



Better utilization of materials' compositions for predicting their properties: Material composition visualization network

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

^a School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China

^b Materials Genome Institute, Shanghai University, Shanghai, 200444, China

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

^d Zhejiang Laboratory, Hangzhou, 311100, Zhejiang, China



ARTICLE INFO

Keywords:

Material composition features densification
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 to comprehensively understand a material's composition–process–structure–performance relationship. Data-driven approaches have been regarded as the fourth paradigm of new materials R&D. However, the complexity of constituent elements in many material datasets leads to very sparse compositional features, posing a tremendous challenge to machine learning models. In this study, a data mapping scheme based on fundamental atomic features was used to visualize the chemical composition characteristics mapped into two-dimensional grayscale image data to solve the problem of sparse material composition matrix. Based on this, a material composition visualization network (MCVN) is proposed and applied to predict the mechanical properties of steel and classify amorphous alloy materials. We compared the MCVN to other machine learning methods. The MCVN had an average R^2 value improvement of 4% 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 it achieved an R^2 of 0.835 on the cross-sectional shrinkage target in the Shanghai Research Institute of Materials's (SRIM's) steel dataset where the other models only had an average R^2 of 0.64. For the unbalanced amorphous alloy material dataset, the MCVN improved the average *Recall* of the small-class crystalline alloy (CRA) from 0.58 to 0.78. The method based on expanding the material chemical composition information is universal and provides a new paradigm for material property prediction.

1. Introduction

The demand for novel materials is increasing; whether for military and defense applications, high-tech equipment, or personal electronics, these materials are essential. In new material development, identifying the relationship between the composition, process, structure, and properties is the basis for discovery. However, owing to the diversity and complexity of materials, a comprehensive understanding and handling of the composition–structure–process–property relationships is extremely challenging. Therefore, data-driven development of advanced materials is central to the Materials Genome Initiative (MGI). Big data and advanced technologies (e.g., artificial intelligence, data mining, and machine learning) are used to accelerate R&D. Big data and machine learning for materials provide a theoretical and methodical 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.

In addition to unimodal learning approaches, multimodal learning has received increasing attention from researchers, with many results emerging from various fields (e.g., multi-objective optimization problems Zhang et al., 2021, emotion recognition Tzirakis et al., 2017, fake news detection Meel and Vishwakarma, 2021, and autonomous driving Politis et al., 2017). Multimodal learning trains models by analyzing data from multiple modalities. It can yield superior results when the research objective involves a combination of modalities. There are many types of heterogeneous data in the field of material genomics, such as text information on material composition and processes that fall under structural data, material imaging under visual data, and its molecular structure under graph data. Fusing the data of various structures to improve the accuracy of material property prediction is an urgent challenge. Therefore, multimodal research has excellent prospects in material genomics.

* Corresponding author at: School of Computer Engineering and Science, Shanghai University, Shanghai, 200444, China.
E-mail address: qqian@shu.edu.cn (Q. Qian).

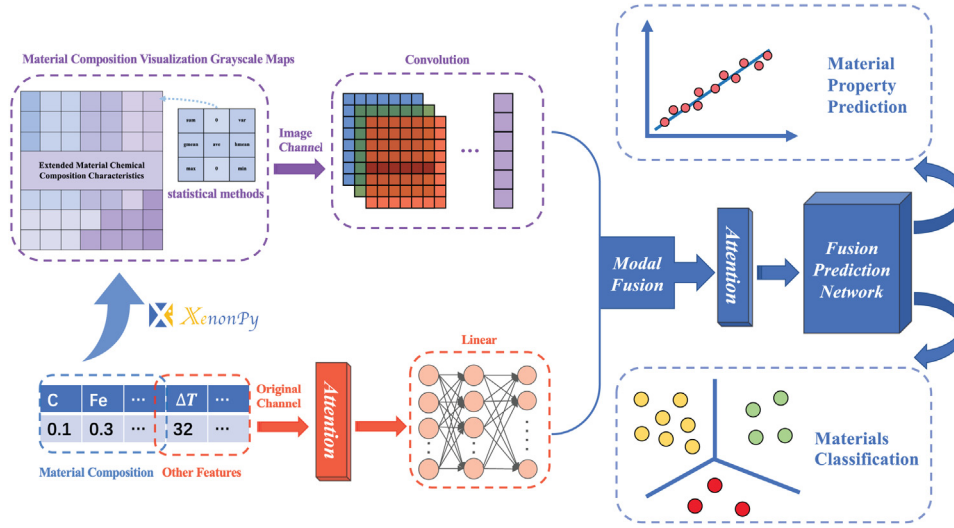


Fig. 1. Better utilization of composition for material property prediction using the MCVN.

Table 1

XenonPy feature statistics calculation method (yoshida-lab, 2019).

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	$f_{ave,i} = w_A f_{A,i} + w_B f_{B,i}$

Multimodal learning can utilize the rich information of different modal data to train models and improve their prediction accuracy. This study focuses on two multimodal aspects: modal representation and modal fusion. Modal representation refers to meaningfully combining data between different modalities, with minimal loss of their respective semantics. Modal representation typically uses joint and coordinated approaches (Ngiam et al., 2011). Joint representation is used to map the data of different modalities to the same feature space and to perform training therein Ma et al. (2018). This method is suitable for 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 representations between modalities are correlated (Soleymani et al., 2018). Co-feature representation is suitable for application scenarios when not all modalities are simultaneously present in the prediction process.

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

In this study, we focused on expanding and visualizing the compositional features of materials. 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, as shown in Fig. 1. Experimental validation was performed to predict the mechanical properties of steel and classify datasets of amorphous alloy materials. The main contributions of this study are summarized as follows:

- An approach using statistical information of the constituent elements is proposed to densify the sparse matrix features of components in materials dataset.
- A method for visualizing elemental extension information based on expert opinion is applied to better utilize material compositional features 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

Below we motivate augmenting the material composition information with atomic statistics and visualizing it as grayscale images for fusion modal learning.

- Machine learning methods rely heavily on feature selection, which is domain-specific (Wu et al., 2022). 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 was used to extend the elemental characteristics of materials and was used to calculate 58 elemental information points (Table 9) using seven statistical formulas (Table 1) to obtain a total of 406 extended dimensional characteristics.
- The 406 high-dimensional features may result in the curse of dimensionality (Bellman, 1957) in traditional machine learning. In Wu et al. (2020b), principal component analysis (PCA) (Jolliffe and Morgan, 1992) was used to reduce the dimensionality of medical images to obtain vectors; therefore, we also used PCA for comparative experiments.
- The 7 × 58-dimensional features have strong or weak correlations with 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,

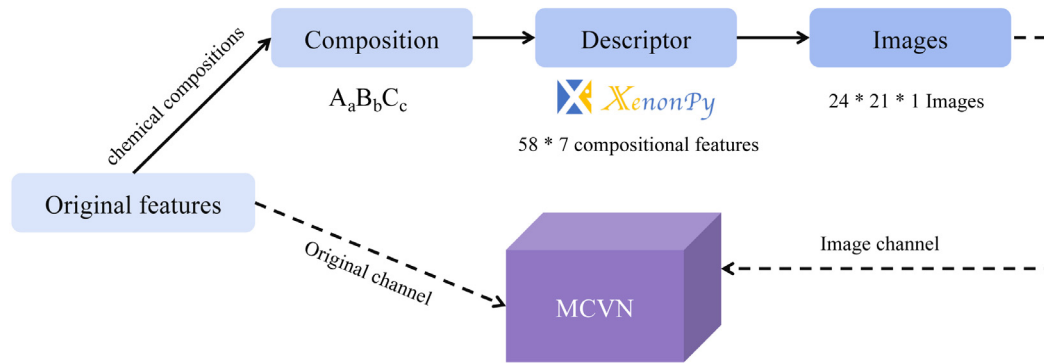


Fig. 2. Process flow of multimodal fusion machine learning with the MCVN.

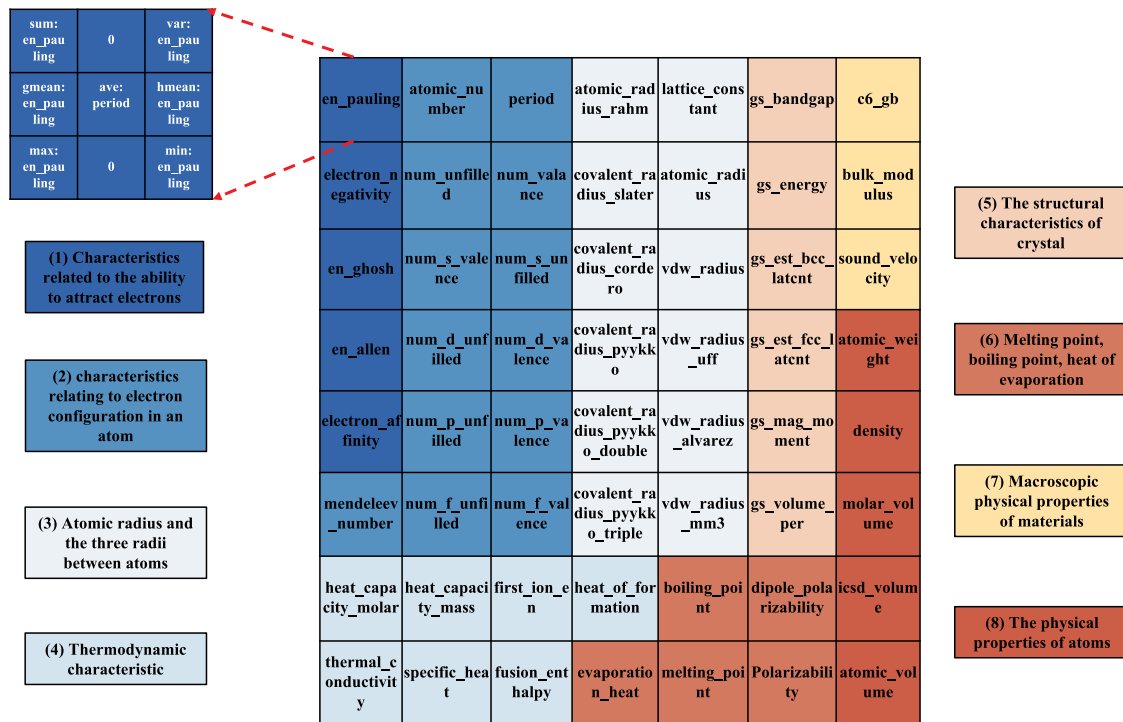


Fig. 3. Arrangement of compositional features. The 56-dimensional features are divided into eight parts: (1) features related to the ability to attract electrons; (2) features related to the arrangement of electrons in atoms; (3) atomic radii and three interatomic radii; (4) thermodynamic features; (5) 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.

which are ideal for extracting information using convolutional neural networks (Fig. 6).

