# Response to the Reviewers' comments:

## Reviewer #1

The paper aims to address the lack of interpretability when using DL methods for ML for materials. This is indeed an important problem. However, based on the submitted work I cannot see if this point was properly addressed or how the work has contributed to solving this issue. The combination of LSTMS for global features and CNNs for local features is an interesting aspect of this paper that the authors could have emphasized more. I fail to understand how the novel CECV distinctly helps to rationalize model behavior. For instance, figure7, shows the bandgap dependence on the composition for different models but the same plots can be generated with Crabnet (as shown in the figures). Finally, the trained embedding vectors are visualized using UMAP which indeed reflects the underlying systems in the way the feature space is clustered. However, UMAP visualizations are to some degree arbitrary in that the distance between clusters - as a reference in the text - has little meaning. A method that truly gives insights into physics should allow extraction of easily interpretable low-dimensional parameters, for instance, a parameter describing pairwise elemental interactions that can be compared between different structural environments.

The main issue of this work is the quality of the presentation.

It is very hard to follow the scientific contribution of this work. For instance, abbreviations should not be introduced in the abstract but explained with references in the introduction. The term "fusion" is also a bit confusing, using the word "combination" would suffice. Another point is the use of many abbreviations without proper explanation "XRD, XPS, SEM..." and the same for the many NN architectures referenced in this work.

I would suggest creating a table that gives a nice overview of the different architectures used, which possibly could be included in Figure 1.

The term "different scales" is first used in the abstract but is not even explained there are "compounds at different scales" - it is not clear to me what is meant by that it is not explained later in the text.

Finally, I encourage the authors to continue working on the interpretability of DL because it is indeed an important issue and I hope the authors are not discouraged based on my rejection. Still based on the submitted work, I did not see a solution to the problem that is claimed to be addressed. To improve the quality of the paper I would recommend more focus on the presentation of the actual novelty of the work and a substantial improvement in terms of structure and form of the text.

Some typos/recommendations:

Figure 4 left: x-axis "training set size", generally too small axis labels and visibility

**Response:** We are appreciate for the referee's questions and criticism of our work. The referee's comments include several questions, which we have organized into six questions and provided responses in a corresponding sequence.

**1.** The paper aims to address the lack of interpretability when using DL methods for ML for materials. This is indeed an important problem. However, based on the submitted work I cannot see if this point was properly addressed or how the work has contributed to solving this issue.

**Response:** Thank you for the referee's questions. We understand your concern about the interpretability of the model. In our work, we did attempt some visualization techniques to explain the relationship between model representations and chemical environments as well as periodic trends. These visualization techniques are aimed at providing an intuitive way to understand the key factors in predicting material properties. Clearly, these visualization techniques can only provide some macroscopic periodic trends and may not meet the requirement you proposed: "A method that truly gives insights into physics should allow extraction of easily interpretable low-dimensional parameters, for instance, a parameter describing pairwise elemental interactions that can be compared between different structural environments." However, it is well known that interpretability is a challenging problem for deep learning models. Here, we provide some tentative explanations. We found that the L-G-DCNN based on the fusion strategy can be decomposed into LSTM and DCNN through a process similar to dismantling building blocks, thereby providing more information about the process of model performance improvement from "poor" to "good". For example, based on the OQMD band gap test dataset, we first filtered out 56 compounds containing Si elements that performed very poorly in the LSTM compared to the L-G-DCNN. As shown in Figure 1, a comparison of the predicted band gap values between LSTM and L-G-DCNN and the corresponding chemical formula labels is provided. Then, we compared the EFM plots of these 56 compounds based on LSTM and L-G-DCNN to obtain (LSTM) - (L-G-DCNN) ΔEFM, as shown in three examples in Figure 2. Through comparison, we can determine which elements interactions are most important for the band gap property. Finally, in Figure 3, we summarize the element interaction pairs that have the greatest impact on the band gap among these 56 compounds. From Figure 3, it can be seen that most of the element interaction pairs contain elements with high electronegativity such as oxygen (O), chlorine (Cl), and fluorine (F), which are very conducive to forming strong covalent bonds, and this has a significant impact on the band gap, which is consistent with our physical intuition, indicating that L-G-DCNN is a high effective deep learning model for predicting material properties from chemical composition data. Table 1 provides the true values of the 56 selected compounds, as well as all the predicted values, differences, and other raw data based on LSTM and L-G-DCNN.

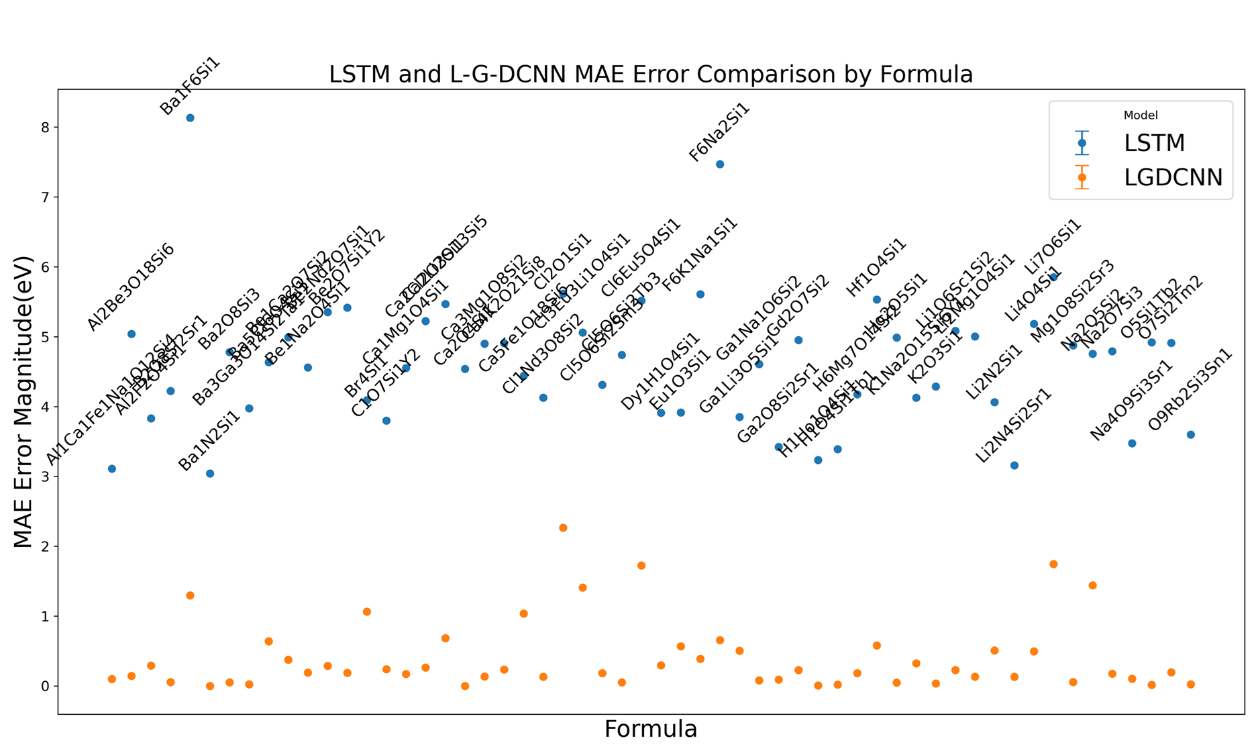


Figure 1 illustrates the comparison of predicted values for 56 Si-containing compounds with significant differences between LSTM and L-GDCNN on the OQMD band gap test dataset, with the chemical formula of each compound labeled in the figure.

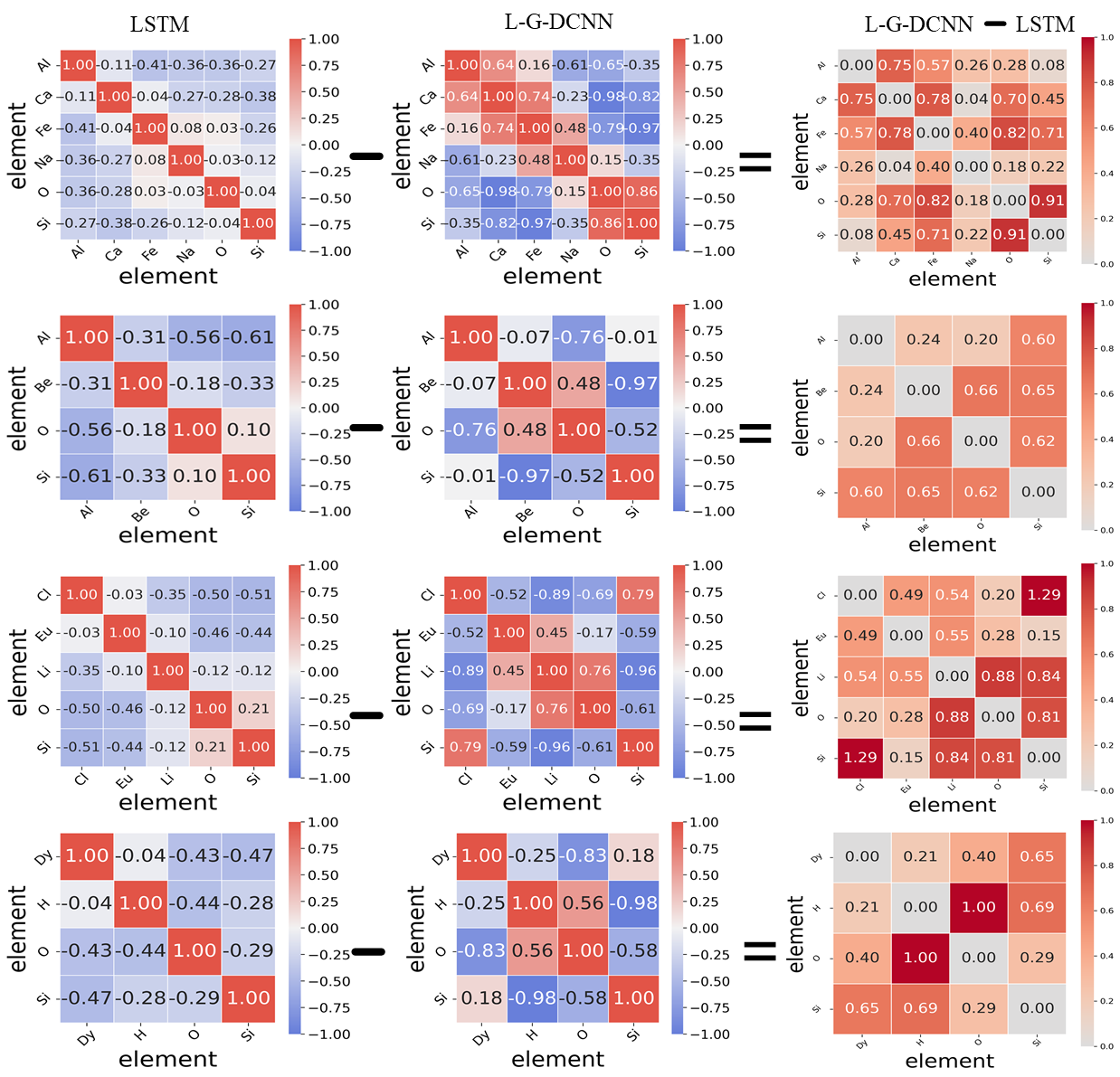


Figure 2 presents a comparison of the EFM plots of LSTM and L-G-DCNN for three example Si-containing compounds, along with the (LSTM)-(L-G-DCNN) ΔEFM. Through ΔEFM, we can determine which element interactions are most important for the band gap. The shading of ΔEFM reflects the significance of the element interactions pairs for the band gap.

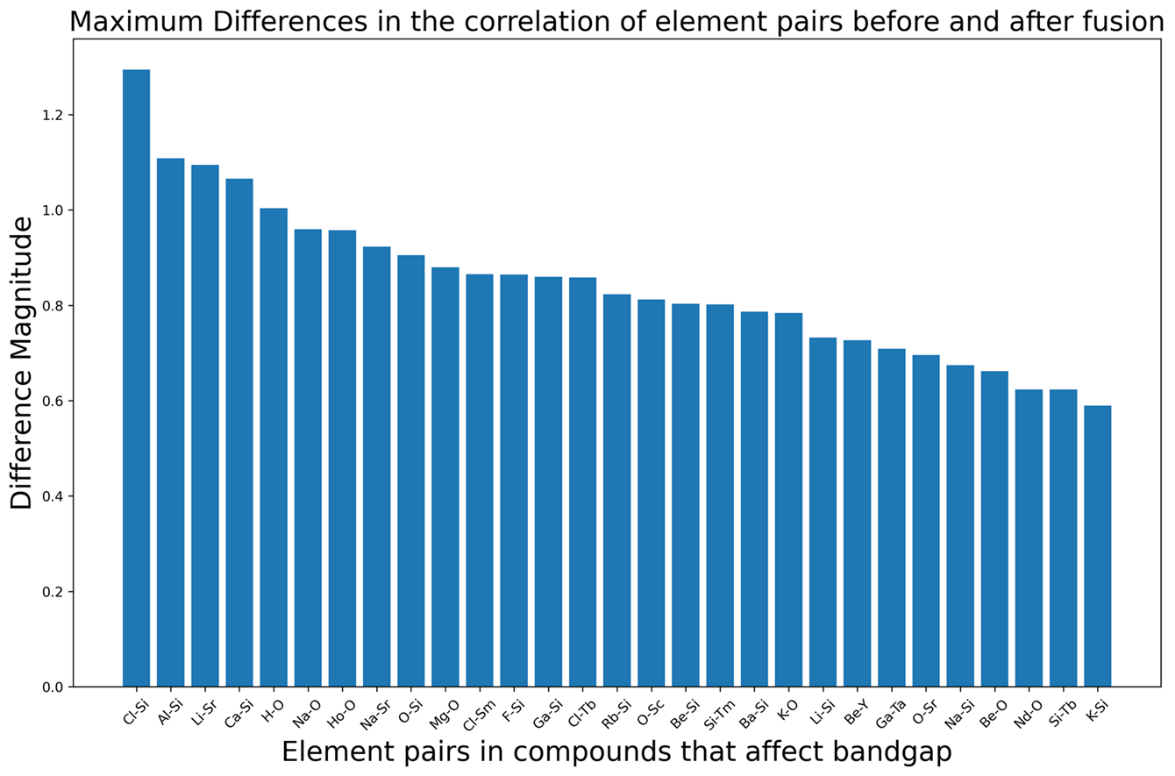


Figure 3 summarizes the most influential element interactions pairs on the band gap among the 56 compounds before and after model fusion. From the figure, it can be observed that most element interactions pairs involve elements with high electronegativity such as O, Cl, and F.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Formula | target | lstm\_pred | lgdcnn\_pred | error\_lstm | error\_lgdcnn |
| Na2O5Si2 | 4.754 | -0.00023 | 3.313114 | 4.754229 | 1.440886 |
| Cl1Nd3O8Si2 | 4.638 | 0.512336 | 4.771503 | 4.125664 | 0.133503 |
| Ca5Fe1O18Si6 | 4.444 | 0.003013 | 3.406679 | 4.440987 | 1.037321 |
| Hf1O4Si1 | 5.601 | 0.068076 | 5.022842 | 5.532924 | 0.578158 |
| F6Na2Si1 | 7.943 | 0.475034 | 8.59861 | 7.467966 | 0.65561 |
| Cl5O6Si2Sm3 | 4.711 | 0.399255 | 4.894914 | 4.311745 | 0.183914 |
| Ca4K2O21Si8 | 4.967 | 0.059111 | 4.729878 | 4.907889 | 0.237122 |
| Ca2Cl2O3Si1 | 5.382 | 0.159295 | 5.117125 | 5.222705 | 0.264875 |
| Cl2O1Si1 | 5.638 | 0.021166 | 3.371808 | 5.616834 | 2.266192 |
| Ga2O8Si2Sr1 | 4.133 | 0.711203 | 4.226184 | 3.421797 | 0.093184 |
| Be1Ca2O7Si2 | 5.221 | 0.228394 | 4.84486 | 4.992606 | 0.37614 |
| Cl6Eu5O4Si1 | 5.534 | 0.017281 | 3.808146 | 5.516719 | 1.725853 |
| Li4O4Si1 | 5.195 | 0.012667 | 4.697077 | 5.182334 | 0.497923 |
| K2O3Si1 | 4.361 | 0.075119 | 4.39543 | 4.285881 | 0.03443 |
| Li2N2Si1 | 4.071 | 0.007891 | 3.563954 | 4.063109 | 0.507046 |
| Mg1O8Si2Sr3 | 5.106 | 0.232572 | 5.049438 | 4.873428 | 0.056562 |
| Be1Na2O4Si1 | 4.872 | 0.313902 | 5.063966 | 4.558098 | 0.191966 |
| Ga1Li3O5Si1 | 3.886 | 0.035373 | 4.389283 | 3.850627 | 0.503283 |
| Ca1Mg1O4Si1 | 4.988 | 0.438792 | 4.817697 | 4.549208 | 0.170303 |
| Na4O9Si3Sr1 | 4.594 | 1.121436 | 4.699292 | 3.472564 | 0.105292 |
| H6Mg7O14Si2 | 5.13 | 0.954281 | 5.314725 | 4.175719 | 0.184725 |
| O7Si2Tm2 | 4.915 | 0.004996 | 5.109752 | 4.910004 | 0.194752 |
| Li7O6Si1 | 5.851 | -0.00028 | 4.106517 | 5.851278 | 1.744483 |
| Ba5Cl6O4Si1 | 4.742 | 0.108709 | 4.101628 | 4.633291 | 0.640372 |
| Ga1Na1O6Si2 | 4.653 | 0.046422 | 4.734563 | 4.606578 | 0.081563 |
| O5Si1Tb2 | 4.921 | -0.00034 | 4.935651 | 4.921344 | 0.014651 |
| Ca2O4Si1 | 4.541 | 0.003101 | 4.540303 | 4.537899 | 0.000697 |
| Ca3Mg1O8Si2 | 5.19 | 0.290707 | 5.052417 | 4.899293 | 0.137583 |
| Ba1F6Si1 | 8.183 | 0.049109 | 6.885648 | 8.13389 | 1.297352 |
| Ba3Ga3O14Si2Ta1 | 4.108 | 0.131982 | 4.08351 | 3.976018 | 0.02449 |
| Ca2Li2O13Si5 | 5.582 | 0.115343 | 4.898315 | 5.466656 | 0.683685 |
| H1Ho1O4Si1 | 4.853 | 1.62065 | 4.862513 | 3.23235 | 0.009513 |
| O9Rb2Si3Sn1 | 4.262 | 0.6632 | 4.237721 | 3.5988 | 0.024279 |
| Li2Mg1O4Si1 | 5.109 | 0.104453 | 5.24047 | 5.004547 | 0.13147 |
| C1O7Si1Y2 | 5.051 | 1.251145 | 4.811564 | 3.799855 | 0.239436 |
| Na2O7Si3 | 4.798 | 0.007625 | 4.621844 | 4.790375 | 0.176156 |
| K1Na2O15Si6Y1 | 5.058 | 0.932117 | 4.734559 | 4.125884 | 0.323441 |
| Cl5O6Si2Tb3 | 4.83 | 0.089065 | 4.777794 | 4.740934 | 0.052206 |
| Ba2O8Si3 | 4.812 | 0.034359 | 4.863309 | 4.777641 | 0.051309 |
| Ba1N2Si1 | 3.042 | 0.002042 | 3.040886 | 3.039958 | 0.001114 |
| Al2Be3O18Si6 | 5.594 | 0.556126 | 5.450961 | 5.037874 | 0.143039 |
| Eu1O3Si1 | 3.915 | 0.001533 | 3.348245 | 3.913467 | 0.566755 |
| Dy1H1O4Si1 | 4.861 | 0.951677 | 5.156366 | 3.909323 | 0.295366 |
| Cl3Eu3Li1O4Si1 | 5.612 | 0.550583 | 4.203979 | 5.061417 | 1.408021 |
| F6K1Na1Si1 | 7.791 | 2.18319 | 8.178376 | 5.60781 | 0.387376 |
| Be2Nd2O7Si1 | 5.367 | 0.013458 | 5.654284 | 5.353542 | 0.287284 |
| Li2N4Si2Sr1 | 3.761 | 0.603332 | 3.628573 | 3.157667 | 0.132427 |
| B2O8Si2Sr1 | 5.959 | 1.734906 | 5.902301 | 4.224094 | 0.056699 |
| H1O4Si1Tb1 | 4.784 | 1.394177 | 4.802793 | 3.389823 | 0.018793 |
| Gd2O7Si2 | 4.95 | -0.00023 | 4.723457 | 4.950231 | 0.226542 |
| Ho2O5Si1 | 4.986 | -0.00051 | 4.936392 | 4.986513 | 0.049608 |
| Al1Ca1Fe1Na1O12Si4 | 3.394 | 0.2837 | 3.492052 | 3.1103 | 0.098052 |
| Li1O6Sc1Si2 | 5.087 | 0.005365 | 4.860103 | 5.081635 | 0.226897 |
| Al2F2O4Si1 | 6.35 | 2.517638 | 6.058063 | 3.832362 | 0.291937 |
| Be2O7Si1Y2 | 5.447 | 0.031457 | 5.258036 | 5.415543 | 0.188964 |

