تکرار را کنترل یا متوقف کرد). برای درک بهتر فرایند بگینگ یا بسته‌سازی، مراحل خلاصه‌شده زیر قابل توجه هستند.

مرحله ۱: ایجاد مجموعه داده‌های متعدد

در فرایند نمونه‌گیری با جایگزینی، داده‌های اصلی چندین بار انتخاب می‌شوند و مجموعه داده‌های جدیدی شکل می‌گیرند. این مجموعه داده‌های جدید می‌توانند شامل زیرمجموعه‌هایی از ردیف‌ها و ستون‌ها باشند که معمولاً به‌عنوان پارامترهای کلیدی در مدل‌های مبتنی بر بگینگ (بسته‌بندی) عمل می‌کنند.

مرحله ۲: ساخت چندگانه

استفاده از نسبت‌های کمتر از یک برای ردیف‌ها و ستون‌ها به ساخت مدل‌های پایدارتر و قابل‌اعتمادتر کمک می‌کند که کمتر در معرض بیش‌برازش قرار دارند. برای هریک از مجموعه داده‌های تولیدشده، یک طبقه‌بندی‌کننده پایه جداگانه آموزش داده و ساخته می‌شود.

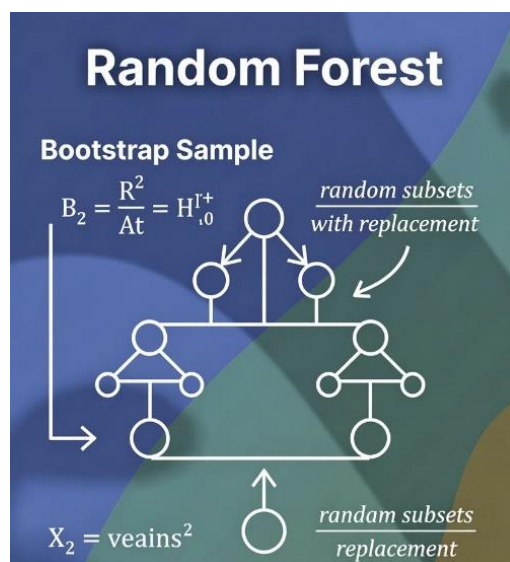
مرحله ۳: ترکیب طبقه‌بندی‌کننده‌ها (شکل ۴)

بسته به نوع مسئله، پیش‌بینی‌های حاصل از تمامی طبقه‌بندی‌کننده‌های پایه با استفاده از میانگین یا حالت (mode) ترکیب می‌شوند. به‌طور کلی، روش میانگین‌گیری در مسائل رگرسیون کاربرد دارد.

چه حد از یکدیگر یا از معیارهای پیشین مستقل هستند. درنهایت، همه‌ی پارامترهای شناخته‌شده در تأثیرگذاری بر توسعه‌ی آلیاژهای آنتروپی بالا با محدودیت‌های خاصی همراه هستند.

۲-۱-۳- طبقه‌بندی جنگل تصادفی

الگوریتم جنگل‌های تصادفی (RF) را نخستین بار توسط لئو بریمن (Breiman, 2001: 39) معرفی کرد. این الگوریتم درواقع نوعی از درخت‌های تصمیم به شمار می‌رود که با ساخت تعداد زیادی درخت و ترکیب نتایج آن‌ها از پدیده بیش‌برازش جلوگیری می‌کند (Liu et al., 2012: 40) و درنهایت یک مجموعه‌ی جنگلی قدرتمند تشکیل می‌دهد. در فرایند تقسیم هر گره، درخت‌های موجود در جنگل تصادفی از نمونه‌های داده‌ای تصادفی (با جایگزینی) یا زیرمجموعه‌ای تصادفی از متغیرهای ورودی استفاده می‌کنند. مدل نهایی، پس از پرورش تعداد قابل‌توجهی درخت، براساس کلاس دارای بیشترین رأی تصمیم‌گیری هستند. یک درخت تصمیم تک‌گانه به مراتب دقیق‌تر از انتخاب تصادفی کلاس عمل می‌کند، اما جنگل‌های تصادفی با ترکیب تعداد زیادی درخت، که هرکدام بر پایه‌ی ورودی‌های تصادفی ساخته شده‌اند، مدلی بسیار مقاوم و قوی تولید می‌کنند.



شکل ۳. الگوریتم جنگل تصادفی

جنگل تصادفی نوعی مدل طبقه‌بندی است که از مجموعه‌ای از درخت‌های تصمیم‌گیری متعدد تشکیل شده است. تصادفی بودن این مدل ناشی از دو اصل اساسی است: نخست،

دیگری ترجیح داده شوند. به طور کلی، فرایند آموزش سریع است، اما مرحله ی پیش بینی پس از آموزش نسبتاً کندتر انجام می شود.

۲-۲- روش الگوریتم های یادگیری گروهی

به طور کلی، یادگیری گروهی مدلی است که براساس چندین مدل مختلف پیش بینی می کند. با ترکیب چندین مدل مختلف، یادگیری گروهی تمایل به انعطاف پذیری بیشتری دارد (جهت گیری کمتر) و حساسیت کمتری به داده ها (واریانس کمتر) دارد. دو روش رایج یادگیری گروهی عبارت هستند از بگینگ یا بسته بندی کردن و بوسستینگ یا تقویت.

Bagging یا Bagging-Boosting Aggregation یک

الگوریتم (ML) یادگیری ماشینی است (شکل ۴) که در آن چندین پیش بینی مستقل با نمونه گیری ساخته می شوند. سپس، نتایج فردی با ابزار (رگرسیون) یا اکثریت رأی (طبقه بندی) با هم ترکیب می شوند تا پیش بینی نهایی انجام شود.

الف) مراحل اجرای الگوریتم بسته بندی (شکل ۵):

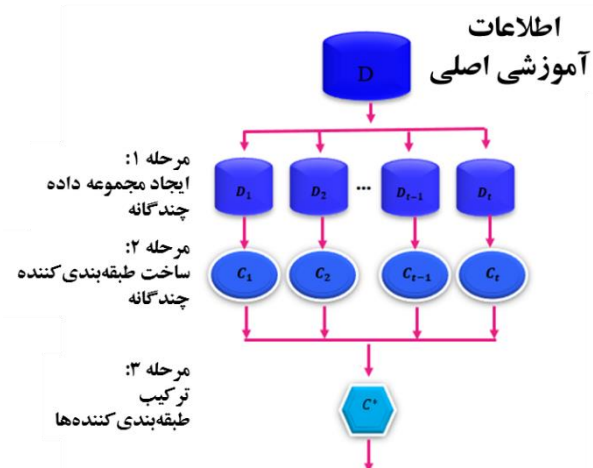
مرحله ۱: چندین زیرمجموعه از مجموعه داده های اصلی با تاپل های مساوی ساخته می شوند و موارد جایگزین را انتخاب می کنند.

مرحله ۲: یک مدل پایه بر روی هریک از این زیرمجموعه ها ساخته می شود.

مرحله ۳: هر مدل به صورت موازی از هر مجموعه ی آموزشی و مستقل از یکدیگر آموزش داده می شود.