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

The remainder 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 the feature extraction and different combination approaches, is explained in Section 3. The experimental results are discussed in Section 4. Finally, we conclude the paper in Section 5 along with discussions and future work.

2. Related work

Due to low computational costs and short development cycles, machine learning coupled with powerful data processing and highly accurate prediction results are widely used for materials testing, materials analysis, and materials design (Wei et al., 2019). Li et al. employed the principle of ensemble learning for time-series forecasting and material fatigue property prognosis (Li et al., 2017). Schutt proposed a

deep-learning-based SchNet network to model atomic systems using continuous filter convolution layers in a specially designed network structure (Schütt et al., 2018). Fernández et al. applied a new gradient-free training algorithm based on approximate Bayesian computation to predict the fatigue in composite materials (Fernández et al., 2022). Li et al. predicted the mechanical properties of heterogeneous materials by using convolutional neural network (CNN) modeling and deep learning (Li et al., 2019). Li et al. proposed a hybrid model based on convolutional networks and long short-term memory networks to predict the critical temperatures of superconductors (Li et al., 2020).

Measured or computed material data combined with various machine learning techniques have been used to solve a myriad of challenging problems (Pilania, 2021). With the development of machine learning methods, increasing amounts of data with different structures can be applied to material machine learning tasks. Dai et al. developed a graph neural network model to obtain an embedding of a polycrystalline microstructure that incorporates not only the physical features of individual grains but also their interactions (Dai

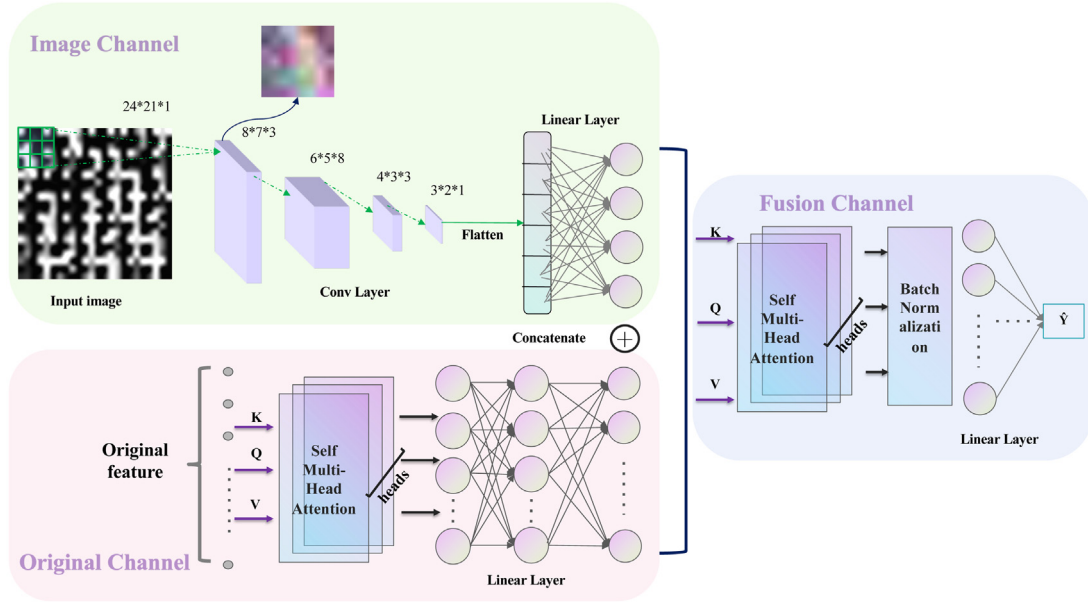


Fig. 4. Material composition visualization network architecture.

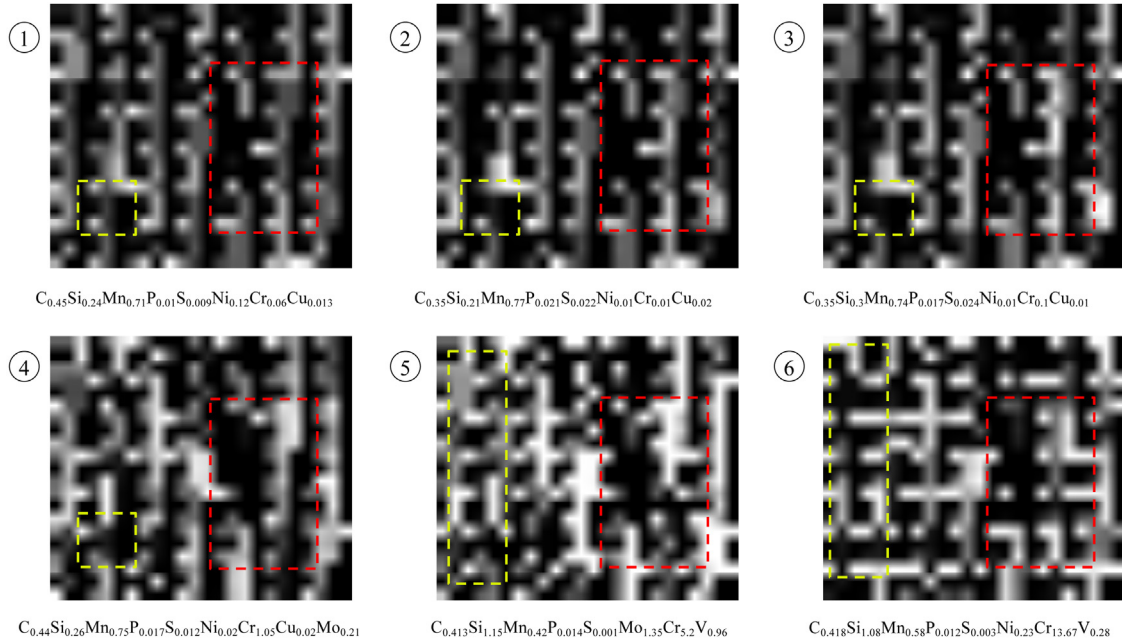


Fig. 5. Element-level description of features to reconstruct texture grayscale maps. The artificial texture grayscale maps of the six differently composed steel sheets are shown in Fig. 5. The steel sheets ①~④ are from the NIMS's steel dataset (Agrawal and Choudhary, 2018). ①, ②, and ③ have the same 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 been added, and the fractured texture at the yellow box has an obvious connection. Steel ⑤ has the element Ni omitted compared with 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.

et al., 2021). Gusenbauer et al. used decision trees to predict simulated three-dimensional microstructures from two-dimensional images (Gusenbauer et al., 2020). Jiang et al. applied atom-specific persistent homology to construct low-dimensional representations of input crystal structures with chemical insights (Jiang et al., 2021).

Schütt et al. developed an efficient deep learning approach that enables spatially and chemically resolved insights into quantum-mechanical observables of molecular systems that could provide size-extensive and uniformly accurate predictions in compositional and configurational chemical spaces for molecules of intermediate size (Schütt et al., 2017).

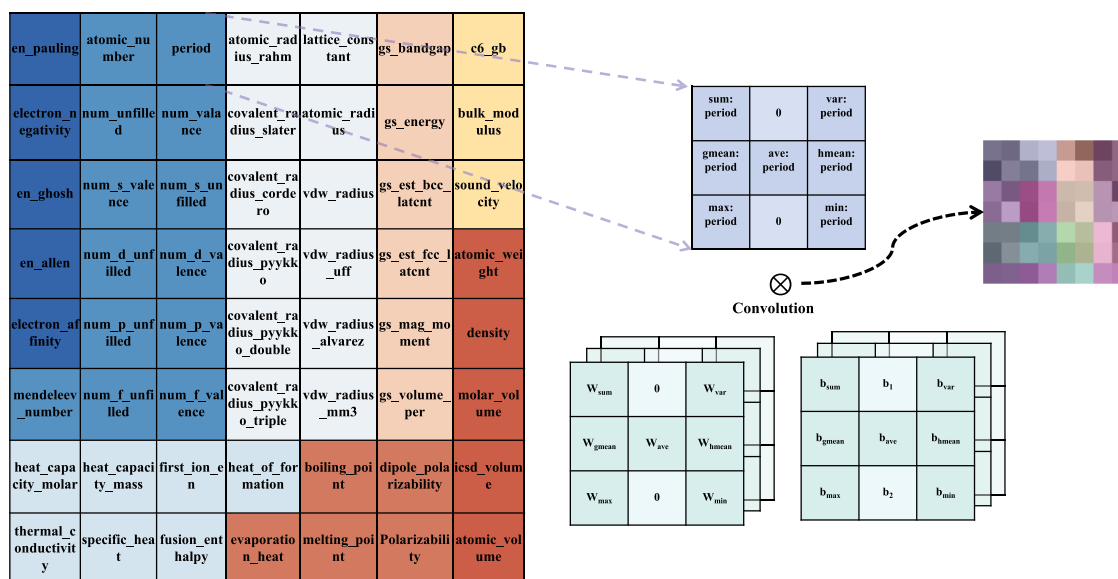


Fig. 6. First convolutional layer to extract element-level statistical feature information.

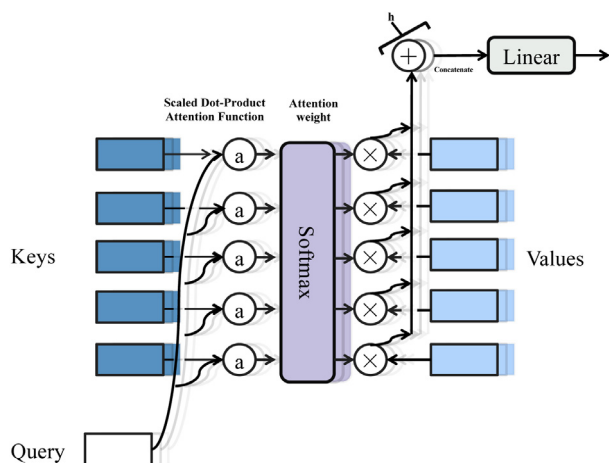


Fig. 7. Multi-head scaled dot-product attention.