Table 1 provides the actual values of the filtered 56 compounds, as well as all the predicted values, differences, and other raw data based on LSTM and L-G-DCNN.

**Manuscript Update:** In our revised manuscript, we have added a new section titled "Identifying critical elemental interactions pairs for band gap prediction" on page 20, and have included Figure 11, which is highlighted in blue font.

**2.** I fail to understand how the novel CECV distinctly helps to rationalize model behavior. For instance, figure 7, shows the bandgap dependence on the composition for different models but the same plots can be generated with Crabnet (as shown in the figures).

**Response:** We appreciate the referee’s questions and criticism on our work. In response to your questions regarding how CECV contribute to a rational explanation of model behavior. We provide a comprehensive explanation here. Regarding the question of how CECV contributes to a rational explanation of model behavior, we will systematically explain the process of EFM and CECV acquisition and their guidance in the iterative design of the model, as illustrated in Figure 4:

**1. Selection and Initialization of Model Components**

Select model components from the Common Model Library, including global models (such as LSTM, GRU1, RNN2) and local models (such as DCNN, DCNN-K1, CNN, TextCNN3, ResNet4), as well as attention and pooling components. Initialize the composition model, such as the fusion of LSTM and DCNN ,(LDCNN), and set linear and activation layers for post-processing.

**2. Model Training and Testing**

Based on the OQMD enthalpy dataset, use the selected model components (such as LSTM, DCNN-K1, DCNN, LDCNN) for training and validation. Divide the dataset into training, validation, and test sets in a 70/15/15 ratio, and train all models under unified parameter settings.

**3. Extraction of Compound EFM**

For 2374 compounds containing silicon elements in the test dataset of OQMD enthalpy (covering binary to heptenary compounds), use the trained models (LSTM, DCNN-K1, DCNN, LDCNN) to extract the element embedding feature matrix (EFM) of the compounds. Each compound corresponds to a 512-dimensional EFM. It is worth noting that we only use compounds containing silicon elements as an example here, but other element compounds can also be set, or compounds with different elements can be randomly selected.

**4. CECV Generation**

Based on the extracted EFM, cluster all Si-containing compounds, set different clustering threshold ranges as [0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4], and then generate a set of CECV, recording the number of clusters corresponding to each threshold under different models. For example, LSTM obtains the clustering vector within the set threshold: [376, 147, 80, 48, 31, 27, 16, 14].

**5. Visualization and Definition of Target Region**

Plot the graph, with the clustering threshold on the x axis and the number of CECVs obtained by each model on the y axis, showing the performance of different models under different clustering thresholds. Mark the LSTM and DCNN-K1 as global and local reference model, defining the target region between the LSTM and DCNN-K1. This step helps to intuitively understand the tendencies and trade-offs of the model in global and local feature extraction. For example, as shown in Figure 5, the two reference models represented by the red and green dashed lines respectively demonstrate the CECVs of LSTM and DCNN-K1 for Si-containing compounds in the OQMD enthalpy dataset under different thresholds. Based on the characteristics of LSTM and DCNN-K1, they have different requirements for CECV at different thresholds. For the global feature extractor LSTM, its performance is close to optimal within the threshold <0.15, as a smaller threshold leads to lower tolerance for distinguishing the chemical environment of compounds, requiring higher demands, causing the model to overly focus on local details and easily overestimate CECV. Therefore, it is necessary to increase the weight of the global feature extractor to adjust the model and achieve more accurate identification of the clustering of compound chemical environments, as indicated by the downward red dashed arrow on the left side of Figure 5. Similarly, for the local feature extractor DCNN-K1, its performance is close to optimal within the threshold >0.15, as a larger threshold leads to higher tolerance for distinguishing the chemical environment of compounds, requiring lower demands, causing the model to overly focus on global information and easily underestimate CECV. Therefore, it is necessary to increase the weight of the local feature extractor to adjust the model, as indicated by the upward red dashed arrow on the right side of Figure 5. Based on the above strategy, we can draw the target region, which is marked as a shaded area in Figure 5. With the understanding of Figure 5, we can refer to the focusing principle of a camera:

1. Global feature extraction is similar to a wide-angle lens of a camera. The global feature extractor (such as LSTM) can capture a broader field of view, i.e., the overall chemical environment of compounds.

2. The local feature extractor (such as DCNN-K1) is similar to a telephoto lens of a camera, focusing on capturing details of local regions.

3. The target region can be seen as the best focus point of a camera, where the image contains rich details while retaining overall clarity.

4. Model adjustment is similar to adjusting the focal length of a camera to obtain the best image. By increasing the weight of global or local model components, the model can be adjusted to better adapt to the recognition requirements of chemical environments.

**6. Iterative Optimization of Model Components**

As shown in Figures 4 and 5, identify the deviation between the current model's CECV and the target region (marked as a shaded area in Figure 5), and decide whether to strengthen the weight of the global feature extractor or the local feature extractor weight, guiding targeted adjustments of model components from the model library to make the CECV of the new model closer to the target region, achieving rational optimization of model design. To understand how CECV guides the rational modification of the model structure, based on Figure 5, we provide further explanation. Compared with the two reference models, the black line of LDCNN performs better in fusion. Within the threshold range <0.15, the CECV of LDCNN is significantly lower than that of DCNN-K1, very close to the shaded area. However, within the threshold range >0.15, its improvement in CECV is limited, with a significant deviation from the shaded area, resulting in mediocre performance of LDCNN. Guided by Figure 5, we determine that the key factors lies in increasing a model component that can increase the weight of the LSTM end in LDCNN. Therefore, a reasonable design 1 is proposed, introducing soft attention5 and residual connections to optimize the LSTM part of LDCNN, resulting in L-ATT-RES-DCNN, as indicated by the cyan arrow in Figure 5. However, L-ATT-RES-DCNN achieves an increase in CECV across the entire threshold range, indicating that introducing soft attention and residual connections to the LSTM part does not effectively increase the weight of the LSTM end. Therefore, a reasonable design 2 is proposed, introducing stacked residual GRU and softmax attention6 to obtain L-G-DCNN. This design significantly reduces the CECV within the threshold range <0.15 and greatly improves the performance of L-G-DCNN, as indicated by the yellow arrow in Figure 5. We can see that within the threshold range <0.15, L-G-DCNN still has room for improvement. Therefore, a reasonable design 3 is proposed, introducing residual connections in L-G-DCNN to obtain L-G-DCNN-V1, as indicated by the green arrow in Figure 5. L-G-DCNN-V1 further reduces the CECV within the threshold range <0.15, entering the target shaded area and achieving optimization. However, there is also a decrease in CECV within the threshold range >0.15. Overall, the target region has not been fully reached. From LDCNN to L-G-DCNN-V1, based on CECV, we have improved the performance of the model through three design steps, achieving a rational and high efficient design.

CECV is a vector representation designed by us to quantify the clustering distribution of compound chemical environments. In materials science, even compounds with the same composition may exhibit significant differences in their physical and chemical properties due to variations in their chemical environments. Through cluster analysis, CECV groups compounds with similar chemical environments into a cluster vector, reflecting the clustering distribution of chemical environments for different compounds. This provides us with a quantitative method for evaluating the model's sensitivity to chemical environments. For example, in Figure 6, we compared the dependency between band gap and composition for L-G-DCNN and Crabnet. By comparing Figures 6(a)(b), and 6(c)(d), it is evident that the predictions of L-G-DCNN align better with DFT calculations. Specifically, there are two larger band gap values between 0 and 0.2 and between 0.2 and 0.4, whereas Crabnet's predictions deviate significantly from the actual computed values between 0 and 0.2, 0.4 and 0.6, and 0.6 and 0.8. This confirms that L-G-DCNN outperforms Crabnet in predicting the composition-band gap dependency. Combining the results in Figure 6, we use the results of CECV to explain the model's predictive behavior. As shown in Figure 7, the CECV obtained from the EFM of Si-containing compounds at different thresholds is displayed. CECV can reflect the clustering distribution of chemical environments for all compounds. If a model can accurately identify the clustering distribution of chemical environments, or in other words, if its CECV can approximate the shaded area in Figure 7, it can more accurately predict band gap values. Based on this, we calculated the CECV for L-G-DCNN and Crabnet and observed their distribution in Figure 7(b). This demonstrates that Crabnet's poor performance in Figure 6 is mainly due to its significant overestimation of the clustering distribution of compound chemical environments across the entire threshold range. The role of CECV is to help us understand how the differences between the predicted band gap and the actual band gap are related to the compound's chemical environment. By analyzing CECV, we can identify in which chemical environments the model performs poorly, allowing us to improve the model in a targeted manner.

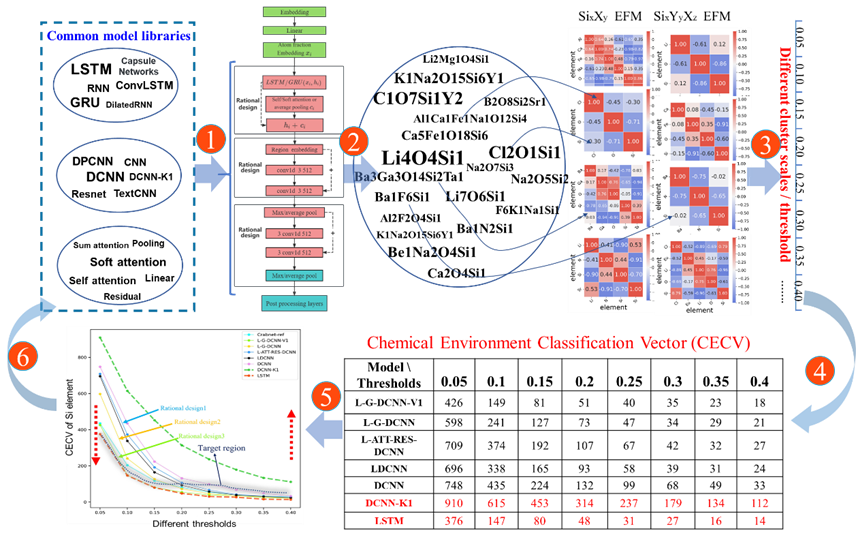


Figure 4 presents the process of obtaining EFM and how, based on this, the CECVs of all models before and after fusion under different thresholds are obtained, thereby achieving rational model design guided by CECV.

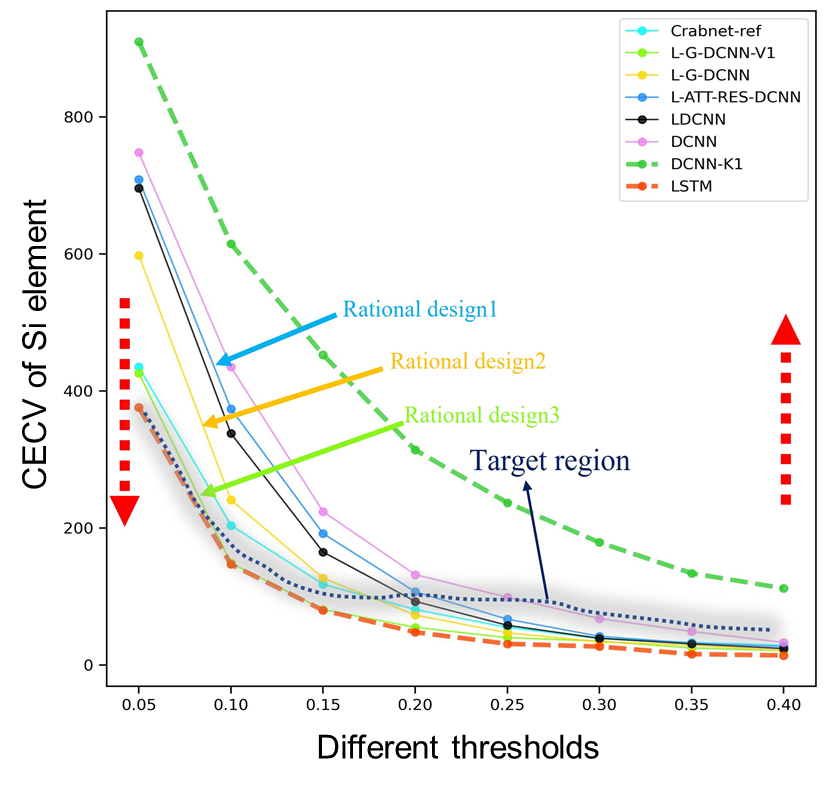


Figure 5 shows the CECV of silicon-containing compounds in the OQMD Formation enthalpytest datasets based on the fusion strategy, using LSTM and DCNN-K1 as global and local reference models. CECV were used as the feedback method to achieve rational model design. The shaded region in the figure marks the CECV of the optimal fusion model. The blue, yellow, and green solid arrows mark the rational design from LDCNN to L-G-DCNN-V1. Two red dashed arrows represent the rational design direction of the model at the LSTM and DCNN-K1 ends. Crabnet-ref is a reference model, which performs best on the OQMD Formation enthalpy.

