

Information Sciences

Better utilization of materials' compositions for predicting their properties: Materials Composition Visualization Network --Manuscript Draft--

Manuscript Number:	
Article Type:	Full length article
Keywords:	visualization of material chemical composition characteristics; multimodal learning; materials property prediction
Corresponding Author:	Quan Qian, PhD Shanghai University CHINA
First Author:	Yeyong Yu
Order of Authors:	Yeyong Yu Xing Wu, PhD Quan Qian, PhD
Abstract:	<p>Materials informatics is the data-driven paradigm in materials research and development and plays the central role in materials genome engineering (MGE). And the core idea is to discover the hidden relationship of material composition-process-structure-performance. However, current methods for chemical composition are limited to the proportions and the chemical characteristics are lost. In this study, a data mapping scheme based on fundamental atomic features was used to visualize the chemical composition characteristics mapped into 2-dimensional grayscale image. Based on this, a material composition visualization network (MCVN) is proposed and applied to predict the steel mechanical properties and the amorphous alloy materials classification. The MCVN had an average R^2 improvement of 3% on the four targets in National Institute for Materials Science dataset, where other models already get an average R^2 of 0.92, and an average R^2 of 0.835 on the cross-sectional shrinkage target on Shanghai Research Institute of Materials Steel Dataset where the other models only had an R^2 of 0.6. For the imbalanced amorphous alloy materials dataset, the MCVN improved the Recall of small-class Crystalline Alloy (CRA) from 0.5 to 0.78. The expanding method of material chemical composition is universal, providing a new paradigm for material property prediction.</p>
Suggested Reviewers:	<p>Hai Jiang, PhD. Prof., Arkansas State University hjiang@astate.edu Expert in the subject</p> <p>Che-Lun Hung, PhD. Prof., National Yang Ming Chiao Tung University clhung@nycu.edu.tw Expert in the subject</p> <p>Dali Zhang, PhD. Prof., Shanghai Jiaotong University: Shanghai Jiao Tong University zhangdl@sjtu.edu.cn Expert in the subject</p> <p>Jianxun Liu, PhD. Prof., Hunan University of Science and Technology ljx529@gmail.com Expert in the subject</p> <p>Jian Cao, PhD. Prof., Shanghai Jiaotong University: Shanghai Jiao Tong University cao-jian@sjtu.edu.cn Expert in the subject</p> <p>Jie Xiong, PhD.</p>

	Harbin Institute of Technology Shenzhen xiongjie@hit.edu.cn Expert in the subject
--	---

April 14, 2022

Information Sciences

Dear Editor:

We wish to submit an article for publication in *Information Sciences*, titled “Better utilization of materials’ compositions for predicting their properties: Materials Composition Visualization Network” authored by *Yeyong Yu & Xing Wu*.

Materials informatics is the data-driven paradigm in materials research and development and plays the central role in materials genome engineering (MGE). It seamlessly integrates materials science and engineering with artificial intelligence and machine learning. In MGE, the chemical composition characteristics of materials are significant features, and the proportions of the material’s compositions do not adequately accentuate them. How to translate the material composition information to another modality and improve the accuracy of materials’ properties prediction by means of multimodal joint representation learning is an urgent problem to be solved.

In this paper, a method for visualizing elemental extension information based on expert opinion is applied to utilize material compositional features better for machine learning. Based on this, a material composition visualization network (MCVN) is proposed to learn the modal fusion of material composition visualization images and raw features to enhance prediction accuracy. The model is validated on regression and classification tasks by three different datasets, and the effectiveness of the modal and network structure modules are verified by ablation experiments.

Further, we believe that this paper will be of interest to the readership of your journal because the method based on expanding material chemical composition information is highly universal and novel, providing a new paradigm for materials’ properties prediction. Besides, due to the extreme heterogeneity of material features, multimodal learning has great application scenarios in material machine learning tasks.

This article has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. There are no conflicts of interest to declare.

Thank you for your consideration. I look forward to hearing from you.

Sincerely,

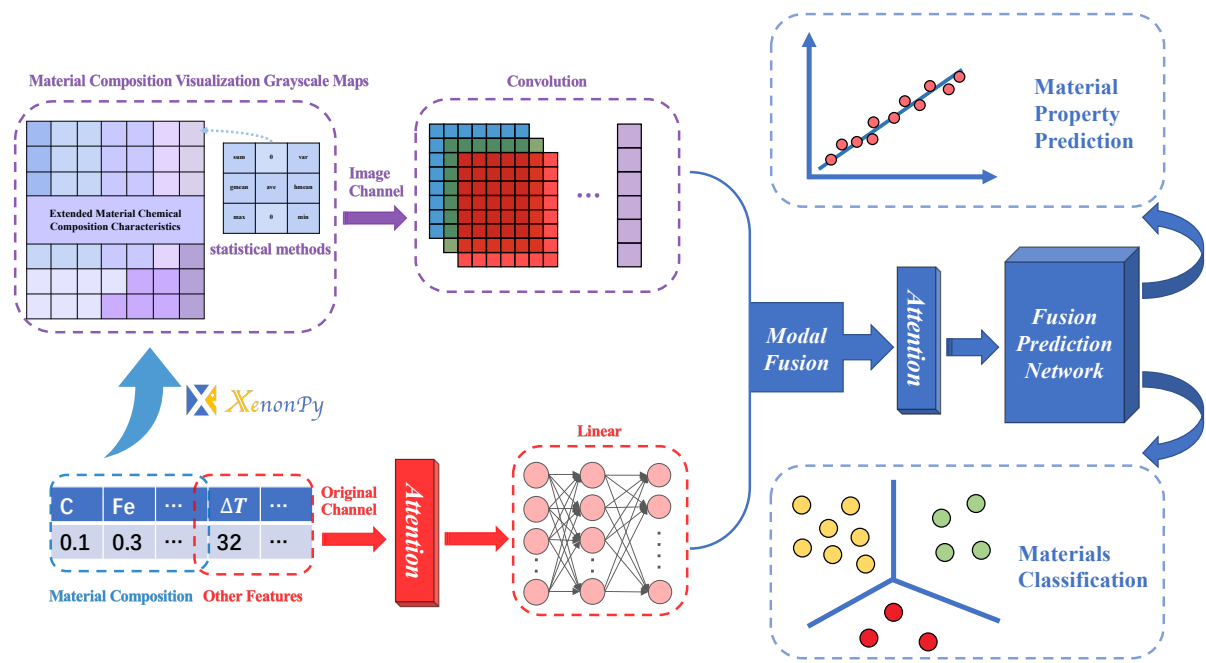
Quan Qian

School of Computer Engineering & Science
Shanghai University
#99 Shangda Rd., Baoshan District, Shanghai, China, 200444
+86-21-66135396
qqian@shu.edu.cn

Graphical Abstract

Better utilization of materials' compositions for predicting their properties: Materials Composition Visualization Network

Yeyong Yu,Xing Wu,Quan Qian



Highlights

Better utilization of materials' compositions for predicting their properties: Materials Composition Visualization Network

Yeyong Yu,Xing Wu,Quan Qian

- The problem in most material machine-learning tasks is identified as using only compositional proportions and losing other properties of the composition.
- A method for visualizing elemental extension information based on expert opinion is applied to utilize compositional features better in material machine-learning tasks.
- A custom multimodal deep-network model is proposed to learn the modal fusion of material composition visualization images and raw features to enhance the model prediction accuracy.
- The experiments results demonstrate the superiority of the proposed method while verifying the efficacy of the modal and network structure modules with ablation experiments.

Better utilization of materials' compositions for predicting their properties: Materials Composition Visualization Network

Yeyong Yu^a, Xing Wu^{a,b,c,d} and Quan Qian^{a,b,c,d,*}

^aSchool of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

^bMaterials Genome Institute, Shanghai University, Shanghai, 200444, China

^cShanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, 200444, China

^dZhejiang Laboratory, Hangzhou, 311100, Zhejiang, China

ARTICLE INFO

Keywords:

visualization of material composition characteristics
multimodal learning
materials property prediction


ABSTRACT

Owing to the complexity and diversity of advanced high-performance materials, it is challenging comprehensively to understand the material composition-process-structure-performance relationship. In the Materials Genome Project, data-driven R&D on new materials, which integrates information technologies such as big data, artificial intelligence, machine learning, and data mining, is considered the fourth paradigm of materials R&D and can achieve double halving of cost and time. However, current material property prediction methods for chemical composition characteristics are limited to the proportions and the chemical characteristics are lost. In this study, a data mapping scheme based on fundamental atomic features was used to visualize the chemical composition characteristics mapped into 2-dimensional grayscale image data for the problem of acquiring single material composition ratio information. Based on this, a material composition visualization network (MCVN) is proposed and applied to predict the mechanical properties of steel and the classification of amorphous alloy materials. We compared the MCVN to other machine learning methods: The MCVN had an average R^2 value improvement of 3% on the four targets in the National Institute for Materials Science's (NIMS's) steel dataset, where other models already get an average R^2 of 0.92, and an R^2 of 0.835 on the cross-sectional shrinkage target on the Shanghai Research Institute of Materials' (SRIM's) steel dataset where the other models only had an R^2 of 0.6. For the unbalanced amorphous alloy materials dataset, the MCVN improved the *Recall* of small-class Crystalline Alloy (CRA) from 0.5 to 0.78. The method based on expanding the material chemical composition information is universal, providing a new paradigm for material property prediction.

1. Introduction

The demand for novel materials is becoming increasingly robust. Whether for military and defense applications, high technology equipment, or personal electronics these materials are essential. In new material development, identifying the relationship between the composition, process, structure, and property is the basis for discovery. However, owing to the diversity and complexity of materials, comprehensively understanding and handling the composition-structure-process-property relationships is extremely challenging. Therefore, the data-driven development of advanced materials is central to the Materials Genome Initiative (MGI). Big data and advanced technologies, such as artificial intelligence, data mining, and machine learning, are used to accelerate R&D. Big data and machine learning for materials provide a theoretical and methodological basis for the data-driven discovery paradigm, which combines material domain knowledge with artificial intelligence techniques to create a new research area in materials informatics. The MGI aims to combine high-throughput experimentation, high-throughput computing, and materials informatics to reduce the financial and time costs of taking advanced materials from discovery to application.

Along with unimodal learning approaches, multimodal learning has received increasing attention from researchers, with many results emerging from fields such as addressing multi-objective optimization problems [41], emotion recognition [32], fake news detection [23], and autonomous driving [26]. Multimodal learning trains models by analyzing data from multiple modalities. It can yield superior results when the research object involves a combination of modalities. Combining different modal data is necessary to ensure that the whole model achieves the best results. There are many heterogeneous data in the field of material genomics, such as the text information on material composition

 yuyeyong@shu.edu.cn (Y. Yu); xingwu@shu.edu.cn (X. Wu); qqian@shu.edu.cn (Q. Qian)
ORCID(s):

and processes that resort under structural data, imaging of the material under visual data, and its molecular structure under graph data. Fusing the data of various structures to improve the accuracy rate of materials' properties prediction is an urgent challenge. Therefore, multimodal research has excellent prospects in the field of material genomics.

Multimodal learning can utilize the rich information of different modal data to train models and improve their prediction accuracies. This study focused on two aspects of multimodal: modal representation and modal fusion. Modal representation refers to combining data between different modalities meaningfully with a minimal loss of their respective semantics. Modal representation commonly uses the joint and coordinated approaches [24]. Joint representation is used to map the data of different modalities to the same feature space and perform training there [22]. This method suits cases in which all modalities must be used for prediction. Collaborative feature representation is a separate mapping representation of multiple modalities with external constraints to ensure that the representation between modalities is correlated [31]. Co-feature representation suits application scenarios when all modalities are not simultaneously present in the prediction process.