مرحله ۴: پیش بینی های نهایی با ترکیب پیش بینی های همه ی مدل ها تعریف می شوند.

تقویت (Boosting) یک الگوریتم یادگیری ماشینی (ML) است که یادگیرندگان ضعیف را به یادگیرندگان قوی تبدیل می کند. یادگیرندگان ضعیف طبقه بندی کننده هایی هستند که بدون توجه به توزیع کلی داده های آموزشی همیشه کمی بهتر از شانس عمل می کنند. در روش تقویت، پیش بینی ها متوالی هستند و هر پیش بینی کننده ی بعدی از خطای پیش بینی کننده ی قبلی درس می گیرد. درختان تقویت کننده ی گرادیان (GBTs) روشی رایج در این دسته است.



شکل ۴. فرایند بسته بندی

۲-۱-۵- زیرمجموعه های تصادفی از ویژگی ها

یکی دیگر از اصول کلیدی الگوریتم جنگل تصادفی محدود کردن بررسی ویژگی ها در زمان تقسیم هر گره از درخت های تصمیم است، به طوری که تنها زیرمجموعه ای تصادفی از کل ویژگی ها ارزیابی می شوند. معمولاً این تعداد به جذر تعداد کل ویژگی ها ($\sqrt{n}$) تقسیم می شود؛ یعنی در هر گره، درخت تصمیم تنها بر پایه ی نمونه ای تصادفی از ویژگی ها، که اندازه ی آن برابر جذر تعداد ویژگی های ورودی است، تقسیم می شود. الگوریتم جنگل تصادفی در شرایط خاصی عملکرد بسیار مناسبی دارد و هنگام ارزیابی آن لازم است مزایا و معایب این روش به دقت در نظر گرفته شود.

۲-۱-۶- مزایا

نیاز به پیش پردازش حداقلی داده های ورودی: جنگل های تصادفی قادر به مدیریت داده های باینری، دسته ای و عددی بدون مقیاس گذاری پیچیده هستند، از جمله سرعت یادگیری بالا و قابلیت پردازش حجم گسترده ای از داده ها، عملکرد عالی به عنوان پیش بینی کننده ویژگی های هدف، کاربرد وسیع در انواع مختلف مسائل مدل سازی، الگوریتمی ساده و بسیار کاربردی برای پیش بینی در کارهای تجربی و آزمایشگاهی.

۲-۱-۷- معایب

دستیابی به دقت بالاتر نیازمند تعداد بیشتری درخت است که این امر زمان آموزش مدل را افزایش می دهد. هرچند جنگل تصادفی در بیشتر کاربردهای عملی سرعت مناسبی دارد، در مواردی که سرعت اجرا حیاتی باشد، ممکن است روش های

بنابراین، شباهت‌های زیادی دارند. این بخش شباهت‌های آن‌ها را توضیح می‌دهد.

۱- هر دو روش مجموعه‌ای برای گرفتن N یادگیرنده از یک یادگیرنده هستند.

۲- هر دو مجموعه داده‌هایی را از طریق نمونه‌گیری تصادفی تولید می‌کنند.

۳- هر دو با میانگین گرفتن از N داده (یا با گرفتن اکثریت آن‌ها، یعنی با رأی اکثریت) تصمیم نهایی را می‌گیرند.

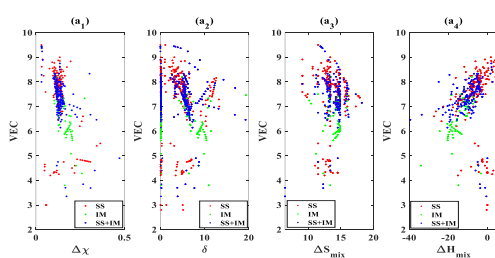
۴- هر دو در کاهش پراکندگی دارای نتیجه‌ی عالی و پایدارتر هستند

تفاوت بین الگوریتم‌های بسته‌بندی و تقویت در جدول ۱ توضیح داده شده است.

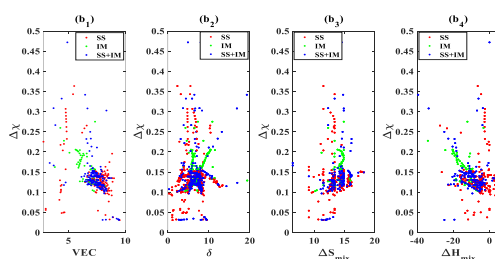
۳- نتایج شبیه‌سازی حاصل از الگوریتم جنگل تصادفی

۳-۱- اهمیت متغیر

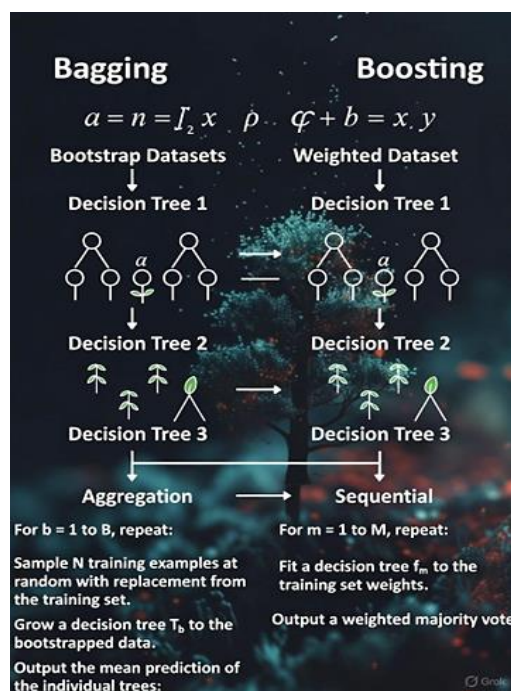
همان‌طور که نشان داده شده است، نمی‌توان، با استفاده از یک ویژگی منفرد، مرزی واضح برای تمایز فازها ترسیم کرد. در نتیجه، هر پنج ویژگی برای گنجاندن در طبقه‌بندی RF استفاده می‌شوند.



شکل ۶. نمودارهای پراکندگی توزیع سه فاز را به‌عنوان تابعی از $(a_1) VEC - \Delta\chi$, $(a_2) VEC - \delta$, $(a_3) VEC - \Delta S_{mix}$, $(a_4) VEC - \Delta H_{mix}$



شکل ۷. نمودارهای پراکندگی که توزیع سه فاز را به‌عنوان تابعی از $(b_1) \Delta\chi - VEC$, $(b_2) \Delta\chi - \delta$, $(b_3) \Delta\chi - \Delta S_{mix}$, $(b_4) \Delta\chi - \Delta H_{mix}$



شکل ۵. بسته‌بندی (پیش‌بینی‌کننده‌های مستقل) در مقابل تقویت (پیش‌بینی‌کننده‌های متوالی)

ب) روش اجرا برای الگوریتم تقویت:

- ۱- مقداردهی اولیه
- ۲- ورودی مدل به‌گونه‌ای آماده می‌شود که نمونه‌های دارای طبقه‌بندی شده اشتباه شناسایی شوند.
- ۳- وزن نقاط داده‌ی طبقه‌بندی اشتباه را افزایش می‌دهد.
- ۴- به مرحله‌ی ۵ از if else رفته (نتیجه‌ی لازم به دست می‌آید).
- ۵- بازگشت به مرحله‌ی ۲
- ۶- پایان

مقایسه‌ی عملکرد این دو روش در کاهش بایاس و واریانس: تقویت (روش بگینگ) با ترکیب تعداد زیادی درخت مستقل و نامرتب در مدل نهایی دارد به کاهش واریانس و افزایش پایداری مدل کمک کند. درمقابل، تقویت (روش بوستینگ) درخت‌ها را به صورت متوالی ایجاد می‌دهد، به‌گونه‌ای که هر درخت جدید بر خطاها و باقی‌مانده‌های درخت‌های قبلی تمرکز دارد. این فرایند با کاهش تدریجی اختلاف میان مقادیر واقعی و مقادیر پیش‌بینی‌شده، سوگیری مدل را کاهش می‌دهد و در نهایت عملکرد و دقت پیش‌بینی آن را بهبود می‌بخشد.