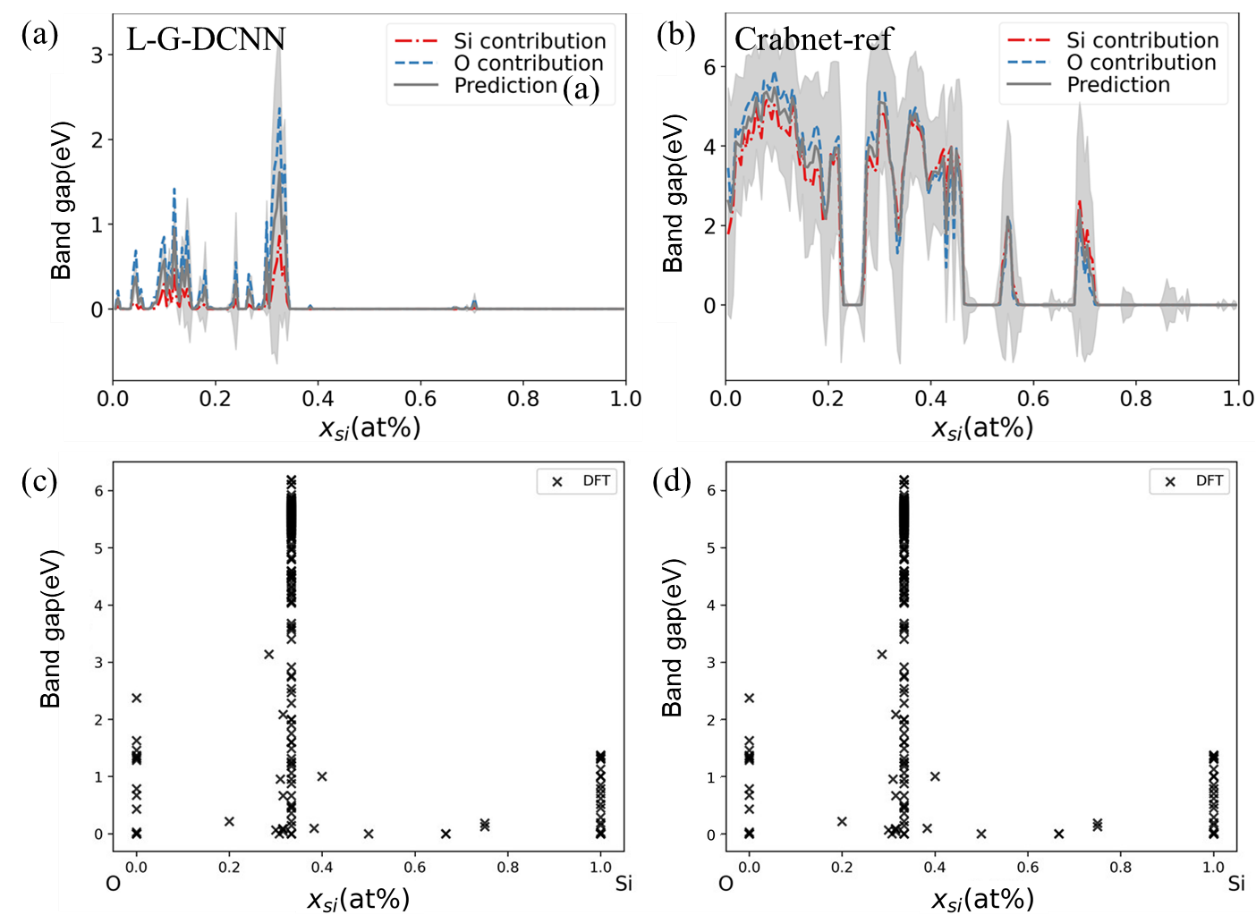
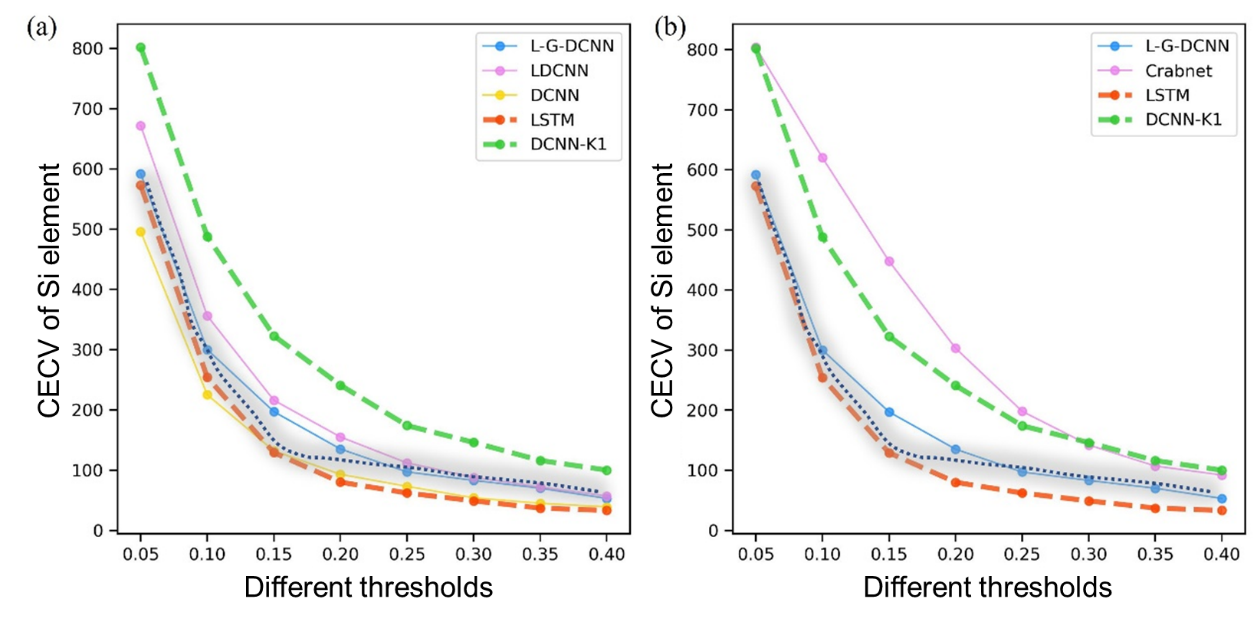


Figure 6 (a) and (b) shows the predictions of L-G-DCNN and Crabnet trained on OQMD band gap data for the Si*x*O1−*x* system. The *x*-axis represents the fractional amount of Si, and the *y*-axis represents the predicted band gap value at a given composition. The blue and red lines represent the individual element contributions to the prediction. The gray shading represents the aleatoric uncertainty for each prediction. (c) and (d) show the DFT calculation results for the Si*x*O1−*x* system queried from the OQMD database.



On the OQMD bandgap dataset, Figure 7(a) shows the CECV of silicon-containing compounds at different thresholds obtained by DCNN-K1, LSTM, DCNN, LDCNN, and L-G-DCNN. (b) CECV of silicon-containing compounds at different thresholds obtained by L-G-DCNN and Crabnet. The shaded region with dashed lines is the target reference region.

**Manuscript Update:** In the "Rational model design" section of our revised manuscript, on pages 4 and 5, we have updated Figure 1 and the explanations to the aforementioned questions, highlighted in blue font.

**3.** Finally, the trained embedding vectors are visualized using UMAP which indeed reflects the underlying systems in the way the feature space is clustered. However, UMAP visualizations are to some degree arbitrary in that the distance between clusters - as a reference in the text - has little meaning.

**Response:** We thank the referee for this question. We understand your concerns about the interpretability of Uniform Manifold Approximation and Projection7 (UMAP) visualizations. You pointed out that “the trained embedding vectors are visualized using UMAP which indeed reflects the underlying systems in the way the feature space is clustered. However, UMAP visualizations are to some degree arbitrary in that the distance between clusters.” In response to your question, we conducted an in-depth analysis and discussion of UMAP visualizations. UMAP is a dimensionality reduction technique that maps high-dimensional data to a lower-dimensional space for visualization. In this process, UMAP tends to preserve both global and local structures of the data, keeping similar data points within clusters close together in the reduced space, while points from different clusters are relatively far apart, reflecting the clustering characteristics of the original data. Although the two- or three-dimensional representations of UMAP cannot directly map back the distances in the original high-dimensional space, they do preserve the relative distances between data points, allowing for inference of potential similarities and differences in the original data based on the relative proximity of clusters. However, this inference is not equivalent to direct physical distances.

We extracted the EFM of compounds containing Si from the L-G-DCNN trained on the OQMD band gap dataset, which already encapsulate the learned feature representations of the L-G-DCNN. If the L-G-DCNN has learned the clustering of chemical environments containing Si compounds during training, these EFMs will reflect these clusters. By applying UMAP to these EFMs for dimensionality reduction, a two-dimensional visualization can be obtained, showcasing the relative positions of data points in the original high-dimensional space and indirectly revealing the clustering results. As shown in Figure 8, with the fusion of models, the predicted results on the OQMD band gap dataset become increasingly closer to the DFT values, with the red circles marking the inaccurate predictions of Si-containing compounds by the LSTM before fusion. Figure 8 demonstrates that with model fusion, the L-G-DCNN not only achieves accurate band gap predictions but also accurately captures the chemical environments of Si-containing compounds, effectively distinguishing between chemical environments with significant band gap differences. Figures 9 and 10 provide a more detailed comparison of the predictions and clustering effects of the LSTM and DCNN layers before and after fusion, showing that with fusion, the LSTM layer's predictions gradually approach the actual DFT values, while also achieving clear clustering of chemical environments with significant band gap differences.

To ensure the reliability of UMAP visualization results, we studied the impact of key UMAP parameters on its visualizations and compared the results with the t-SNE dimensionality reduction method. As shown in Figure 11, we compared the effects of two key parameters influencing UMAP visualizations (n\_neighbors and min\_dist) and found that adjusting these parameters does not change the clustering results. Figure 12 shows a comparison with the results of the t-SNE dimensionality reduction method, revealing that on the OQMD band gap test dataset, the two-dimensional projections of EFM for silicon elements by DCNN-K1, LSTM, DCNN, LDCNN, and L-G-DCNN are similar between t-SNE8 and UMAP methods, confirming the stability of the UMAP visualization of EFM.

Furthermore, we provided additional explanations of the meanings of the two key parameters of UMAP. n\_neighbors: This parameter affects how UMAP constructs the neighborhood of each data point, thereby influencing the local structure of clustering. A smaller n\_neighbors value may lead to more localized clustering, making cluster boundaries more distinct but potentially sacrificing the continuity of the overall structure. Larger values may promote smoother global structures but could blur local boundaries, making clusters appear more coherent but less refined. min\_dist: Controls the minimum distance between points in the reduced dimension, affecting the compactness and separability of clustering. A higher min\_dist value encourages larger gaps between clusters, making clusters more dispersed and aiding in distinguishing different clusters; a lower value may result in tighter clustering but could blur the boundaries between clusters.

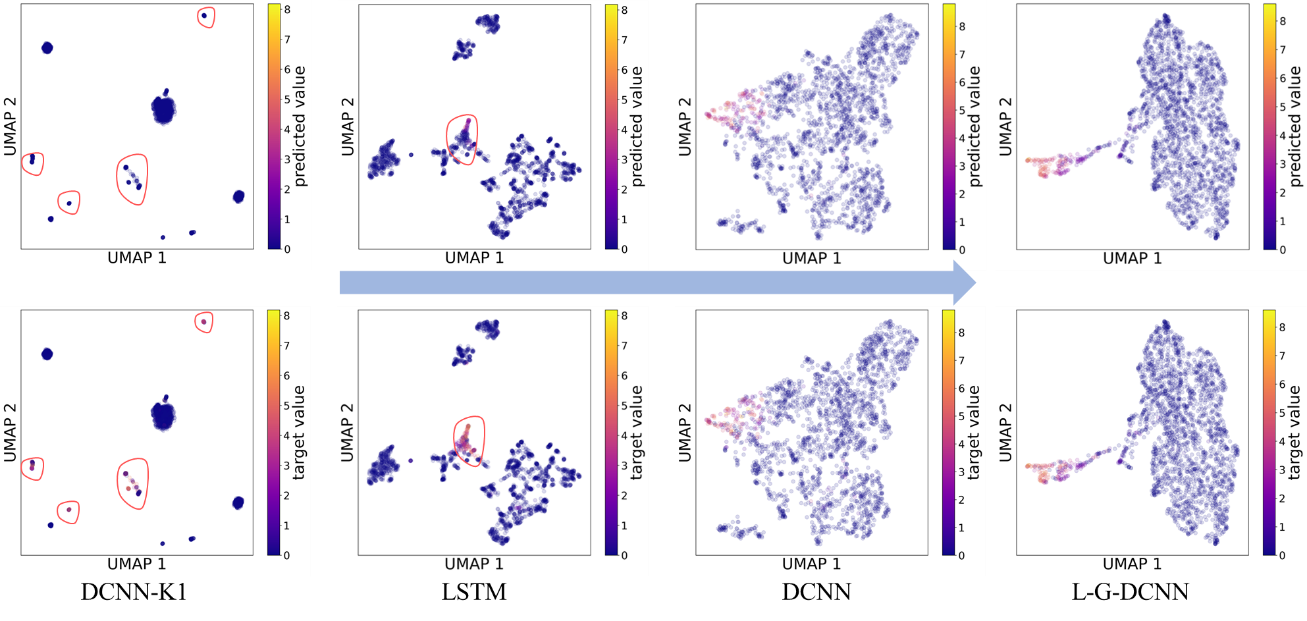


Figure 8 displays the two-dimensional projections of the EFM for silicon-containing compounds in 2374 different chemical environments, based on UMAP, using the OQMD band gap test dataset. Each point in the figure is colored based on the predicted and computed band gap values. The points with significant differences between predicted and computed values in DCNN-K1 and LSTM are marked with red circles.

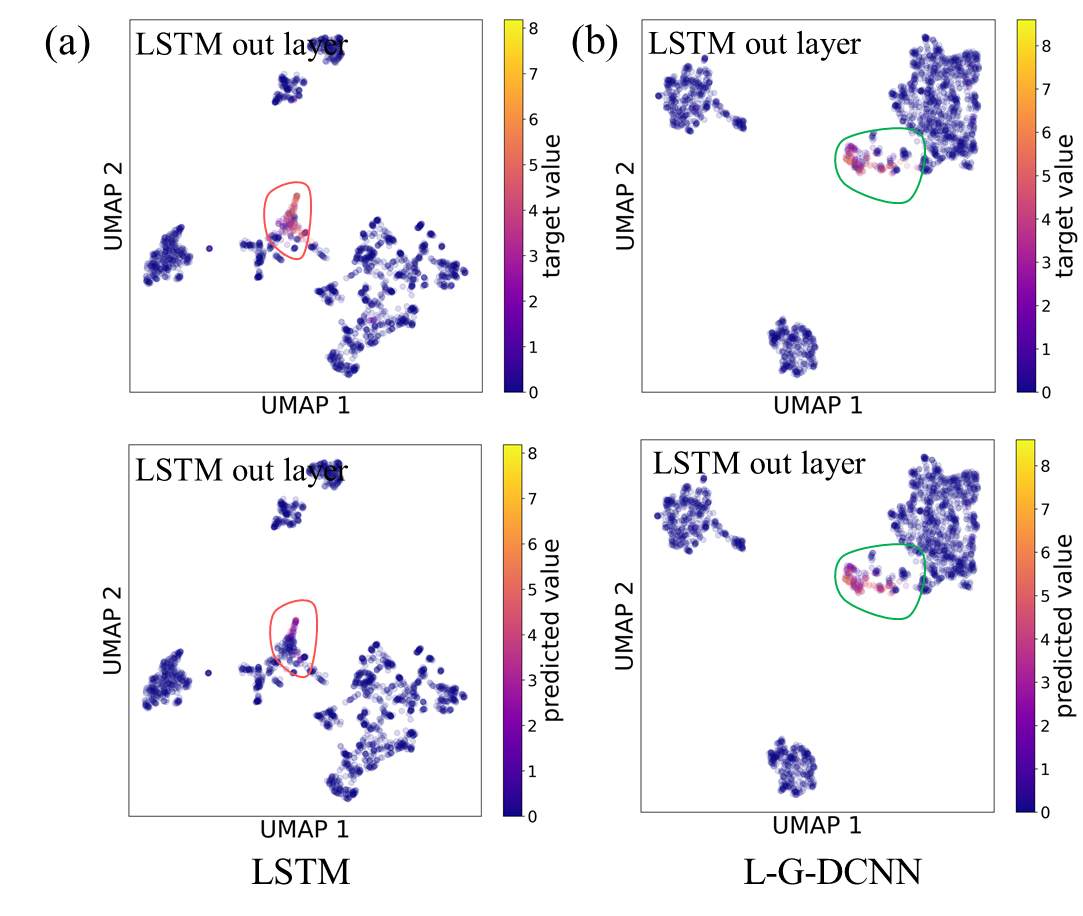


Figure 9 (a) and (b), illustrates the comparison between the predicted and actual values of the UMAP two-dimensional projections of the EFM from the LSTM output layer for all 2374 Si-containing compounds before and after model fusion. The points with significant differences between predicted and actual values are marked with red circles, while the ones with noticeable improvement in differences are denoted by green circles.

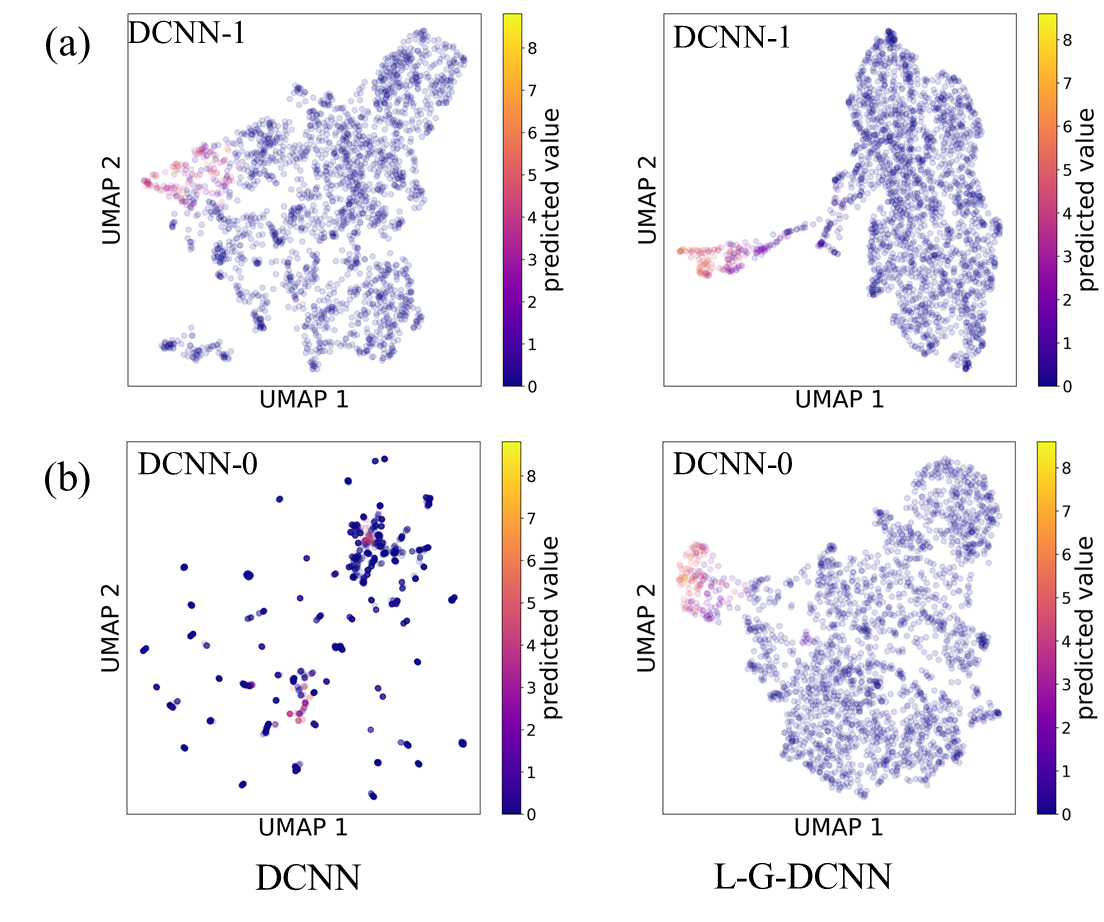


Figure 10 (a) and (b) presents the comparison of the UMAP two-dimensional projections of the EFM obtained from the DCNN-0 and DCNN-1 layers for all 2374 Si-containing compounds before and after fusion. The points are colored based on the predicted band gap values, demonstrating the L-G-DCNN's ability to clearly distinguish chemical environments with significant band gap differences.