This study focuses on multimodal research with practical applications that incorporate the characteristics of data from the material genomics field. Specifically, chemical composition is an essential feature in material genomics. However, when simply relying on the compositional proportions, the model has difficulty in capturing information about the intrinsic characteristics of the materials' chemical compositions. The expansion, mapping, and enhancement of the chemical composition information of materials to make more significant differences between different chemical composition samples and how the model can learn the enhanced information of the chemical composition of materials with the original information are urgent problems that must be solved.

In this paper, we focus on expanding and visualizing material compositional features. A material composition visualization network (MCVN), based on the extended mapping of image modalities using chemical composition information for cross-modal joint learning, was designed to realize the extraction and fusion of different modal data features in Fig.1. Experimental validation was performed to predict the mechanical properties of steel and the classification datasets of amorphous alloy materials. The main contributions of this study are as follows:

- The problem is identified of using only compositional proportions in most material machine-learning tasks and losing other properties.
- A method for visualizing elemental extension information based on expert opinion is applied to utilize material compositional features better for machine learning.
- A novel multimodal deep network, represented in Fig.4, is proposed to learn the modal fusion of material composition visualization images and raw features to enhance prediction accuracy.

1.1. Motivation

The motivation for augmenting material composition information with atomic statistics and visualizing it as grayscale images for fusion modal learning is described thus.

- Machine learning methods rely heavily on feature selection, which is domain-specific [37]. In material genomics, the chemical composition characteristics of materials are significant features, and the proportions of the material's composition do not adequately accentuate them.
- Most of the dataset's material compositional features are represented by elements with atomic information that is omitted from the dataset but can be obtained from other sources. Elemental information, such as the atomic mass and radius, can be calculated from statistical information based on the compositional proportions. XenonPy is used to extend the elemental characteristics of materials and can calculate 58 elemental information points (Table9) using seven statistical formulas (Table1) to obtain a total of 406 extended dimensional characteristics.
- The 406 high-dimensional features may cause curse of dimensionality [4] in traditional machine learning. In [38], principal component analysis (PCA) [17] was used to reduce the dimensionality of medical images to obtain vectors; therefore, we also used PCA for extended for comparative experiments.
- The 7×58-dimensional features have strong or weak correlations among the 58 elements (Fig.3), and the statistics of each element are identically calculated, for the high-dimensional features that, following experts' opinions are converted into grayscale images, which are ideal for extracting information using convolutional neural networks (Fig.6).

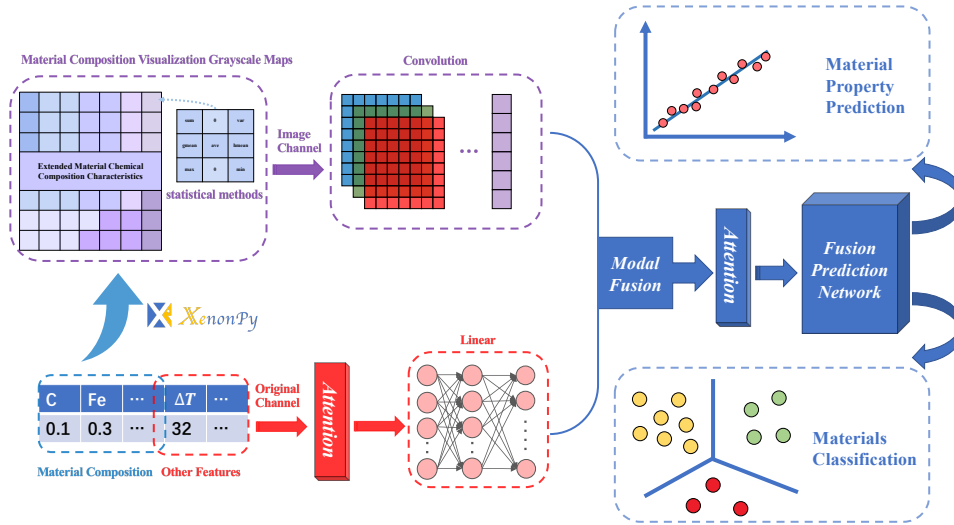


Figure 1: Better utilization of composition for materials' properties prediction with the MCVN

- A multimodal neural network must be designed to perform modal fusion for learning the image modal and the original modal.

Rest of the paper is organized as follows: In Section 2, a literature review in the field of material genomics is presented. The proposed multimodal framework along with feature extraction and different combination approaches are explained in Section 3. The experimental results are discussed in Section 4. Finally, we conclude in Section 5 along with discussions and future works.

2. Related Work

The most commonly used methods for material genomics currently focus on traditional machine learning algorithms such as support vector regression (SVR) [7], random forest (RF) [6], and XGBoost [8]. Xiong et al. used SVR and random forest algorithms to predict the mechanical properties of steel, such as fatigue strength and fracture strength [40]; Liam et al. used gradient boosting decision trees (GBDT) to predict the grain boundaries of polycrystalline materials [14]; Zhu et al. used evolutionary algorithms to search for grain boundary structures to predict grain boundary phase transitions [43]; Dai et al. used decision trees to identify nanomagnetic conductors [10]; Gusenbauer et al. used decision trees to predict simulated three-dimensional microstructures from two-dimensional images [13]; Bartel et al. predicted the stability of composites by extreme gradient boosting (XGBoost) [3]. Choudhary et al. used the GBDT to predict the properties of inorganic materials [9].

Deep learning has a wide range of applications in the field of material genomics. Schutt proposed a deep-learning-based SchNet network to model atomic systems using continuous filter convolution layers in a specially designed network structure [29]. For polishing, Wang et al. proposed a deep belief network (DBN) based on the relationship between the material removal rate and the process parameters. Using a particle swarm algorithm, they studied the effects of the network structure and learning rate on the prediction accuracy [34]. Li et al. predicted the mechanical properties of heterogeneous materials using image modeling and deep learning [20]. Sigmund et al. proposed the use of neural networks to predict the adsorption effect of carbon-containing substances on organic pollutants [30]. Jha et al. used migration learning to accelerate density functional theory (DFT) calculations to predict the formation energies of materials [16]. Dong et al. used deep learning to implement bandgap prediction in boron-nitride graphene [11]. Li et al. proposed a hybrid model based on convolutional networks and long short-term memory (LSTM) networks for predicting the critical temperatures of superconductors [19].

Multimodal techniques have contributed significantly to advances in material classification and property prediction. Zheng et al. developed a visual-tactile cross-modal retrieval framework for perceptual estimation by associating tactile

information with the visual information of materials' surfaces [42]. Erickson et al. released a model that learns a compact multimodal representation of spectral measurements and texture images for material classification [12]. Wei et al. proposed a model that can learn material representations from auditory and multi-tactile sources [35].

However, the chemical composition characteristics of the materials in the methods and models mentioned above are limited to compositional proportions, and the elemental information is lost. The chemical characteristics of the material's composition are critical in machine-learning tasks. Simultaneously, the multimodal information of the material is obtained through sensor detection and relies only on the dataset. Therefore, we undertook to design a method to improve the accuracy of material property prediction by extending the element information of material components and performing multimodal learning that includes the original modes.

3. Proposed Method

3.1. XenonPy expands and visualizes materials' compositional features

XenonPy [18] is a Python library that implements comprehensive machine learning tools for material informatics. XenonPy provides an interface for public material databases and contains a comprehensive library of material descriptors (composition, structure, and molecular). In addition, it provides pre-trained models that contain 35 properties of small molecules, polymers, and inorganic compounds and an interface for pre-trained models to implement migration learning. iQSPR-X, an inverse molecular design algorithm based on Bayesian inference, was proposed by Wu et al. for XenonPy. A custom molecular design algorithm can be built using predefined modules and a pre-trained model library in XenonPy [36]. C. Liu et al. used XenonPy's material descriptor library to extend the chemical compositional features of materials to classify quasicrystals [21]. XenonPy provides a rich set of tools for applying material informatics to various tasks where the descriptor generator can compute multiple types of numerical descriptors from the composition structures. XenonPy's built-in descriptor generator can generate 58 element-level descriptor properties (see Appendix Table9) for 94 elements (from H to Pu) and uses seven statistics to obtain seven statistical features (see Table1) for each element-level descriptor. Thus, by entering the chemical composition information of the material into the descriptor module (descriptor), a matrix of element-level descriptor features ($58 * 7$ dimensions) was obtained (compositional features).

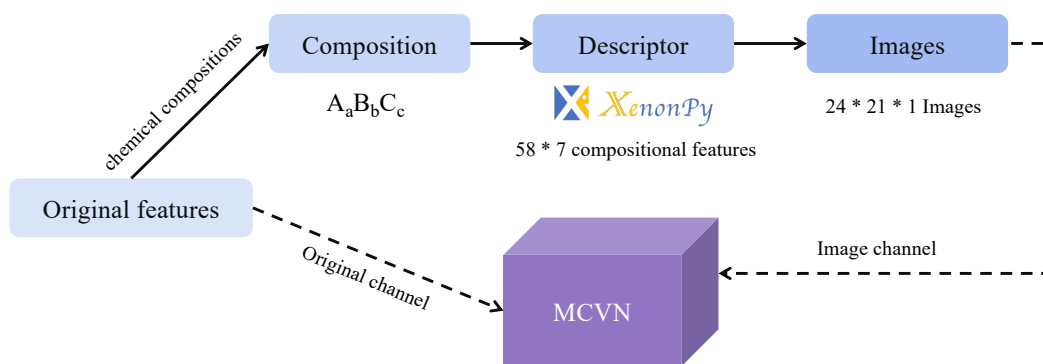


Figure 2: Process flow of multimodal fusion machine learning with MCVN

As shown in Fig.2, the chemical composition information of the material was first extracted from the original features and input into the XenonPy's descriptor module to calculate 58 descriptive attributes. The output of 58-dimensional features is roughly divided into atomic chemical features, atomic physical features, thermodynamic features, and the Herfindahl-Hirschman index [28]. The Herfindahl-Index and Hirschman-Index latter, which are economic indicators, are used to measure industrial concentrations and don't correlate with the other 56-dimensional material features; therefore, gray-scale graph mapping is not considered. After obtaining the 56-dimensional element-level features, XenonPy provides seven statistics (see Table1) for calculating the statistical features corresponding to the 56 element-level attributes of chemical components. For example, a binary composition combination $A_{w_A} B_{w_B}$, with its element-level features denoted as $f_{A,i}$ and $f_{B,i}$ ($i = 1, 2, \dots, 56$), calculates a seven-dimensional statistic for each element-level feature, where w_A^* and w_B^* denote the normalized composition content of elements A and B. After

Table 1

Xenonpy feature statistics calculation method [18]

Feature transformation method	Feature calculation formula
Weighted average	$f_{ave,i} = w_A^* f_{A,i} + w_B^* f_{B,i}$
Weighted variance	$f_{var,i} = w_A^* (f_{A,i} - f_{ave,i})^2 + w_B^* (f_{B,i} - f_{ave,i})^2$
Geometric mean	$f_{gmean,i} = \sqrt[w_A + w_B]{f_{A,i}^{w_A} * f_{B,i}^{w_B}}$
Harmonic mean	$f_{hmean,i} = \frac{w_A + w_B}{\frac{w_A}{f_{A,i}} + \frac{w_B}{f_{B,i}}}$
Max-pooling	$f_{max,i} = \max(f_{A,i}, f_{B,i})$
Min-pooling	$f_{min,i} = \min(f_{A,i}, f_{B,i})$
Weighted sum	$f_{ave,i} = w_A f_{A,i} + w_B f_{B,i}$

the calculation, 56×7 high-dimensional statistical element level descriptive features were obtained. Subsequently, the element-level defining features of each dimension were expanded into a 3×3 matrix block, as shown in Fig.3, and the remaining two values were occupied with 0. The statistical element-level descriptive features are re-arranged according to certain correlations and visualized into a composition of 24×21 single-channel grayscale images as the modal image input of the MCVN.