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 material surfaces (Zheng et al., 2019). Erickson et al. developed a model that learns a compact multimodal representation of spectral measurements and texture images for material classification (Erickson et al., 2020). Wei et al. proposed a model that can learn material representations from auditory and multi-tactile sources (Wei et al., 2021).

However, the complexity of the constituent elements in many material datasets leads to very sparse compositional features, posing a tremendous challenge to machine learning models. The chemical characteristics of the material composition are critical in machine-learning tasks. Simultaneously, multimodal information of the material is obtained through sensor detection and relies only on the dataset. Therefore, we designed a method to improve the accuracy of material property prediction by extending the element information of the material components and performing multimodal learning that includes the original modes.

3. Proposed method

3.1. Expanding and visualizing material compositional features using XenonPy

XenonPy (yoshida-lab, 2019) 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 descriptors). 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 (Wu et al., 2020a). Liu et al. used XenonPy's material descriptor library to extend the chemical compositional features of materials to classify quasicrystals (Liu et al., 2021).

In most material datasets, the original composition features are very sparse due to numerous elements in the material construction. XenonPy provides a rich set of tools for applying material informatics to various tasks, in which the descriptor generator can compute multiple types of numerical descriptors from the composition of the material. XenonPy's built-in descriptor generator can generate 58 element-level descriptor properties (see Table 9 in the appendix) for 94 elements (from H to Pu) and uses seven statistics to obtain seven statistical features (see Table 1) for each element-level descriptor. Thus, by entering the chemical composition information of the material into the descriptor module, a matrix of element-level descriptor features (58×7 dimensions) was obtained.

As shown in Fig. 2, the chemical composition information of the material was first extracted from the original features and input into 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 (Rhoades, 1995). The Herfindahl and Hirschman indices, which are economic indicators, are used to measure industrial concentrations and do not 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 Table 1) to calculate

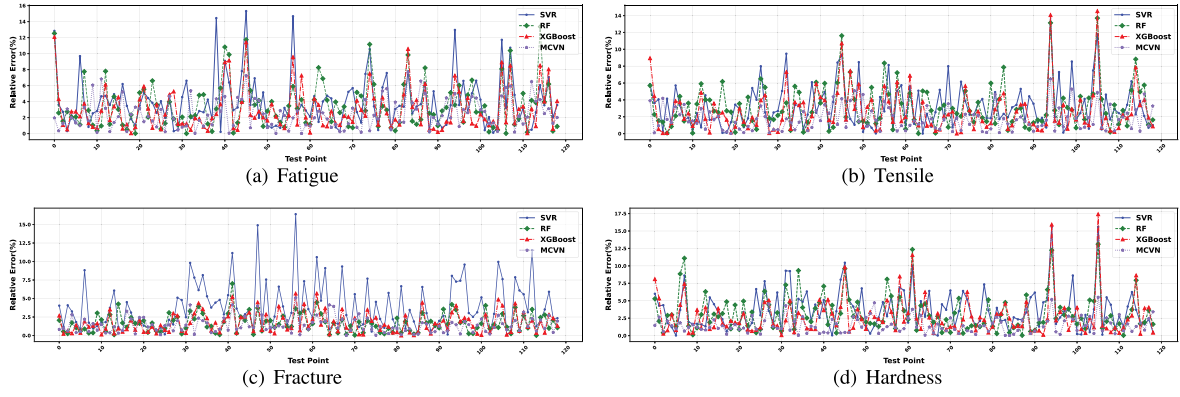


Fig. 8. Relative error of each point on the four target values in the test set in the NIMS's steel dataset. For the four target values, the relative error of the MCVN is smoother and closer to 0 than those of the other three models. The MCVN is more accurate in predicting the mechanical properties of 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 exceed 6%, while the relative errors of the MCVN are less than 1%. At the 95th and 105th test points of the tensile target, the relative errors of the remaining three models are approximately 12%, while the relative errors of the MCVN are 5% ~ 6%. At the 32nd and 95th test points of the Fracture target value, the relative error of the remaining three models exceed 2%, while the relative error of the MCVN approaches 0%. Finally, at the 94th and 105th test points of the hardness target value, the relative errors of the remaining three models exceed 12.5%, while the relative errors of the MCVN approach 5%.

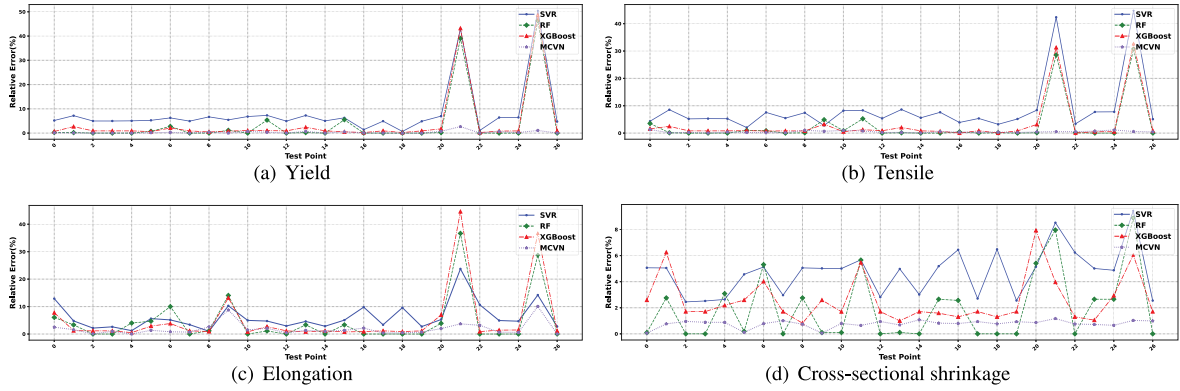


Fig. 9. Relative error of each point on the four target values in the test set in the SRIM steel dataset. For the four objectives, the average relative errors of the MCVN are all less than 5%. In contrast, the relative error curves of the remaining three models fluctuate more and are less effective. 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 is 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 statistical features corresponding to the 56 element-level attributes of the 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 $f_{B,i}$ ($i = 1, 2, \dots, 56$), is used to calculate 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, respectively. After 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 0, which could reduce the two w parameters in the first CNN layer. The statistical element-level descriptive features were rearranged according to certain correlations and visualized into a composition of 24×21 single-channel grayscale images as the modal image input of the MCVN. The generated grayscale images are fuzzy due to pixel limitation, Versaci et al. (2015) presents a fuzzy image preprocessor based on geometric considerations that have led to the formulation of an effective and efficient tool with reduced computational load, particularly useful for real-time applications, which can be included in the consideration of image pre-processing tools.

The concept of image mapping is based on the following four points for the positional arrangement of the broad categories shown in Fig. 2:

- The characteristics of electron arrangement in atoms, characteristics related to the ability to attract electrons, and the 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 (e.g., 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.

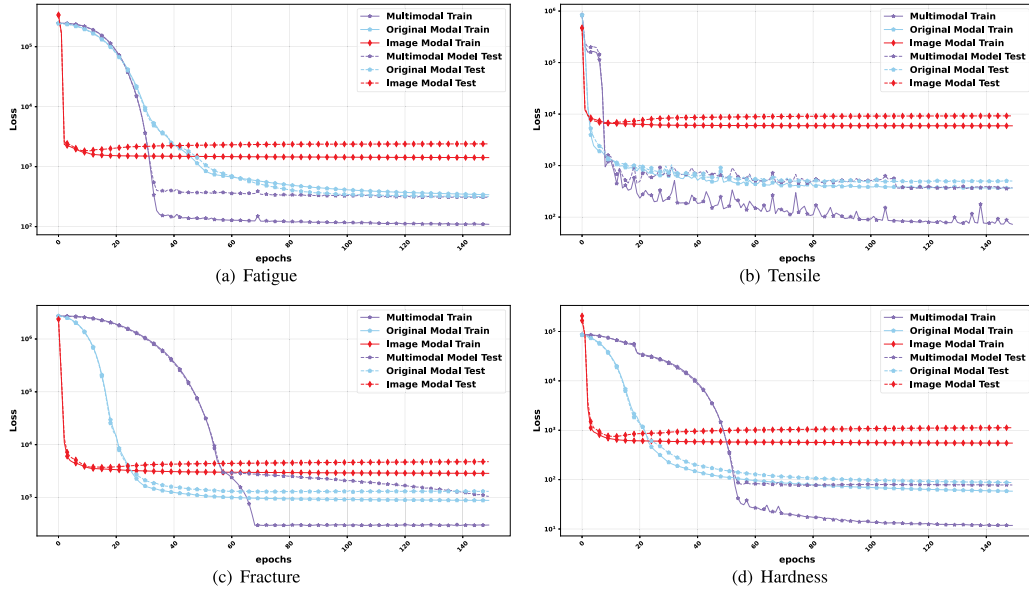


Fig. 10. Modal ablation experimental MSE loss curves for the NIMS steel dataset.

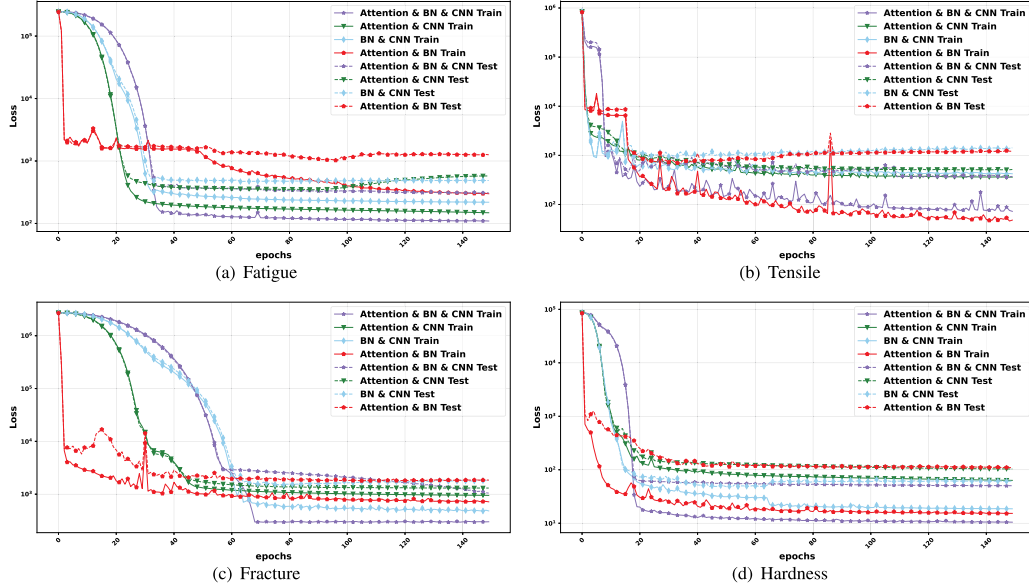


Fig. 11. Module ablation experimental MSE loss curves for the NIMS steel dataset.

- Significant categories related to thermodynamic properties are the physical properties of atoms, characteristics related to the electron arrangement outside the nucleus of atoms, and melting/boiling points. According to the thermodynamics equations, because some thermodynamic properties (i.e., 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 (e.g., 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

The MCVN (Fig. 4) has two modality inputs. One is the original features, including the material 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 layer captures the internal correlations of the data or features and the fully connected layer enhances the representation of the network. The other input is the output from XenonPy, that is, the statistical element-level descriptive features re-constituted into grayscale image information. According to unified rules, the element-level descriptive features are arranged by correlation between atomic structure features, atomic physical properties, crystal structure features, and grayscale mapping of statistical information, forming an image with texture features (Fig. 3). CNN efficiently extract texture feature information in parallel, and the obtained deep feature information of the image 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. However, the features of the image modal range between 0 ~ 255 after grayscale mapping, whereas the original features are

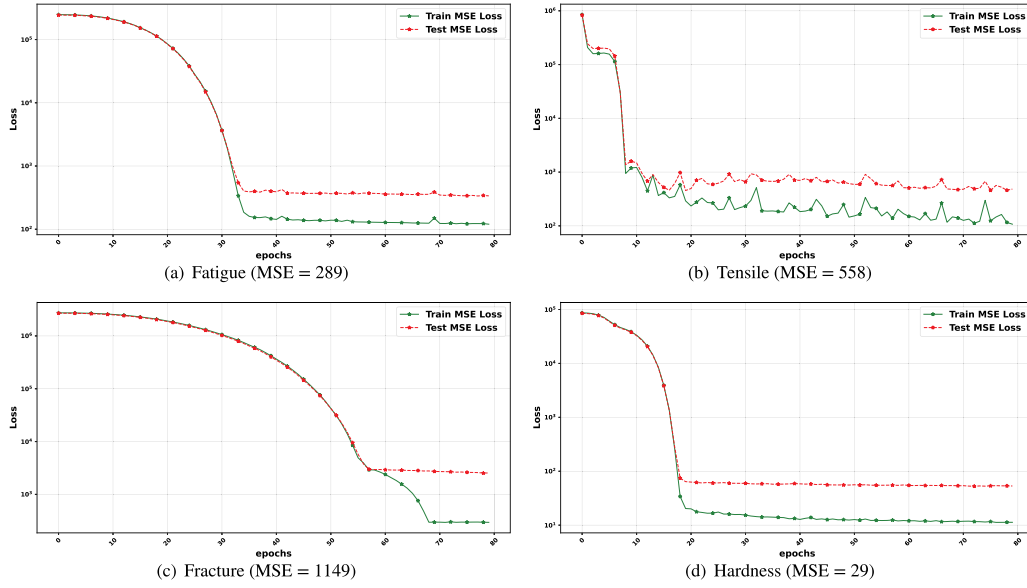


Fig. 12. Model learning MSE loss curves for the NIMS steel dataset.

distributed between 0 ~ 1 after normalization. This implies that the gaps between the feature values of the same sample are too large and affect the training of the hidden layer. After modal fusion, due to numerical variability between the two modals, batch normalization (BN) is required to reset the modal's fused hidden layer distribution to a standard normal distribution with a mean of 0 and variance of 1. Finally, the output of the BN layer is input to the fully connected layer to predict the target value.

3.2.1. Multimodal learning

The composition information of materials is an essential feature in materials informatics. However, traditional machine learning methods treat material composition information equally with other feature information and input it directly into the machine learning model for training. This retains only the proportional information of the components, whereas the basic information of the components is lost. XenonPy can calculate the composition information of materials based on their element-level descriptive features, which significantly augments the semantic information of the material components. However, the high-dimensional element-level defining features obtained by XenonPy, based on seven statistical formulas, introduce the curse of dimensionality (Bellman, 1957), where, as the dimensionality increases, the space volume increases rapidly and sparsifies the available data. This decrease is known as the Hughes phenomenon (Oommen et al., 2008). 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.

CNNs can manage high-dimensional image information owing to their local perception and parameter sharing. In this study, based on material composition information, the statistical element-level descriptive features output by XenonPy were reconstructed as grayscale images. Element-level statistical information textures were constructed manually according to relevance ranking, and the CNN extracted 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 mapped to the vector feature space to

train the model (Eq. (1)).

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

where x_{image} denotes the input of the image channel; $y_{\text{image channel}}$ denotes the output of the image channel; x_{original} denotes the input of the original channel; $y_{\text{original channel}}$ 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, BN layer, and vector stitching operation, respectively. In addition, $y_{\text{Multimodal}}$ 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 chemical composition information of the material. It has been demonstrated that using multimodal data can promote information complementarity (Baltrušaitis et al., 2018).

3.2.2. CNN extracting element-level descriptors

The input of the image modality was an array of statistical element-level descriptors, and each element-level descriptor adopted the same statistical method and order (Fig. 6), while the element-level descriptors with high relevance 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 of the components of a material. We arrange the 7-dimensional statistics in a 3*3 window and use a convolutional kernel with 3 strides to better exploit the translational invariance of CNN and arrange the features with higher correlation closer together to better exploit the local connectivity of CNN. Eq. (2) represents the process of extracting statistical information from the element-level descriptors in the first convolutional layer.

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

where $y_{\text{Conv-1}}$ 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

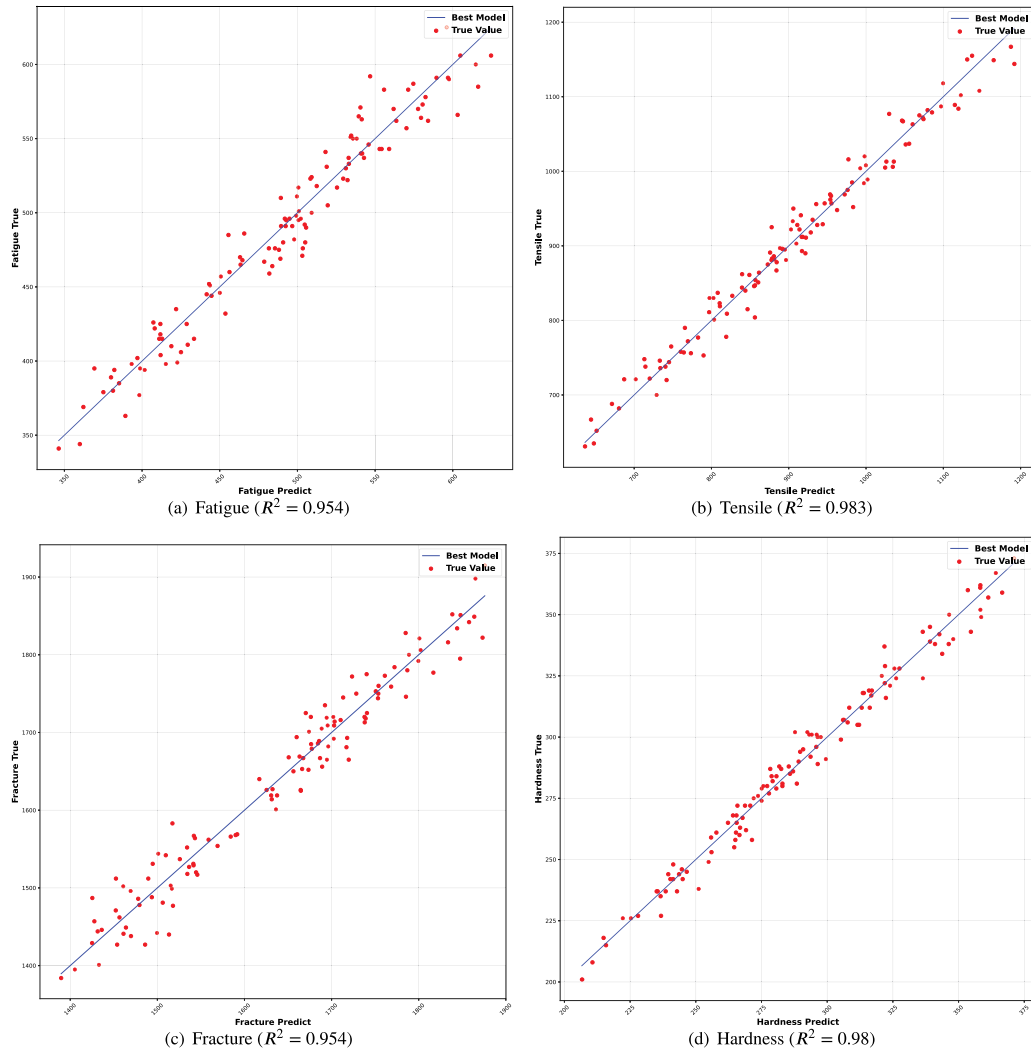


Fig. 13. Predicted vs. true values for the NIMS steel dataset.

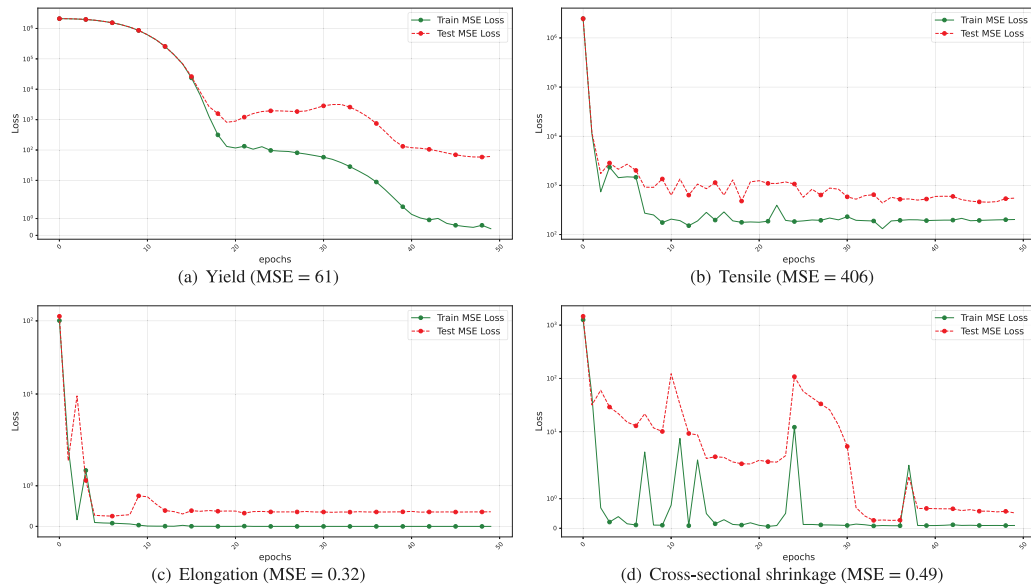


Fig. 14. Model learning MSE loss curves for the SRIM steel dataset.

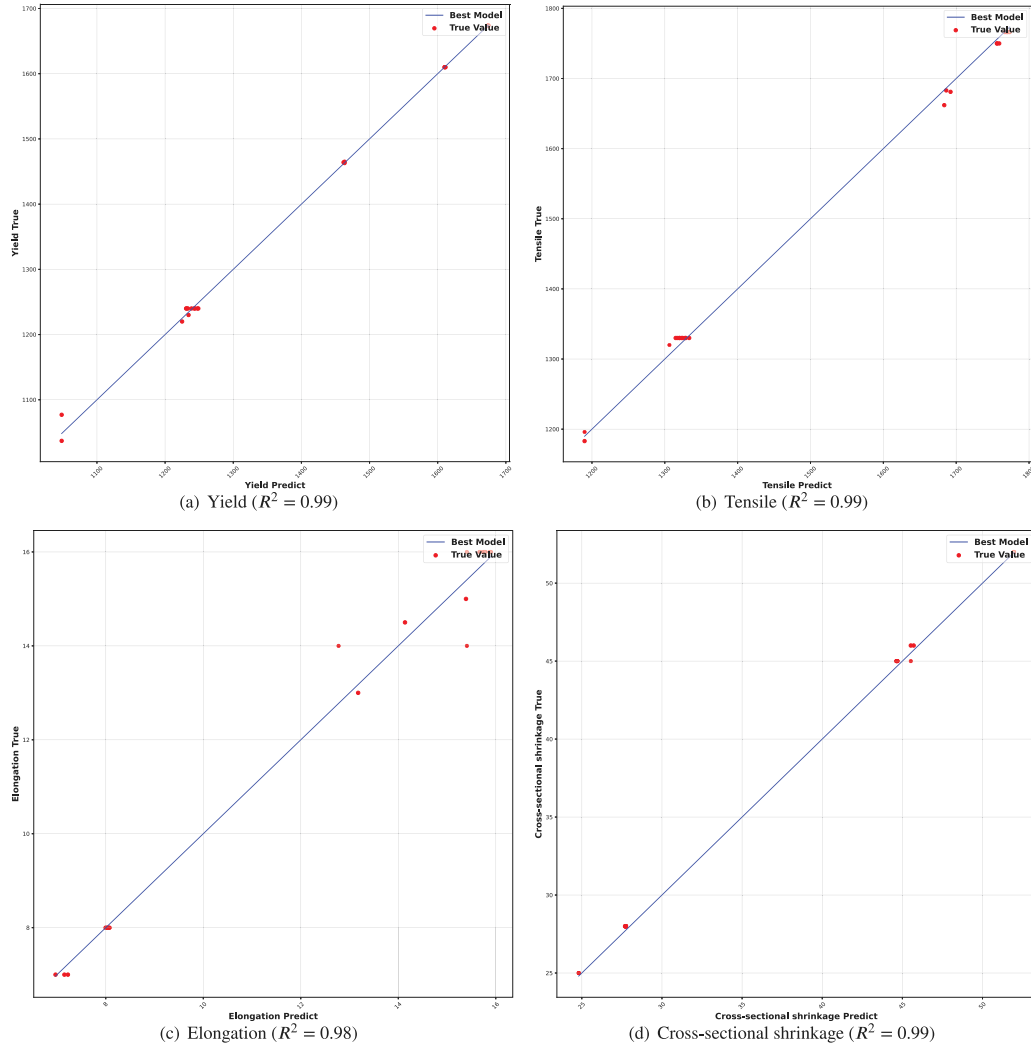


Fig. 15. Predicted vs. true values for the SRIM steel dataset.

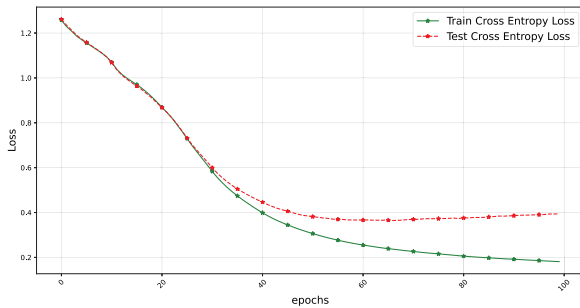


Fig. 16. Model learning cross entropy loss curves for the BMG-RMG-CRA dataset. (F1-Score = 0.857, Accuracy = 0.858, Cross Entropy Loss = 0.39).

statistical descriptor, and 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 dynamically learn the convolutional kernel parameters. Second, because the neurons on the same feature mapping surface have equal weights, the network can learn in parallel, which is a significant advantage of convolutional networks over fully connected networks. Because the artificially constructed image has fewer pixels compared with 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. Multiple convolution kernels allow the feature channel dimensions 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, whereas 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 affects the training of the hidden layers after modal fusion. In 2015, Ioffe et al. proposed BN, which is a data normalization method for deep neural networks (Ioffe and Szegedy, 2015). Typically, before the activation layer in a deep neural network, the BN layer calculates data from a batch input of size m (Eq. (3) Ioffe and Szegedy, 2015). The purpose of BN is to pull the value toward the linear region of the subsequent nonlinear transformation to be performed, increase the derivative values, enhance information backpropagation, and accelerate training convergence. However, this leads to a decrease in the 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

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

network (Ioffe and Szegedy, 2015).

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

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

Utilizing a BN 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, aid regularization, and reduce overfitting (Ioffe and Szegedy, 2015). In this study, BN 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 target values.

The 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 similarities between other samples. Similarity was used as the weight for multiplying with the corresponding sample value. Finally, critical information regarding the sample being scrutinized was obtained. 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* (Vaswani et al., 2017). The most common attention functions are additive attention and scaled dot product attention (Eq. (4) Vaswani et al., 2017). The attention function used in this study was dot product

attention. First, Q and K of the same dimension d_k were obtained using the attention function for the attention score. The attention score was compressed between 0 and 1 by the softmax activation function, and the output value of the object in this calculation was obtained by multiplying it with the original V .

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

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 (Vaswani et al., 2017), which is well suited for parallel GPU operations.

The single-headed attention model overly focuses on one-sided information when encoding information at the current location. Fig. 7 shows the multiheaded dot product attention model. The multiheaded attention model can operate in parallel because the multiheaded attention mechanism is parallel to single-headed attention. The output of the parallel single-headed attention function is conjoined and fed into a fully connected layer to obtain the output of the multiheaded attention, (as shown in Eq. (5) Vaswani et al., 2017). The benefit of the multiheaded attention mechanism is that it allows the model to focus jointly on information from different representation subspaces at different locations (Vaswani et al., 2017), enhancing the representational power of the model.

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

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

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 study, we employed $h = 2$ parallel attention heads.

In material data without time-series information, the similarity and correlation between samples are more critical than the location of the data samples, that is, the model must attend more to the changes in each sample relative to other 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 the data or features. Self-attention is the initialization of Q , K , and V for sampling X . All three have precisely the same value and dimensionality, which is ideal for the dot product attention function. Compared with 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 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 multiheaded self-attentiveness mechanism to capture information in the original feature space of the material and 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 (Agrawal and Choudhary, 2018) and SRIM steel dataset contained steel composition and processing information. The composition information was mapped to grayscale maps, and the original modal information for modal fusion was used to

Table 3

Experimental R^2 results of the steel material property prediction model for the NIMS's steel dataset. Here, 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 original features..

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$
Fatigue	0.954	0.937	0.926	0.939	0.915	0.949	0.913	0.909	0.913	0.908	0.89	0.780	0.882	0.903	0.886	0.904	0.920	0.931	0.921
Tensile	0.983	0.966	0.958	0.968	0.936	0.973	0.952	0.939	0.945	0.938	0.925	0.785	0.916	0.919	0.919	0.90	0.942	0.951	0.928
Fracture	0.954	0.937	0.942	0.947	0.869	0.920	0.864	0.894	0.915	0.905	0.905	0.859	0.907	0.913	0.941	0.929	0.925	0.930	0.936
Hardness	0.980	0.965	0.954	0.968	0.923	0.960	0.926	0.927	0.940	0.929	0.907	0.783	0.899	0.927	0.934	0.904	0.935	0.948	0.931

Table 4

Statistical information of the characteristics of the SRIM 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

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)) and mean absolute percentage error (MAPE) (Eq. (8)). The loss function was the mean squared error (MSE) (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, the higher the accuracy of the model prediction.

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

$$MAPE = \frac{\sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{\hat{y}_i} \right)^2}{n} \quad (8)$$

where y_i and \hat{y}_i are the predicted value by the model and the ground truth, respectively, for each point i in training set N .

For the classification task on the amorphous alloy material dataset (Xiong et al., 2020), 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 the different types of amorphous alloys. We verified the *Precision*, *Recall*, and *F1 - Score* (Eq. (9)) for the classification task: Cross-entropy was used as the loss function (Eq. (10)).

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 - Score &= \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \end{aligned} \quad (9)$$

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.

$$\text{Cross-entropy}(y, \hat{y}) = - \sum_k \hat{y}_k \log(y_k) \quad (10)$$

where y_k and \hat{y}_k are the predicted score by the model and the ground truth, respectively, 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 and 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: $A(12, 2) - BN - L(12, 8) - L(8, 4) - L(4, 1)$

Attention Network (contrast model): $A(N, 2) - L(N, 8) - L(8, 1)$

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; $A(i, h)$ represents the multi-head attention layers with i -dimension input and h heads; and N denotes the numbers of input features. For the regression task, we used ReLU and LogSoftmax activation functions for the classification task.

The first CNN layer in the Image Channel has a $3*3$ convolutional kernel and stride of 3, which was deliberately designed. The $3*3$ window size allows for the extraction of seven-dimensional statistical information for each extended feature. Meanwhile, a stride of 3 allows the convolution kernel to extract information for each feature independently.

Train-Test Split. The following datasets were split into training and test sets using a 2 : 1 ratio using the shuffle module in the Sk-learn toolkit (OlivierGrisel, 2022).

Comparison Models. The experimental results of MCVN are compared with the following typical machine learning algorithms.

- Neural network based on back-propagation: neural networks with attention mechanisms(Attention Network) (Vaswani et al., 2017)
- Neural network based on randomization: The new ensemble deep Random Vector Functional Link Network with re-normalization (edRVFL_N) (Shi et al., 2022, 2021) .
- Linear regression: Ridge regression (Bottou, 2010)
- Support vector machine: SVM (Chang and Lin, 2011)
- Bagging tree model: Random forest (RF) (Breiman, 2015)
- Boosting model: XGBoost (Chen and Guestrin, 2016)

Table 5Experimental R^2 results of the steel property prediction model on the SRIM steel dataset. O , E , and $E\&P$ are described in Table 3.

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$	O	E	$E\&P$
Yield	0.974	0.932	0.912	0.938	0.928	0.947	0.940	0.923	0.949	0.928	0.891	0.899	0.880	0.926	0.947	0.944	0.908	0.926	0.922
Tensile	0.98	0.942	0.924	0.956	0.957	0.970	0.956	0.944	0.959	0.942	0.918	0.905	0.878	0.94	0.958	0.933	0.915	0.948	0.95
Elongation	0.943	0.912	0.897	0.927	0.843	0.869	0.872	0.882	0.77	0.858	0.875	0.858	0.859	0.859	0.860	0.830	0.823	0.814	0.793
Cross-sectional shrinkage	0.835	0.638	0.593	0.723	0.729	0.701	0.755	0.624	0.596	0.654	0.544	0.617	0.582	0.668	0.556	0.658	0.656	0.411	0.605

Table 6Experimental results of amorphous alloy material classification for the BMG—RMG—CRA dataset. O , E , and $E\&P$ are detailed in Table 3.

Algorithm	Category		Precision	Recall	F1-score	Support
RF (Breiman, 2015)	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 (Chen and Guestrin, 2016)	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	
Attention Network (Vaswani et al., 2017)	BMG	<i>O</i>	0.842	0.875	0.858	208
		<i>E</i>	0.895	0.860	0.877	
		<i>E&P</i>	0.899	0.861	0.879	
	RMG	<i>O</i>	0.871	0.905	0.888	1241
		<i>E</i>	0.890	0.905	0.898	
		<i>E&P</i>	0.885	0.913	0.899	
	CRA	<i>O</i>	0.786	0.583	0.684	510
		<i>E</i>	0.794	0.562	0.678	
		<i>E&P</i>	0.806	0.596	0.70	
edRVFL_N (Shi et al., 2022)	BMG	<i>O</i>	0.936	0.567	0.706	208
		<i>E</i>	0.967	0.860	0.910	
		<i>E&P</i>	0.937	0.581	0.718	
	RMG	<i>O</i>	0.770	0.923	0.840	1241
		<i>E</i>	0.847	0.930	0.887	
		<i>E&P</i>	0.785	0.930	0.851	
	CRA	<i>O</i>	0.742	0.503	0.600	510
		<i>E</i>	0.798	0.643	0.712	
		<i>E&P</i>	0.777	0.547	0.642	
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

Table 7

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

Modal	Model	Fatigue	Tensile	Fracture	Hardness
Image Channel	CNN	0.604	0.602	0.797	0.583
Original Channel	Attention Network (Vaswani et al., 2017)	0.927	0.932	0.925	0.945
Multimodal Channels	MCVN	0.954	0.983	0.954	0.98

Table 8

Module ablation experimental results of the steel material property prediction model for the NIMS dataset. Attention is the multi-headed self-attention module in the MCVN structure, BN is the BN layer in the MCVN structure during modal fusion, CNN denotes the convolutional layers in the image channel in the MCVN structure, ✓ implies that the module is present in the structure, and × indicates that it is not. A 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

4.2. Predicting the mechanical properties of steel materials

Steel is the most widely used structural material in engineering. High-strength steel materials have significant advantages in terms of structural stress performance, building use function, and socioeconomic and environmental benefits. Therefore, it is important to develop new steel materials that can meet these requirements in complex environments. Studying the mechanical properties of steel is important for the development of high-strength steel (Gang et al., 2013).

Current structural materials demonstrate high strength and high toughness. The stress properties of high-strength steel structures, in terms of stability, fatigue, and fracture, have been investigated in the literature (Ban et al., 2008). Owing to the large variation in its chemical composition, the fatigue crack development mechanism can also differ. Yield strength, tensile strength, elongation, and cross-sectional shrinkage are indicators that describe the plastic properties of steel materials. The results of this study (Ban et al., 2011) show that high-strength structural steels have excellent toughness, relatively high yield strength/tensile strength, and low elongation.

Therefore, among steels, fatigue resistance, fracture strength, hardness, yield strength, tensile strength, elongation, and cross-sectional shrinkage as representatives of their mechanical properties, it is necessary to investigate in depth the relationship between the mechanical property indices of steel and structural safety, and high-precision predictions are essential for the safe use of the material.

4.2.1. NIMS's steel dataset

This dataset (Agrawal and Choudhary, 2018) is from the National Institute for Materials Science (NIMS), Japan. It contains the chemical composition, processing conditions, and mechanical properties of materials, such as fatigue strength, tensile strength, fracture strength, and hardness. In this dataset, the effects of rotational bending fatigue strength (hereinafter 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 fatigue life cycles. The material contained nine alloying elements: carbon (C), silicon (Si), manganese (Mn), phosphorus (P), sulfur (S), nickel (Ni), chromium (Cr), copper (Cu), and Molybdenum (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. Meanwhile, three types of heat treatments are included: normalizing, quenching, and tempering. After 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 (the statistical details of each feature are presented in Table 2), and four target properties (i.e., fatigue strength, tensile strength, fracture strength, and hardness).

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

The parameters of all models in the experiment are shown in Table 10, and the learning curve of MCVN is shown in Fig. 12. The experimental results are presented in Table 3. Other machine learning models 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 value 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. To better represent the prediction accuracy of the MCVN for the four target values of the NIMS steel dataset, Fig. 8 presents the relative errors of the MCVN and the other machine learning models, SVR, RF, and XGBoost, for each test point of the NIMS steel dataset. The MCVN exhibited insignificant relative errors compared with the other three models for all four test points where most of the target values were found. The experimental results of MAPE and MSE are presented in Tables 11 and 12. The xy -plot of the predicted versus true values is presented in Fig. 13.

4.2.2. SRIM steel dataset

This dataset contained the chemical composition, processing conditions, and properties of the materials. The predicted mechanical properties in the dataset included yield strength, tensile strength, elongation, and cross-sectional shrinkage. The materials contain eleven 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, as described in natural language; therefore, one-hot encoding was used. The dataset contains 81 data samples with 16-dimensional features, including 11 alloying elements, 5 process parameters (the statistical details of each feature are presented in Table 4), and 4 target properties (i.e., yield strength, tensile strength, elongation, and cross-sectional shrinkage).

The evaluation index in this experiment was identical to that of the NIMS steel dataset. However, the dataset had fewer samples than the NIMS steel dataset, and the unbalanced distribution due to a single random division would have affected the accuracy of evaluating the model. Therefore, 10 independent experiments with different random seeds were performed to calculate the average R^2 value and enhance the evaluation accuracy. The parameters of all models in the experiment are shown in Table 13. We chose an independent experiment with superior overall performance to present its learning curve (Fig. 14).

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 conducting the 10 independent experiments, the MCVN improved the accuracy of prediction compared to other models on the yield strength, tensile strength and elongation target values (the experimental results are shown in Table 5). The MCVN achieved an R^2 of 0.835 for the target value of cross-section shrinkage, whereas all other machine learning models performed poorly and only achieved an average score of 0.64. The relative errors of the MCVN and the other three machine learning models, SVR, RF, and XGBoost, are shown in Fig. 9 for each test point of the SRIM steel dataset. The experimental results of MAPE and MSE are presented in Tables 14 and 15. We chose an independent experiment with superior overall performance to present the xy -plot of the predicted versus true values (Fig. 15).

4.3. Classification of amorphous alloy materials

Amorphous alloys are receiving increased attention for their excellent physical and chemical properties and are promising materials for engineering applications. Amorphous alloy development differs from that of traditional materials, and the rapid determination of whether a given material is an amorphous alloy is critical in practical applications.

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	mendelev_number	Atom number in mendelev'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
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 295 K
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=0 K ground state
28	gs_energy	DFT energy per atom (raw VASP value) of T=0 K ground state
29	gs_est_bcc_latcnt	Estimated BCC lattice parameter based on the DFT volume
30	gs_est_fcc_latcnt	Estimated FCC lattice parameter based on the DFT volume
31	gs_mag_moment	DFT magnetic moment of T=0 K ground state
32	gs_volume_per	DFT volume per atom of T=0 K 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 20 °C
36	first_ion_en	First ionization 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

In this paper, the MCVN is proposed to predict the glass-forming ability of amorphous materials. The dataset was taken from the literature (Xiong et al., 2020), where BMG denotes bulk metallic glasses, RMG denotes ribbon metallic glass, and CRA denotes a crystalline alloy. The original characteristics in the dataset only include chemical composition characteristics of amorphous alloys, which contain 56 elements, such as Cu, Zr, and Ni. Because amorphous alloys are generally composed of only 4–5 elements, the composition data for each are very

sparse. The unbalanced classification dataset contains 6471 records, among which BMG have 1211 records, CRA has 1552 records, and RMG has 3708 records. The sample size of RMG was much larger than those of CRA and BMG.

The MCVN's experimental results for the amorphous alloy samples were compared with 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 parameters of all models

Table 10

Specific parameters of the models for the NIMS steel dataset regression experiments. The MCVN and Attention Network are built using pytorch-1.11. edRVFL_N is the model in (Shi et al., 2022). SVR, RF, Ridge are the models in scikit-learn-1.0.2. XGBoost is the model in xgboost-1.5.2.

Model	Parameters
MCVN	MCVN(learning_rate=0.008, epoch=80000, batch_size=256)
Attention Network	NN(learning_rate=0.01, epoch=80000, batch_size=256)
edRVFL_N	EnsembleDeepRVFL(n_nodes=100, lam_reg, w_random_vec_range=[-1, 1], b_random_vec_range=[0,1], activation='relu', n_layer=10, gamma_BN=1, alpha_BN=0, same_feature=False, task_type='regression')
Ridge	Ridge(alpha=0.1, max_iter=1000)
SVR	SVR(kernel='rbf', degree = 3, gamma='auto', C = 1.0)
RF	RandomForestRegressor(n_estimators=100)
XGBoost	XGBRegressor(max_depth=10, learning_rate=0.1, n_estimators=100)

Table 11

Experimental MAPE results of the steel material property prediction model for the NIMS steel dataset. *O*, *E*, and *E&P* are detailed in Table 3.

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		<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	3.82%	5.23%	6.14%	5.06%	7.06%	7.22%	8.55%	7.55%	7.22%	7.64%	9.13%	18.27%	9.80%	8.05%	9.47%	7.97%	6.64%	5.73%	6.56%
Tensile	1.41%	2.82%	3.48%	2.65%	5.31%	2.24%	3.98%	5.06%	4.56%	5.15%	6.23%	17.86%	6.97%	6.72%	6.72%	8.30%	4.81%	4.07%	5.98%
Fracture	3.82%	5.23%	4.81%	4.40%	10.88%	6.64%	11.29%	8.80%	7.06%	7.89%	7.89%	11.71%	7.72%	7.22%	4.90%	5.89%	6.23%	5.81%	5.31%
Hardness	1.66%	2.90%	3.82%	2.65%	6.39%	3.32%	6.14%	6.06%	4.98%	5.89%	7.72%	18.02%	8.39%	6.06%	5.48%	7.97%	5.39%	4.31%	5.73%

Table 12

Experimental MSE results of the steel material property prediction model for the NIMS steel dataset. *O*, *E*, and *E&P* are detailed in Table 3.

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		<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	289.98	543.92	750.44	509.93	990.13	1037.27	1453.88	1134.85	1037.27	1159.93	1658.22	6632.88	1908.18	1289.43	1781.01	1262.98	877.07	652.46	855.28
Tensile	558.96	2235.87	3411.83	1980.56	7922.25	1409.99	4456.26	7196.95	5850.78	7434.85	10879.5	89405.8	13647.3	12689.9	12689.9	19341.4	6506.46	4643.88	10026.6
Fracture	1149.14	2155.45	1826.89	1525.49	9319.68	3475.67	10044.6	6101.97	3923.70	4901.23	4901.23	10796.8	4697.04	4110.52	1890.43	2737.63	3054.78	2661.06	2224.42
Hardness	29.81	91.31	157.73	76.33	441.96	119.26	408.19	397.23	268.35	375.76	644.71	3510.13	760.40	397.23	324.70	686.98	314.94	201.56	354.89

Table 13

Specific parameters of the models for the SRIM steel dataset regression experiments. The MCVN and Attention Network are built using pytorch-1.11. edRVFL_N is the model in (Shi et al., 2022). SVR, RF, Ridge are the models in scikit-learn-1.0.2. XGBoost is the model in xgboost-1.5.2.

Model	Parameters
MCVN	MCVN(learning_rate=0.008, epoch=50000, batch_size=256)
Attention Network	NN(learning_rate=0.005, epoch=50000, batch_size=256)
edRVFL_N	EnsembleDeepRVFL(n_nodes=100, lam_reg, w_random_vec_range=[-1, 1], b_random_vec_range=[0,1], activation='relu', n_layer=10, gamma_BN=1, alpha_BN=0, same_feature=False, task_type='regression')
Ridge	Ridge(alpha=0.08, max_iter=1000)
SVR	SVR(kernel='rbf', degree = 3, gamma='auto', C = 1.0)
RF	RandomForestRegressor(n_estimators=100)
XGBoost	XGBRegressor(max_depth=15, learning_rate=0.08, n_estimators=100)

Table 14

Experimental MAPE results of the steel property prediction model on the SRIM steel dataset. *O*, *E*, and *E&P* are detailed in Table 3.

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		<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	2.16%	5.65%	7.31%	5.15%	5.98%	4.40%	4.98%	6.39%	4.23%	5.98%	9.05%	8.39%	9.97%	6.15%	4.40%	4.65%	7.64%	6.15%	6.48%
Tensile	1.66%	4.82%	6.31%	3.65%	3.57%	2.49%	3.65%	4.65%	3.40%	4.82%	6.81%	7.89%	10.13%	4.98%	3.49%	5.56%	7.06%	4.32%	4.15%
Elongation	4.73%	7.31%	8.56%	6.06%	13.04%	10.88%	10.63%	9.80%	19.11%	11.80%	10.38%	11.80%	11.71%	11.71%	11.63%	14.12%	14.71%	15.45%	17.20%
Cross-sectional shrinkage	13.71%	30.08%	33.82%	23.02%	22.52%	24.85%	20.36%	31.25%	33.57%	28.75%	37.89%	31.83%	34.74%	27.59%	36.90%	28.42%	28.59%	48.95%	32.82%

in the experiment are shown in Table 16, and the learning curve of the MCVN is shown in Fig. 16. The original sample ratio was retained in the test set, and the RMG dataset remained larger than those for CRA and BMG. The experimental results are presented in Table 6. XGBoost performed relatively well in terms of *Accuracy* but only achieved a *Recall* of 0.57 and *F1 - Score* of 0.68 on the CRA samples and did not distinguish well between them and the other two types of amorphous alloys. The MCVN achieved highly superior performance overall, obtaining the highest *F1 - Score* on the BMG and RMG sample size and an *F1 - Score* of 0.82 and *Recall* of 0.78 in the

classification of CRA. It distinguished well between the three types of amorphous alloy materials. This verifies that, in the classification task, the multimodal data are fused and complementary, allowing the MCVN to better classify each category of materials.

4.4. Ablation experiments

To demonstrate the efficacy of the MCVN's network structure and find the best possible architectural configuration, this study conducted

Table 15Experimental MSE results of the steel property prediction model on the SRIM steel dataset. *O*, *E*, and *E&P* are detailed in Table 3.

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFLN (Shi et al., 2022)			Ridge (Bottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestin, 2016)		
		<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	877.31	6001.03	10050.1	4988.74	6727.80	3645.52	4672.08	7694.66	3375.58	6727.80	15419.1	13238.8	18688.3	7106.75	3645.52	4069.90	10984.5	7106.75	7895.82
Tensile	566.76	4766.48	8184.06	2743.13	2619.86	1275.21	2743.13	4443.42	2381.82	4766.48	9527.29	12787.6	21089.2	5100.87	2499.42	6360.50	10237.1	3831.32	3542.27
Elongation	0.30	0.73	1.00	0.50	2.33	1.62	1.54	1.31	5.00	1.90	1.47	1.90	1.88	1.88	1.85	2.73	2.96	3.27	4.05
Cross-sectional shrinkage	26.02	125.25	158.33	73.34	70.19	85.45	57.37	135.13	156.01	114.43	198.75	140.21	167.01	105.35	188.43	111.80	113.11	331.60	149.13

Table 16

Specific parameters of the models for the BMG-RMG-CRA dataset classification experiments. MCVN and Attention Network are built by pytorch-1.11. edRVFLN is the model in Shi et al. (2022). RF is the bagging models in scikit-learn-1.0.2. XGBoost is the boost model in xgboost-1.5.2.

Model	Parameters
MCVN	MCVN(learning_rate=0.01, epoch=10000, batch_size=256)
Attention Network	NN(learning_rate=0.01, epoch=10000, batch_size=256)
edRVFLN	EnsembleDeepRVFL(n_nodes=100, lam_reg, w_random_vec_range=[-1, 1], b_random_vec_range=[0,1], activation='relu', n_layer=10, gamma_BN=1, alpha_BN=0, same_feature=False, task_type='classification')
RF	RandomForestClassifier(n_estimators=10)
XGBoost	XGBClassifier(max_depth=50, learning_rate=0.01, n_estimators=50, objective='reg:gamma')

modal and modular ablation experiments on the NIMS steel dataset with an evaluation metric of R^2 .

4.4.1. Modal ablation experiment

The results of the modal ablation experiment for the R^2 values are shown in Table 7. Here, the image channel indicates the image input using only the visualization of the materials' compositional features; the original channel indicates the original input of the dataset, including material compositions and preparation process; and the multimodal channel fuses material composition visualization images and original features. Note that the data of the image modal are obtained from the mapping of the component features in the original modal and do not contain information on heat treatment and process. Thus, the results obtained by using CNN to extract the information of image modalities are not satisfactory.

Fig. 10 presents the loss curves of the three network structures for learning optimization on the four target values in the NIMS steel dataset. The image channel model with image input converges the fastest. However, the accuracy rate decreases significantly owing to feature loss during the preparation process. Because of 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 of the four target values.

4.4.2. Module ablation experiment

The results of the R^2 module ablation experiment are listed in Table 8. Fig. 11 presents the loss curves of the three network structures for learning optimization on the four target values in the NIMS steel dataset.

Table 8 and Fig. 11 show the efficacy of the multi-headed self-attentive and BN modules 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 required for model convergence. However, adding BN may cause the model to oscillate owing to shrinkage of the gradient scale. Adding the multi-headed self-attentive module may significantly retard the model's training speed owing to increased computational effort.

5. Conclusions and future studies

This paper proposed a multimodal deep learning network structure (i.e., MCVN) that can better utilize material compositions to predict their properties. Sparse material compositional features are commonly

observed in material databases. Thus, the use of elemental statistical information to densify the sparse composition feature matrix has strong generalization. First, XenonPy converted material compositional features into statistical element-level features. According to the grayscale feature mapping scheme, a single modal dataset is augmented to be multimodal. 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. This is then extracted by using a multi-headed self-attentive model and a fully connected layer, followed by modal fusion with the modal image as the output. Due to the widespread availability of compositional features in the material dataset, the MCVN has excellent generalization capabilities.

Ablation experiments verified the necessity of having two modal and two network module structures to improve the accuracy of the MCVN 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 larger data volumes and manages more complexity.

5.1. Regression task

The regression task of our study involves the actual versus predicted regression. Traditional machine learning algorithms and the MCVN were compared on the NIMS steel dataset steel mechanical property prediction task. The MCVN achieved a notable improvement despite an average R^2 of 0.92 obtained by other models. We achieved R^2 values of 0.954 for fatigue strength, 0.983 for tensile strength, 0.954 for fracture strength, and 0.98 for hardness.

In the steel mechanical property prediction task of SRIM, the MCVN underwent 10 independent experiments, obtaining R^2 values of 0.974, 0.98, 0.943, and 0.835 for yield strength, tensile strength, elongation, and cross-section shrinkage, respectively, which is a 3% to 4% improvement in yield strength, tensile strength and elongation target values where other machine learning models performed well. In addition, the MCVN improved the R^2 result by 12% for the section-shrinkage target value, on which the other 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.

Table 17

Image arrangement shuffle experimental results of the steel materials' property prediction model for NIMS. Shuffle Image(shuffle block) indicates that the 56 blocks of the image arrangement (Fig. 3) are disrupted, and the generated images are arranged in the order of Table 9. Shuffle Image(shuffle zeros) indicates that the ordering of the statistical information in the 3*3 window is disturbed and arranged in the order of {0, sum, var, gmean, ave, hmean, max, min, 0}.

Image arrangement	Fatigue	Tensile	Fracture	Hardness
Shuffle Image(shuffle block)	0.937	0.966	0.937	0.965
Shuffle Image(shuffle zeros)	0.958	0.973	0.948	0.973
Schedule Image	0.954	0.983	0.954	0.98

5.2. Classification task

The experiments based on amorphous alloy data show that the $F1 - Score$ of the MCVN is significantly improved compared with the other machine learning models in the classification task. The predicted $F1 - Score$ reached 0.917 for BMG, 0.921 for RMG, and 0.818 for 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 0.778 *Recall* in CRA which is the most critical category in the dataset. This experiment verified that, by visualizing the elemental-level features of material chemistry, multimodal data can complement the enhanced unimodal data using only chemical compositional features.

5.3. Future work

Because of the universality of material composition information and its expansion flexibility, the MCVN has excellent scalability and generalization capabilities:

- The arrangement of the expanded image is very flexible. The effect of image arrangement on the experimental results is discussed in Table 17. Evidently, the arrangement of 56 blocks with expert knowledge has a positive effect on the experimental results. The introduction of mapping rules with more prior knowledge of the material can be considered in the future; with more information and density, the mapped data is expected to achieve better prediction performance.
- The information expanded from the compositions can be classified according to self-defined rules, divided into vectors of multiple channels, and fed into the model for learning, eventually designing a fusion network to predict material properties.
- In addition to 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. The combination of other material data (e.g., material microstructure and texture images) can be considered with gray-scale map data.
- The molecular structure and spatial structure of materials can also be used as graph structure data, which, combined with graph neural networks, will enhance the prediction capability of multimodal deep neural networks.

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.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported 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. 202102AB080019-3, 202002AB080001-2), and Key Project of Shanghai Zhangjiang National Independent Innovation Demonstration Zone (No. ZJ2021-ZD-006). The authors gratefully appreciate the anonymous reviewers' valuable comments.

Appendix

See Figs. 12–16 and Tables 9–17.

References

- Agrawal, A., Choudhary, A., 2018. An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. *Int. J. Fatigue* 113, 389–400.
- Baltrušaitis, T., Ahuja, C., Morency, L.-P., 2018. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (2), 423–443.
- Ban, H., Shi, G., Shi, Y., Wang, Y., 2008. Research on the force performance of steel structures made of ultra-high strength steel(in Chinese). In: *Research on Steel Structural Engineering(Volume 7)–Proceeding of 2008 Colloquium Institute of Structural Stability and Fatigue*.
- Ban, H.Y., Shi, G., Shi, Y.J., Wang, Y.Q., 2011. Research progress on the mechanical property of high strength structural steels. *Adv. Mater. Res.* 250, 640–648.
- Bellman, R., 1957. *Dynamic Programming*, Princeton Univ. Press Princeton, New Jersey.
- Botou, L., 2010. Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. Springer, pp. 177–186.
- Breiman, L., 2015. Random forests leo breiman and adele cutler. In: *Random Forests-Classification Description*. Vol. 106.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 1–27.
- 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.
- Dai, M., Demirel, M.F., Liang, Y., Hu, J.-M., 2021. Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *Npj Comput. Mater.* 7 (1), 1–9.
- 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.
- Fernández, J., Chiachío, M., Chiachío, J., Muñoz, R., Herrera, F., 2022. Uncertainty quantification in neural networks by approximate Bayesian computation: Application to fatigue in composite materials. *Eng. Appl. Artif. Intell.* 107, 104511.
- Gang, S., Hui-yong, B., Yong-jiu, S., Yuan-qing, W., 2013. Overview of research progress for high strength steel structures. *Eng. Mech.* 30 (1), 1–13.
- 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 Comput. Mater.* 6 (1), 1–10.
- 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.
- Jiang, Y., Chen, D., Chen, X., Li, T., Wei, G.-W., Pan, F., 2021. Topological representations of crystalline compounds for the machine-learning prediction of materials properties. *Npj Comput. Mater.* 7 (1), 1–8.
- Jolliffe, I.T., Morgan, B., 1992. Principal component analysis and exploratory factor analysis. *Stat. Methods Med. Res.* 1 (1), 69–95.
- 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 (2), 262.
- 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. *Comput. Methods Appl. Mech. Engrg.* 347, 735–753.
- Li, D.Z., Wang, W., Ismail, F., 2017. A regulated boosting technique for material fatigue property prognostics. *Eng. Appl. Artif. Intell.* 57, 153–159.
- 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. *Adv. Mater.* 33 (36), 2102507.

- Ma, M., Sun, C., Chen, X., 2018. Deep coupling autoencoder for fault diagnosis with multimodal sensory data. *IEEE Trans. Ind. Inform.* 14 (3), 1137–1145.
- Meel, P., Vishwakarma, D.K., 2021. HAN, image captioning, and forensics ensemble multimodal fake news detection. *Inform. Sci.* 567, 23–41.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., Ng, A.Y., 2011. Multimodal deep learning. In: *ICML*.
- OlivierGrisel, 2022. *Scikit-learn*. <https://scikit-learn.org>.
- 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. *Math. Geosci.* 40 (4), 409–424.
- Pilania, G., 2021. Machine learning in materials science: From explainable predictions to autonomous design. *Comput. Mater. Sci.* 193, 110360.
- Politis, I., Brewster, S., Pollick, F., 2017. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. *Int. J. Mob. Hum. Comput. Interact. (IJMHCI)* 9 (3), 1–16.
- Rhoades, S.A., 1995. Market share inequality, the HHI, and other measures of the firm-composition of a market. *Rev. Ind. Organ.* 10 (6), 657–674.
- Schütt, K.T., Arbabzadah, F., Chmiela, S., Müller, K.R., Tkatchenko, A., 2017. Quantum-chemical insights from deep tensor neural networks. *Nature Commun.* 8 (1), 1–8.
- Schütt, K.T., Sauceda, H.E., Kindermans, P.-J., Tkatchenko, A., Müller, K.-R., 2018. SchNet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* 148 (24), 241722.
- Shi, Q., Hu, M., Suganthan, P.N., Katuwal, R., 2022. Weighting and pruning based ensemble deep random vector functional link network for tabular data classification. *Pattern Recognit.* 132, 108879.
- Shi, Q., Katuwal, R., Suganthan, P.N., Tanveer, M., 2021. Random vector functional link neural network based ensemble deep learning. *Pattern Recognit.* 117, 107978.
- 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.
- Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W., Zafeiriou, S., 2017. End-to-end multimodal emotion recognition using deep neural networks. *IEEE J. Sel. Top. Sign. Proces.* 11 (8), 1301–1309.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Versaci, M., Calcagno, S., Morabito, F.C., 2015. Image contrast enhancement by distances among points in fuzzy hyper-cubes. In: *International Conference on Computer Analysis of Images and Patterns*. Springer, pp. 494–505.
- Wei, J., Chu, X., Sun, X.-Y., Xu, K., Deng, H.-X., Chen, J., Wei, Z., Lei, M., 2019. Machine learning in materials science. *InfoMat* 1 (3), 338–358.
- 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 Trans. Ind. Inform.*.
- 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. *Inform. Sci.* 593, 385–397.
- Wu, S., Lambard, G., Liu, C., Yamada, H., Yoshida, R., 2020a. iQSPR in xenonpy: a bayesian molecular design algorithm. *Molecular Inform.* 39 (1–2), 1900107.
- Wu, X., Zhong, M., Guo, Y., Fujita, H., 2020b. The assessment of small bowel motility with attentive deformable neural network. *Inform. Sci.* 508, 22–32.
- Xiong, J., Shi, S.-Q., Zhang, T.-Y., 2020. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. *Mater. Des.* 187, 108378.
- yoshida-lab, 2019. *Xenonpy*. <https://xenonpy.readthedocs.io>.
- 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. *Inform. Sci.* 581, 304–326.
- Zheng, W., Liu, H., Wang, B., Sun, F., 2019. Cross-modal surface material retrieval using discriminant adversarial learning. *IEEE Trans. Ind. Inform.* 15 (9), 4978–4987.