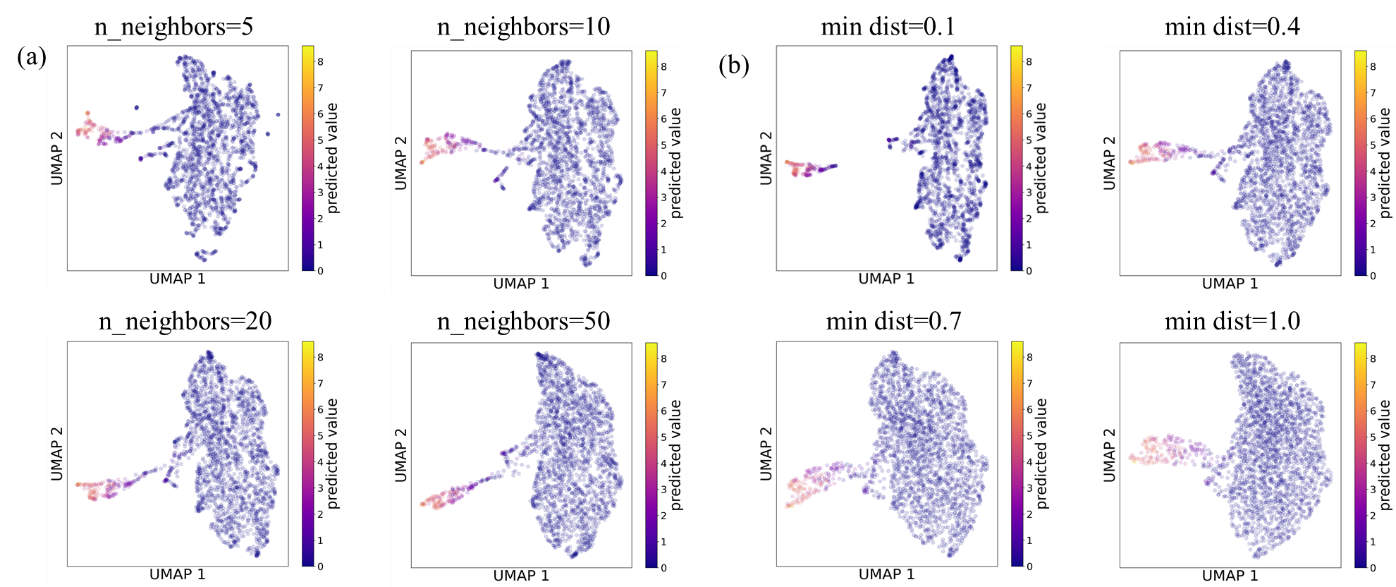


Figure 11 (a) and (b) presents the impact of UMAP's two core parameters, n\_neighbors and min\_dist, on the dimensionality reduction visualization results of the EFM for all 2374 Si-containing compounds.

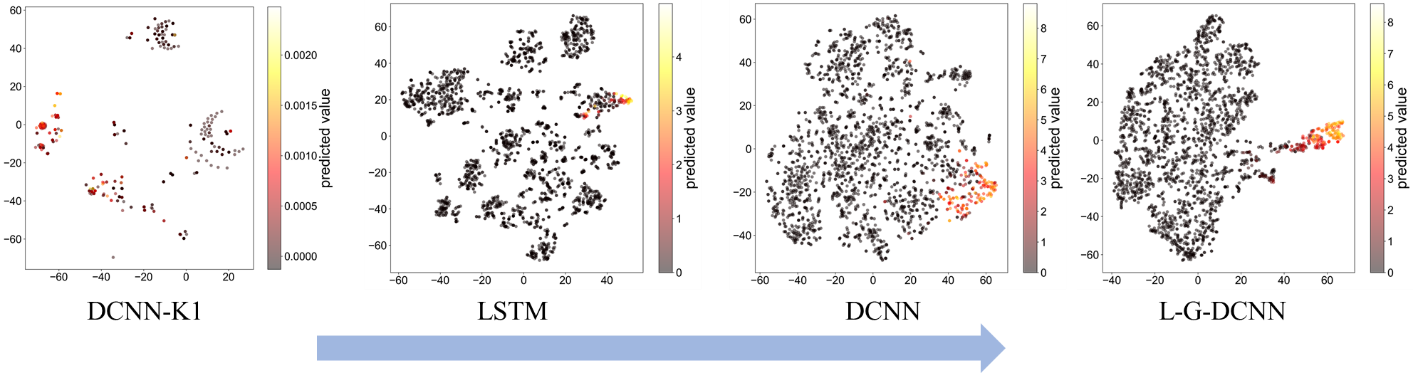


Figure 12 displays the two-dimensional t-SNE projection results of the EFM for silicon-containing compounds in 2374 different chemical environments, based on the OQMD band gap test dataset, showing similar results to the UMAP in Figure 2.

**Manuscript Update:** In the "Visualization of the evolution process of compound EFM before and after fusion" section of our revised manuscript, on pages 18 and 19, we have updated Figure 9, and provided explanatory notes and textual refinements for Figure 9, highlighted in blue font.

**4.** The main issue of this work is the quality of the presentation. It is very hard to follow the scientific contribution of this work. For instance, abbreviations should not be introduced in the abstract but explained with references in the introduction. The term "fusion" is also a bit confusing, using the word "combination" would suffice. Another point is the use of many abbreviations without proper explanation "XRD, XPS, SEM..." and the same for the many NN architectures referenced in this work. The term "different scales" is first used in the abstract but is not even explained there are "compounds at different scales" - it is not clear to me what is meant by that it is not explained later in the text.

**Response:** We appreciate the referee for raising this question. In the revised version, we have indeed removed the abbreviations from the abstract and retained their explanations and references in the introduction. For abbreviations lacking proper explanations, we have included their full forms, such as X-ray Diffraction (XRD), X-ray Photoelectron Spectroscopy (XPS), Scanning Electron Microscopy (SEM), and Transmission Electron Microscopy (TEM). Additionally, for the problems of inadequately explained abbreviations about the neural network architectures, we have provided explanations for the abbreviated, including Deep Convolutional Neural Network with a 1x1 Kernel Size (DCNN-K1), Representation Learning from Stoichiometry9 (Roost), Compositionally Restricted Attention-Based network10 (CrabNet), and Formula graph self-attention network for materials discovery11 (Finder). Once again, we thank the referee for the question and hope that our revisions meet the referee's requirements, enhancing the quality and readability of the paper.

Furthermore, we have taken note of the referee's suggestion regarding the use of the term "fusion," and we appreciate this input. While we believe that using the term "combination" is feasible, we also wish to explain why we chose to use "fusion." In our work, L-G-DCNN appears to be a combination of LSTM12 and DCNN13, but we think that "fusion" more accurately describes the characteristics and effects of L-G-DCNN. Fusion typically implies combining the advantages of different models to achieve better performance. Similar to building blocks, while each block may not have many distinguishing features locally, unexpected results can emerge when observed as a whole. Through the additional Table 2, we compared the averages of LSTM+DCNN and L-G-DCNN, and observed that the MAE of L-G-DCNN relative to the average of LSTM+DCNN is smaller, with a decline ranging from 0.087% to 51.22%. Furthermore, in response to question 4, we analyzed the UMAP two-dimensional projections of the EFM of 2374 silicon-containing compounds. These EFM were obtained from the DCNN-0 and DCNN-1 layers before and after fusion. By coloring these projections based on the predicted band gap values, we observed significant changes in the clustering of the DCNN-0 and DCNN-1 layers after fusion, demonstrating the strong capability of L-G-DCNN in distinguishing chemical environments with significant band gap differences. These results further support the rationale behind L-G-DCNN as a fusion strategy. While we understand the referee's suggestion to use the term "composition," we believe that "fusion" better reflects the effectiveness of L-G-DCNN. In the revised manuscript, we have optimized the overall wording to ensure the accuracy and readability of the paper.

Regarding the term "different scales," we clarify its meaning here. “Different scales ” refer to the thresholds set for obtaining the chemical environment clustering of all compounds at different thresholds for each model. In cluster analysis, we explore the clustering of chemical environments of compounds by setting different thresholds. We acknowledge that the specific meaning of "different scales" was not promptly explained in the manuscripts, which may have caused confusion. We will revise the manuscript to replace "scale" with "threshold" for a clearer understanding of the paper.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Benchmark properties** | **LSTM** | **DCNN** | **L-G-DCNN** | **MAE decline percentage %** |
| **AFLOW Bulk modulus** | 8.93 | 9.23 | 8.221 | 9.460 |
| **AFLOW Debye temperature** | 34.4 | 34.8 | 32.210 | 6.908 |
| **AFLOW Shear modulus** | 9.96 | 9.22 | 8.946 | 6.715 |
| **AFLOW Thermal conductivity** | 2.27 | 2.26 | 2.197 | 3.002 |
| **AFLOW Thermal expansion** | 4.18e-06 | 4.21e-06 | 3.61e-06 | 13.945 |
| **AFLOW Band gap** | 0.377 | 0.348 | 0.310 | 14.483 |
| **AFLOW Energy per atom** | 0.104 | 0.115 | 0.090 | 17.808 |
| **Bartel Decomposition (Ed)** | 0.0801 | 0.0717 | 0.062 | 18.314 |
| **Bartel Formation (Ef)** | 0.0837 | 0.0754 | 0.057 | 28.347 |
| **MP Bulk modulus** | 11.3 | 11.8 | 10.612 | 8.121 |
| **MP Energy above convex hull** | 0.13 | 0.0913 | 0.083 | 24.989 |
| **MP Elastic anisotropy** | 8.01 | 8.04 | 8.018 | 0.087 |
| **MP Magnetic moment** | 2.55 | 2.07 | 2.002 | 13.333 |
| **MP Shear modulus** | 12.8 | 12.5 | 11.701 | 7.502 |
| **OQMD Band gap** | 0.115 | 0.0531 | 0.041 | 51.220 |
| **OQMD Energy per atom** | 0.0466 | 0.0436 | 0.035 | 22.395 |
| **OQMD Formation enthalpy** | 0.0471 | 0.0402 | 0.033 | 24.399 |
| **OQMD Volume per atom** | 0.365 | 0.345 | 0.299 | 15.775 |

Table 2 presents a comparison of the Mean Absolute Error (MAE) on Benchmark datasets between the pre-fusion LSTM, DCNN, and L-G-DCNN. The "MAE decline percentage %" showcases the decrease in MAE of L-G-DCNN relative to the average MAE of LSTM+DCNN, effectively demonstrating the advantage of the fusion strategy employed by L-G-DCNN. These models are compared on a held-out test dataset. All models are evaluated using the same training, validation, and test datasets.

**5.** I would suggest creating a table that gives a nice overview of the different architectures used, which possibly could be included in Figure 1.

**Response:** We thank the referee for this good question. We fully acknowledge the importance of clearly outlining different architectures, as this will help readers better understand our research. We conducted more in-depth literature research and carried out a comprehensive model evaluation on the matbench official website. Currently, models that only use composition information without manual feature engineering, in chronological order, mainly include Crabnet (2021)10 and Finder (2022)11. Based on this, we created Table 1, which details the parameter comparison of the models:

1. L-GDCNN based on the fusion architecture, our latest composition-only model proposal.

2. CrabNet based on the Transformer5 architecture, a composition-only model proposed in 2021.

3. Finder based on the Graph Neural Network14 (GNN) architecture, a composition-only model proposed in 2022. In Figure 13, we compared the architectures of three models.

1. The CrabNet architecture, as depicted in Figure 13(a), involves several components, including the input EDM, self-attention layers, updated and final element representations (EDM′ and EDM″), residual network, and the final model output. The calculation steps for element contributions and prediction of targets and uncertainties are also outlined. After the Transformer encoder updates the element representations, each EDM″ passes through a fully connected residual network, transforming the EDMs into the shape (nelements, nelements, 3). The final three vectors, namely element-proto-contributions 𝑝′, element-uncertainties 𝑢′, and element-logits, are defined. The element scaling factor s is obtained by applying the sigmoid function to the element-logits. The element-contributions are then derived by multiplying the element-proto-contributions 𝑝′ by their respective scaling factor s, resulting in element contributions 𝑦′. The mean of the element contributions is then calculated and output as the predicted property value for each compound.
2. The architecture of Finder, as depicted in Figure 13(b), involves the processing of a formula graph through multiple message passing layers followed by a post-processing neural network. Each message passing layer is coupled with a global attention pooling layer to enable residual connections to subsequent layers. The core operations of the architecture are executed during the message phase, which involves predicting directional edge attributes, allowing information to cascade from neighboring nodes to the edges. These edge features, along with end-node attributes, contribute to a message vector. Each message vector is weighted by a self-attention mechanism that quantifies the importance of other nodes for the current message vector. The aggregate step summarizes all messages around a given node via a local pooling function. Finally, at the update step, the aggregated message vector is added to the initial node attribute, completing one cycle of information flow.
3. The architecture of L-G-DCNN, as depicted in Figure 13(c). The L-G-DCNN is obtained by fusing the global and local models (LSTM||GRU and DCNN, where LSTM||GRU represents the stack of LSTM and GRU.). The element embedding learned by L-G-DCNN from the data will perceive the types of composing elements and the element context. Chemical compositions are inputted by their constituent element's atomic number and fraction. Atomic numbers are used to retrieve element representations, which can be achieved through mat2vec. Therefore, we use mat2vec as the default source of chemical information for each element. Mat2vec is based on using unsupervised word embedding methods to encode knowledge in materials science literature into information-dense word embeddings. These embeddings can capture latent knowledge in materials science literature, including the basic structure of elements in the periodic table and the relationship between material properties, and can be used to better understand and predict material properties. To achieve dimension matching, we pass the obtained element embeddings through a fully connected network, with the default dimension of 512. Chemical stoichiometric information is used to obtain atomic fraction embeddings. The chemical stoichiometry of each element is represented by two fraction embeddings, with the first part representing stoichiometry and the second part using logarithmic scaling to map stoichiometry. This logarithmic transformation preserves small fraction amounts and better adapts to doped systems. The L-G-DCNN takes a three-dimensional tensor as input, generated by adding atomic fraction embeddings and element embedding matrices element-wise. Each row corresponds to an element and the columns contain element embeddings. The chemical compositions are batched on the third dimension.

Through the presentation in Figure 13 and Table 3, we aim to provide readers with a clear view of the differences and connections between different models, thereby enhancing their understanding of our research contributions. We believe that this detailed comparison will improve the clarity and persuasiveness of the paper. Once again, we thank you for your suggestion, as it has greatly helped us enhance the quality of the manuscript.