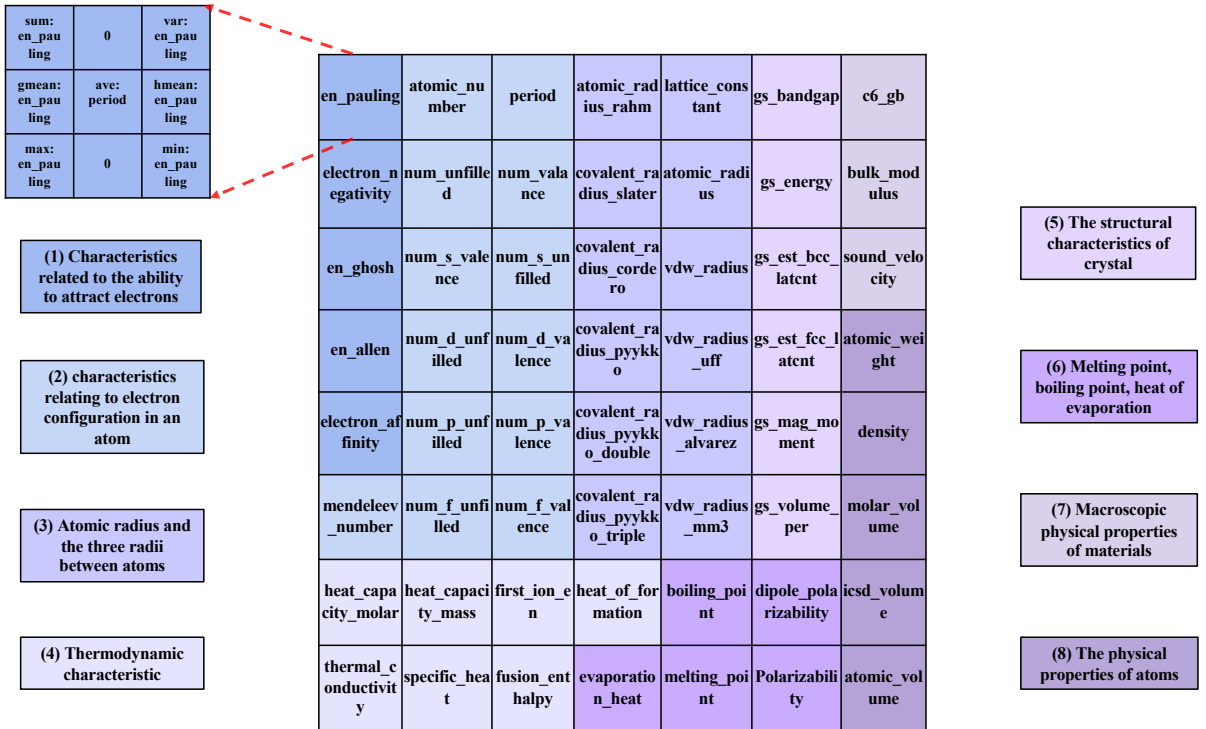


Figure 3: Arrangement of compositional features. The 56-dimensional features are divided into eight parts, namely: (1) features related to the ability to attract electrons; (2) features related to the arrangement of electrons in atoms; (3) atomic radii and the three interatomic radii; (4) thermodynamic features; (5) structural features of crystals; (6) melting point, boiling point, and enthalpy of vaporization; (7) macroscopic physical properties of materials; and (8) physical properties of atoms.

For the positional arrangement of the broad categories in Figure 2, the concept of image mapping is based on the following four points:

- The characteristics of electron arrangement in atoms, characteristics related to the ability to attract electrons, and three types of atomic radius characteristics are arranged in adjacent positions in the image. Because the ability of an atom to attract electrons is mainly determined by the atomic structure, the smaller the atomic radius, the higher the number of outermost electrons in the electron arrangement, and the stronger the ability of an atom to attract electrons.
- The crystal structure features are arranged with the atomic radius-related features because the lattice constant in the crystal structure indicates the edge length of the crystal unit cell, and the change in the lattice constant reflects the change in the composition and force state inside the crystal. The lattice constant and atomic radii can be calculated for general crystal structures.
- The main factors affecting the physical properties of atoms include at least two aspects: atomic structure and internal structure of the macroscopic material. The macroscopic physical properties of a material, such as the bulk modulus, reflect the material's resistance to external homogeneous compression in the elastic regime and are more related to the physical properties of the atoms. In contrast, the crystal structure, such as the lattice constant, reflects the relationship between atomic structure and physical properties. Therefore, the physical properties of atoms, macroscopic physical properties of materials, and characteristics of the crystal structure are arranged in adjacent positions in the diagram.
- The significant categories related to the thermodynamic properties are the physical properties of atoms, characteristics related to the electron arrangement outside the nucleus of atoms, and the melting/boiling points. According to the thermodynamics equations, because some thermodynamic properties, such as the melting/boiling points are influenced by the environment, changes in the physical properties of atoms reflect changes in the thermodynamic properties. In addition, some thermodynamic properties, such as specific heat capacity, are peculiar to the substance and consequently, are unique to the substance's atomic composition.

3.2. Material Composition Visualization Network structure

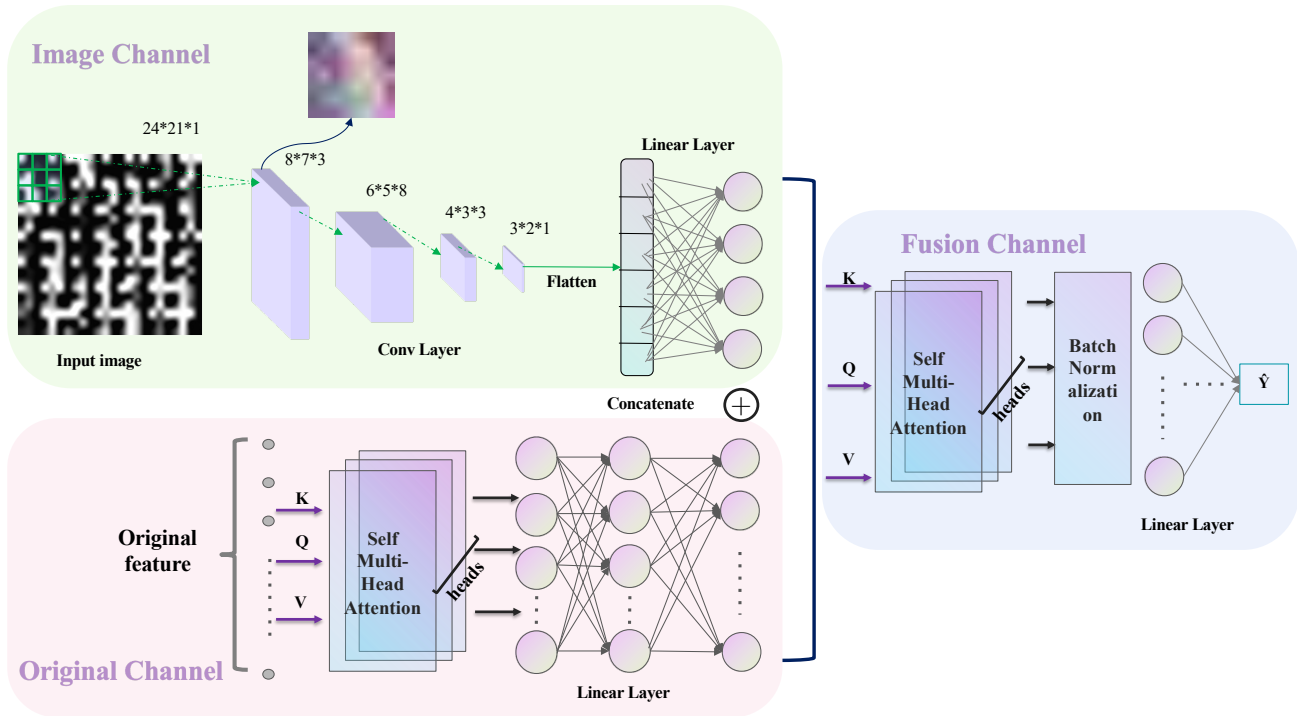


Figure 4: Material Composition Visualization Network architecture

The MCVN (Fig.4) had two modality inputs. One is the original features, including the material's composition information, how it is processed, and the experimental environment. These are directly passed through the multi-headed self-attention(MHSA) layer and the fully connected layer, the MHSA captures the internal correlation of data or features, and the fully connected layer enhances the representation of the network. The other input is the output from XenonPy of statistical element-level descriptive features reconstituted into grayscale image information. According to unified rules, the element-level descriptive features are sorted by correlation and grayscale mapping of statistical information, forming an image with texture features. Convolutional neural networks (CNNs) efficiently extract texture feature information in parallel, and the obtained image deep feature information is input to the fully connected layer for splicing with the output of the original feature modality to obtain the joint feature representation after modal fusion. The MHSA captures the connection of the joint feature representation between different samples, but the features of the image modal range between 0 ~ 255 after grayscale mapping, while the original features are distributed between 0 ~ 1 after normalization. This means that the gaps between the feature values of the same sample are too large and will affect the training of the hidden layer. Batch normalization is required to input the distribution of the hidden layer after modal fusion is drawn back to the standard normal distribution with a mean of 0 and a variance of 1 and is finally input to the fully connected layer to predict the target values.

3.2.1. Multimodal learning

The composition information of materials is an essential feature in materials informatics. However, traditional machine learning methods treat the material composition information equally with other feature information and input it directly into the machine learning model for training. This retains only the components' proportional information, whereas the components' basic information is lost. Xenonpy can calculate the composition information of materials based on its element-level descriptive features, which significantly augments the material components' semantic information. However, the high-dimensional element-level defining features obtained by Xenonpy based on seven statistical formulas introduce the curse of dimensionality [4]; by which, as the dimensionality increases, the space volume increases rapidly, and sparsifies the available data. This decrease is known as the Hughes phenomenon [25]. If the sample size is not sufficiently large and the dimensionality is very high, the machine learning model becomes prone to overfitting, and the ability of the model to generalize is significantly reduced.

Convolutional neural networks can manage high-dimensional images information because of their local perception and parameter sharing. In this paper, based on material composition information the statistical element-level descriptive features output by XenonPy are reconstructed into grayscale images, element-level statistical information textures are constructed manually according to relevance ranking, and the CNN extracts these texture features.

Fig.5 illustrates that the artificially constructed texture grayscale map can distinguish the different components of the material to some extent and becomes an important feature of the material. In this study, the original modal features and manually constructed image modalities were jointly represented, and both were mapped to the vector feature space to train the model (Eq.(1)).

$$\begin{aligned} y_{image\ channel} &= Linear(Conv(x_{image})) \\ y_{original\ channel} &= Linear(Attention(x_{original})) \\ y_{Multimodal} &= Linear(BN(Attention(y_{image\ channel} \oplus y_{original\ channel}))) \end{aligned} \quad (1)$$

where x_{image} denotes the input of the image channel, $y_{imagechannel}$ denotes the output of the image channel, $x_{original}$ denotes the input of the original channel, $y_{originalchannel}$ denotes the output of the original channel, $Conv$, $Linear$, $Attention$, BN and \oplus denote the convolutional layer, fully connected layer, multi-headed self-attentive layer, batch normalization layer, and vector stitching operation, respectively. Besides, $y_{Multimodal}$ denotes the result of modal fusion for target prediction.

In this study, modal fusion was performed to improve the network's prediction accuracy by visualizing the image modal and the original material characteristic modal from the material's chemical composition information. Modal fusion refers to the fusion of data from multiple modalities that contributes to target value prediction, modal fusion is intended to predict better the target value, and multi-modal data can engender information complementarity [2].