ج) شباهت بین بسته‌بندی و تقویت

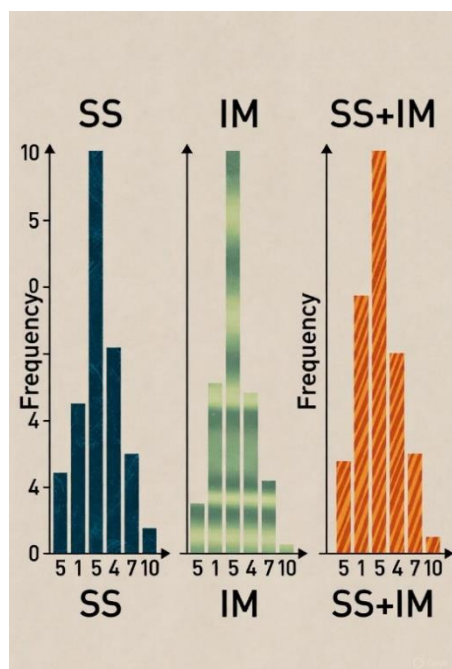
متداول‌ترین روش‌ها، بسته‌بندی و تقویت، به‌عنوان روش‌های الگوریتم یادگیری گروهی طبقه‌بندی می‌شوند.

گرفته شوند تا ابتدا فازهای SS+IM از IM و سپس فازهای SS از SS+IM متمایز شوند. برای نشان دادن کاربرد عملی مدل های یادگیری ماشین، این روش بر روی آلیاژهای آنتروپی بالای TiZrNbFeCr ، TiZrNbFeV و TiZrNbCrV اعمال شد. نتایج حاصل از الگوریتم RF حاکی از آن است که آلیاژهای TiZrNbFeCr و TiZrNbFeV عمدتاً دارای فاز محلول جامد هستند، درحالی که آلیاژ TiZrNbCrV شامل فاز محلول جامد همراه با فاز بین فلزی است.

۴- نتایج حاصل از الگوریتم های یادگیری گروهی

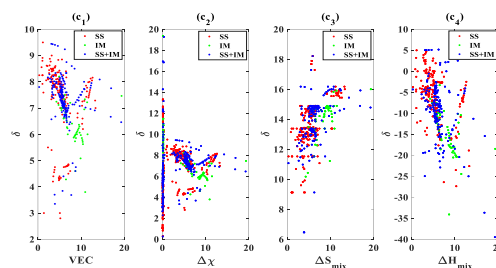
۴-۱- اهمیت متغیر

همان طور که در شکل ۱۱ مشاهده می شود، تجزیه و تحلیل داده های آماری بر روی پایگاه داده انجام می شود. در قطر تصویر، فازهای آلیاژهای آنتروپی بالا را توصیف می کنند که توسط بیش از یک مشخصه ورودی تعیین می شوند و نمی توان از یک نشان برای تعیین مرز مشخص بین فازها استفاده کرد.

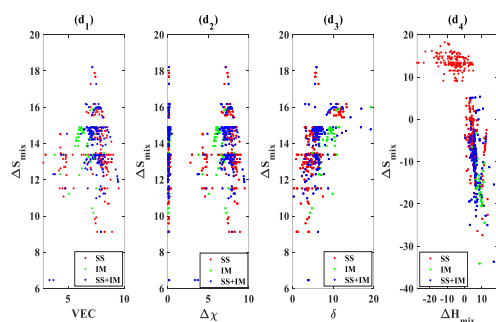


شکل ۱۱. نمایش هیستوگرام های توزیع فازهای SS، IM و SS+IM

نمودارهای پراکندگی توزیع سه فاز را به عنوان تابعی از یک جفت از پنج مشخصه در خارج از قطر تصویر نشان می دهند. هر پنج مشخصه در طبقه بندی توسط الگوریتم های یادگیری گروهی استفاده می شوند. قطر تصویر ۹ هیستوگرام های توزیع فازهای SS، IM و SS+IM را در مقادیر متغیر یکی از پنج مشخصه نشان می دهد.

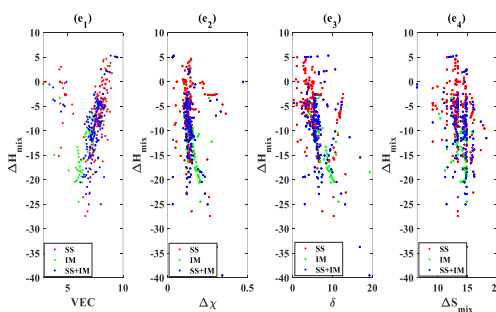


شکل ۸. نمودارهای پراکندگی که توزیع سه فاز را به عنوان تابعی از $(c_1)\Delta\chi - \text{VEC}$ ، $(c_2)\Delta\chi - \delta$ ، $(c_3)\Delta\chi - \Delta S_{\text{mix}}$ ، $(c_4)\Delta\chi - \Delta H_{\text{mix}}$



شکل ۹. نمودارهای پراکندگی که توزیع سه فاز را به عنوان تابعی از

$$(d_1)\Delta S_{\text{mix}} - \text{VEC}, (d_2)\Delta S_{\text{mix}} - \Delta\chi, (d_3)\Delta S_{\text{mix}} - \delta, (d_4)\Delta S_{\text{mix}} - \Delta H_{\text{mix}}$$

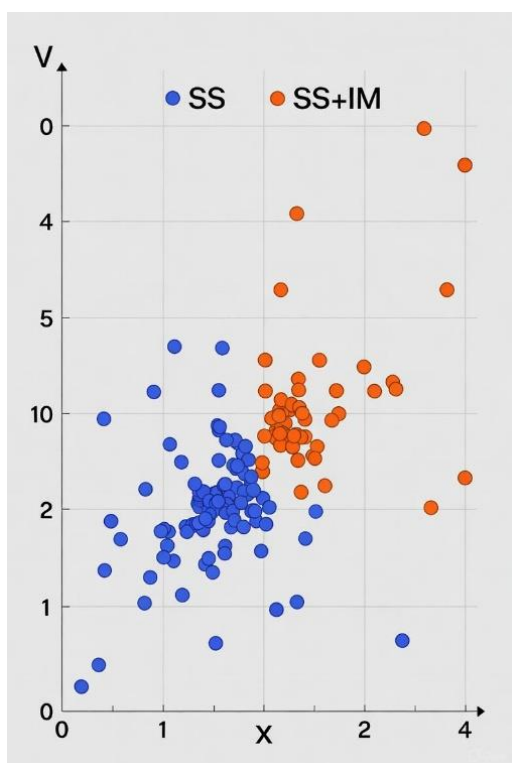


شکل ۱۰. نمودارهای پراکندگی توزیع سه فاز را به عنوان تابعی از

$$(e_1)\Delta H_{\text{mix}} - \text{VEC}, (e_2)\Delta H_{\text{mix}} - \Delta\chi, (e_3)\Delta H_{\text{mix}} - \delta, (e_4)\Delta H_{\text{mix}} - \Delta S_{\text{mix}}$$