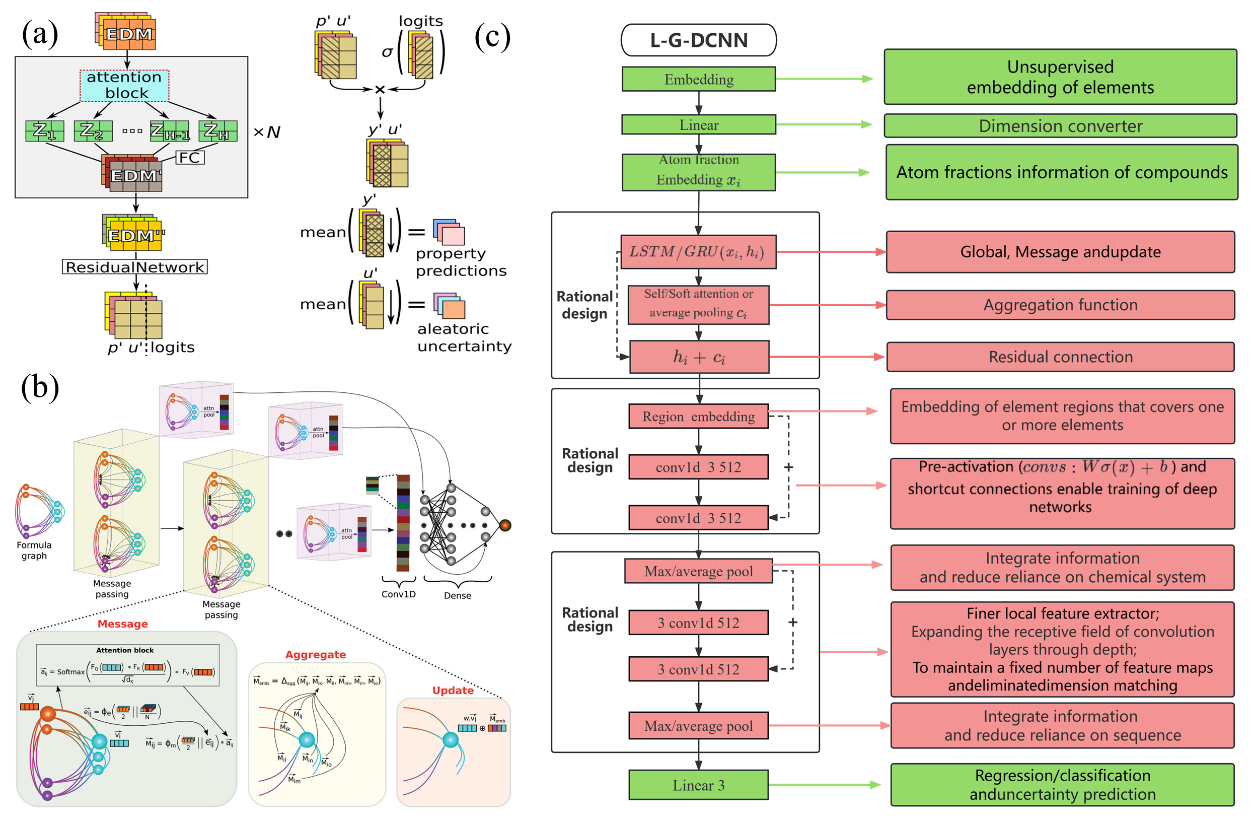


Figure 13 presents an overview of three distinct model architectures: (a) Crabnet, (b) Finder, and (c) L-G-DCNN.

|  |  |
| --- | --- |
| **L-G-DCNN architecture details and parameters** | |
| Embedding type and dimension | mat2vec or one hot, 512 |
| Stoichiometry embedding dimension | 512 |
| LSTM and projection layer | The Bidirectional LSTM hidden features is 512;  Number of recurrent layers is 1;  projection linear layer input size is 512, output size is 1 |
| Softmax attention and residual layer | Size of each input sample is 512;  Size of each output sample is 1 |
| GRU and projection layer | The Bidirectional GRU hidden features is 512;  Number of recurrent layers is 1; |
| Region embedding 1D convolution layer | Number of channels in the input is 512;  Number of channels in the output is 512;  The convolving kernel size is 3, padding is 1; |
| Pre-activation 1D convolution layer | Repeat numbers is 1;  Number of channels in the input is 512;  Number of channels in the output is 512;  The convolving kernel size is 3, padding is 1; |
| Max/avergage pooling, repeat pre-activation 1D convolution layer and residual layer | Repeat numbers is 2;  Number of channels in the input is 512;  Number of channels in the output is 512;  The convolving kernel size is 3, padding is 1;  Max pooling kernel size and stride are 1 |
| Full connection network | Size of each input sample is 512;  Size of each output sample is element numbers |
| Optimiser, Batchsize, Epochs, Loss function ,Learning rate | Adam, 128, 300, L1 robust loss, learning rate scheduler that cycles between a minimum of 1e-4 and a maximum of 6e-3, adjusting within a 30000 step range without momentum cycling. |
| Total Parameters | 8635911 |

|  |  |
| --- | --- |
| **Crabnet architecture details and parameters** | |
| Embedding type and dimension | mat2vec, embeddings numbers is 119,  Embedding dimension is 200 |
| Linear transformer layers | Input and output size are 512 |
| Fractional encoder dimension | 512 |
| Transformer encoder layers | out\_dims:3, d\_model:512, N:3, heads :4 |
| Self attention and projection layer | Attention layers is 3;  In: 512, Out: 512, Bias: True |
| Linear layers(x6) | In: 512, Out: 2048, Bias: True;  In: 2048, Out: 512, Bias: True |
| Dropout layers | P: 0.1 |
| LayerNorm layers | Shape: (512,), Eps: 1e-05, Affine: True |
| Post processing layers | 1024,512,256,128, element numbers |
| Optimiser, Batchsize, Epochs, Loss function ,Learning rate | Adam, 128, 300, L1 robust loss, learning rate scheduler that cycles between a minimum of 1e-4 and a maximum of 6e-3, adjusting within a 30000 step range without momentum cycling. |
| Total Parameters | 11987206 |

|  |  |
| --- | --- |
| **Finder architecture details and parameters** | |
| Embedding type and dimension, Internal dimension | Matscholar, 200, 200 |
| Distance expansion | Gaussian basis, 20 equidistant points between 0 and 5 with a std of 0.5 |
| message passing layers and Pooling operators | 2 layers, ‘element-wise mean’ for local pooling within message passing layer. ‘Soft-attention + element-wise sum’ for global pooling |
| Self-attention layer parameters | dK=dQ=dV=200, QNet, KNet and VNet each has a single layer with 200 units |
| Output layer | Single unit or 3000 units for dielectric function prediction |
| Batch normalisation | Only at message passing output layer. mean=0, var=0.5, gamma=1, beta=0 |
| Post processing neural network  parameters | One convolutional-1D layer with 64 filters. 4 dense  layers with 512, 1024, 1024 and 256 units. L2  regularisation (10-6) used to avoid overfitting. |
| Optimiser, Batch size , Epochs, Loss function, Learning rate, | Adam, 128, 500 for structure-agnostic Finder and 1200 for structure based Finder, L1 robust loss, 3x10-4, reduced by a factor of 0.999 at every epoch |
| Total Parameters | 21651288 |

Table 3 presents a parameter comparison of three models: L-GDCNN based on the fusion architecture, CrabNet based on the Transformer architecture, and Finder based on the GNN architecture.

**Manuscript Update:** In the "Methods" section of our revised manuscript, on pages 22 and 23, we have added an explanation and comparative analysis of the model architectures of Crabnet and Finder, and have included an overview of the different architectures of Crabnet, Finder, and L-G-DCNN in the Supplementary TableS8, highlighted in blue font.

**6.**Still based on the submitted work, I did not see a solution to the problem that is claimed to be addressed. To improve the quality of the paper I would recommend more focus on the presentation of the actual novelty of the work and a substantial improvement in terms of structure and form of the text.

**Response:** We thank the referee for this good question. We have made the following optimizations to the overall structure of the paper, as illustrated in Figure 14: We have specifically adjusted the content of the “Abstract,” “Introduction,” “Rational model design,” “The contribution of elements in property prediction as a function of composition before a5nd after fusion,” “Comparison of the distribution of overall element contributions differences in property prediction before and after fusion,” “Visualization of the evolution process of compound EFM before and after fusion,” and “Methods” sections. The specific details of the adjustments made to each section are as follows:

1. With regard to the “Abstract and Introduction” sections, we have emphasized the rational guidance based on Chemical Environment Clustering Vectors (CECV), the use of LSTM for global feature extraction, and DCNN for local feature extraction, leading to the fusion design of L-G-DCNN. This approach effectively captures and leverages physical insights, promoting the rational design of deep learning networks. We explicitly state that the core focus of this paper is the rational design of L-G-DCNN based on the CECV strategy, and a deep understanding of the model’s fusion process through visualization techniques. Through benchmark testing, we have demonstrated the superiority of L-G-DCNN in predicting material properties from chemical composition data.
2. In the “Rational model design” section, in addition to explaining CECV and providing insights into how the chemical environment of elements influences material properties, we have added a detailed description of the process for obtaining EFM and illustrations (see Figure 4 in the response to question 2), as well as how CECV is used to optimize and efficiently adjust the model architecture.
3. For the sections “The contribution of elements in property prediction as a function of composition before and after fusion,” “Comparison of the distribution of overall element contributions differences in property prediction before and after fusion,” and “Visualization of the evolution process of compound EFM before and after fusion,” we emphasize that these contributions are not directly aimed at the interpretability of deep learning models, but rather serve to visualize the understanding of the model fusion process. These three sections each provide visual evidence of the understanding of the model fusion process through the following aspects: first, the visualization comparison of the contribution of elements in property prediction as a function of composition before and after fusion; second, the comparison of the distribution of overall element contribution differences in property prediction before and after fusion; and third, the visualization comparison of the evolution process of compound EFM before and after fusion. Additionally, similar to the feature importance analysis of traditional machine learning models based on known feature engineering, such as SHAP and LIME, we attempted to explore the interpretability of the L-G-DCNN model. By systematically deconstructing L-G-DCNN into LSTM and DCNN, we were able to track the improvement process of the model’s performance. Taking the OQMD band gap dataset as an example, we first identified 56 silicon-containing compounds with poor performance in the LSTM model. Subsequently, we compared the EFM plots of these 56 compounds, calculated ΔEFM between LSTM and L-G-DCNN, and revealed the elemental interactions pairs with the greatest impact on band gap properties. Finally, we summarized the elemental interactions pairs with the greatest impact on band gap and provided intuitive insights for understanding property prediction (see Figure 3 in the response to question 1).
4. In the “Methods” section, in addition to introducing key points of the L-G-DCNN architecture, inputs, intermediate processing, and final outputs, we also compared the architectures and parameters of three models, including Finder based on the GNN architecture, Crabnet based on the Transformer architecture, and L-G-DCNN based on the fusion architecture.

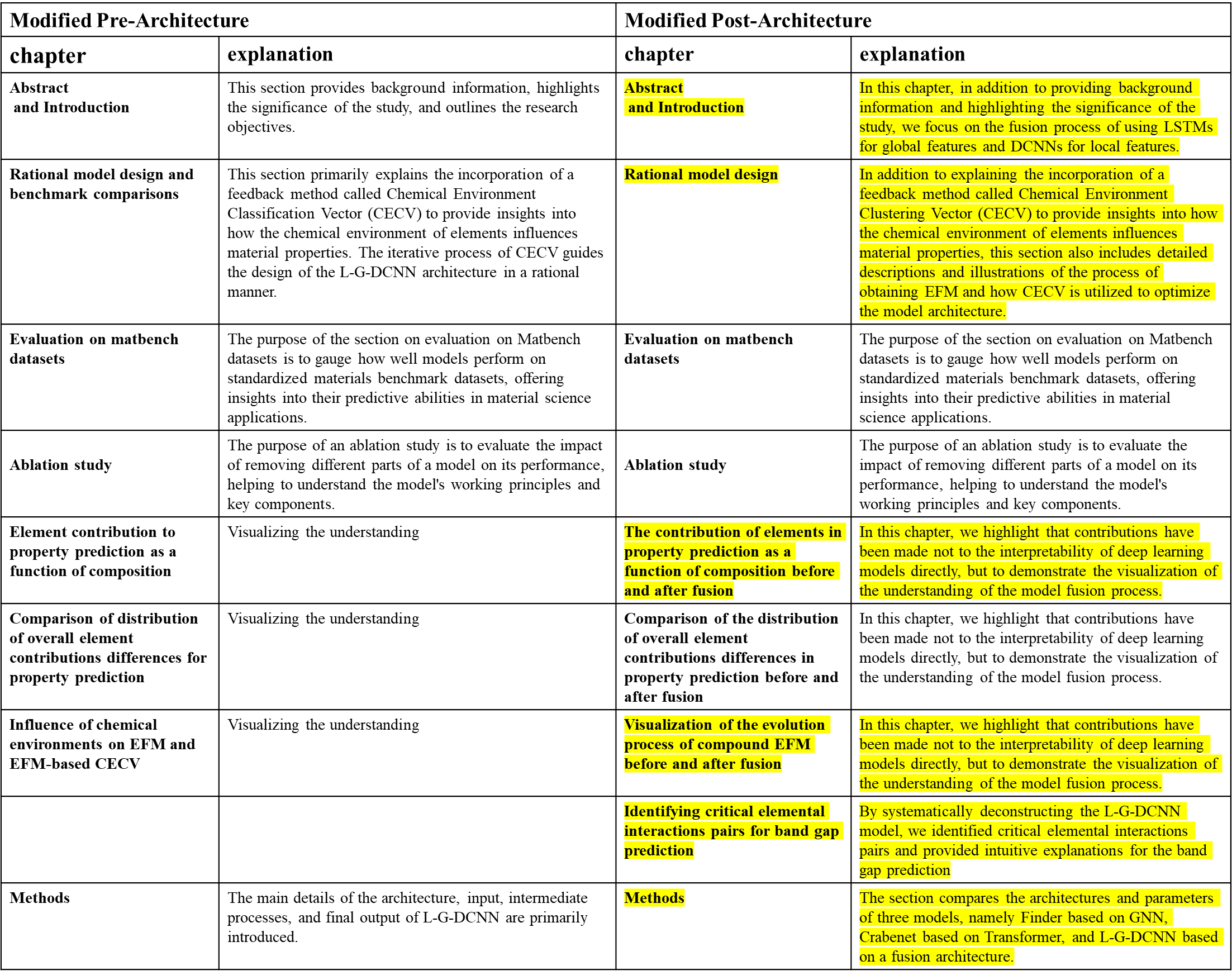


Figure 14 illustrates the comparison of changes in each section before and after the manuscript revision, along with explanations, where the sections highlighted in yellow represent the main revised sections and explanations.

## Reviewer #2

The article presents a comprehensive study on the development and application of a deep learning model, L-G-DCNN, which incorporates a novel feedback mechanism, the CECV, to enhance the prediction of materials properties. The integration of CECV into the model design provide a method to capture and utilize physical insights effectively, prompting the rational design of deep learning networks. Overall, the article is well crafted by presenting the context and implications of the study, supported by the relevant background, methodology, results, and conclusions.

**Response:** We appreciate the referee's useful comments and positive recognition of our work. We have revised the manuscript following the referee’s suggestions.

**1.** The clarity of the graphical illustrations needs to be improved. For instance, in Figure 5c, which displays the element feature matrices, the numerical values are quite challenging to discern. It will be highly beneficial if the resolution of this figure, along with any other similar instances throughout the document, could be enhanced to ensure that all figures are legible and effectively communicate the intended information.

**Response:** We greatly appreciate the referee for pointing out the problems in our figures. In response to your suggestions, we have made modifications to Figures 4 and 5 in the manuscript. Specifically, for Figure 4, on the left side, we have increased the axis labels for "training set size" to ensure improved clarity and readability. Addressing the problem of unclear EFM plots display in Figure 5, we have enhanced the image resolution to ensure better discernibility of the values. We have also made similar adjustments to figures with similar problems throughout the manuscript, increasing the pixel density of the images to 300-500 dpi, while enlarging the size of the x and y-axes as well as labels for Figures 2 to 10. We are grateful for the referee 's suggestions and have comprehensively revised the relevant content to ensure the clarity and readability of the figures.

**2.**The article selects Roost and Crabnet as baseline models and uses LSTM|GRU and DCNN as the proposed method's backbone, which does not represent a strong baseline. Many models, including those that do not utilize three-dimensional information, have achieved better results, for example, on the Matbench dataset. It would make the contributions of the method more convincing if stronger baselines were included.

**Response:** We appreciate the referee’s questions and criticism on our work. You pointed out that the benchmark models Roost and Crabnet we selected, as well as the L-G-DCNN architectures used, may not have constituted a strong comparative benchmark. We understand your point and agree that there are indeed some models that have achieved more outstanding results on the Matbench dataset15.

In response to your questions, we conducted a more in-depth literature review and further evaluated the models on the Matbench official website(https://matbench.materialsproject.org/). We found that currently, models that only utilize compositional information without manual feature engineering, in chronological order, mainly include Roost (2020), Crabnet (2021), and Finder (2022).

Based on this, we compared L-G-DCNN with these models on 18 different benchmark datasets and 10 small datasets containing computational and experimental data. The comparison results, summarized in Figure 1 and Table 1, show that L-G-DCNN outperformed in 15 out of 18 benchmark datasets and 7 out of 10 small datasets, demonstrating the superiority of L-G-DCNN in terms of advancement and sampling efficiency. These comparisons cover datasets of different scales (ranging from 293 to 3582 samples) and various material properties such as Thermoelectric, Stability, ElectricPotential, Conductivity, and Superconductivity temperature. We ensured that all comparisons were conducted on a unified dataset, training strategy, and parameter settings, including learning rate, optimizer, loss function, evaluation function, as well as batch size and number of epochs, all using their respective optimal parameters. Additionally, Table 2 provides the full names and abbreviations of all 10 small datasets, as well as detailed information on data sources and classifications.

It is worth mentioning MODNET (2021)16, as shown in Figure 2, in the design of MODNET (2021), the representation of structural or compositional features is achieved through a series of descriptors involving physical, chemical, and geometric properties, including atomic mass, electronegativity, space group, and local environment, allowing the model to more directly reveal the unknown connections between the compositional features and target properties. In contrast, Crabnet, Finder and L-G-DCNN can directly learn structural or compositional feature representations from the data. The input representation of MODNET differs significantly from that of Finder, Crabnet, Roost, and our L-G-DCNN. In MODNET, a substantial amount of learning occurs before neural network training, undoubtedly enhancing its predictive performance. Therefore, including MODNET would introduce unfairness in comparative analysis. Additionally, it is important to note that MODNET has not yet demonstrated the ability to automatically learn under a one-hot input representation lacking chemical information. In contrast, Crabnet and our L-G-DCNN have already demonstrated their capability under one-hot input conditions. This difference is crucial for evaluating the performance of models in terms of generalization ability and reliance on prior chemical knowledge.

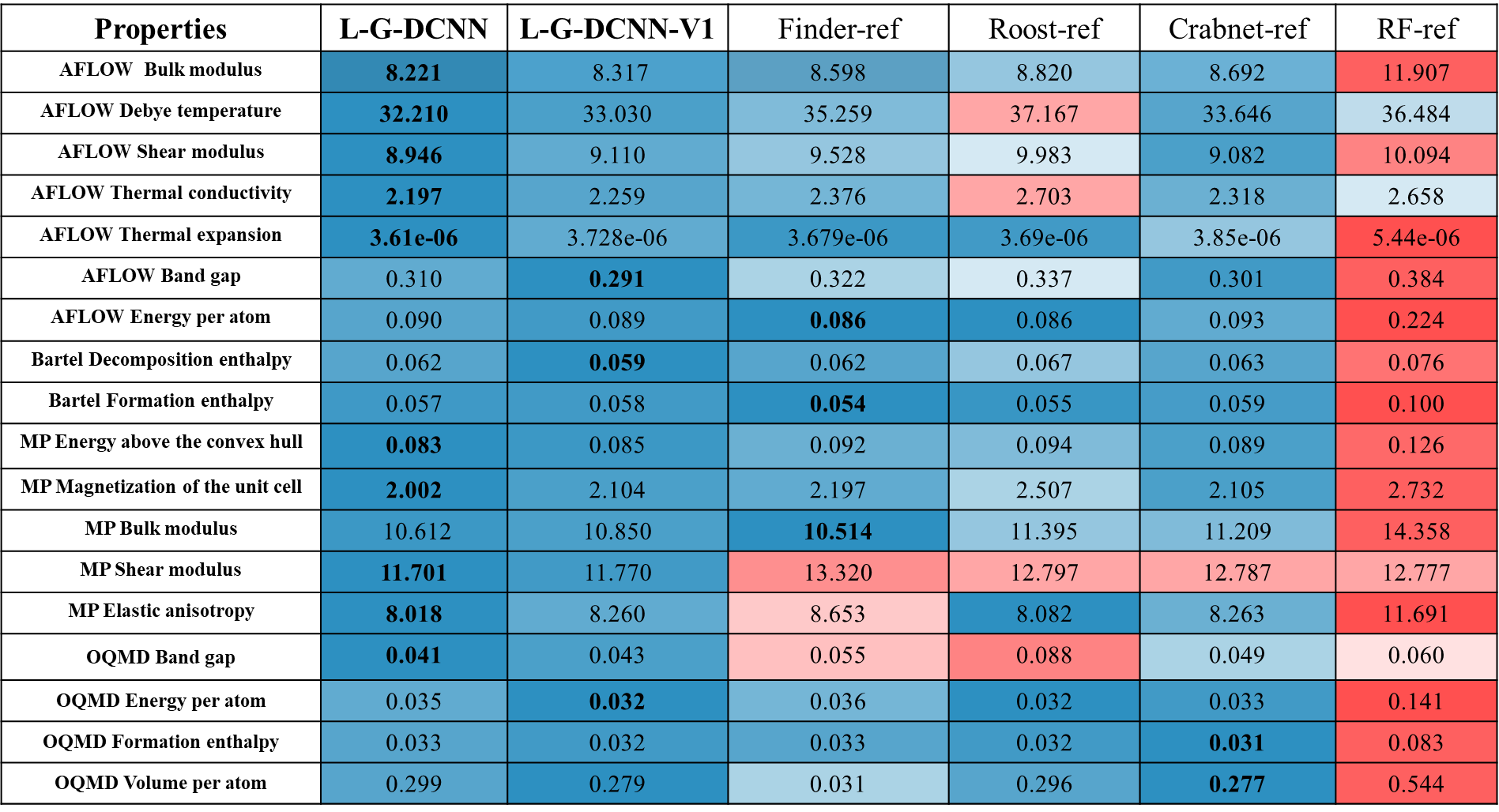


Figure 1 compares the MAE values of the L-G-DCNN series models with the reference models Roost, Crabnet, and Finder on the benchmark test dataset. Cells are colored based on the relative MAE within each row (blue indicating better performance, red indicating worse performance). The best MAE values are highlighted in bold.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets | Property | L-G-DCNN | Finder-ref | Roost-ref | Crabnet-ref | RF-ref | Size |
| Jarvis DFT | Exfoli | 48.200 | 59.584 | 54.593 | 45.500 | **42.823** | 557 |
| Jarvis DFT | piezo\_max\_di | **4.523** | 4.629 | 5.349 | 4.882 | 5.053 | 2126 |
| Jarvis DFT | piezo\_max\_dij | **36.838** | 45.324 | 64.786 | 37.267 | 71.074 | 689 |
| Jarvis DFT | piezo\_max\_eij | 0.310 | **0.298** | 0.443 | 0.301 | 0.412 | 1123 |
| Experiment | ETC | **13.450** | 16.263 | 27.318 | 14.567 | 27.133 | 342 |
| Experiment | ESC | **2.665** | 3.261 | 3.377 | 2.761 | 3.000 | 3582 |
| Experiment | EFE | **0.065** | \* | 0.077 | 0.067 | 0.0714 | 1061 |
| Experiment | ESC | 59.194 | 56.838 | 62.320 | 66.617 | **55.563** | 293 |
| Experiment | EBG | **0.276** | 0.324 | 0.381 | 0.301 | 0.378 | 5805 |
| Experiment | EEC | **206.057** | 300.806 | 223.459 | 215.003 | 281.502 | 293 |

Table 1 presents the MAE performance of L-G-DCNN and the reference models on small computational and experimental datasets of varying data sizes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Property | Abbreviation | Data Source | Category |
| Jarvis DFT | ExfoliationEnergy (eV/atom) | Exfoli | Jarvis17 | MechanicalEnergyCost |
| Jarvis DFT | Dfpt Piezo Max Dielectric(ε11) | piezo\_max\_di | Jarvis | ElectricPotential |
| Jarvis DFT | DfptPiezoMaxDij(cm2) | piezo\_max\_dij | Jarvis | ElectricPotential |
| Jarvis DFT | DfptPiezoMaxEij(cm2) | piezo\_max\_eij | Jarvis | ElectricPotential |
| Experiment | Expt thermal conductivity(W/mK) | ETC | matminer15 | Thermoelectric |
| Experiment | Expt super conductivity(K) | ESC | matminer | Super conductivity |
| Experiment | Expt formation enthalpy(eV/atom) | EFE | matminer | Stability |
| Experiment | Exp seebeck coefficient(muV/K) | ESC | matminer | Thermoelectric |
| Experiment | Exp band gap(eV) | EBG | matminer | Conductivity |
| Experiment | Expt Electrical conductivity(S/cm) | EEC | matminer | Thermoelectric |

Table 2 presents a comparison of the full names and abbreviations of material properties for all 10 computational and experimental small datasets, along with the sources and classifications of the data.

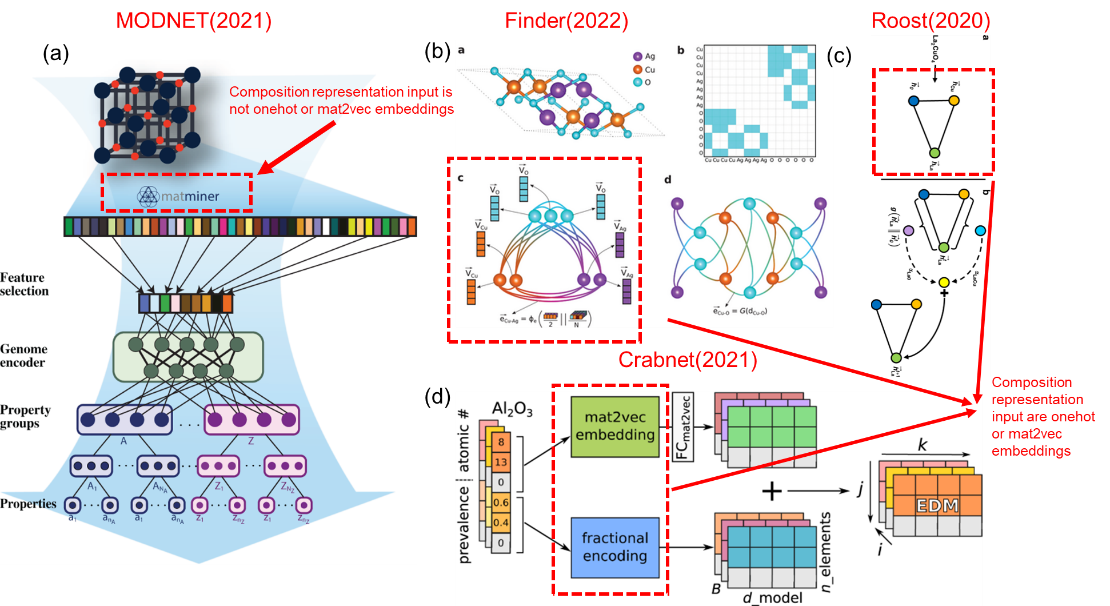


Figure2 illustrates a comparison of the input representations between the newer MODNET (2021) and the models Finder (2022), Crabnet (2021), and Roost (2020) that we selected in our paper.

**Manuscript Update:** In the "Rational model design" section of our revised manuscript, on pages 9 and 10, we have added the MAE results of the current state-of-the-art structure-agnostic model Finder (2022) on the extended dataset, representing a stronger baseline, and compared it with our L-G-DCNN. We have also updated the comparative analysis description, highlighted in red font.

**3.**The CECV method use EFM to get the feedback of whether fusion mode is good or not, which is one of the most important innovation in this article. But the acquisition process of this matrix and how to modify model structure have not been introduce and analysis quantitative. It will be clearer if it can be illustrated in detail.

**Response:** Thank you for the meticulous review and valuable questions from the reviewers. In response to your inquiries regarding the acquisition process of EFM and how to adjust the model according to the guidance of CECV, we provide a detailed explanation here. Below, we explain the acquisition of EFM and CECV, as well as the process guiding the iterative design of the model, as shown in Figure 3:

**1. Selection and Initialization of Model Components**

Select model components from the Common Model Library, including global models (such as LSTM, GRU, RNN) and local models (such as DCNN, DCNN-K1, CNN, TextCNN, ResNet), as well as attention and pooling components. Initialize the composition model, such as the fusion of LSTM and DCNN (LDCNN), and set linear and activation layers for post-processing.

**2. Model Training and Testing**

Based on the OQMD enthalpy dataset, use the selected model components (such as LSTM, DCNN-K1, DCNN, LDCNN) for training and validation. Divide the dataset into training, validation, and test sets in a 70/15/15 ratio, and train all models under unified parameter settings.

**3. Extraction of Compound EFM**

For 2374 compounds containing silicon elements in the test dataset of OQMD enthalpy (covering binary to heptenary compounds), use the trained models (LSTM, DCNN-K1, DCNN, LDCNN) to extract the element embedding feature matrix (EFM) of the compounds. Each compound corresponds to a 512-dimensional EFM. It is worth noting that we only use compounds containing silicon elements as an example here, but other element compounds can also be set, or compounds with different elements can be randomly selected.

**4. CECV Generation**

Based on the extracted EFM, cluster all Si-containing compounds, set different clustering threshold ranges as [0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4], and then generate a set of CECV, recording the number of clusters corresponding to each threshold under different models. For example, LSTM obtains the clustering vector within the set threshold: [376, 147, 80, 48, 31, 27, 16, 14].

**5. Visualization and Definition of Target Region**

Plot the graph, with the clustering threshold on the x axis and the number of CECVs obtained by each model on the y axis, showing the performance of different models under different clustering thresholds. Mark the LSTM and DCNN-K1 as global and local reference model, defining the target region between the LSTM and DCNN-K1. This step helps to intuitively understand the tendencies and trade-offs of the model in global and local feature extraction. For example, as shown in Figure 4, the two reference models represented by the red and green dashed lines respectively demonstrate the CECVs of LSTM and DCNN-K1 for Si-containing compounds in the OQMD enthalpy dataset under different thresholds. Based on the characteristics of LSTM and DCNN-K1, they have different requirements for CECV at different thresholds. For the global feature extractor LSTM, its performance is close to optimal within the threshold <0.15, as a smaller threshold leads to lower tolerance for distinguishing the chemical environment of compounds, requiring higher demands, causing the model to overly focus on local details and easily overestimate CECV. Therefore, it is necessary to increase the weight of the global feature extractor to adjust the model and achieve more accurate identification of the clustering of compound chemical environments, as indicated by the downward red dashed arrow on the left side of Figure 4. Similarly, for the local feature extractor DCNN-K1, its performance is close to optimal within the threshold >0.15, as a larger threshold leads to higher tolerance for distinguishing the chemical environment of compounds, requiring lower demands, causing the model to overly focus on global information and easily underestimate CECV. Therefore, it is necessary to increase the weight of the local feature extractor to adjust the model, as indicated by the upward red dashed arrow on the right side of Figure 4. Based on the above strategy, we can draw the target region, which is marked as a shaded area in Figure 4. With the understanding of Figure 4, we can refer to the focusing principle of a camera:

1. Global feature extraction is similar to a wide-angle lens of a camera. The global feature extractor (such as LSTM) can capture a broader field of view, i.e., the overall chemical environment of compounds.

2. The local feature extractor (such as DCNN-K1) is similar to a telephoto lens of a camera, focusing on capturing details of local regions.

3. The target region can be seen as the best focus point of a camera, where the image contains rich details while retaining overall clarity.

4. Model adjustment is similar to adjusting the focal length of a camera to obtain the best image. By increasing the weight of global or local model components, the model can be adjusted to better adapt to the recognition requirements of chemical environments.

**6. Iterative Optimization of Model Components**

As shown in Figures 3 and 4, identify the deviation between the current model's CECV and the target region (marked as a shaded area in Figure 4), and decide whether to strengthen the weight of the global feature extractor or the local feature extractor weight, guiding targeted adjustments of model components from the model library to make the CECV of the new model closer to the target region, achieving rational optimization of model design. To understand how CECV guides the rational modification of the model structure, based on Figure 4, we provide further explanation. Compared with the two reference models, the black line of LDCNN performs better in fusion. Within the threshold range <0.15, the CECV of LDCNN is significantly lower than that of DCNN-K1, very close to the shaded area. However, within the threshold range >0.15, its improvement in CECV is limited, with a significant deviation from the shaded area, resulting in mediocre performance of LDCNN. Guided by Figure 4, we determine that the key factors lies in increasing a model component that can increase the weight of the LSTM end in LDCNN. Therefore, a reasonable design 1 is proposed, introducing soft attention and residual connections to optimize the LSTM part of LDCNN, resulting in L-ATT-RES-DCNN, as indicated by the cyan arrow in Figure 4. However, L-ATT-RES-DCNN achieves an increase in CECV across the entire threshold range, indicating that introducing soft attention and residual connections to the LSTM part does not effectively increase the weight of the LSTM end. Therefore, a reasonable design 2 is proposed, introducing stacked residual GRU and softmax attention to obtain L-G-DCNN. This design significantly reduces the CECV within the threshold range <0.15 and greatly improves the performance of L-G-DCNN, as indicated by the yellow arrow in Figure 8. We can see that within the threshold range <0.15, L-G-DCNN still has room for improvement. Therefore, a reasonable design 3 is proposed, introducing residual connections in L-G-DCNN to obtain L-G-DCNN-V1, as indicated by the green arrow in Figure 4. L-G-DCNN-V1 further reduces the CECV within the threshold range <0.15, entering the target shaded area and achieving optimization. However, there is also a decrease in CECV within the threshold range >0.15. Overall, the target region has not been fully reached. From LDCNN to L-G-DCNN-V1, based on CECV, we have improved the performance of the model through three design steps, achieving a rational and high efficient design.

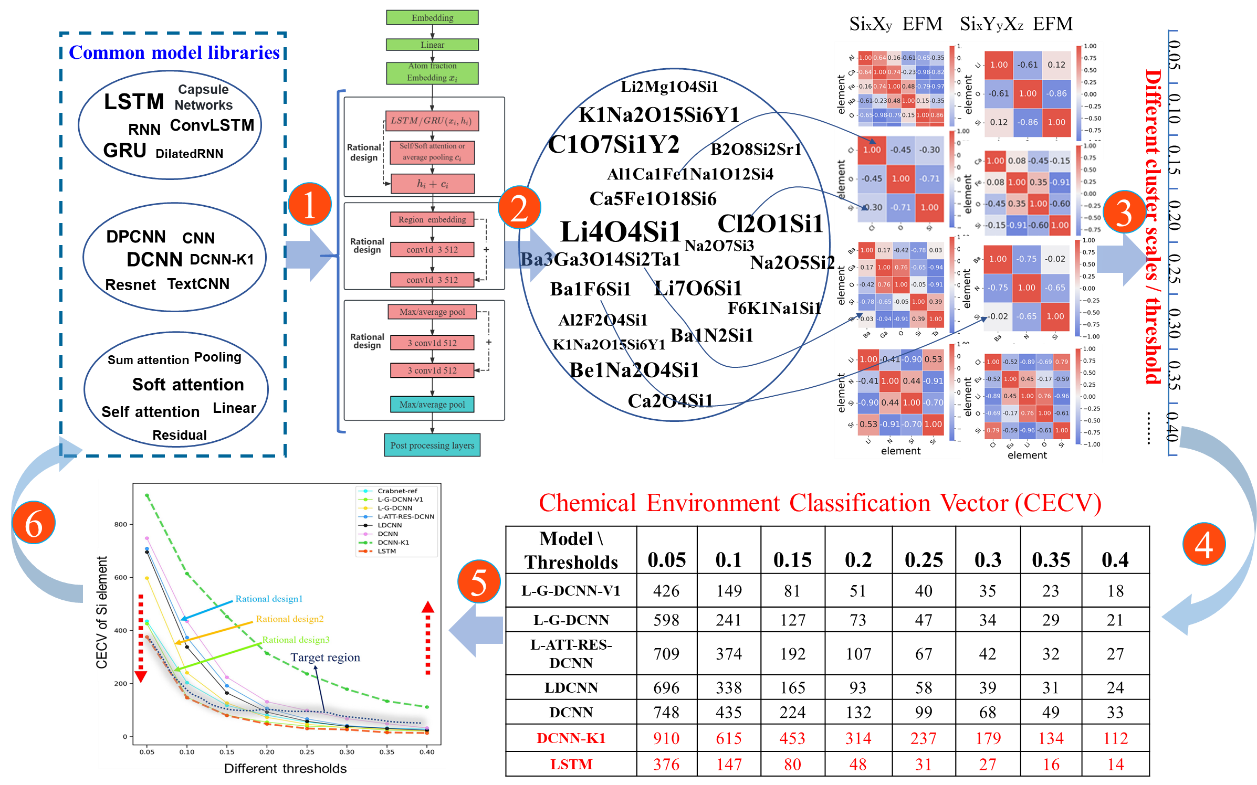


Figure 3 illustrates the process of obtaining EFM and how, based on this, the CECVs of all models before and after fusion under different thresholds are obtained, thereby achieving rational model design guided by CECV.