3.2.2. The CNN extracts element-level descriptors

The input of the image modality is the array of statistical element-level descriptors, and each element-level descriptor adopts the same statistical method and order (Fig.6), while the element-level descriptors with high relevance

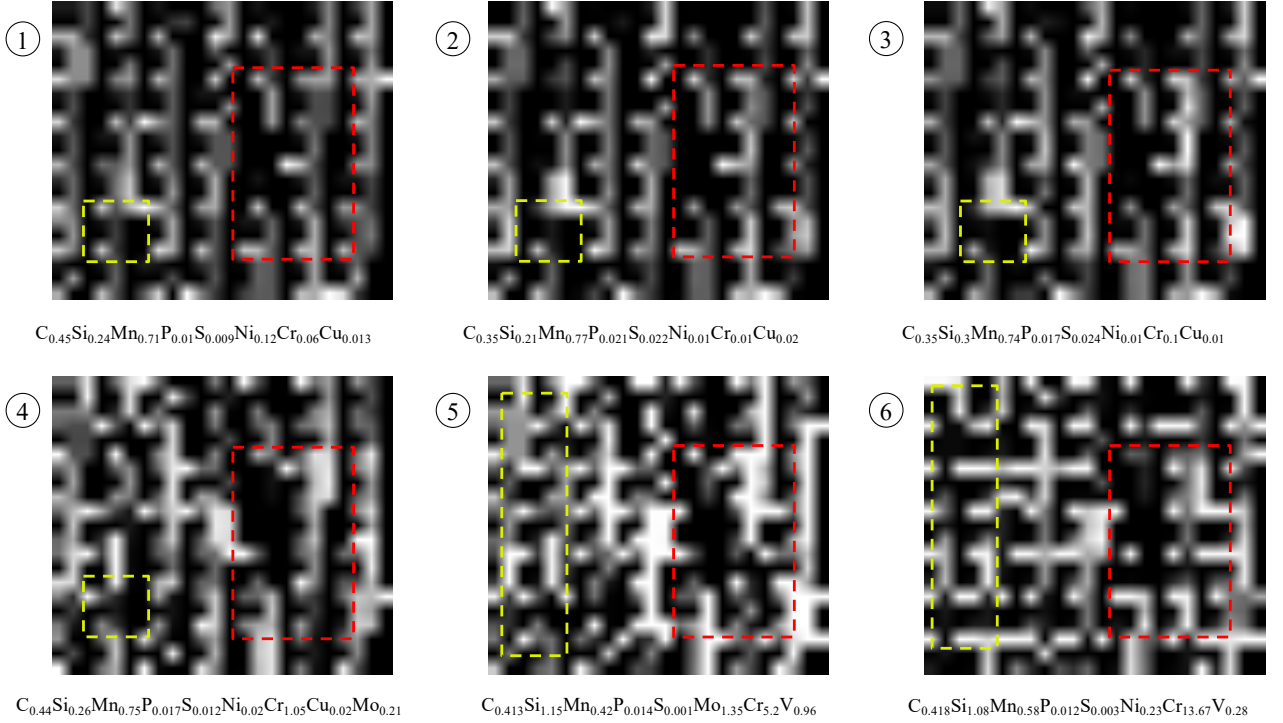


Figure 5: Element-level description of features to reconstruct texture grayscale maps. The artificial texture grayscale maps of the six differently composed steel sheets are shown in Fig.5. The steel sheets ①~④ are from the NIMS's steel dataset [1]. ①, ②, and ③ have the same composition containing C, Si, Mn, P, S, Ni, Cr, and Cu. The differences are the compositional proportions, as such, the texture of the grayscale maps obtained from the visualization of the element-level description features is highly similar. There is no significant difference in the texture of the red and yellow boxes. There are only variations in texture brightness, with the texture's brightness around the red box increasing stepwise and the texture around the yellow box decreasing stepwise in the grayscale maps of ①, ②, and ③. However, steel ④, which also belongs to NIMS's steel dataset, has the new element Mo added, thus the textures in the red and yellow boxes differ significantly from those in the grayscale diagrams of ①~③. Steels ⑤ and ⑥ are from the steel dataset of the SRIM. The composition of steel ⑥ has the new element V added and has Cu omitted compared with that of ①~③, therefore, the textures in the red boxes in the grayscale diagrams are significantly changed, and the original fractured textures are significantly connected. The texture around the red box in the gray diagram has changed significantly, and the original fractured texture has an obvious connection. Steel ⑤ has the element Ni omitted compared to ⑥, into which the new element Mo has been added, and the fractured texture at the yellow box has an obvious connection. Steel ④ also contains Mo, therefore, the same connected texture appears at the yellow box position in the gray diagram of ⑤.

are also arranged adjacently. The essential features of a CNN are local perception and parameter sharing, which are ideal for extracting artificially constructed statistical information on a material's components. Equation (2) represents the process of extracting statistical information from element-level descriptors in the first convolutional layer.

$$y_{Conv-1} = w_{sum} \times x_{sum} + w_{var} \times x_{var} + w_{gmean} \times x_{gmean} + w_{ave} \times x_{ave} + w_{hmean} \times x_{hmean} + w_{max} \times x_{max} + w_{min} \times x_{min} + \mathbf{b} \quad (2)$$

where y_{Conv-1} denotes the output of the convolution kernel of the first convolution layer, x denotes any one of the 56 element-level descriptors, w denotes the convolutional trainable weights for each statistical descriptor, \mathbf{b} denotes the bias of the first convolutional layer.

First, the feature extraction layer of the CNN learns implicitly from the training data, avoiding explicit feature extraction to learn the convolutional kernel parameters dynamically. Second, because the neurons on the same feature mapping surface have exact weights, the network can learn in parallel, which is a significant advantage convolutional

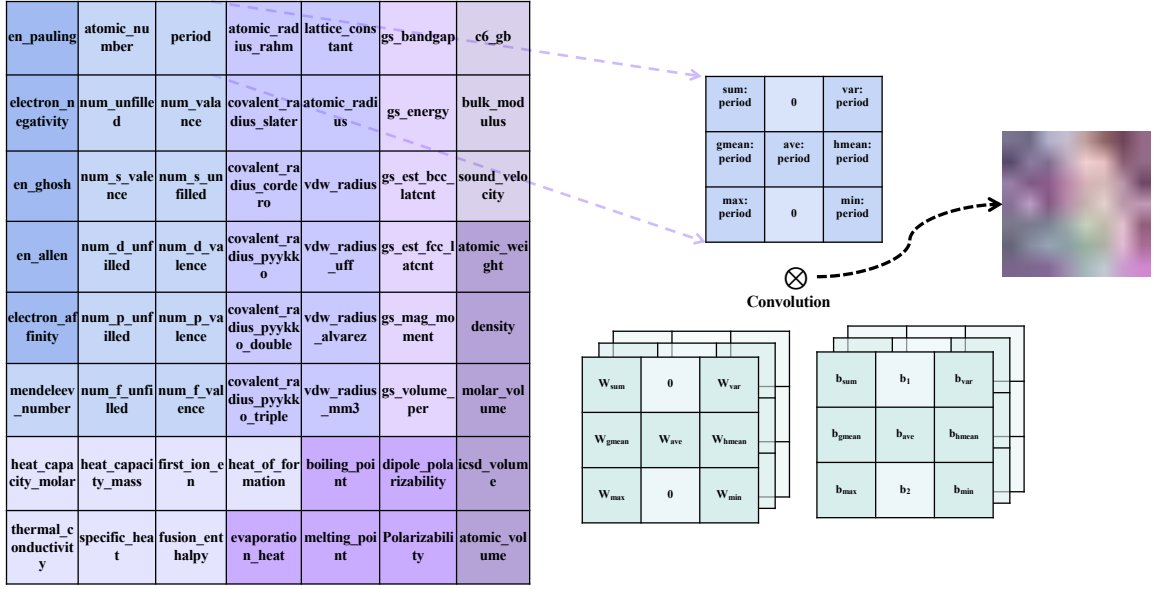


Figure 6: Convolutional layer to extract element-level statistical feature information

networks have over fully connected networks. Because the artificially constructed image has fewer pixels compared to the actual image, no sampling layer (pooling layer) is added after the convolutional layer to reduce the dimensionality of the image, and there is no padding boundary during convolution; the dimensionality of the image also decreases because the convolutional layers are stacked. The multiple convolution kernels allow the feature channels' dimension to increase and enrich the information extracted from the image.

3.2.3. Batch Normalization

When the MCVN performs modal fusion, after normalization the feature input range of the image channels is mapped between 0 and 255, while the original feature input channels are normalized(0 ~ 1). The output range of the two channels is too large, and the scale of the features is not uniform, which, after modal fusion, affects the hidden layers' training. In 2015, Ioffe et al. proposed batch normalization (BN), which is a data normalization method for deep neural networks [15]. Typically, before the activation layer in a deep neural network, the BN layer calculates the data from a batch input of size m (Eq. (3)). The purpose of BN is to pull the value toward the linear region of the subsequent nonlinear transformation that is to be performed, increase the derivative values, enhance information backpropagation, and accelerate training convergence. However, this leads to a decrease in network expressiveness. To prevent this, two reconstruction parameters, scale γ , and shift β , which are learned from training are added to each neuron and used to inverse transform the transformed activation and enhance the expressiveness of the network [15].

$$\begin{aligned}
 \mu_B &\leftarrow \frac{1}{m} \sum_{i=1}^m x_i \\
 \sigma_B^2 &\leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \\
 \hat{x}_i &\leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\
 y_i &\leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i)
 \end{aligned} \tag{3}$$

Mini-batch mean μ_B and mini-batch variance σ_B^2 are calculated and based on the μ_B and σ_B^2 normalizes the input x , where ϵ is a minimal positive number used to avoid the division by 0 anomalies caused by encountering an input with a variance of 0. After normalization, the activation x of the network layer forms a customarily distributed data \hat{x} with mean 0 and variance 1.

Setting a batch normalization layer in a deep neural network can accelerate the convergence speed of model training, stabilize the model training process, attenuate gradient explosions or gradient vanishing, render the model less sensitive to the initialization of weights, help regularization, and reduce overfitting [15]. In this study, batch normalization is performed after modal fusion, which can standardize the features of the image modality and the original data modality after connecting them, guaranteeing a balanced input distribution and uniform feature scale of the fully connected layer after modal fusion, and accelerating the MCVN's convergence speed.

3.2.4. Multiple self-attention mechanisms

Owing to the prohibitive cost of material experiments, material datasets are usually characterized as follows: the amount of data is small, and the experimental data are generated mainly by orthogonal experimental designs. Therefore, specific correlations are found between different samples in the material dataset. This correlation positively affects the accuracy of predicting the target values.

The multi-headed self-attention(MHSA) mechanism is an algorithm that calculates the similarity between samples at high speed and in parallel. The MHSA function was used for each sample to calculate the similarity between it and the others. Similarity is used as a weight for multiplying with the corresponding sample value. Finally, critical information was obtained about the sample being scrutinized. The attention function can be described as mapping a query and a set of key-value pairs to an output, where the *Query*, *Key*, *Value*, and output are all vectors. The output is computed as a weighted sum of values, where the weight assigned to each *Value* is computed from the similarity function of the *Query* to the corresponding *Key* [33]. The most common attention functions are additive attention and scaled dot product attention (Eq. (4)) [33], respectively. The attention function used in this study was dot product attention. First, Q and K of the same dimension d_k are obtained using the attention function for the attention score. The attention score is compressed to between 0 and 1 by the softmax activation function, and finally, the output value of the object in this calculation is obtained by multiplying it with the original V .

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where $Value(V)$ represents the object's value to be noticed, $Key(K)$ represents the critical information of each object, $Query(Q)$ represents the other information of each attention object, and $\frac{1}{\sqrt{d_k}}$ represents the scaling factor.

In practical applications, scaled dot product attention is much faster and more spatially efficient than other attention functions. Moreover, it can be implemented in highly optimized matrix multiplication code [33], which is well suited for GPU parallel operations.