۳-۲- نتیجه ی تخمین جنگل تصادفی

براساس جدول ۱، تمایز بین فازهای SS و IM و همچنین بین فازهای SS+IM و IM در روش جنگل تصادفی حتی با تعداد کمتری درخت به راحتی قابل دستیابی است. در مقابل، جداسازی فازهای SS از SS+IM چالش بیشتری دارد. با توجه به دقت بالای به دست آمده در طبقه بندی های دودویی، امکان طبقه بندی غیرمستقیم سه فاز نیز فراهم می شود. برای مثال، مدل های آموزش دیده می توانند در رویکردی دودویی به کار



شکل ۱۲. نمودار پراکندگی توزیع دو کلاس SS و SS+IM را هنگام پیش‌بینی در سیستم مختصات X-V نشان می‌دهد

نمودار شکل ۱۳ در اینجا نشان می‌دهد که کدام نمونه‌ها به اشتباه طبقه‌بندی شده‌اند. از آنجایی که روش RF به عنوان الگوریتم بسته‌بندی مجموعه‌ای کارایی بیشتری دارد، مجموعه داده‌ها توسط این الگوریتم با دقت بیشتری تجزیه و تحلیل می‌شوند. در شکل ۱۳، مقادیر بالاترین، میانگین و حداقل دقت برای سه وظیفه‌ی طبقه‌بندی با روش RF نشان داده شده است.

با توجه به جدول ۱، الگوریتم بگینگ یا بسته‌بندی گروهی در مقایسه با بوس‌تینگ یا درختان تقویت‌شده دقت بیشتری دارد. علاوه بر این، هر دو الگوریتم ارائه‌شده کارایی بیشتری در طبقه‌بندی باینری دارند.

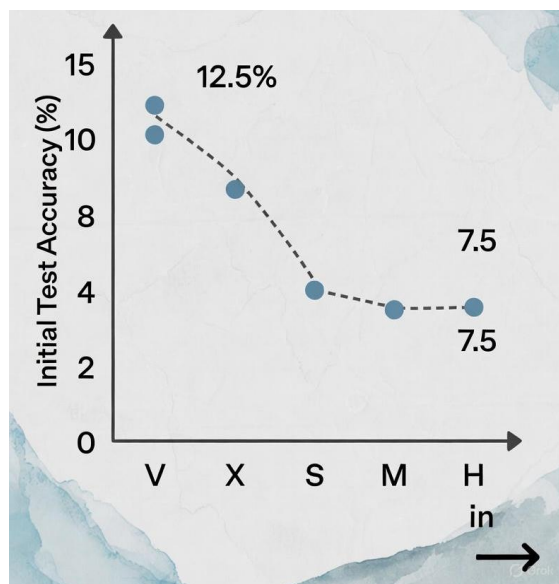
جدول ۱. دقت طبقه‌بندی کلی هر طبقه‌بندی‌کننده

بوس‌تینگ	بگینگ	فاز
دقت متوسط (%)	دقت متوسط (%)	
۷۰/۴	۷۵/۳	SS, IM, SS+IM
۷۸/۱	۹۰	SS, IM
۷۱/۲	۷۳/۱	SS, SS+IM
۷۹/۱	۸۵/۲	SS+IM IM,

علاوه بر این، این الگوریتم‌ها می‌توانند تفاوت بین فازهای SS و IM و همچنین فازهای SS+IM و IM را مشخص کنند، اما تمایز کلاس‌های SS و SS+IM دشوار است (شکل ۱۰ را ببینید). مرز بین فازهای SS و SS+IM، همان‌طور که توسط هر جفت از پنج ویژگی مشخص می‌شود، زمانی که در یک صفحه نمایش داده می‌شود، مبهم است و حتی برخی طبقه‌بندی‌ها با هم هم‌پوشانی دارند.

زمانی‌ای پیروی می‌کند که هر سه فاز را هم‌زمان دسته‌بندی می‌کند.

علاوه بر این، اهمیت نسبی پنج ویژگی ورودی در دقت آزمایش بررسی می‌شود، یعنی طبقه‌بندی مجموعه داده را براساس چهار ویژگی بررسی و کاهش دقت را محاسبه می‌کنیم. با توجه به شکل ۱۵، اهمیت V و X در مقایسه با سایر مشخصات اهمیت بالایی در دقت آزمایش دارند.



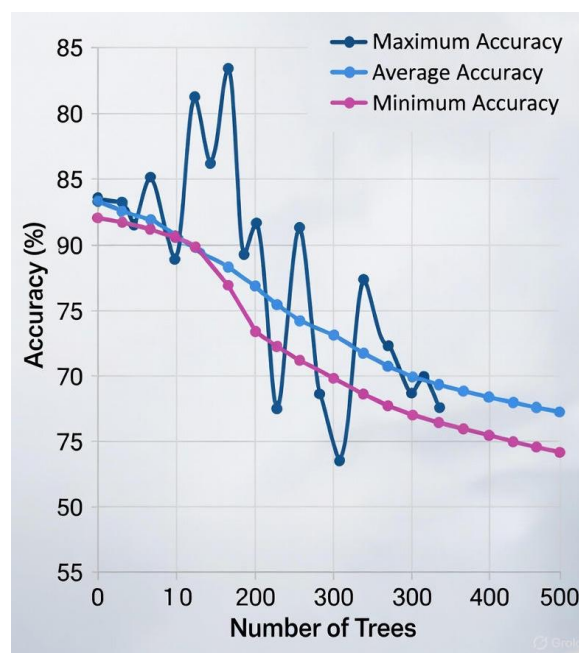
شکل ۱۵. تأثیر حذف یکی از پنج ویژگی بر دقت آزمایش اولیه هنگام طبقه‌بندی هر سه فاز با روش RF

مطابق شکل ۱۵، هرچه دقت بیشتر کاهش یابد، اهمیت این ویژگی را برای طبقه‌بندی فازها در آلیاژهای با آنتروپی بالا نشان می‌دهد. در جدول ۲، نتیجه‌ی تجربی روش RF با K-نزدیک‌ترین همسایه (KNN) و ماشین بردار پشتیبان (SVM) مقایسه شده است (Huang et al., 2019: 7).

جدول ۲. دقت هر طبقه‌بندی‌کننده

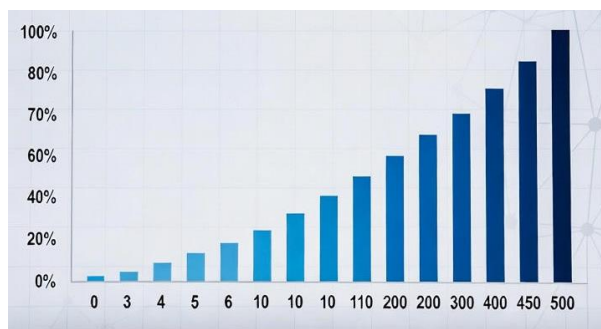
دقت (%)	طبقه‌بندی‌کننده
۷۵/۳	RF
۶۴/۳	SVM
۶۸/۶	KNN

با توجه به جدول ۲، دقت روش RF به‌عنوان الگوریتم بسته‌بندی مجموعه‌ای کارآمدتر از SVM و KNN برای



شکل ۱۳. نمایش نمودار حداکثر، میانگین و حداقل دقت آزمایش که از طبقه‌بندی‌کننده‌ی RF در مجموعه داده‌ی آلیاژهای آنتروپی بالا به دست آمده است

در شکل ۱۳، منحنی بالا و پایین به ترتیب حداکثر دقت آزمایش و کمترین دقت را نشان می‌دهد. نمودار دقت متوسط بین نمودارهای حداکثر و حداقل قرار می‌گیرد. در جنگل تصادفی به‌عنوان الگوریتم بسته‌بندی گروهی، افزایش تعداد درختان دقت روش را افزایش می‌دهد. مطابق شکل ۱۳، از یک عدد به بعد، افزایش تعداد درختان تأثیر قابل‌توجهی در دقت ندارد و به دلیل حجم عملیاتی باعث کاهش سرعت روش می‌شود.



شکل ۱۴. نقش تعداد درختان رشد یافته بر دقت در روش جنگل تصادفی

ارتباط نسبی پنج مشخصه‌ی ورودی، علاوه بر این، همان‌طور که در شکل ۱۲ نشان داده شده است، تجزیه و تحلیل می‌شود. مشاهده می‌شود که در هنگام طبقه‌بندی فازهای IM+IM و IM، ترتیب مرتبط بودن ویژگی‌ها از همان الگوی

۵- نتایج حاصل از پیش بینی جنگل تصادفی

در پژوهش حاضر، از الگوریتم جنگل تصادفی برای پیش بینی فازهای آلیاژهای آنتروپی بالا بهره گرفته شد. این الگوریتم بر روی یک مجموعه داده ی تجربی گسترده شامل ۴۰۱ آلیاژ آنتروپی بالای متمایز متشکل از ۱۷۴ نمونه با فاز SS، ۵۴ نمونه با فاز IM و ۱۷۳ نمونه با فاز SS+IM اجرا شد. دقت متوسط به دست آمده برای طبقه بندی هم زمان سه فاز ۷۴/۳ درصد بود. در ادامه، تمرکز بر طبقه بندی دودویی سه جفت فاز قرار گرفت و دقت های حاصل از روش RF برای تمایز فازهای SS از IM، SS+IM از SS و SS+IM از SS به ترتیب ۸۶/۷، ۹۴/۳ و ۷۸/۹ درصد ثبت شد. یافته های تجربی با سایر روش های یادگیری ماشین مقایسه شدند. نتایج برتری دقت روش پیشنهادی از تحقیقات پیشین را نشان می دهند. دو مزیت کلیدی این رویکرد در طبقه بندی آلیاژهای آنتروپی بالا عبارت هستند از: نخست، کاهش قابل توجه بیش برآزش در درخت های تصمیم که به ارتقای دقت کلی کمک می کند و، دوم، مدیریت خودکار داده های گم شده ی موجود در مجموعه داده. این مطالعه زمینه ساز روشی نوین در طراحی محاسباتی آلیاژهای آنتروپی بالا است که می تواند فرایند کشف آلیاژهای فلزی دیگر را برای کاربردهای مهندسی پیشرفته تسریع کند. براساس نتایج این روش، آلیاژ TiZrNbCrV عمدتاً فاز محلول جامد را همراه با ترکیبات بین فلزی تشکیل می دهد که این پیش بینی با نتایج حاصل از نمونه ی تفجوشی شده همخوانی دارد.

۶- نتایج حاصل از پیش بینی الگوریتم یادگیری

گروهی

در این پژوهش، از الگوریتم های یادگیری گروهی^۱ برای پیش بینی فازهای آلیاژهای آنتروپی بالا استفاده شد. اصل بنیادین این الگوریتم ها ترکیب چندین مدل پایه ی متفاوت است که به افزایش انعطاف پذیری (کاهش بایاس) و کاهش حساسیت به نوسانات داده ها (کاهش واریانس) منجر می شود. دو رویکرد رایج در یادگیری گروهی بگینگ و بوستینگ هستند. این الگوریتم ها بر روی مجموعه داده ی تجربی گسترده ای شامل ۴۰۱ آلیاژ آنتروپی بالای متمایز مشتمل بر ۱۷۴ نمونه با فاز SS، ۵۴ نمونه با فاز IM و ۱۷۳ نمونه با فاز SS+IM اجرا شدند. دقت

طبقه بندی (آلیاژهای آنتروپی بالا) است. فشردگی محاسباتی کمتر و عدم نیاز به استفاده از GPU برای تکمیل آموزش از مزایای اصلی روش RF است. طبقه بندی غیرمستقیم سه فاز به دلیل دقت آزمایش برتر در این طبقه بندی های باینری ممکن است. برای مثال، برای متمایز ساختن فازهای SS و SS+IM از فازهای IM و فازهای SS+IM از فاز IM، مدل های آموزش دیده ی به کار گرفته شده در اینجا را می توان در رویکردی دومرحله ای اعمال کرد.

به منظور نمایش کاربردهای مدل های یادگیری ماشین، از این روش برای شناسایی فازهای آلیاژهای TiZrNbCrV، TiZrNbFeCr و TiZrNbFeV استفاده شد. همان طور که در جدول ۵ مشاهده می شود، فازهای سه آلیاژ پیش بینی شده و، بر این اساس، تمام آلیاژها حاوی فاز محلول جامد هستند.

جدول ۳. مرحله ی طبقه بندی کلی هر طبقه بندی کننده برای آلیاژهای TiZrNbCrV، TiZrNbFeCr و TiZrNbFeV با آنتروپی بالا

آلیاژها	Bagging trees	Boosted trees
TiZrNbCrV TiZrNbFeCr TiZrNbFeV	SS	SS

برای بررسی صحت نتایج یادگیری ماشین، نتایج XRD سه آلیاژ انتخاب شده ارائه شده است. تمام نمونه های TiZrNbCrV، TiZrNbFeCr و TiZrNbFeV در حالت دو فاز محلول جامد بدون هیچ فاز بین فلزی هستند. محلول جامد فعلی برای هر سه آلیاژ دوفازی است. فاز اصلی BCC (مکعب مرکزپر) است، که در آن BCC1 بخش اصلی و BCC2 بخش فرعی با کسرهای کوچک هستند. وجود دو فاز به جای یک فاز می تواند به نحوه ی تولید آن ها مربوط باشد. علاوه بر این، نتایج عملی نشان می دهد که الگوریتم های یادگیری مجموعه ای سازگاری مناسبی در شرایط واقعی دارند و می توانند کمک بزرگی برای طراحی آلیاژهای جدید با آنتروپی بالا باشند (Vaidya et al., 2019: 41). یکی از مسائل موجود در حوزه ی پژوهش درباره ی آلیاژهای آنتروپی بالا استفاده از روش های هوش مصنوعی یا داده کاوی به منظور کاهش تعداد آزمایش ها و پیش بینی نتایج است، که در این زمینه بحث پیش بینی فاز نهایی توسط یادگیری عمیق از ۴۰۱ داده ی موجود توسط دو الگوریتم جنگل تصادفی و یادگیری گروهی (بگینگ و بوستینگ) بررسی شد.

براساس نتایج RF، آلیاژهای TiZrNbFeV و TiZrNbFeCr عمدتاً فاز محلول جامد تشکیل می دهند، درحالی که آلیاژ TiZrNbCrV شامل فاز محلول جامد همراه با فاز بین فلزی است.

دقت متوسط روش های یادگیری گروهی برای طبقه بندی سه فاز (محلول جامد، بین فلزی و ترکیب هر دو) در روش بگینگ ۷۵/۳ درصد و در روش بوستینگ ۷۰/۴ درصد به دست آمد. برای نشان دادن کاربرد عملی مدل های یادگیری ماشین، این رویکردها بر روی آلیاژهای TiZrNbCrV ، TiZrNbFeCr و TiZrNbFeV اعمال شدند و نتایج پیش بینی نشان داد که تمامی این آلیاژها حاوی فاز محلول جامد هستند (Yousefi, 2025a: 42-45).

براساس توسعه کارهای پیشین (Yousefi, 2024b: 42-46)، بررسی مقایسه ای جامعی بر عملکرد الگوریتم های یادگیری ماشین مختلف در پیش بینی فازهای آلیاژهای آنتروپی بالا، به ویژه در سری آلیاژهای پایه TiZrNb ارائه شده است. در تحقیقات پیشین، تمرکز عمدتاً بر کاربرد الگوریتم های خاصی مانند جنگل تصادفی^۱ و یادگیری گروهی بگینگ و بوستینگ بوده است که نتایج قابل توجهی در دقت طبقه بندی فازهای SS، IM و SS+IM نشان داده اند. در تحقیق حاضر، با بهره گیری از همان پایگاه داده ی تجربی گسترده (شامل ۴۰۱ آلیاژ آنتروپی بالا)، عملکرد این الگوریتم ها مجدداً ارزیابی و مقایسه شد تا نقاط قوت و محدودیت های هر کدام در شرایط یکسان روشن شود.

علاوه بر این، یکی از اهداف کلیدی این مطالعه برجسته کردن پتانسیل و توانمندی های روش های یادگیری ماشین پیشرفته تر است که تا کنون در حوزه ی آلیاژهای آنتروپی بالا کمتر مورد توجه قرار گرفته اند. الگوریتم هایی نظیر شبکه های عصبی عمیق، یادگیری تقویتی^۲ مدل های مبتنی بر ترانسفورمر^۳ یا رویکردهای هیبرید (مانند ترکیب یادگیری ماشین با محاسبات کوانتومی یا شبیه سازی های مولکولی) می توانند دقت پیش بینی را بهبود بخشند، داده های پیچیده تر را مدیریت کنند و حتی به کشف ترکیبات شیمیایی جدید کمک کنند. این روش های نوین با قابلیت پردازش ویژگی های چندبعدی و روابط غیرخطی

متوسط طبقه بندی هم زمان سه فاز با روش های بوستینگ و بگینگ به ترتیب ۷۰/۴ و ۷۵/۳ درصد به دست آمد. در ادامه، تمرکز بر طبقه بندی دودویی قرار گرفت و دقت های حاصل از الگوریتم بگینگ برای تمایز فازهای SS از IM، SS از SS+IM و IM از SS+IM به ترتیب ۹۰، ۷۳/۱ و ۸۵/۲ درصد بود. مقادیر مربوط برای الگوریتم بوستینگ نیز به ترتیب ۷۸/۱، ۷۱/۲ و ۷۹/۱ درصد ثبت شد. براساس این نتایج عددی، الگوریتم های مبتنی بر بگینگ کارایی مؤثرتری در طبقه بندی فازهای آلیاژهای آنتروپی بالا از خود نشان دادند.

دقت روش پیشنهادی حدود ۱۰ درصد بالاتر از الگوریتم های SVM و KNN در طبقه بندی آلیاژهای آنتروپی بالا است. مرز جداکننده بین فازهای SS و SS+IM هنگام ترسیم براساس جفت های مختلف از پنج ویژگی اغلب نامشخص است و در مواردی هم پوشانی مشاهده می شود. روش پیشنهادی دو برتری اصلی در این حوزه دارد: نخست، کاهش بیش برآزش در درخت های تصمیم که به بهبود دقت کمک می کند و، دوم، مدیریت خودکار داده های گم شده در مجموعه داده. این مطالعه رویکردی نوین برای طراحی محاسباتی آلیاژهای آنتروپی بالا ارائه می دهد که می تواند فرایند کشف آلیاژهای فلزی دیگر را برای کاربردهای مهندسی پیشرفته از جمله مواد زیست سازگار یا صنعتی تسریع کند. پیش بینی های حاصل از این روش نشان می دهد که تمامی نمونه ها عمدتاً حاوی فاز محلول جامد هستند که این نتیجه با یافته های آلیاژسازی مکانیکی و نمونه های تفجوشی شده آلیاژ TiZrNbFeV همخوانی دارد.

نتایج تجربی حاکی از آن است که الگوریتم های یادگیری گروهی سازگاری مناسبی در شرایط واقعی دارند و می توانند نقش مهمی در طراحی آلیاژهای جدید آنتروپی بالا ایفا کنند. در الگوریتم جنگل تصادفی، افزایش تعداد درخت ها به بهبود دقت منجر می شود. اما، پس از یک حد مشخص، افزودن درخت های بیشتر تأثیر ناچیزی در دقت دارد و تنها سرعت محاسبات را به دلیل حجم عملیاتی بالاتر کاهش می دهد. دقت نهایی پیش بینی فازها با روش جنگل تصادفی ۷۵/۳ درصد بود که بر روش های SVM با دقت ۶۴/۳ درصد، KNN با ۶۸/۶ دقت درصد و ANN با دقت ۷۴/۳ درصد برتری دارد.

تعمیم‌پذیری مدل‌ها افزایش یابد و کشف ترکیبات نوین تسهیل شود.

سپاسگزاری

از کمک‌های آقای مهندس مجید عقاب‌نشین در تحلیل داده‌ها سپاسگزاریم.

منابع

- Banfield, R. E., Hall, L. O., Bowyer, K. W., & Kegelmeyer, W. P. (2007). A comparison of decision tree ensemble creation techniques. *IEEE Trans Pattern Anal Mach Intell.*, 29(1), 173-80. <https://doi.org/10.1109/tpami.2007.250609>. PMID: 17108393
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Cantor, B. (2014). Multicomponent and high entropy alloys. *Entropy*, 16(9), 4749-4768. <https://doi.org/10.3390/e16094749>
- Feng, R., Liaw, P. K., Gao, M. C. et al. (2017). First-principles prediction of high-entropy-alloy stability. *Npj Comput Mater*, 3, 50. <https://doi.org/10.1038/s41524-017-0049-4>
- Fontana, M. G., & Greene, N. D. (2018). *Corrosion engineering*. McGraw-Hill. <https://repository.um-palembang.ac.id/id/eprint/9139/1/Mars%20Fontana-Corrosion%20Engineering%28www.iranidata.com%29.pdf>
- Haase, C., Tang, F., Wilms, M. B., Weisheit, A., & Hallstedt, B. (2017). Combining thermodynamic modeling and 3D printing of elemental powder blends for high-throughput investigation of high-entropy alloys – Towards rapid alloy screening and design. *Materials Science and Engineering: A*, 688, 180-189, ISSN 0921-5093. <https://doi.org/10.1016/j.msea.2017.01.099>
- Han, L., Luo, S., Yu, J., Pan, L., & Chen, S. (2015). Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. *IEEE J Biomed Health Inform*, 19(2), 728-734. <https://doi.org/10.1109/JBHI.2014.2325615>. Epub 2014 May 19. PMID: 24860043.
- Harrysson, O. L. A., & Cormier, D. R. (2005). Direct fabrication of custom orthopedic implants using electron beam melting technology. In *Advanced Manufacturing Technology for Medical Applications* (pp. 191-206). <https://scispace.com/pdf/direct-fabrication-of-custom-orthopedic-implants-using-2mtipatlwj.pdf>
- Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 832-844. <http://dx.doi.org/10.1109/34.709601>
- Hu, J., Zhang, J., Xiao, H., Xie, L., Shen, H., Li, P., ... & Zu, X. (2020). A density functional theory study of the hydrogen absorption in high entropy alloy TiZrHfMoNb. *Inorganic Chemistry*, 59(14), 9774-9782. <https://doi.org/10.1021/acs.inorgchem.0c00989>
- Huang, W., Martin, P., & Zhuang, H. L. (2019). Machine-learning phase prediction of high-entropy alloys. *Acta Materialia*, 169, 225-236. <https://doi.org/10.1016/j.actamat.2019.03.012>
- Huhn, W. P., & Widom, M. (2013). Prediction of A2 to B2 phase transition in the high-entropy alloy Mo-Nb-Ta-W. *JOM*, 65, 1772-1779. <https://doi.org/10.1007/s11837-013-0772-3>
- King, D. J. M., Middleburgh, S. C., McGregor, A. G., & Cortie, M. B. (2016). Predicting the formation and stability of single phase high-entropy alloys. *Acta Materialia*, 104, 172-179. ISSN 1359-6454. <https://doi.org/10.1016/j.actamat.2015.11.040>
- Krishna, Y. V., Jaiswal, U. K., & Rahul, M. R. (2021). Machine learning approach to predict new multiphase high entropy alloys.

پیچیده‌تر می‌توانند مکمل رویکردهای فعلی باشند و راه را برای طراحی هوشمندتر و سریع‌تر آلیاژهای آنتروپی بالا به‌ویژه در کاربردهای زیست‌پزشکی هموار سازند.

این پژوهش نه تنها خلأهای موجود در مقایسه‌ی نظام‌مند روش‌ها را پر می‌کند، بلکه زمینه‌ساز تحقیقات آتی برای ادغام تکنیک‌های پیشرفته‌تر یادگیری ماشین در این حوزه خواهد بود. که در آینده‌ی نزدیک مقالات پژوهشی آن‌ها نیز منتشر خواهد شد.

۷- نتیجه‌گیری

این پژوهش، به‌عنوان مطالعه‌ای مقایسه‌ای، عملکرد الگوریتم‌های یادگیری ماشین مختلف از جمله جنگل تصادفی و یادگیری گروهی بگینگ و بوستینگ را در پیش‌بینی فازهای آلیاژهای آنتروپی بالا ارزیابی کرد و، بر پایه‌ی پایگاه داده‌ی تجربی مشترک شامل حدود ۴۰۰ آلیاژ، مقایسه‌ای نظام‌مند در شرایط یکسان ارائه داد. نتایج این مقایسه نشان داد که روش بگینگ، با دقت متوسط بیش از ۷۵ درصد در طبقه‌بندی سه‌فازی، کارایی برتری از بوستینگ و تکنیک‌های پایه مانند SVM و KNN دارد، درحالی‌که جنگل تصادفی در مدیریت بیش‌برازش و محاسبات سبک (بدون نیاز به GPU) عملکرد بهتری از خود نشان داد.

این بررسی مقایسه‌ای نه تنها نتایج کارهای پیشین را تکمیل و تأیید کرد (Yousefi et al., 2024a; Yousefi et al., 2024c; Yousefi et al., 2025b: 45; 43)، بلکه با پرکردن خلأ عدم ارزیابی مستقیم این الگوریتم‌ها درک عمیق‌تری از نقاط قوت و محدودیت‌های هر روش فراهم آورد. کاربرد عملی مدل‌ها بر روی آلیاژهای پایه TiZrNbX نیز با نتایج تجربی XRD سازگاری مناسبی نشان داد و پتانسیل بالای این رویکردها در پیش‌بینی واقعی فازها (عمدتاً محلول جامد) را تأیید کرد.

درنهایت، این مطالعه‌ی مقایسه‌ای تأکید دارد که الگوریتم‌های جنگل تصادفی و یادگیری گروهی ابزارهای مؤثر، سریع و کم‌هزینه‌ای برای تسریع طراحی آلیاژهای آنتروپی بالا به‌ویژه در کاربردهای زیست‌پزشکی مانند ایمپلنت‌های ارتوپدی هستند. پیشنهاد می‌شود تحقیقات آینده بر ادغام روش‌های پیشرفته‌تر مانند شبکه‌های عصبی عمیق یا مدل‌های هیبرید تمرکز کنند تا، با پایگاه‌های داده‌ی بزرگ‌تر، دقت و قابلیت

- Materials Research*, 34(5), 664–686. <https://doi.org/10.1557/jmr.2019.37>
33. Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2(1), 16028. <https://doi.org/10.1038/npjcompumats.2016.28>
 34. Yang, B. S., Di, X., & Han, T. (2008). Random forests classifier for machine fault diagnosis. *Journal of Mechanical Science and Technology*, 22, 1716–1725. <https://doi.org/10.1007/s12206-008-0603-6>
 35. Ye, Y. F., Wang, Q., Lu, J., Liu, C. T., & Yang, Y. J. M. T. (2016). High-entropy alloy: challenges and prospects. *Materials Today*, 19(6), 349–362. <https://doi.org/10.1016/j.mattod.2015.11.026>
 36. Yeh, J. W. (2016). Overview of high-entropy alloys. In *High-entropy alloys: Fundamentals and applications* (pp. 1–19). doi: <https://doi.org/10.1007/978-3-319-27013-5>
 37. Yousefi, M., Rajabi, M., Reyhani, A., & Asgari, N. (2024a). Corrosion and biocompatibility properties of TiZrNbCrV, TiZrNbFeCr, and TiZrNbFeV. *Journal of Materials Engineering and Performance*, 33. <https://doi.org/10.1007/s11665-023-08990-3>
 38. Yousefi, M., Rahmani, K., Rajabi, M., Reyhani, A., & Moudi, M. (2024b). Random forest classifier for high entropy alloys phase diagnosis. *Afrika Matematika*, 35(3), 57. <https://doi.org/10.1007/s13370-024-01198-1>
 39. Yousefi, M., Rahmani, K., & Rajabi, M., (2024c). Investigating X-ray diffraction and microstructure analysis of new high entropy alloys TiZrNbXX. *Advanced Processes in Materials Engineering*, 18, 69. <https://jnm.marvdasht.iau.ir/?lang=fa>
 40. Yousefi, M., Rajabi, M., Reyhani, A., Asgari, N., & Rahmani, K. (2025a). Investigation of Microstructure of Pressed and Sintered TiZrNbCrV, TiZrNbFeCr, and TiZrNbFeV High Entropy Alloys Produced by Powder Metallurgy and Mechanical Alloying Methods. *Metal Science and Heat Treatment*, 1-7. <https://doi.org/10.1007/s11041-025-01165-y>
 41. Yousefi, M., Rahmani, K., Rajabi, M., Reyhani, A., & Asgari, N. (2025b). The ensemble learning algorithms for prediction high entropy alloys phases. *Afrika Matematika*, 36(3), 117. <https://doi.org/10.1007/s13370-025-01334-5>
 42. Zhang, Y., Zhou, Y. J., Lin, J. P., Chen, G. L., & Liaw, P. K. (2008). Solid-solution phase formation rules for multi-component alloys. *Advanced engineering materials*, 10(6), 534–538. <https://doi.org/10.1002/adem.200700240>
 43. Zhang, Y. (2010). Mechanical properties and structures of high entropy alloys and bulk metallic glasses composites. In *Materials Science Forum* (Vol. 654). <https://doi.org/10.4028/www.scientific.net/MSF.654-656.1058>
 44. Zhang, C. H., Lin, M. H., Wu, B., Ye, G. X., Zhang, L. K., Chen, T., ... & Wang, C. (2011). Explore the possibility of forming fcc high entropy alloys in equal-atomic systems CoFeMnNi M and CoFeMnNiSm M. *Journal of Shanghai Jiaotong University (Science)*, 16(2), 173–179. <https://doi.org/10.1007/s12204-011-1113-8>
 45. Zhang, B., Qi, S., Monkam, P., Li, C., Yang, F., Yao, Y. D., & Qian, W. (2019). Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. *IEEE Access*, 7, 110358–110371. <https://doi.org/10.1109/ACCESS.2019.2933670>
 - Scripta Materialia*, 197, 113804. <https://doi.org/10.1016/j.scriptamat.2021.113804>
 15. Lee, S. Y., Byeon, S., Kim, H. S., Jin, H., & Lee, S. (2021). Deep learning-based phase prediction of high-entropy alloys: Optimization, generation, and explanation. *Materials & Design*, 197, 109260. <https://doi.org/10.1016/j.matdes.2020.109260>
 16. Leon, F., Floria, S. A., & Bădică, C. (2017). Evaluating the effect of voting methods on ensemble-based classification. In *2017 IEEE International Symposium on Intelligent Signal Processing (INISTA)*. <https://doi.org/10.1109/INISTA.2017.8001122>
 17. Lind, A. P., & Anderson, P. C. (2019). Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PLOS ONE*, 14(7), e0219774. <https://doi.org/10.1371/journal.pone.0219774>
 18. Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random Forest. In *ICICA 2012*. Springer. https://doi.org/10.1007/978-3-642-34062-8_32
 19. Machaka, R. (2021). Machine learning-based prediction of phases in high-entropy alloys. *Computational Materials Science*, 188, 110244. <https://doi.org/10.1016/j.commatsci.2020.110244>
 20. Miracle, D. B., & Senkov, O. N. (2017). A critical review of high entropy alloys and related concepts. *Acta Materialia*, 122, 448–511. <https://doi.org/10.1016/j.actamat.2016.08.081>
 21. Qiao, L., Liu, Y., & Zhu, J. (2021). A focused review on machine learning aided high-throughput methods in high entropy alloy. *Journal of Alloys and Compounds*, 877, 160295. <https://doi.org/10.1016/j.jallcom.2021.160295>
 22. Rickman, J. M., Balasubramanian, G., Marvel, C. J., Chan, H. M., & Burton, M. T. (2020). Machine learning strategies for high-entropy alloys. *Journal of Applied Physics*, 128(22). <https://doi.org/10.1063/5.0030367>
 23. Roy, A., & Balasubramanian, G. (2021). Predictive descriptors in machine learning and data-enabled explorations of high-entropy alloys. *Computational Materials Science*, 193, 110381. <https://doi.org/10.1016/j.commatsci.2021.110381>
 24. Ruta, D., Gabrys, B., & Lemke, C. (2010). A generic multilevel architecture for time series prediction. *IEEE Transactions on Knowledge and Data Engineering*, 23(3), 350–359. <https://doi.org/10.1109/TKDE.2010.137>
 25. Schapire, R. E., & Singer, Y. (1998). Improved boosting algorithms using confidence-rated predictions. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT)*. <https://doi.org/10.1145/279943.279960>
 26. Singh, S. K., & Singh, V. K. (2021). *Design and development of high entropy alloys using artificial intelligence*. IntechOpen. <https://doi.org/10.5772/intechopen.96761>
 27. Smallman, R. E., & Ngan, A. H. W. (2011). *Physical metallurgy and advanced materials*. Elsevier. <https://shop.elsevier.com/books/physical-metallurgy-and-advanced-materials/smallman/978-0-7506-6906-1>
 28. Sun, W., Huang, X., & Luo, A. A. (2017). Phase formations in low density high entropy alloys. *Calphad*, 56, 19–28. <https://doi.org/10.1016/j.calphad.2016.11.002>
 29. Talebi, H., Peeters, L. J., Otto, A., & Tolosana-Delgado, R. (2022). A truly spatial random forests algorithm for geoscience data analysis and modelling. *Mathematical Geosciences*, 54(1), 1–22. <https://doi.org/10.1007/s11004-021-09946-w>
 30. Tokar, O., Vovk, O., Kolyasa, L., Havryliuk, S., & Korol, M. (2018, September). Using the Random Forest Classification for Land Cover Interpretation of Landsat Images in the Prykarpattya Region of Ukraine. In *CSIT (1)* (pp. 241–244). <https://doi.org/10.1109/STC-CSIT.2018.8526646>
 31. Tsai, M. H., & Yeh, J. W. (2014). High-entropy alloys: A critical review. *Materials Research Letters*, 2(3), 107–123. <https://doi.org/10.1080/21663831.2014.912690>
 32. Vaidya, M., Muralikrishna, G. M., & Murty, B. S. (2019). High-entropy alloys by mechanical alloying: A review. *Journal of*