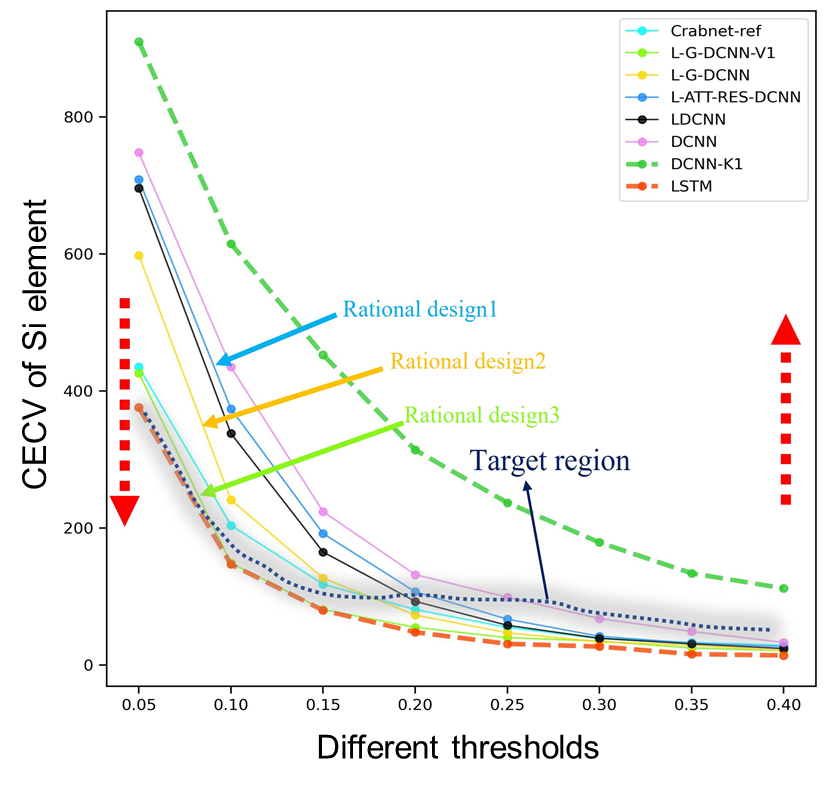


Figure 4 shows the CECV of silicon-containing compounds in the OQMD Formation enthalpytest datasets based on the fusion strategy, using LSTM and DCNN-K1 as global and local reference models. CECV were used as the feedback method to achieve rational model design. The shaded region in the figure marks the CECV of the optimal fusion model. The blue, yellow, and green solid arrows mark the rational design from LDCNN to L-G-DCNN-V1. Two red dashed arrows represent the rational design direction of the model at the LSTM and DCNN-K1 ends. Crabnet-ref is a reference model, which performs best on the OQMD Formation enthalpy.

**Manuscript Update:** In the "Rational model design" section of our revised manuscript, on pages 4 and 5, we have updated Figure 1 and the explanations to the aforementioned questions, highlighted in blue font.

**4.** The model's performance compared with many SOTA models in the field is compared, while it lacks sufficient explanation of how the methods develop further enhance this field. Some descriptions from the domain perspective, especailly how the new algorithms could be adopted are helpful.

**Response:** We appreciate the referee’s reasonable questions and constructive criticism on our work. This prompted us to delve deeper into how L-G-DCNN brings advancements in the field of materials science and to provide a detailed explanation of its advantages. In response to your inquiries, the following points provide a specific description of how L-G-DCNN can further enhance this field:

**1.** **Innovative framework for expanding deep learning model design:** The innovation of L-G-DCNN lies in its unique fusion architecture design, which differs from previous models such as Roost and Finder, which relied on graph neural networks, or the CrabNet model based on Transformer. By fusing the global feature extraction capability of LSTM and the local feature capture capability of DCNN, L-G-DCNN not only transcends the limitations of a single model but also opens up a new design philosophy for deep learning applications in materials science. This fusion framework emphasizes the comprehensive analysis of global and local features, providing a more holistic perspective for understanding and predicting material properties. This shift from singular to comprehensive in model design promotes the deep application of deep learning models in the field of materials.

**2. Transparent demonstration of model performance evolution:** The modular design of L-G-DCNN, akin to "building blocks," allows us to visually track the gradual improvement of the model from discrete components (LSTM and DCNN) to the fusion model (L-G-DCNN). This step-by-step demonstration not only reflects the transition from "poor" to "good," but more importantly, it reveals the more precise capture of interactions between elements during the model evolution process. For example, by comparing the differences in prediction results for specific compounds between LSTM and L-G-DCNN, we can directly identify which key interactions of elements pairs are more accurately modeled, thereby understanding how these interactions of elements pairs affect material properties. This process deepens our understanding of the mapping relationship between material properties and their chemical environment and provides important guidance for the design of new materials with specific properties.

**3. Efficient and rational optimization strategy based on CECV:** One of the core design principles of L-G-DCNN is based on CECV. CECV introduces the chemical environment, providing a new method for optimizing surrogate models in small datasets. It not only enriches the physical basis of model design, making model predictions more consistent with the basic principles of materials science, but also guides rational model design through feedback mechanisms. For example, in alloy component design, which typically involves fitting surrogate models (e.g., the commonly used Gaussian process18 (GP) model) based on available data to guide adaptive experimental synthesis and characterization. For sequential design, whether autonomous or manual synthesis, a bottleneck arises as the number of alloy elements and their compositions exceed a threshold, making the search space difficult to handle. Although research19 has shown that the active learning framework combining reinforcement learning (RL)20 with GP surrogate models may be a better solution, in many applications, such as high-entropy alloys or multi-component alloy scenarios, available alloy data is often limited, severely affecting the predictive accuracy of GP surrogate models. If the GP prediction is inaccurate, the calculation of the reward value obtained by RL is inaccurate, leading to an inaccurate loss function guiding RL training, and ultimately making it difficult for RL to access regions with higher reward values. We analyze the main reasons for the inaccuracy of GP predictions, which are primarily due to the limited available experimental data and the lack of clear guidance for evaluating the difference between the current prediction and the true level, and cannot tell GP how to adjust to improve predictive accuracy. Therefore, common strategies can only ensemble more models or perform transfer learning, more complex feature engineering, etc., which cannot fundamentally solve the problem of GP under data scarcity. In contrast, after training on the same small dataset, L-G-DCNN can calculate the CECV of LSTM, DCNN, and L-G-DCNN, not only evaluating the difference from the target region (based on the results calculated by LSTM and DCNN-K1), but also knowing specifically whether the difference lies in the LSTM end or the DCNN-K1 end, thus knowing how to adjust. For example, this adjustment may involve introducing model components that enhance the global chemical environment recognition capability or increasing model components that enhance the local chemical environment recognition capability, which can achieve efficient adjustment of L-G-DCNN. This is also a form of ensemble strategy, but it becomes more rational and efficient under the guidance of CECV. In an active learning framework, CECV can guide model design, select the most informative experiments to supplement the dataset, and thus efficiently improve model performance. This strategy is particularly crucial under conditions of limited resources and scarce data, as it ensures that each experiment adds the maximum value to model learning, accelerating the transformation from data to knowledge.

L-G-DCNN, through its unique structural design (fusing LSTM and DCNN), demonstrates high predictive ability even under conditions of limited data. LSTM captures the global chemical environment, while DCNN focuses on the detailed local chemical environment. This complementarity allows the model to more fully utilize information in limited data and improve prediction accuracy. To validate the advantages of L-G-DCNN on small experimental datasets, we conducted extensive testing and comparisons on 11 different small datasets. We compared it with the current state-of-the-art models, including the CrabNet based on the Transformer architecture, Roost based on graph neural networks, and traditional random forest models. These comparisons covered datasets of different data sizes (ranging from 312 to 3518 samples) and various material properties, such as Thermoelectric, Stability, Elasticity, Conductivity, and Superconductivity temperature, among others. As shown in Table 3, L-G-DCNN performed the best overall, leading in 9 out of the 11 datasets. Additionally, Table 4 provides the full names and abbreviations of all 11 experimental datasets, as well as detailed information on their sources and classifications.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Datasets | Property | L-G-DCNN | Finder-ref | Roost-ref | Crabnet-ref | RF-ref | Size |
| Experiment | SS | **100.656** | \* | 194.200 | 119.950 | 110.884 | 312 |
| Experiment | EFE | 0.0606 | **\*** | **0.057** | 0.068 | 0.0656 | 1196 |
| Experiment | ESC | 50.840 | **\*** | 47.062 | 53.676 | **48.891** | 1093 |
| Experiment | EER | **3.198** | **\*** | 3.802 | 4.421 | 33.324 | 1093 |
| Experiment | EPF | **0.00029** | **\*** | 0.00030 | 0.00031 | 0.00032 | 1093 |
| Experiment | EFM | **0.125** | **\*** | 0.143 | 0.133 | 0.183 | 1093 |
| Experiment | ETC | **1.496** | **\*** | 2.130 | 1.503 | 1.528 | 714 |
| Experiment | EEC | **579.591** | **\*** | 704.123 | 756.250 | 892.204 | 1093 |
| Experiment | ECT | **121.262** | **\*** | 130.557 | 123.926 | 130.707 | 3518 |
| Experiment | EBG | **0.250** | **\*** | 0.323 | 0.266 | 0.367 | 6354 |
| Experiment | ESC | **5.212** | **\*** | 7.909 | 5.620 | 5.774 | 16317 |

Table 3 presents the comparison of MAE between L-G-DCNN and reference models on 11 small experimental material property datasets of different sizes, where \* denotes that the Finder does not support compounds with non-integer stoichiometry.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Property | Abbreviation | Data Source | Category |
| Experiment | Steel strength (MPa) | SS | matminer | Elasticity |
| Experiment | Expt formation enthalpy(eV/atom) | EFE | matminer | Stability |
| Experiment | Exp seebeck coefficient(muV/K) | ESC | matminer | Thermoelectric |
| Experiment | Exp Electrical resistivity(ohm.cm) | EER | matminer | Thermoelectric |
| Experiment | Exp power factors(W/mK^2) | EPF | matminer | Thermoelectric |
| Experiment | Exp figure of merit(unitless) | EFM | matminer | Thermoelectric |
| Experiment | Expt thermal conductivity(W/mK) | ETC | matminer | Thermoelectric |
| Experiment | Expt Electrical conductivity(S/cm) | EEC | matminer | Thermoelectric |
| Experiment | Exp curie temperature (K) | ECT | matminer | Magnetism |
| Experiment | Exp band gap(eV) | EBG | matminer | Conductivity |
| Experiment | Expt super conductivity T(K) | ESC | matminer | Super conductivity T |

Table 4 illustrates the comparison of the full names and abbreviations of all 11 experimental datasets, along with the sources and classifications of the data.

**Manuscript Update:** In the "Discussion" section of our revised manuscript, on page 21, we have added how the new algorithms could be adopted and enhance the field of materials, highlighted in red font.

## Reviewer #3

This paper presents a new deep learning model called L-G-DCNN for predicting material properties from chemical composition data. The model architecture fuses global (LSTM/GRU) and local (DCNN) feature extractors to capture interactions between elements in compounds. A key aspect is the introduction of a feedback method called the Chemical Environment Classification Vector (CECV) that aims to incorporate physical insights about how the chemical environments of elements influence material properties. The CECV is used to rationally guide the design of the L-G-DCNN architecture through an iterative process. The authors benchmark their model against several existing approaches like Roost, Crabnet, and random forest regression across 28 materials datasets. They also propose visualization techniques to interpret how the model representations relate to chemical environments and periodic trends. Overall, the work explores a fusion strategy with feedback from physical insights for designing neural networks tailored to the materials domain.

**Response:** We greatly appreciate the referee for her/his time to review our manuscript and give us the above extremely encouraging comments. We have revised our manuscript following the referee’s suggestions.

**1.** The work looks interesting and well-written with good benchmarking but my main concern is that the work should include explainability in the form of Shapley or lime plots. For Shapley-based explanations, this could include feature importance plots and summary plots as well as individual force plots.

**Response:** We appreciate the recognition and suggestions from the referee regarding our work. We understand your concern about model interpretability, especially the need for interpretability in the form of SHAP or LIME. In our work, we have indeed attempted some visualization techniques to explain the relationship between model representations and chemical environments as well as periodic trends. These visualization techniques aim to provide an intuitive way to comprehend the key factors in the model's prediction of material properties.

SHAP and LIME are two commonly used interpretability methods for explaining the predictions of machine learning models. In material model interpretation, SHAP values measure the contribution of each feature to the property predictions, helping to determine the impact of features on material property. LIME, on the other hand, is a local interpretability method that explains the property prediction of individual samples by generating similar samples and training an explanatory model, aiding in understanding the impact of key features in the property prediction of specific material samples. These methods offer a deep understanding of material property and feature impacts, facilitating the material design and optimization process. However, interpretability of features is generally a challenge problems for deep learning models. Although methods like SHAP and LIME can help identify which features play a crucial role in property predictions, for deep learning models like L-G-DCNN, the features represented by element embedding vectors often lack clear physical meanings, thus hindering clear explanations.

However, based on the fusion strategy of L-G-DCNN, we can decompose L-G-DCNN into LSTM and DCNN through a process similar to dismantling building blocks, providing more information on the model performance improvement process. For example, based on the OQMD band gap test dataset, we first filtered out 56 compounds containing Si elements that performed poorly in LSTM compared to L-G-DCNN. As shown in Figure 1, we compared the predicted band gap values of LSTM and L-G-DCNN along with the corresponding chemical formula labels. Subsequently, we compared and analyzed the EFM plots of the 56 compounds for LSTM and L-G-DCNN, obtaining (LSTM)-(L-G-DCNN) ΔEFM as shown in Figure 2. Through comparison, we can determine which pairwise elements interactions are most important for the band gap. Finally, in Figure 3, we summarized the most influential element interactions pairs on the band gap among these 56 compounds, revealing that most of the element pairs include elements with higher electronegativity such as oxygen (O), chlorine (Cl), and fluorine (F), which are favorable for forming strong covalent bonds and thus significantly impacting the band gap property. This aligns with our physical intuition, indicating that L-G-DCNN is an effective deep learning model for predicting material properties from chemical composition. Table 1 provides the actual values of the filtered 56 compounds, as well as all the predicted values, differences, and other raw data based on LSTM and L-G-DCNN.

Therefore, although conducting SHAP and LIME analyses on deep learning models may encounter challenges in feature interpretability, we believe that Figure 3 can clearly reflect which pairwise element interactions have a significant impact on the band gap, similar to the feature importance analysis of traditional machine learning models based on known feature engineering. The analysis in Figure 3 will help us understand the contribution of interactions between elements to material properties. In conclusion, we greatly appreciate and acknowledge the referee's questions regarding more interpretability analysis in the form of SHAP or LIME, and we have provided a similar explanation based on the fusion strategy and made corresponding modifications in the main manuscript.

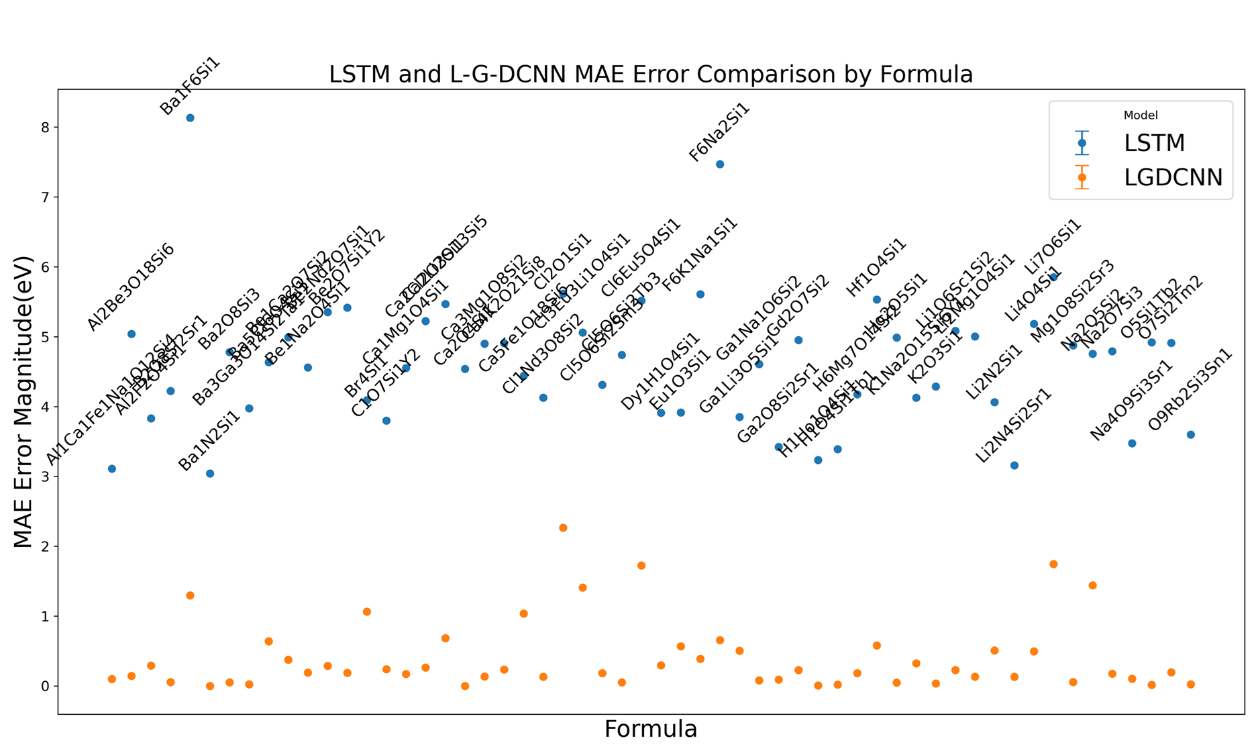


Figure 1 illustrates the comparison of predicted values for 56 Si-containing compounds with significant differences between LSTM and L-GDCNN on the OQMD band gap test dataset, with the chemical formula of each compound labeled in the figure.

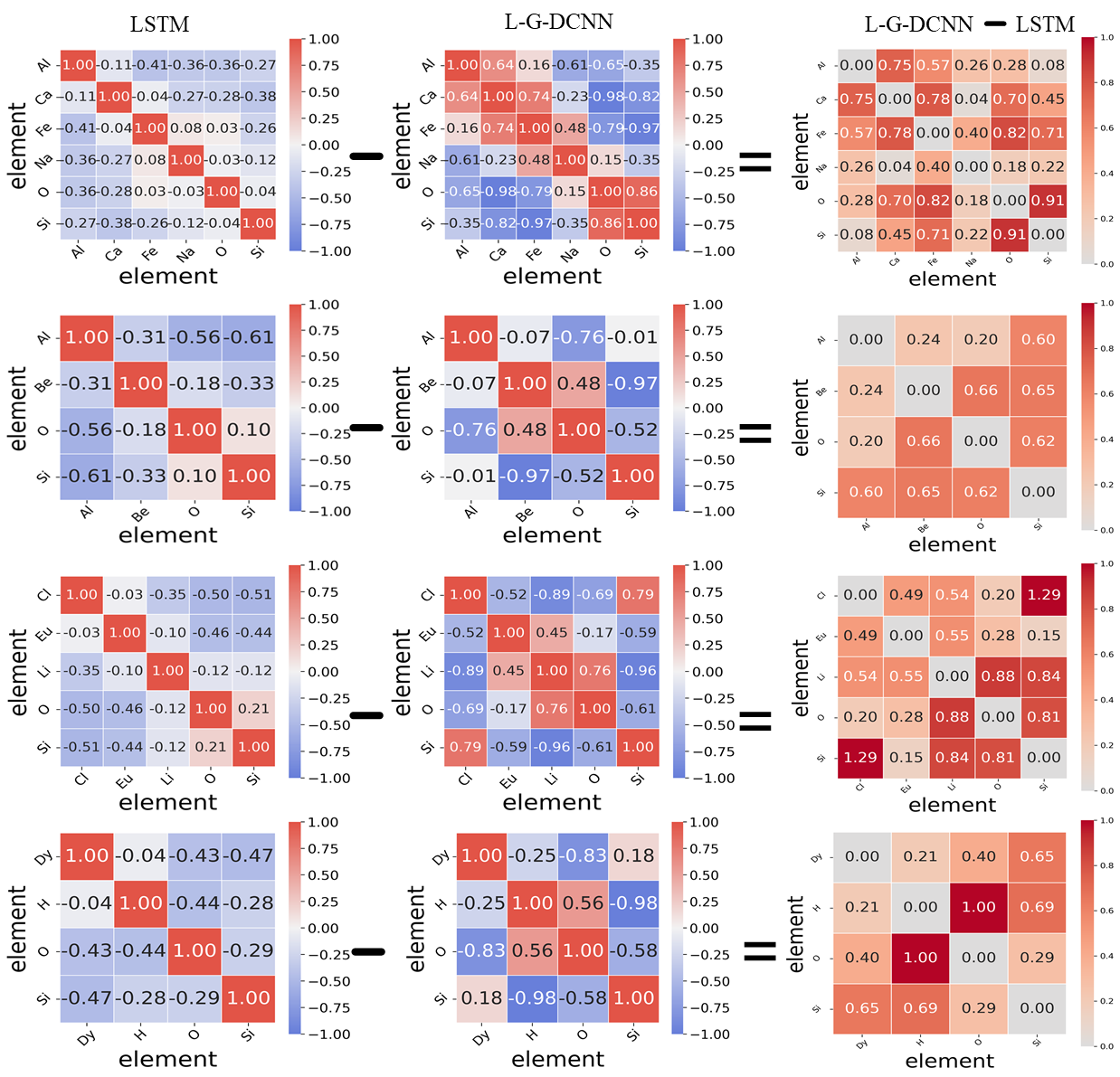


Figure 2 presents a comparison of the EFM plots of LSTM and L-G-DCNN for three example Si-containing compounds, along with the (LSTM)-(L-G-DCNN) ΔEFM. Through ΔEFM, we can determine which element interactions are most important for the band gap. The shading of ΔEFM reflects the significance of the element interactions pairs for the band gap.

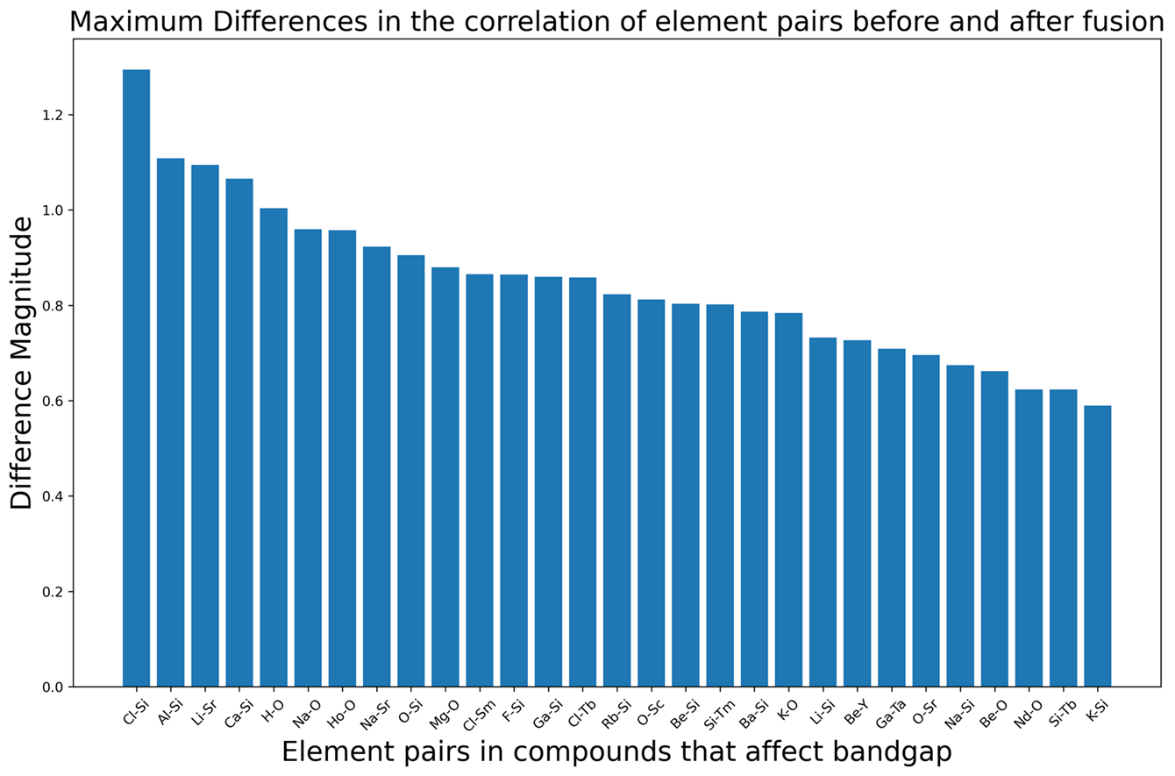


Figure 3 summarizes the most influential element interactions pairs on the band gap among the 56 compounds before and after model fusion. From the figure, it can be observed that most element interactions pairs involve elements with high electronegativity such as O, Cl, and F.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Formula | target | lstm\_pred | lgdcnn\_pred | error\_lstm | error\_lgdcnn |
| Na2O5Si2 | 4.754 | -0.00023 | 3.313114 | 4.754229 | 1.440886 |
| Cl1Nd3O8Si2 | 4.638 | 0.512336 | 4.771503 | 4.125664 | 0.133503 |
| Ca5Fe1O18Si6 | 4.444 | 0.003013 | 3.406679 | 4.440987 | 1.037321 |
| Hf1O4Si1 | 5.601 | 0.068076 | 5.022842 | 5.532924 | 0.578158 |
| F6Na2Si1 | 7.943 | 0.475034 | 8.59861 | 7.467966 | 0.65561 |
| Cl5O6Si2Sm3 | 4.711 | 0.399255 | 4.894914 | 4.311745 | 0.183914 |
| Ca4K2O21Si8 | 4.967 | 0.059111 | 4.729878 | 4.907889 | 0.237122 |
| Ca2Cl2O3Si1 | 5.382 | 0.159295 | 5.117125 | 5.222705 | 0.264875 |
| Cl2O1Si1 | 5.638 | 0.021166 | 3.371808 | 5.616834 | 2.266192 |
| Ga2O8Si2Sr1 | 4.133 | 0.711203 | 4.226184 | 3.421797 | 0.093184 |
| Be1Ca2O7Si2 | 5.221 | 0.228394 | 4.84486 | 4.992606 | 0.37614 |
| Cl6Eu5O4Si1 | 5.534 | 0.017281 | 3.808146 | 5.516719 | 1.725853 |
| Li4O4Si1 | 5.195 | 0.012667 | 4.697077 | 5.182334 | 0.497923 |
| K2O3Si1 | 4.361 | 0.075119 | 4.39543 | 4.285881 | 0.03443 |
| Li2N2Si1 | 4.071 | 0.007891 | 3.563954 | 4.063109 | 0.507046 |
| Mg1O8Si2Sr3 | 5.106 | 0.232572 | 5.049438 | 4.873428 | 0.056562 |
| Be1Na2O4Si1 | 4.872 | 0.313902 | 5.063966 | 4.558098 | 0.191966 |
| Ga1Li3O5Si1 | 3.886 | 0.035373 | 4.389283 | 3.850627 | 0.503283 |
| Ca1Mg1O4Si1 | 4.988 | 0.438792 | 4.817697 | 4.549208 | 0.170303 |
| Na4O9Si3Sr1 | 4.594 | 1.121436 | 4.699292 | 3.472564 | 0.105292 |
| H6Mg7O14Si2 | 5.13 | 0.954281 | 5.314725 | 4.175719 | 0.184725 |
| O7Si2Tm2 | 4.915 | 0.004996 | 5.109752 | 4.910004 | 0.194752 |
| Li7O6Si1 | 5.851 | -0.00028 | 4.106517 | 5.851278 | 1.744483 |
| Ba5Cl6O4Si1 | 4.742 | 0.108709 | 4.101628 | 4.633291 | 0.640372 |
| Ga1Na1O6Si2 | 4.653 | 0.046422 | 4.734563 | 4.606578 | 0.081563 |
| O5Si1Tb2 | 4.921 | -0.00034 | 4.935651 | 4.921344 | 0.014651 |
| Ca2O4Si1 | 4.541 | 0.003101 | 4.540303 | 4.537899 | 0.000697 |
| Ca3Mg1O8Si2 | 5.19 | 0.290707 | 5.052417 | 4.899293 | 0.137583 |
| Ba1F6Si1 | 8.183 | 0.049109 | 6.885648 | 8.13389 | 1.297352 |
| Ba3Ga3O14Si2Ta1 | 4.108 | 0.131982 | 4.08351 | 3.976018 | 0.02449 |
| Ca2Li2O13Si5 | 5.582 | 0.115343 | 4.898315 | 5.466656 | 0.683685 |
| H1Ho1O4Si1 | 4.853 | 1.62065 | 4.862513 | 3.23235 | 0.009513 |
| O9Rb2Si3Sn1 | 4.262 | 0.6632 | 4.237721 | 3.5988 | 0.024279 |
| Li2Mg1O4Si1 | 5.109 | 0.104453 | 5.24047 | 5.004547 | 0.13147 |
| C1O7Si1Y2 | 5.051 | 1.251145 | 4.811564 | 3.799855 | 0.239436 |
| Na2O7Si3 | 4.798 | 0.007625 | 4.621844 | 4.790375 | 0.176156 |
| K1Na2O15Si6Y1 | 5.058 | 0.932117 | 4.734559 | 4.125884 | 0.323441 |
| Cl5O6Si2Tb3 | 4.83 | 0.089065 | 4.777794 | 4.740934 | 0.052206 |
| Ba2O8Si3 | 4.812 | 0.034359 | 4.863309 | 4.777641 | 0.051309 |
| Ba1N2Si1 | 3.042 | 0.002042 | 3.040886 | 3.039958 | 0.001114 |
| Al2Be3O18Si6 | 5.594 | 0.556126 | 5.450961 | 5.037874 | 0.143039 |
| Eu1O3Si1 | 3.915 | 0.001533 | 3.348245 | 3.913467 | 0.566755 |
| Dy1H1O4Si1 | 4.861 | 0.951677 | 5.156366 | 3.909323 | 0.295366 |
| Cl3Eu3Li1O4Si1 | 5.612 | 0.550583 | 4.203979 | 5.061417 | 1.408021 |
| F6K1Na1Si1 | 7.791 | 2.18319 | 8.178376 | 5.60781 | 0.387376 |
| Be2Nd2O7Si1 | 5.367 | 0.013458 | 5.654284 | 5.353542 | 0.287284 |
| Li2N4Si2Sr1 | 3.761 | 0.603332 | 3.628573 | 3.157667 | 0.132427 |
| B2O8Si2Sr1 | 5.959 | 1.734906 | 5.902301 | 4.224094 | 0.056699 |
| H1O4Si1Tb1 | 4.784 | 1.394177 | 4.802793 | 3.389823 | 0.018793 |
| Gd2O7Si2 | 4.95 | -0.00023 | 4.723457 | 4.950231 | 0.226542 |
| Ho2O5Si1 | 4.986 | -0.00051 | 4.936392 | 4.986513 | 0.049608 |
| Al1Ca1Fe1Na1O12Si4 | 3.394 | 0.2837 | 3.492052 | 3.1103 | 0.098052 |
| Li1O6Sc1Si2 | 5.087 | 0.005365 | 4.860103 | 5.081635 | 0.226897 |
| Al2F2O4Si1 | 6.35 | 2.517638 | 6.058063 | 3.832362 | 0.291937 |
| Be2O7Si1Y2 | 5.447 | 0.031457 | 5.258036 | 5.415543 | 0.188964 |

Table 1 provides the actual values of the filtered 56 compounds, as well as all the predicted values, differences, and other raw data based on LSTM and L-G-DCNN.

**Manuscript Update:** In our revised manuscript, we have added a new section titled "Identifying critical elemental interactions pairs for band gap prediction" on page 20, and have included Figure 11, which is highlighted in blue font.

## REFERENCES

(1) Ballakur, A. A.; Arya, A. Empirical Evaluation of Gated Recurrent Neural Network Architectures in Aviation Delay Prediction. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 14-16 Oct. 2020, 2020; pp 1-7. DOI: 10.1109/ICCCS49678.2020.9276855.

(2) Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *ArXiv* **2014**, *abs/1409.2329*.

(3) Chen, Y. Convolutional Neural Network for Sentence Classification. 2015.

(4) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2015**, 770-778.

(5) Vaswani, A.; Shazeer, N. M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Neural Information Processing Systems*, 2017.

(6) Zhang, Y.; Li, K.; Li, K.; Zhong, B.; Fu, Y. R. Residual Non-local Attention Networks for Image Restoration. *ArXiv* **2019**, *abs/1903.10082*.

(7) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **2018**, *3* (29). DOI: 10.21105/joss.00861.

(8) Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9* (11).

(9) Goodall, R. E. A.; Lee, A. A. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. *Nature Communications* **2020**, *11* (1), 6280. DOI: 10.1038/s41467-020-19964-7.

(10) Wang, A. Y.-T.; Kauwe, S. K.; Murdock, R. J.; Sparks, T. D. Compositionally restricted attention-based network for materials property predictions. *npj Computational Materials* **2021**, *7* (1), 77. DOI: 10.1038/s41524-021-00545-1.

(11) Ihalage, A.; Hao, Y. Formula Graph Self-Attention Network for Representation-Domain Independent Materials Discovery. *Advanced Science* **2022**, *9* (18), 2200164. DOI: <https://doi.org/10.1002/advs.202200164> (acccessed 2023/12/26).

(12) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9* (8), 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.

(13) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **2017**, *60* (6), 84-90. DOI: 10.1145/3065386.

(14) Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Transactions on Neural Networks* **2009**, *20* (1), 61-80. DOI: 10.1109/TNN.2008.2005605.

(15) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. *npj Computational Materials* **2020**, *6* (1), 138. DOI: 10.1038/s41524-020-00406-3.

(16) De Breuck, P.-P.; Evans, M. L.; Rignanese, G.-M. Robust model benchmarking and bias-imbalance in data-driven materials science: a case study on MODNet. *Journal of Physics: Condensed Matter* **2021**, *33* (40), 404002. DOI: 10.1088/1361-648X/ac1280.

(17) Choudhary, K.; Garrity, K. F.; Reid, A. C. E.; DeCost, B.; Biacchi, A. J.; Walker, A. H. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.; Centrone, A.; et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *NPJ COMPUTATIONAL MATERIALS* **2020**, *6* (1). DOI: 10.1038/s41524-020-00440-1.

(18) Seeger, M. W. Gaussian Processes For Machine Learning. *International journal of neural systems* **2004**, *14 2*, 69-106.

(19) Xian, Y.; Dang, P.; Tian, Y.; Jiang, X.; Zhou, Y.; Ding, X.; Sun, J.; Lookman, T.; Xue, D. Compositional design of multicomponent alloys using reinforcement learning. *Acta Materialia* **2024**, *274*, 120017. DOI: <https://doi.org/10.1016/j.actamat.2024.120017>.

(20) Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. A. Playing Atari with Deep Reinforcement Learning. *ArXiv* **2013**, *abs/1312.5602*.