The single-headed attention model overly focuses on one-sided information when encoding information at a current location. Figure 7 shows the multi-headed dot product attention model. Because the multi-headed attention mechanism is parallel to single-headed attention, the multi-headed attention model can operate in parallel. The output of the parallel single-headed attention function is conjoined and fed into a fully connected layer to obtain the output of the multi-headed attention(as shown in Eq. (5)). The benefit of the multi-headed attention mechanism is that it allows the model to focus jointly on information from different representation subspaces at different locations [33], enhancing the representational power of the model.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (5)$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$. In this work we employ $h = 2$ parallel attention heads.

In material data without time-series information, the similarity and correlation between samples are more critical than the data samples' location, that is, the model must attend more to the changes in each sample relative to other

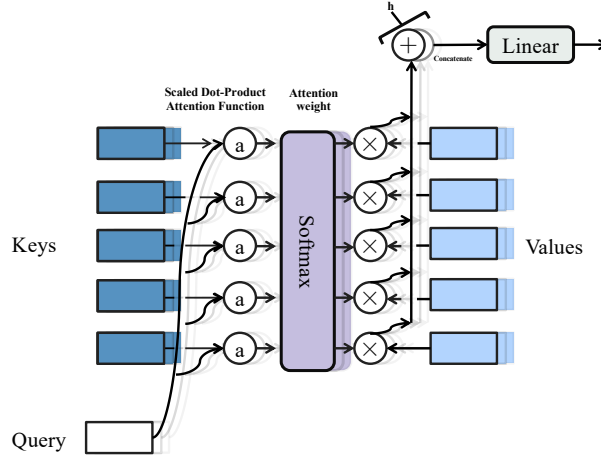


Figure 7: Multi-Head Scaled Dot-Product Attention

samples. The self-attention mechanism is a variation of the attention mechanism that reduces reliance on external information and better captures the internal relevance of data or features. Self-attention is the initialization of Q , K , and V to sample X . All three have precisely the same value and dimensionality, which is ideal for the dot product attention function. Compared to recurrent neural networks (RNNs) and CNNs, the correlation between input samples can be captured in parallel. Although the relative location information between samples is lost, more attention is paid to the similarity and correlation between the samples. Therefore, self-attentiveness is well suited for capturing the correlation between different samples in material data without temporal information.

In summary, this study adopts a multi-headed self-attentiveness mechanism to capture the information in the original feature space of the material and the information in the multimodal feature space during modal fusion.

4. Experiments and Analysis

4.1. Implementation Details

In this study, the superiority of the MCVN for material property prediction was verified on regression and classification tasks.

The regression task on the NIMS's dataset [1] and SRIM's Steel dataset contained steel composition information and processing information. The composition information was mapped into grayscale maps and the original modal information for modal fusion was used to predict the mechanical properties of the steel and verify the accuracy of the regression prediction. The evaluation metric of the algorithm was the coefficient of determination R^2 (Eq. (6)). The loss function was the mean-squared loss (Eq. (7)).

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (6)$$

where \bar{y} is the average of the target values, \hat{y} are the predicted values of the target, y are the actual values of the target, and R^2 limits the evaluation of the prediction results to less than 1. The closer R^2 is to 1 means the higher the accuracy of the model prediction.

$$L(y, \hat{y}) = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (7)$$

where y_i and \hat{y}_i are the predicted value by the model and the groundtruth for each point i in N train set.

For the classification task on the amorphous alloy materials dataset [39], only the composition information of the amorphous alloy was available. Only the composition information was input into the MCVN for modal fusion after grayscale mapping to classify different types of amorphous alloys. We verified the *Precision*, *Recall*, and *F1 – Score*(Eq. (8)) for the classification task: Cross-entropy was used as the loss function (Eq. (9)).

$$\begin{aligned} \text{Precise} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ \text{F1 – Score} &= \frac{2}{\frac{1}{\text{Precise}} + \frac{1}{\text{Recall}}} = 2 \cdot \frac{\text{Precise} \cdot \text{Recall}}{\text{Precise} + \text{Recall}} \end{aligned} \quad (8)$$

where *TP* indicates correctly predicting positive samples as positive; *FN* indicates incorrectly predicting positive samples as negative; *FP* indicates incorrectly predicting negative samples as positive, and *TN* indicates correctly predicting negative samples as negative. Comparing *Accuracy*, *Precision*, *Recall*, and *F1 – Score* on an unbalanced classification dataset can evaluate the accuracy of the classification model more accurately.

$$L(y, \hat{y}) = - \sum_k^N \hat{y}_k \log(y_k) \quad (9)$$

where y_k and \hat{y}_k are the predicted score by the model and the groundtruth for each class k in N categories.

The hardware platform was an Intel Xeon Gold 5118 CPU with 12 computing cores at 2.3 GHz, the NVIDIA® V100 Tensor Core GPU with 16 GB video memory and 900 GB/s bandwidth; the code implementation used Python version 3.8, and the neural network was built using PyTorch-1.11.

Network architecture details. The MCVN consists of the following layers for each module:

- Image Channel: $C(3, 3, 3) - C(8, 3, 1) - C(4, 3, 1) - C(1, 2, 1) - L(6, 2)$
- Original Channel: $A(N, 2) - L(N, 8)$
- Fusion Channel: $A(12, 2) - BN - L(12, 8) - L(8, 4) - L(4, 1)$

where $C(c, k, s)$ denotes a standard convolutional layer with $k * k$ filters, c channels, and s stride, $L(i, o)$ denotes a fully-connected (linear) layer mapping features from i -dimensions to o -dimensions, $A(i, h)$ means multi-head attention layers with i -dimension input and h heads, N denotes the numbers of input features. For the regression task, we used the ReLU and LogSoftmax activation functions for the classification task.

4.2. Predicting the mechanical properties of steel materials

Steel is the most widely used structural material in engineering. Developing new steel materials that meet such requirements and face complex environments is of great significance. With no early macroscopic deformation and severe structural integrity loss after damage, abrupt fatigue and fracturing are the main types of structural failure in these materials. This, therefore, has been the focus of attention in the engineering field. Fatigue strength is the critical value at, or below which fatigue failure will not occur within the applied stress range for a given fatigue cycle (fatigue life). Steel's toughness is crucial for evaluating engineering materials, and current structural materials tend to have high strength and toughness. Yield strength, tensile strength, elongation, and cross-sectional shrinkage are indicators that describe the plastic properties of steel materials. Yield strength refers to the strength of the material when yielding occurs, which is usually specified to produce a 0.2% residual deformation of the stress value for its yield limit. If the external force exceeds this limit, it cannot recover. If the yield strength of steel is insufficient, there will be deformation expansion, deepening cracks, and possibly severe accidents.

Tensile strength measures a material's resistance under tension until the stress value is reached when a fracture occurs. The maximum load-bearing capacity of the steel under static tensile conditions affects its fracture resistance. The greater the ratio of yield strength to tensile strength the higher the reliability of the structural part. Plasticity is the ability of a material to undergo permanent deformation without damage, in reaction to an external force. Elongation δ and cross-sectional shrinkage ψ are engineering metrics that characterize the plasticity of materials. The plasticity

Table 2

Statistical information on the characteristics of the NIMS's steel dataset

Feature	Description	Min	Max	Mean	Std
NT	Normalizing temperature	825	990	865.6	17.37
QT	Quenching temperature	825	865	848.2	9.86
TT	Tempering temperature	550	680	605	42.4
C	Carbon content	0.28	0.57	0.407	0.061
Si	Silicon content	0.16	0.35	0.258	0.034
Mn	Manganese content	0.37	1.3	0.849	0.294
P	Phosphorus content	0.007	0.031	0.016	0.005
S	Sulfur content	0.003	0.03	0.014	0.006
Ni	Nickel content	0.01	2.78	0.548	0.899
Cr	Chromium content	0.01	1.12	0.556	0.419
Cu	Copper content	0.01	0.22	0.064	0.045
Mo	Molybdenum content	0	0.24	0.066	0.089
RR	Depression rate	420	5530	971.2	601.4
dA	Plastic inclusions	0	0.13	0.047	0.032
dB	Intermittent inclusions	0	0.05	0.003	0.009
dC	Isolated inclusions	0	0.04	0.008	0.01

index in engineering can satisfy the material's bending and ductility performance requirements. In addition to strength, high plasticity that enhances resistance to brittleness and cracking is also vital.

Therefore, in steels, fatigue resistance, fracture strength, hardness, yield strength, tensile strength, elongation, and cross-sectional shrinkage are representative of its mechanical properties, and high-precision prediction is important for the material's safe use.

4.2.1. The National Institute for Materials Science's(NIMS's) Steel Dataset

This dataset [1] is from the National Institute for Materials Science (NIMS), Japan, and contains the chemical composition, processing conditions, and mechanical properties of the material, such as the fatigue strength, tensile strength, fracture strength, and hardness. In this dataset, the effects of the rotational bending fatigue strength (hereafter referred to as fatigue strength) of the material, fatigue test conditions (e.g., loading frequency and profile), test temperature and environment, and specimen size on the fatigue behavior were measured over 107 cycles of fatigue life. The material contained nine alloying elements: carbon (C), silicon (Si), manganese (Mn), phosphorus (P), sulfur (S), nickel (Ni), chromium (Cr), copper (Cu), and (Mo). The remaining parameters are the area fractions of non-metallic inclusions, such as inclusions formed by plastic processing, discontinuously arranged inclusions, and isolated inclusions. In terms of heat treatment, three types of heat treatments are included: normalizing, quenching, and tempering. After the heat treatment, the specimens were cooled to room temperature, and fatigue tests were performed. In this dataset, there are 360 data samples with 16-dimensional features, including nine alloying elements, one depression rate, three heat-treatment temperatures, three inclusions, and four target properties (fatigue strength, tensile strength, fracture strength, and hardness). The specific feature descriptions and information on the distribution of the eigenvalues are shown in Table2.

The compositional features were obtained as statistical element-level descriptors using the XenonPy computational library, converted into images, and then fed into the MCVN along with the original features.

The dataset was divided into training and testing sets following a 2 : 1 ratio. The MCVN's experimental results were compared with the currently popular machine learning regression models, and with the support vector machine regression (SVR) [7], linear regression ridge regression with L2 regularization (Ridge) [5], stochastic gradient descent (SGD) [27], decision tree a (DT) [27], integrated learning of the tree models random forest (RF) [6], and XGBoost [8] algorithms. The experimental results are specified in Table3. The machine learning models such as XGBoost, SVR, and RF, with the original 16-dimensional features as inputs, obtained R^2 mean values above 0.9. These models are already excellent, and XGBoost, a robust integrated learning algorithm developed in recent years, achieved an R^2 of 0.92 for all four target values. However, the MCVN improved each target value by 3% ~ 4% compared to XGBoost and achieved high R^2 experimental results of 0.98 for the tensile and hardness targets. Better to represent the prediction

Table 3

Experimental results of steel material property prediction model for the NIMS's steel dataset. *O* indicates that the input of the algorithm is the *Original* features, *E* indicates that the input of the algorithm is the vector of high-dimensional features obtained by *Expanding* with XenonPy and splicing with the original features, and *E&P* indicates that the input of the algorithm is a vector obtained by stitching the vector after dimensionality reduction using PCA with the original features.

Target	MCVN	SVR			RF			Ridge			SGD			DT			XGBoost		
		<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>
Fatigue	0.954	0.89	0.780	0.882	0.903	0.886	0.904	0.909	0.913	0.908	0.897	0.846	0.899	0.795	0.803	0.798	0.920	0.931	0.921
Tensile	0.983	0.925	0.785	0.916	0.919	0.919	0.90	0.939	0.945	0.938	0.921	0.771	0.925	0.819	0.884	0.832	0.942	0.951	0.928
Fracture	0.954	0.905	0.859	0.907	0.913	0.941	0.929	0.894	0.915	0.905	0.887	0.63	0.903	0.856	0.894	0.895	0.925	0.930	0.936
Hardness	0.980	0.907	0.783	0.899	0.927	0.934	0.904	0.927	0.940	0.929	0.910	0.921	0.915	0.881	0.873	0.826	0.935	0.948	0.931

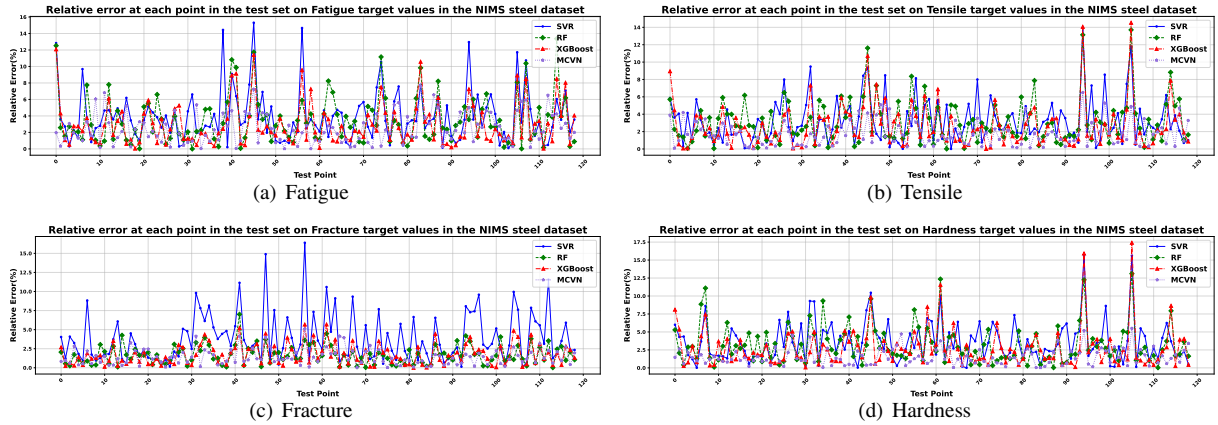


Figure 8: Relative error of each point on the 4 target values in the test set in the NIMS's steel dataset. At the four target values, the relative error of the MCNV is smoother and closer to 0 than the other three models, and the MCNV is more accurate in predicting the mechanical properties of steel at the test points where the other three models did not perform well. For example, at the 41st and 74th test points of the Fatigue target, the relative errors of the remaining three models exceeded 6%, while the relative errors of the MCNV were less than 1%; at the 95th and 105th test points of the Tensile target, the relative errors of the remaining three models were approximately 12%, while the relative errors of the MCNV were 5% ~ 6%; at the 32nd and 95th test points of the Fracture target value, the relative error of the remaining three models exceeded 2% while the relative error of the MCNV approximated 0%; at the 94th and 105th test points of the Hardness target value, the relative errors of the remaining three models exceeded 12.5%, while the relative errors of the MCNV approached 5%.

accuracy of the MCNV on the four target values of the NIMS's steel dataset, Fig.8 presents the relative errors of the MCNV and the other three machine learning models, SVR, RF, and XGBoost, for each test point of the NIMS's steel dataset. The MCNV exhibited insignificant relative errors than the other three models at all four test points where most of the target values were found.

The experimental results validate that multiple modal data can fuse and complement each other in the MCNV to obtain improved regression models.

4.2.2. Shanghai Research Institute of Materials'(SRIM's) Steel Dataset

This dataset contains the materials' chemical compositions, processing conditions, and properties. The predicted mechanical properties in the dataset included the yield strength, tensile strength, elongation, and cross-sectional shrinkage. The materials contain nine alloying elements: carbon (C), silicon (Si), manganese (Mn), phosphorus (P), sulfur (S), nickel (Ni), chromium (Cr), copper (Cu), molybdenum (Mo), vanadium (V), and aluminum (Al). The remaining process parameters are the melting method, thermal processing method, heat treatment process, and heat treatment status, described in natural language; therefore, one-hot encoding is used. The dataset contains 81 data samples with 16-dimensional features, including 11 alloying elements, five process parameters, and four target properties (yield strength, tensile strength, elongation, and cross-sectional shrinkage). The compositional features and distributions are described in Table4.

Table 4

Statistical information on the characteristics of the SRIM's steel dataset

Feature	Description	Min	Max	Mean	Std
C	Carbon content	0.138	0.52	0.307	0.13
Si	Silicon content	0.13	1.15	0.679	0.412
Mn	Manganese content	0.13	1.74	1.016	0.538
P	Phosphorus content	0	0.024	0.01	0.003
S	Sulfur content	0	0.009	0.0025	0.0013
Ni	Nickel content	0	3.32	1.486	1.464
Cr	Chromium content	0	13.92	6.81	6.48
Cu	Copper content	0	0.678	0.268	0.318
Mo	Molybdenum content	0	1.39	0.243	0.368
Al	Aluminum content	0	0.8	0.3	0.36
V	Vanadium content	0	1	0.21	0.26

Table 5Experimental results of the steel property prediction model on the SRIM's steel dataset. *O*, *E* and *E&P* are described in Table3.

Target	MCVN	SVR			RF			Ridge			SGD			DT			XGBoost		
		<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>	<i>O</i>	<i>E</i>	<i>E&P</i>
Yield	0.974	0.891	0.899	0.880	0.926	0.947	0.944	0.923	0.949	0.928	0.874	< -10	0.898	0.851	0.868	0.910	0.908	0.926	0.922
Tensile	0.98	0.918	0.905	0.878	0.94	0.958	0.933	0.944	0.959	0.942	0.892	< -10	0.909	0.958	0.955	0.954	0.915	0.948	0.95
Elongation	0.943	0.875	0.858	0.859	0.859	0.860	0.830	0.882	0.77	0.858	0.862	< -10	0.823	0.776	0.779	0.773	0.823	0.814	0.793
Cross-sectional shrinkage	0.835	0.544	0.617	0.582	0.668	0.556	0.658	0.624	0.596	0.654	0.557	< -10	0.563	0.598	0.422	0.579	0.656	0.411	0.605

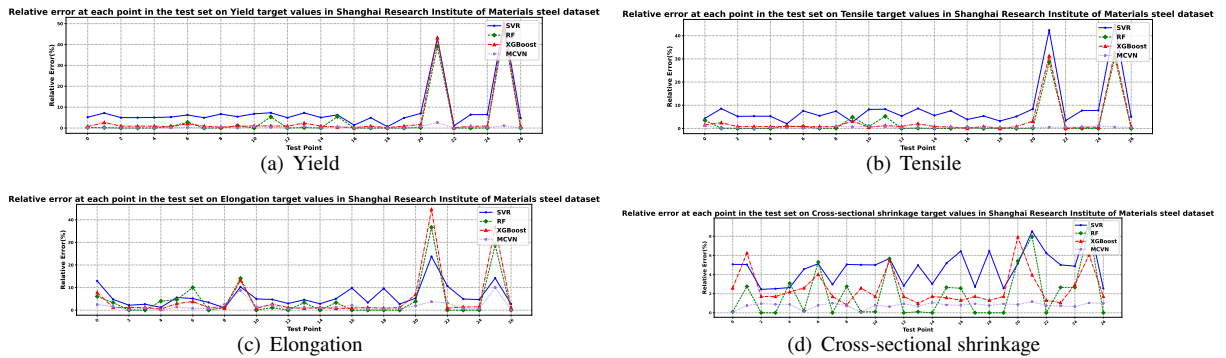


Figure 9: Relative error of each point on the 4 target values in the test set in the SRIM's steel dataset. Above the four objectives, the average relative errors of the MCVN are all less than 5%. In contrast, the relative error curves of the remaining three models are fluctuating more and are less effective. At the 21st and 25th test points of the Yield and Tensile target values, the relative errors of the remaining three models exceed 30%, while the relative errors of the MCVN are less than 5%; at the 21st test point of the Elongation target value, the relative errors of the remaining three models are in the range of 20% ~ 40%, while the relative error of the MCVN was below 5%. At the cross-sectional shrinkage target value, the relative error curve of the MCVN was significantly smoother than that of the other three models.

The evaluation index of this experiment was identical to the NIMS's steel dataset. However, the dataset has fewer samples than the NIMS's steel dataset and the unbalanced distribution due to a single random division would have affected the accuracy of evaluating the model. Therefore, 10-fold cross-validation was repeatedly used to divide each randomly generated fold into training and testing sets according to a 2:1 ratio to calculate the average R^2 value and enhance the evaluation accuracy.

This dataset contains only 81 data points. Because deep neural networks have many parameters and powerful fitting abilities, they are more prone to overfitting on small datasets than traditional machine learning. However, after ten-fold cross-validation, the MCVN and the traditional machine learning models improved the yield strength and the tensile strength target values (the experimental results are shown in Table5) while the elongation improved from 0.882 to

Table 6

Experimental results of amorphous alloy material classification for the BMG-RMG-CRA dataset. *O*, *E* and *E&P* are described in Table3.

Algorithm	Category		Precision	Recall	F1-score	Support
SVM	BMG	<i>O</i>	0.88	0.56	0.68	208
		<i>E</i>	0.90	0.74	0.80	
		<i>E&P</i>	0.81	0.67	0.73	
	RMG	<i>O</i>	0.75	0.92	0.83	1241
		<i>E</i>	0.78	0.95	0.86	
		<i>E&P</i>	0.78	0.89	0.74	
	CRA	<i>O</i>	0.71	0.45	0.55	510
		<i>E</i>	0.81	0.44	0.57	
		<i>E&P</i>	0.73	0.53	0.61	
RF	BMG	<i>O</i>	0.83	0.94	0.87	208
		<i>E</i>	0.72	0.93	0.82	
		<i>E&P</i>	0.77	0.92	0.84	
	RMG	<i>O</i>	0.82	0.94	0.87	1241
		<i>E</i>	0.83	0.92	0.87	
		<i>E&P</i>	0.81	0.92	0.86	
	CRA	<i>O</i>	0.86	0.52	0.65	510
		<i>E</i>	0.88	0.54	0.67	
		<i>E&P</i>	0.83	0.50	0.62	
XGBoost	BMG	<i>O</i>	0.92	0.85	0.89	208
		<i>E</i>	0.92	0.87	0.90	
		<i>E&P</i>	0.92	0.80	0.86	
	RMG	<i>O</i>	0.83	0.95	0.89	1241
		<i>E</i>	0.86	0.94	0.90	
		<i>E&P</i>	0.81	0.95	0.87	
	CRA	<i>O</i>	0.83	0.57	0.68	510
		<i>E</i>	0.84	0.67	0.75	
		<i>E&P</i>	0.83	0.54	0.66	
MCVN	BMG		0.931	0.904	0.917	208
	RMG		0.901	0.941	0.921	1241
	CRA		0.861	0.778	0.818	510

0.943 for the R^2 value. The MCVN achieved an R^2 of 0.835 for the target value of cross-section shrinkage, where all traditional machine learning models performed poorly and only achieved an average score of 0.6. The relative errors of the MCVN and the other three traditional machine learning models, SVR, RF, and XGBoost, are presented in Fig.9 for each test point of the SRIM's steel dataset.

4.3. Classification of Amorphous Alloy Materials

The amorphous alloys now receiving increased attention for their excellent physical and chemical properties are promising materials for engineering practice. Amorphous alloy development is different from that of traditional materials, and the rapid determination of whether a given material is an amorphous alloy is critical in practical applications. In this paper, the MCVN is proposed to predict the glass-forming ability of amorphous materials. The dataset is taken from the literature [39], where BMG denotes bulk metallic glasses, RMG denotes ribbon metallic glass, and CRA denotes a crystalline alloy. The original characteristics in the dataset are only the chemical composition characteristics of amorphous alloys, which contain 56 elements, such as Cu, Zr, Ni, etc. Because amorphous alloys are generally composed of only 4-5 elements, the composition data of each is very sparse. The unbalanced classification dataset contains 6471 records, among which BMG has 1211 records, CRA has 1552 records, and RMG has 3708 records. The sample size of RMG is much larger than that of CRA and BMG.

The MCVN's experimental results for the amorphous alloy samples were compared with the currently popular machine learning classification models using the same random seeds and dividing the training and testing sets according to the 2 : 1 ratio. The comparison algorithms were the support vector machine (SVM) [7], random forest (RF) [6], and

Table 7

Modal ablation experimental results of the steel materials' property prediction model for NIMS

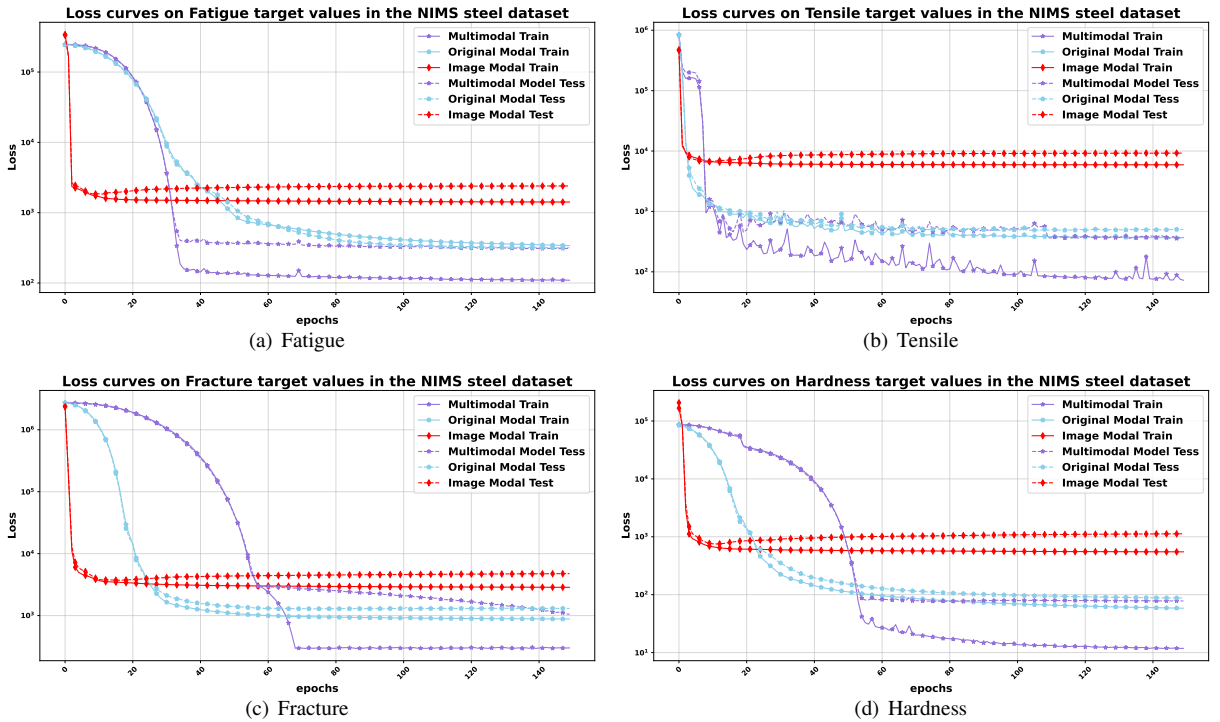
Modal	Fatigue	Tensile	Fracture	Hardness
Image Channel	0.604	0.602	0.797	0.583
Original Channel	0.927	0.932	0.925	0.945
Multimodal Channels	0.954	0.983	0.954	0.98

XGBoost [8]. The original sample distribution was retained in the test set, and the RMG dataset remained larger than those for CRA and BMG. The experimental results are given in Table6. XGBoost performs relatively well regarding *Accuracy*, but only achieves an *Recall* of 0.57 and a 0.68 *F1 – Score* on the CRA samples, and does not distinguish well between them and the other two types of amorphous alloys. The MCVN produces much superior performance overall, achieving the highest *F1 – Score* on the BMG and RMG sample size and achieving an *F1 – Score* of 0.82 and *Recall* of 0.78 in the classification of CRA. It well distinguishes between the three types of amorphous alloy materials, thus verifying that in the classification task, the multiple modal data is fused and complementary allowing the MCVN better to classify each category of materials.

4.4. Ablation Experiments

To demonstrate the efficacy of the MCVN's network structure and to find the best possible architectural configuration, this study conducted modal ablation and modular ablation experiments on the NIMS's steel dataset, which was divided into training and test sets using the 2 : 1 ratio, with an evaluation metric of R^2 .

4.4.1. Modal Ablation Experiment

**Figure 10:** Modal ablation experimental loss curves for the NIMS's steel dataset

The results of the modal ablation experiment for the R^2 values are shown in Table7, where the image channel indicates the image input using only the visualization of the materials' compositional features, the original channel

Table 8

Module ablation experimental results of steel material property prediction model for the NIMS's dataset. Attention is the multi-headed self-attention module in the MCVN structure, BN is the batch normalization layer in the MCVN structure during modal fusion, CNN denotes the convolutional layers in the Image Channel in the MCVN structure, ✓ means that the module is present in the structure, and × means it is not. (Missing CNN module indicates that PCA is used instead of CNN for reducing the dimensionality of the extended features).

Attention	BN	CNN	Fatigue	Tensile	Fracture	Hardness
×	✓	✓	0.93	0.92	0.931	0.938
✓	×	✓	0.946	0.956	0.948	0.961
✓	✓	×	0.905	0.923	0.923	0.932
✓	✓	✓	0.954	0.983	0.954	0.98

indicates the original input of the dataset including the material compositions and the preparation process, while the multimodal channel fuses material composition visualization images and original features.

Figure 10 presents the loss curves of the three network structures for learning optimization on the four target values in the NIMS's steel dataset. The image channel model with image input converges fastest, but the accuracy rate decreases significantly owing to feature loss during the preparation process. Owing to its complex network structure, the Multimodal channel model converges less rapidly than the original channel model. However, the MCVN outperformed the other two models in training the four target values.

4.4.2. Module Ablation Experiment

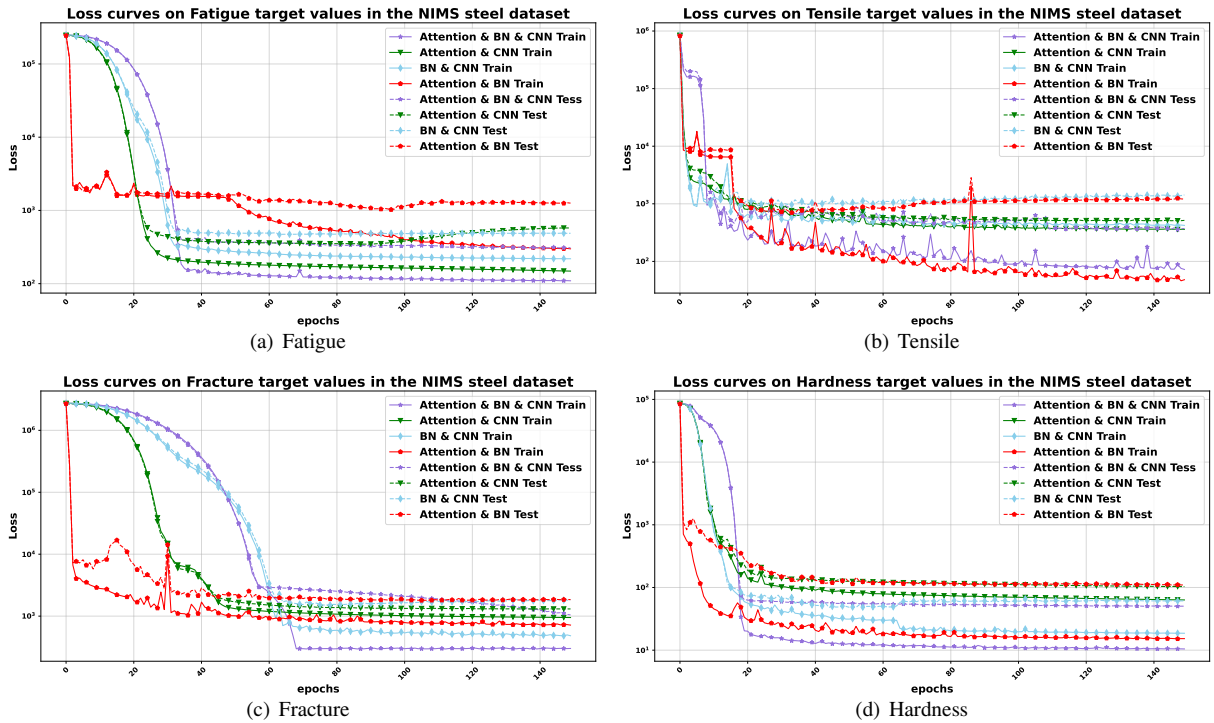


Figure 11: Module ablation experimental loss curves for the NIMS's steel dataset

The results of the module ablation experiment R^2 are presented in Table8. Figure 11 presents the loss curves of the three network structures for learning optimization on the four target values in the NIMS's steel dataset.

Both Table8 and Fig.11 show the efficacy of the multi-headed self-attentive module and the BN module in improving the model's prediction ability. The addition of the BN module accelerates the training speed of the model

and reduces the number of rounds for model convergence. However, adding BN may cause the model to oscillate owing to the shrinkage of the gradient scale. Adding the multi-headed self-attentive module may significantly retard the model's training speed owing to the increased computational effort.

5. Conclusions and Future Studies

This study proposes a multimodal deep learning network structure (MCVN) that can better utilization of materials' compositions for predicting their properties. First, XenonPy converts material compositional features into statistical element-level features. According to the grayscale feature mapping scheme, a single modal dataset is augmented to be multimodal. The implicit information contained in the image's modal features is then mapped as textures in the grayscale map, which is processed by a CNN to filter noise efficiently and extract critical information, which is then extracted by a multi-headed self-attentive model and a fully connected layer, followed by modal fusion with the modal image being the output.

Ablation experiments verified the necessity of having two modal and two network module structures to improve the MCVN's accuracy in predicting material properties. The MCVN also has limitations, mainly because it is based on a deep neural network architecture. The MCVN is easy to overfit, and incurs higher computational effort. However, it is suitable for greater data volumes, and manages more complex.

5.1. Regression Task

Traditional machine learning algorithms and the MCVN were compared on the NIMS's steel dataset steel mechanical property prediction task. We achieved an R^2 result of 0.954 for fatigue strength, 0.983 for tensile strength, 0.954 for fracture strength, and 0.98 for hardness, despite the excellent performance of traditional machine learning models. In the steel mechanical property prediction task of SRIM, the MCVN underwent ten-fold cross-validation to have an R^2 result of 0.974 for yield strength, with 0.98, 0.943, and 0.835 respectively for tensile strength, elongation, and cross-section shrinkage, which is a 3% to 4% improvement in the yield strength and tensile strength target values where traditional machine learning models performed well, as well as a 6% R^2 improvement in elongation. In addition, the MCVN improved the R^2 result by 18% for the section-shrinkage target value, on which the traditional machine learning models performed poorly. These two experiments on the MCVN verified that by visualizing the elemental-level features of material chemistry, multiple modal data can be fused and complemented with the original modal features and can significantly improve its regression-prediction accuracy.

5.2. Classification Task

The experiments based on amorphous alloy data show that the $F1 - Score$ of the MCVN significantly improves compared with the traditional machine learning models in the classification task. The predicted $F1 - Score$ reached 0.917 for bulk metallic glasses (BMG), 0.921 for ribbon metallic glass (RMG), and 0.818 for crystalline alloy (CRA). In the unbalanced dataset, the MCVN well-classified CRA and BMG with a small sample size from RMG with a large sample size and achieved high prediction accuracy in each category and high recall in the small sample category. This experiment verified that by visualizing the elemental-level features of material chemistry, multimodal data can complement the enhanced unimodal data by only using chemical compositional features.

5.3. Future Work

Besides the element-level feature mapping mentioned in this paper, other features from the multimodal data can be introduced for mapping, and relevant material image data can also be directly introduced. Therefore, in future work, the combination of other material data (such as material microstructure image data) can be considered with gray-scale map data, and the gray-scale graph mapping scheme mentioned in this paper would follow a single order for mapping. The introduction of mapping rules with more prior knowledge of the material can be considered in the future, with more information and information density, the mapped data is expected to achieve better prediction performance. The molecular structure and spatial structure of materials can also be used as graph structure data, that, combined with graph neural networks, will enhance the prediction capability of multimodal deep neural networks.

6. Acknowledgments

This work was sponsored by the National Key Research and Development Program of China (No. 2018YFB0704400), Key Research Project of Zhejiang Laboratory (No.2021PE0AC02), Key Program of Science and Technology of Yunnan

Province (No. 202002AB080001-2, 202102AB080019-3), Key Project of Shanghai Zhangjiang National Independent Innovation Demonstration Zone(No. ZJ2021-ZD-006). The authors gratefully appreciate the anonymous reviewers for their valuable comments.

7. Data and code availability

The data and source code that support the findings of this study are available at <https://github.com/yuyouyu32/MCVN>.

8. Conflict of 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.

CRedit authorship contribution statement

Yeyong Yu: Writing - Original draft, Data curation, Software, Implementation, Investigation, Formal analysis, Visualization. **Xing Wu:** Supervision, Writing - Review & Editing. **Quan Qian:** Conceptualization, Methodology, Funding acquisition, Project administration, Supervision, Writing - Review & Editing.

References

- [1] Agrawal, A., Choudhary, A., 2018. An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. *International Journal of Fatigue* 113, 389–400.
- [2] Baltrušaitis, T., Ahuja, C., Morency, L.P., 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 423–443.
- [3] Bartel, C.J., Trewartha, A., Wang, Q., Dunn, A., Jain, A., Ceder, G., 2020. A critical examination of compound stability predictions from machine-learned formation energies. *npj Computational Materials* 6, 1–11.
- [4] Bellman, R., 1957. *Dynamic programming*, Princeton univ. Press Princeton, New Jersey.
- [5] Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent, in: *Proceedings of COMPSTAT'2010*. Springer, pp. 177–186.
- [6] Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32.
- [7] Chang, C.C., Lin, C.J., 2011. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 1–27.
- [8] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.
- [9] Choudhary, K., Garrity, K.F., Sharma, V., Biacchi, A.J., Hight Walker, A.R., Tavazza, F., 2020. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *NPJ Computational Materials* 6, 1–13.
- [10] Dai, M., Hu, J.M., 2020. Field-free spin-orbit torque perpendicular magnetization switching in ultrathin nanostructures. *npj Computational Materials* 6, 1–10.
- [11] Dong, Y., Wu, C., Zhang, C., Liu, Y., Cheng, J., Lin, J., 2019. Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride. *npj Computational Materials* 5, 1–8.
- [12] Erickson, Z., Xing, E., Srirangam, B., Chernova, S., Kemp, C.C., 2020. Multimodal material classification for robots using spectroscopy and high resolution texture imaging, in: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE. pp. 10452–10459.
- [13] Gusenbauer, M., Oezelt, H., Fischbacher, J., Kovacs, A., Zhao, P., Woodcock, T.G., Schrefl, T., 2020. Extracting local nucleation fields in permanent magnets using machine learning. *npj Computational Materials* 6, 1–10.
- [14] Huber, L., Hadian, R., Grabowski, B., Neugebauer, J., 2018. A machine learning approach to model solute grain boundary segregation. *npj Computational Materials* 4, 1–8.
- [15] Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: *International conference on machine learning*, PMLR. pp. 448–456.
- [16] Jha, D., Choudhary, K., Tavazza, F., Liao, W.k., Choudhary, A., Campbell, C., Agrawal, A., 2019. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nature communications* 10, 1–12.
- [17] Jolliffe, I.T., Morgan, B., 1992. Principal component analysis and exploratory factor analysis. *Statistical methods in medical research* 1, 69–95.
- [18] yoshida lab., Xenonpy. <https://xenonpy.readthedocs.io>. 2019.
- [19] Li, S., Dan, Y., Li, X., Hu, T., Dong, R., Cao, Z., Hu, J., 2020. Critical temperature prediction of superconductors based on atomic vectors and deep learning. *Symmetry* 12, 262.
- [20] Li, X., Liu, Z., Cui, S., Luo, C., Li, C., Zhuang, Z., 2019. Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning. *Computer Methods in Applied Mechanics and Engineering* 347, 735–753.
- [21] Liu, C., Fujita, E., Katsura, Y., Inada, Y., Ishikawa, A., Tamura, R., Kimura, K., Yoshida, R., 2021. Machine learning to predict quasicrystals from chemical compositions. *Advanced Materials* 33, 2102507.

- [22] Ma, M., Sun, C., Chen, X., 2018. Deep coupling autoencoder for fault diagnosis with multimodal sensory data. *IEEE Transactions on Industrial Informatics* 14, 1137–1145.
- [23] Meel, P., Vishwakarma, D.K., 2021. Han, image captioning, and forensics ensemble multimodal fake news detection. *Information Sciences* 567, 23–41.
- [24] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning, in: *ICML*.
- [25] Oommen, T., Misra, D., Twarakavi, N.K., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. *Mathematical geosciences* 40, 409–424.
- [26] Politis, I., Brewster, S., Pollick, F., 2017. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 9, 1–16.
- [27] Quinlan, J.R., 2014. *C4. 5: programs for machine learning*. Elsevier.
- [28] Rhoades, S.A., 1995. Market share inequality, the hhi, and other measures of the firm-composition of a market. *Review of industrial organization* 10, 657–674.
- [29] Schütt, K.T., Sauceda, H.E., Kindermans, P.J., Tkatchenko, A., Müller, K.R., 2018. Schnet—a deep learning architecture for molecules and materials. *The Journal of Chemical Physics* 148, 241722.
- [30] Sigmund, G., Gharasoo, M., Hüffer, T., Hofmann, T., 2020. Deep learning neural network approach for predicting the sorption of ionizable and polar organic pollutants to a wide range of carbonaceous materials. *Environmental science & technology* 54, 4583–4591.
- [31] Soleymani, S., Dabouei, A., Kazemi, H., Dawson, J., Nasrabadi, N.M., 2018. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification, in: *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE. pp. 3469–3476.
- [32] Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W., Zafeiriou, S., 2017. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing* 11, 1301–1309.
- [33] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems* 30.
- [34] Wang, P., Gao, R.X., Yan, R., 2017. A deep learning-based approach to material removal rate prediction in polishing. *CIRP Annals* 66, 429–432.
- [35] Wei, J., Shaowei, C., Hao, P., Hu, J., Wang, S., Lou, Z., 2021. Multimodal unknown surface material classification and its application to physical reasoning. *IEEE Transactions on Industrial Informatics* .
- [36] Wu, S., Lambard, G., Liu, C., Yamada, H., Yoshida, R., 2020a. iqspr in xenonpy: a bayesian molecular design algorithm. *Molecular informatics* 39, 1900107.
- [37] Wu, X., Chen, C., Li, P., Zhong, M., Wang, J., Qian, Q., Ding, P., Yao, J., Guo, Y., 2022. Ftap: Feature transferring autonomous machine learning pipeline. *Information Sciences* 593, 385–397.
- [38] Wu, X., Zhong, M., Guo, Y., Fujita, H., 2020b. The assessment of small bowel motility with attentive deformable neural network. *Information Sciences* 508, 22–32.
- [39] Xiong, J., Shi, S.Q., Zhang, T.Y., 2020a. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. *Materials & Design* 187, 108378.
- [40] Xiong, J., Zhang, T., Shi, S., 2020b. Machine learning of mechanical properties of steels. *Science China Technological Sciences* 63, 1247–1255.
- [41] Zhang, W., Zhang, N., Zhang, W., Yen, G.G., Li, G., 2021. A cluster-based immune-inspired algorithm using manifold learning for multimodal multi-objective optimization. *Information Sciences* 581, 304–326.
- [42] Zheng, W., Liu, H., Wang, B., Sun, F., 2019. Cross-modal surface material retrieval using discriminant adversarial learning. *IEEE transactions on industrial informatics* 15, 4978–4987.
- [43] Zhu, Q., Samanta, A., Li, B., Rudd, R.E., Frolov, T., 2018. Predicting phase behavior of grain boundaries with evolutionary search and machine learning. *Nature communications* 9, 1–9.

Table 9: XenonPy element-level properties

Index	feature	description
1	period	Period in the periodic table
2	atomic_number	Number of protons found in the nucleus of an atom
3	mendeleviev_number	Atom number in mendeleviev’s periodic table
4	atomic_radius	Atomic radius
5	atomic_radius_rahm	Atomic radius by Rahm et al
6	atomic_volume	Atomic volume
7	atomic_weight	The mass of an atom
8	icsd_volume	Atom volume in ICSD database
9	lattice_constant	Physical dimension of unit cells in a crystal lattice
10	vdw_radius	Van der Waals radius
11	vdw_radius_alvarez	Van der Waals radius according to Alvarez
12	vdw_radius_mm3	Van der Waals radius from the MM3 FF
13	vdw_radius_uff	Van der Waals radius from the UFF

Table 9: XenonPy element-level properties

Index	feature	description
14	covalent_radius_cordero	Covalent radius by Cordero et al
15	covalent_radius_pyykko	Single bond covalent radius by Pyykko et al
16	covalent_radius_pyykko_double	Double bond covalent radius by Pyykko et al
17	covalent_radius_pyykko_triple	Triple bond covalent radius by Pyykko et al
18	covalent_radius_slater	Covalent radius by Slater
19	c6_gb	C_6 dispersion coefficient in a.u
20	density	Density at 295K
21	dipole_polarizability	Dipole polarizability
22	electron_affinity	Electron affinity
23	electron_negativity	Tendency of an atom to attract a shared pair of electrons
24	en_allen	Allen's scale of electronegativity
25	en_ghosh	Ghosh's scale of electronegativity
26	en_pauling	Pauling's scale of electronegativity
27	gs_bandgap	DFT bandgap energy of T=0K ground state
28	gs_energy	DFT energy per atom (raw VASP value) of T=0K ground state
29	gs_est_bcc_latent	Estimated BCC lattice parameter based on the DFT volume
30	gs_est_fcc_latent	Estimated FCC lattice parameter based on the DFT volume
31	gs_mag_moment	DFT magnetic moment of T=0K ground state
32	gs_volume_per	DFT volume per atom of T=0K ground state
33	hhi_p	Herfindahl-Hirschman Index (HHI) production values
34	hhi_r	Herfindahl-Hirschman Index (HHI) reserves values
35	specific_heat	Specific heat at 20oC
36	first_ion_en	First ionisation energy
37	fusion_enthalpy	Fusion heat
38	heat_of_formation	Heat of formation
39	heat_capacity_mass	Mass specific heat capacity
40	heat_capacity_molar	Molar specific heat capacity
41	evaporation_heat	Evaporation heat
42	boiling_point	Boiling temperature
43	bulk_modulus	Bulk modulus
44	melting_point	Melting point
45	thermal_conductivity	Thermal conductivity at 25 C
46	sound_velocity	Speed of sound
47	Polarizability	Ability to form instantaneous dipoles
48	molar_volume	Molar volume
49	num_unfilled	Total unfilled electron
50	num_valance	Total valance electron
51	num_d_unfilled	Unfilled electron in d shell
52	num_d_valance	Valance electron in d shell
53	num_f_unfilled	Unfilled electron in f shell
54	num_f_valance	Valance electron in f shell
55	num_p_unfilled	Unfilled electron in p shell
56	num_p_valance	Valance electron in p shell
57	num_s_unfilled	Unfilled electron in s shell
58	num_s_valance	Valance electron in s shell

Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: