JNTM JOURNAL of NEW TECHNOLOGY and MATERIALS

Journal of New Technology and Materials

Vol. 15, No. 1 (2025), 9–22

RESEARCH ARTICLE

Concordance-Gated Multimodal Routing for Foundation-Model Selection in Materials Discovery

Simon Weinstock[1,*]

[1]Carnegie Mellon Software Engineering Institute

*Correspondence: wsimon@sei.cmu.edu

Received date: November 11, 2024; Accepted date: March 12, 2025

Abstract

Materials discovery becomes more dependent on the use of foundation models that must process scientific language, molecular graphs, crystal and atomistic structures, spectra, images, and reaction sequences. An essential challenge is not only achieving sufficient accuracy upon training, but also the prior choice of a model architecture to employ for each individual materials-discovery task. This study proposes *concordance-gated multimodal routing* (CGMR), an interpretable approach to architecture selection based on quantification of a small matrix representing each task's data type and modalities as well as available model architectures. Four types of materials-discovery operations are considered: data extraction, property prediction, molecule generation, and synthesis prediction. In addition to their modality vectors, all models are represented by binary architecture vectors. Three indices are computed: Modality Breadth Index, Architecture Coupling Index, and task–architecture Concordance Score. These indices form the CGMR score that reflects the routing difficulty of each task and maintains a meaningful scientific distinction among recognition, prediction, generation, and sequence-to-sequence translation. The proposed methodology provides a direct answer to the main question posed by the paper: a compact descriptor table can serve as the basis for quantitative architecture selection in foundation modeling provided modality breadth is taken into account alongside architecture direction. For the four material-discovery tasks, property prediction achieves the highest score (0.79), since it employs multiple types of input data from several model types and involves both encoder and encoder-decoder translation pathways. Data extraction and molecule generation receive identical scores (0.47); the former entails encoder-based recognition, whereas the latter relies on decoder-centered generation. Synthesis prediction gets the smallest score (0.36); this does not imply its simplicity in chemical terms, but rather the narrowness of its descriptor pathway aligned with encoder-decoder routing.

Keywords: materials informatics, foundation models, multimodal learning, architecture routing, property prediction, molecular generation, synthesis planning, interpretable model selection

1 Introduction

Material discovery has been accelerated by the development of computational approaches to materials informatics and engineering. Modern research projects combine high-throughput simulations, experimental libraries, chemical databases, synthesis records, spectroscopy collections, patents, and laboratory images. This information is heterogeneous in its format and scientific meaning. A crystalline structure conveys different information compared to a synthesis description, Raman spectrum, molecular graph, or property table. That is the reason why the challenge is not to insert more data into a bigger computational architecture. The challenge is to select a model architecture that matches the scientific operation carried out: recognition, inference, generation, or translation.

Machine learning is a well-established technique in modern materials informatics research. Data-driven works have already proved that compositional features, crystalline structure, and descriptors could serve as predictors for formation energy, band gap, mechanical properties, molecular properties, and relations between processes, structure, and properties provided that a good representation is chosen [1, 2]. Tools like the Materials Project and pymatgen allowed the development of high-throughput materials analysis and made it computationally tractable [3, 4]. Later benchmarking studies demonstrated the importance of task-specificity in materials model comparison [5]. All these advances play a significant

role in the current study since architecture selection is also a task-specific procedure. A model that performs well at one material task might be misaligned to another one if the input representation, output form, and scientific operation vary.

Architecture and representation cannot be decoupled since they are tightly related. Crystal graph convolutions were proposed as an architecture that allowed treating crystals as graph networks for property predictions [6]. SchNet has shown that continuous-filter convolution allows distance-based interaction learning to become a central concept in atomistic and molecular learning [7]. MEGNet introduced the idea of connecting atomistic environment with state properties, while ALIGNN has demonstrated the importance of using angular information in line graph message passing [8, 9]. Finally, M3GNet proved the efficiency of applying graph deep learning for potential building and materials screening [10]. All these architectures reflect the idea that scientific hypotheses should be considered when constructing a model, since a property-dependent feature should also be taken into account.

Scientific text extraction and chemical information extraction is another relevant line of works. The increasing number of scientific publications in chemistry made automatic chemical and materials information extraction necessary because a full manual curation process is impossible [11]. Pretrained embeddings were used to demonstrate that latent chemical properties can be extracted from text. Domain-specific language models have shown that they allow achieving a higher level of accuracy in materials information extraction [12, 13]. This background is especially relevant since the materials information that can be extracted is not exclusively textual or visual but includes information in tables, paragraphs, figures, and captions. This makes chemical information extraction a recognition and association task rather than a free-form generation task.

Chemical molecule generation and reaction prediction form another branch of materials informatics research. The language for chemical strings like SMILES became a starting point for chemical information representation and generation [14, 15]. Generative and latent variable models demonstrated that molecule search is possible in continuous or conditioned design spaces [16]. However, reaction prediction and retrosynthesis are much closer to translation tasks since there is no generation but mapping from one chemical state to another [17, 18]. Such a distinction is essential since decoder-centric generative architectures and encoder-decoder translators perform a different computational job even if both provide molecule representations as output. A molecule generator suggests candidate chemical compounds. A synthesis predictor finds reaction-like operations given target or reactant chemical states.

Foundation models intensify architecture challenges. Although their main advantage lies in a unified representation and reuse, they can blur the boundaries of task specificity [19]. It is not rare to see cases when a model is selected according to its popularity or language capabilities rather than due to material problem requirements. The risk of architecture misalignment is thus high. A text-only approach may ignore the presence of crystal structures or spectroscopy. A generator may propose chemically valid molecules but ignore their synthesis feasibility. A translator may include unnecessary layers in order to be used in encoding-decoding. Multimodal models may involve irrelevant input or output streams.

This paper seeks to address the issue of architecture misalignment through answering a concrete research question. The goal is to determine *whether a small tabular format of tasks, modalities, and architectures could be translated into a quantifiable and interpretable model selection methodology for foundation-model-based materials discovery.* The answer is Concordance-Gated Multimodal Routing (CGMR) method. It quantifies how wide is the evidence basis, how many architecture families should be used, and whether at least one of the selected architectures matches the direction of the operational flow. This makes CGMR a planning method rather than a training framework. The aim of CGMR is to guide the early design process of materials discovery models prior to actual pretraining and finetuning.

The contribution of this paper lies in the following aspects. First, it transforms four tasks of materials discovery into a normalized task–modality–architecture matrix. Second, it introduces interpretative indices like Modality Breadth Index, Architecture Coupling Index, and Concordance Score. Third, it combines these indices into a CGMR score and analyzes differences between data extraction, property prediction, molecular generation, and synthesis prediction. Fourth, it discusses all results and conclusions as materials-informatics model design choices instead of just numbers.

2 Materials and Methodology

2.1 Task descriptor construction and normalization

Categorical task descriptors used in this study include four material discovery tasks: data extraction, property prediction, molecular generation, and synthesis prediction. In each of the tasks, a unique combination of active evidence modalities, an architectural family, and an application is presented. Task descriptors have been designed with minimalism in mind because this work aims at determining if a concise task descriptor set can be transformed into a comprehensible routing method. Laboratory experiments are not performed since this work focuses on routing methods, not material synthesis or property measurement.

Table 1 is a descriptor set that serves as the empirical basis for CGMR calculation. Its importance goes beyond listing four task names. It highlights the scientific use of each task separately from the surface modality name. Text, for instance, occurs in all four tasks but fulfills very different purposes: it is involved in carrying entities and relations for data extraction, metadata for property prediction, conditioning information for molecular generation, and reaction language

for synthesis prediction. This interpretation ensures that the task matrix will not simply quantify the number of evidence channels.

Table 1. Normalized task–modality–architecture matrix for CGMR.

Task	Modalities	Architecture family	Representative materials-discovery use
Data extraction	Text, image	Encoder	Extraction of molecular names, property values, processing conditions, and document-level relations from articles, patents, tables, or figures.
Property prediction	Text, graph, structural, spectroscopic	Encoder; encoder–decoder	Prediction and contextual explanation of materials or molecular properties from structured and unstructured scientific evidence.
Molecular generation	Text, graph	Decoder	Generation of chemically valid candidate structures conditioned on target function, composition, property goal, or design constraint.
Synthesis prediction	Text	Encoder–decoder	Translation from targets, precursors, reaction statements, or procedure descriptions into synthesis steps, products, or retrosynthetic actions.

2.2 Modality and architecture encoding

Define the modality universe as

$$\mathcal{M} = \{\text{text}, \text{image}, \text{graph}, \text{structural}, \text{spectroscopic}\}. \tag{1}$$

Eq. (1) enumerates evidence modalities that were used in the current CGMR task analysis. While this list might seem arbitrary and overly narrow (e.g., not including diffraction or process log data), it was designed in a way to facilitate the subsequent analysis. Namely, Eq. (1) includes only evidence channels observed in the four considered tasks.

Modality vector for task t_i is

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{i|\mathcal{M}|}], \qquad x_{ij} = \begin{cases} 1, & \text{if modality } m_j \text{ is used by task } t_i, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Eq. (2) represents the description of each task using a binary vector. It can now be said that two tasks have different applications if they require different evidence modalities, even if the outputs are dissimilar. As an example, data extraction and molecule generation both use two evidence channels: data extraction uses text and image for recognition, whereas molecule generation uses text and graph for production.

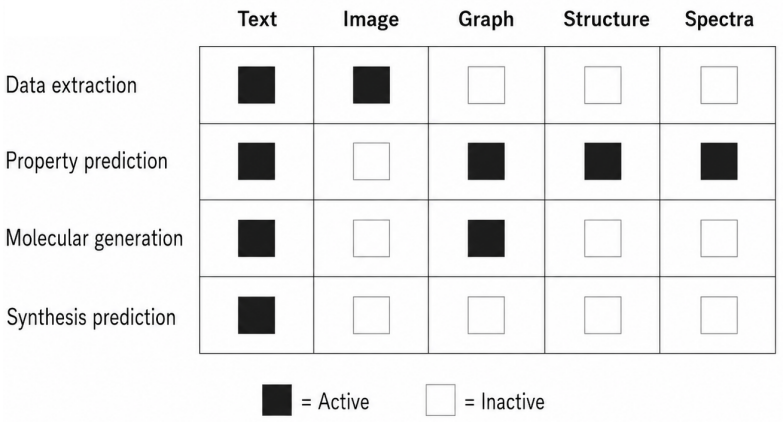

Figure 1. Binary activation matrix of modality vectors for four tasks. Active evidence channels are represented by dark cells.

The visualization of task evidence channels can be seen in Figure 1. One may conclude from this matrix that there are universal and specialized evidence channels for a task descriptor. Text is the universally active channel in all the tasks, while graph is shared between property and molecule prediction. The remaining modalities are used in more specialized tasks. This observation is important in relation to designing materials models because it implies that the model should not incorporate additional channels unless they are scientifically relevant for the task.

$$\mathcal{A} = \{\text{encoder}, \text{decoder}, \text{encoder–decoder}\}. \tag{3}$$

Eq. (3) introduces architecture families according to the computational direction. Activation of an encoder means recognition and/or representation learning. Activation of a decoder denotes output generation (in the generative sense). Activation of an encoder–decoder pair denotes translation from an input state to an output state. The choice of the architecture direction is crucial because the representation of materials may depend on whether a particular task is recognition, inference, or generation-related.

Architecture vector for task t_i is

$$\mathbf{z}_i = [z_{i1}, z_{i2}, z_{i3}], \qquad z_{ik} = \begin{cases} 1, & \text{if architecture } a_k \text{ is assigned to task } t_i, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Eq. (4) specifies the computational direction required by a particular task. Although some model variants may implement a particular task better than others, Eq. (4) reflects the first-order computation direction. For instance, property prediction requires both encoder learning and encoder–decoder translation: an encoder learns materials-property representations, and encoder–decoder generates explanations.

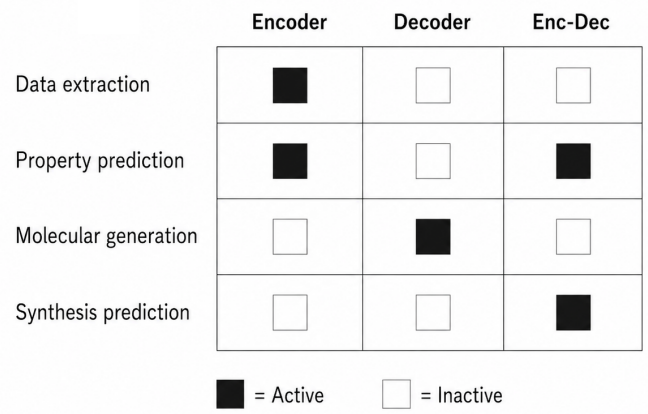

Figure 2. Activation matrix of architecture families for four tasks. Active computational families for each task are shown by dark cells.

It becomes evident from Figure 2 that each of the four tasks requires an exclusive direction of computing. Data extraction corresponds to encoder recognition, molecule generation to decoder production, and synthesis prediction to encoder–decoder translation. The property prediction family is placed in-between the other three because materials-property inference starts with representation but may also involve generation/translation from the representation. The figure therefore provides the architectural counterpart to the modality matrix in Figure 1.

2.3 Indices of CGMR

First, the Modality Breadth Index can be defined as:

$$\mathrm{MBI}(t_i) = \frac{\sum_{j=1}^{|\mathcal{M}|} x_{ij}}{|\mathcal{M}|}. \tag{5}$$

According to Eq. (5), MBI shows the proportion of active modalities against the entire modality space. High MBI indicates the need to collect evidence via multiple sources, and the route implementation may require modality-specific encoders, fusion layers, alignment losses, and cross-modal quality control. However, low MBI reflects that the route is narrower in evidence breadth, although it does not necessarily imply that the problem is easier from a scientific standpoint.

For instance, synthesis prediction has low MBI in the given matrix since it is implemented as text-based translation, yet chemically valid route planning is still quite challenging.

Second, the Architecture Coupling Index equals:

$$\mathrm{ACI}(t_i) = \frac{\sum_{k=1}^{|\mathcal{A}|} z_{ik}}{|\mathcal{A}|}. \tag{6}$$

According to Eq. (6), ACI shows how many architecture families are engaged in a specific task. If ACI is high, this means that the route implementation may involve more than one computational direction, e.g., representation learning and conditional output production. In the current matrix, the property prediction task has the highest value of ACI because this task involves both encoder and encoder-decoder routes. However, high ACI does not necessarily suggest that the task is hard to solve on a scientific basis.

Third, the task–architecture Concordance Score shows the consistency of routing between the computational route design and its purpose. In this study, each task and architecture is associated with a particular role: recognition/prediction = 0, translation = 1, generation = 2. Therefore, concordance can be calculated as follows:

$$\mathrm{CS}(t_i) = 1 - \frac{\min_{a_k : z_{ik}=1} |r(t_i) - r(a_k)|}{2}. \tag{7}$$

In Eq. (7), we evaluate whether there exists at least one selected architecture which supports the scientific operation of a particular task. Value one implies perfect routing, while lower values could reflect the misdirections of a chosen architecture; for example, the usage of solely generation route for a recognition task or translation task being interpreted as simple regression analysis. Such an inclusion is intended deliberately because, even with proper modality coverage, an inappropriate route cannot support a scientifically sensible solution.

Finally, the routing score itself:

$$\mathrm{CGMR}(t_i) = 0.55\,\mathrm{MBI}(t_i) + 0.30\,\mathrm{ACI}(t_i) + 0.15\,\mathrm{CS}(t_i). \tag{8}$$

As seen from Eq. (8), CGMR reflects all aspects of computing route design: evidence breadth, architecture coupling, and task-architecture consistency. MBI accounts for the largest weight because multimodal evidence integration is typically the main computational challenge in the development of materials foundation models. Next, ACI has a comparatively large weight due to additional challenges related to the implementation of hybrid routes. Last but not least, CS has the smallest weight in terms of numbers, yet it plays an important role: any routing is scientifically invalid if it does not coincide with the direction of a specific problem.

2.4 Protocol of computational evaluation

The proposed protocol consists of three descriptors' analysis. First, MBI, ACI, CS, and CGMR have to be calculated for each of the analyzed tasks. Second, ablation analysis of modality is executed by eliminating modalities sequentially and tracking corresponding tasks. Third, architecture–routing recommendations have to be made based on previously obtained data. This protocol has been deliberately created as a low-budget and interpretive solution to enable early planning before costly training, fine-tuning, deployment, and autonomous laboratory integration.

3 Results and Discussion

3.1 Encoded descriptor matrix

Normalized task records have been encoded as modality and architecture binaries. Table 2.

Table 2. Binary task encoding used for CGMR calculation.

Task	Text	Image	Graph	Structural	Spectroscopic	Enc.	Dec.	Enc.–Dec.
Data extraction	1	1	0	0	0	1	0	0
Property prediction	1	0	1	1	1	1	0	1
Molecular generation	1	0	1	0	0	0	1	0
Synthesis prediction	1	0	0	0	0	0	0	1

Table 2 offers the quantitative backbone of the paper's solution to the research problem. This table demonstrates that

task descriptors from Table 1 are sufficiently structured for quantitative analysis. As seen, the encoded pattern explains the high burden of property prediction route due to the activation of four evidence modalities and two architectural families. The synthesis prediction route is activated only once per each of the two criteria.

Based on the pattern encoded in Table 2 and visualized as Figure 1 and Figure 2, text turns out to be the only evidence modality activated in all four tasks. This universality does not mean that all materials discovery tasks belong to text-mining. Text is the carrier of scientific information with different functions depending on the type of task. In data extraction, it serves as a source of entities and relations; in prediction - as a source of properties or metadata; in generation - as an input of constraints or prompts; finally, in synthesis prediction - as a carrier of procedures or reactions.

Graphical evidence occurs in tasks of property prediction and molecular generation. The importance of this coincidence lies in the fact that graphs are the interface between forward inference and inverse design. In the former, graphs are used to define the connectivity and local chemical environment; in the latter, for validation of structure and candidate management. Spectral and structural evidence are found exclusively in property prediction tasks and thus represent the most multimodal of tasks presented here. Finally, image evidence is used in data extraction because materials-related documents contain a lot of information in visual form such as plots, drawings, micrographs, phase diagrams, etc.

3.2 CGMR score calculation and task ranking

The calculated MBI, ACI, CS, and CGMR values are reported in Table 3. The concordance score is one for all four tasks because the assigned architecture families are directionally consistent with their scientific roles. The ranking is therefore driven by modality breadth and architecture coupling.

Table 3. CGMR scores for the four materials-discovery tasks.

Task	MBI	ACI	CS	CGMR score
Data extraction	0.40	0.33	1.00	0.47
Property prediction	0.80	0.67	1.00	0.79
Molecular generation	0.40	0.33	1.00	0.47
Synthesis prediction	0.20	0.33	1.00	0.36

The scores in Table 3 yield the numerical conclusion of the study. Property prediction yields the highest CGMR score of 0.79 since it is the task that involves the broadest descriptor base together with the largest architectural coupling. Extraction and generation are both assigned 0.47, but this does not mean that the same architecture needs to be used. Instead, it shows that both tasks have equally wide descriptor bases. Meanwhile, their architectural directions need to be different. Synthesis prediction obtains 0.36, pointing to a narrow and well-defined route rather than a low-value scientific problem.

Property prediction receives the highest score because of the multiplicity of factors controlling a material's property. A molecular property can depend on chemical bonds and functional groups. A crystalline property depends on periodicity, defects, and symmetries. Measured responses can depend on the presence of particular spectroscopic signals and the specifics of processing conditions. Hence, the CGMR score reflects the need to consider property prediction through a fusion strategy. A combination of graph, structural, spectroscopic, and textual features cannot be collapsed into a single regressor since each of them is scientifically relevant.

The cases of data extraction and molecular generation illustrate the necessity to consider the concordance term. They both have an MBI of 0.40 and ACI of 0.33, but data extraction entails encoder-centered entity recognition, whereas molecular generation requires decoder-centered candidate identification. Without taking into account the architecture direction, these two tasks could look like each other. CGMR avoids such misconceptions by accounting for the difference in the architecture directions: Extraction involves entities, property values, relations between documents, and conditions; generation needs to generate candidate structures with respect to validity, novelty, and property-conditioned routes.

Finally, synthesis prediction receives the lowest score in terms of descriptor width because of its text-only encoder–decoder architecture. At the same time, a synthesis route is highly complex from the perspective of chemistry as it includes aspects of route feasibility, precursor availability, selectivity, reactions' mutual compatibility, stoichiometry, and procedural correctness. Therefore, one needs to interpret the results carefully: The low score reflects the specificity of synthesis as a chemical route rather than its value as a scientific endeavor.

Figure 3 clarifies that CGMR is a decision layer rather than a replacement for model training. The figure should be read from left to right: a task descriptor table is encoded, the indices are calculated, and the resulting scores guide route selection. This sequence is the core methodological contribution because it creates a transparent bridge between a qualitative materials-discovery problem and an implementable model-design plan. The workflow also shows where expert judgement enters the process: the scientific validity of the descriptors determines the usefulness of the routing output.

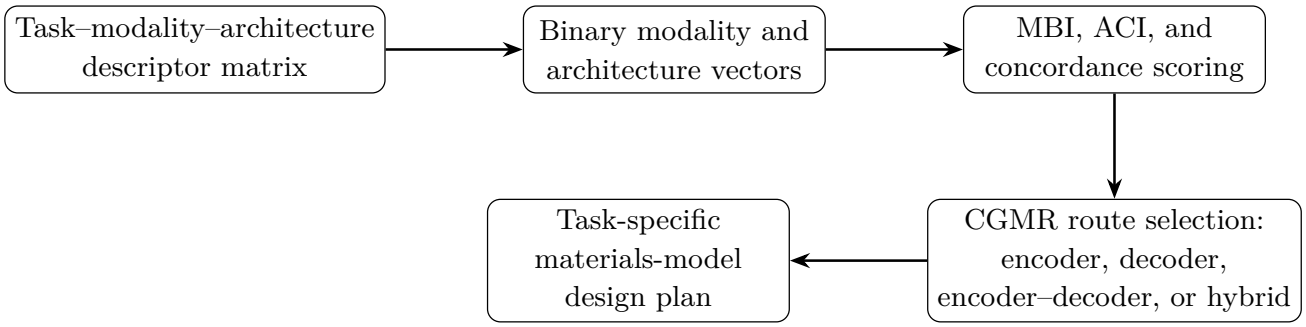

Figure 3. Workflow of the Concordance-Gated Multimodal Routing method. The method transforms task descriptors into interpretable routing scores and then uses the scores to guide model-family selection for materials discovery.

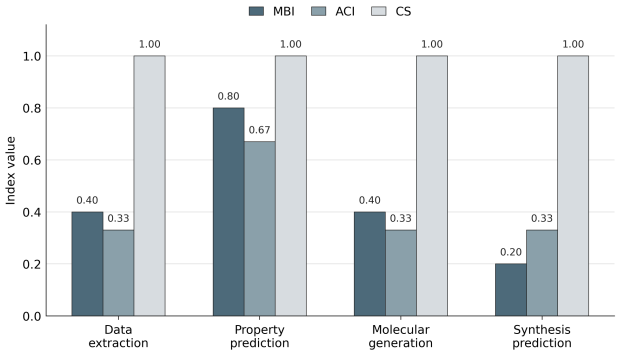

Figure 4. Component-index values used in CGMR calculation. MBI denotes Modality Breadth Index, ACI denotes Architecture Coupling Index, and CS denotes task–architecture Concordance Score. Property prediction has the largest component values for MBI and ACI, while all tasks retain full concordance.

Figure 4 separates the components before weighting. This matters because the final CGMR ranking is not a black-box score. Property prediction is high because both MBI and ACI are high. Synthesis prediction is low because MBI is low, not because the architecture is mismatched. The full concordance values across all tasks confirm that the descriptor assignments are internally consistent: each task has at least one route that agrees with its operational direction.

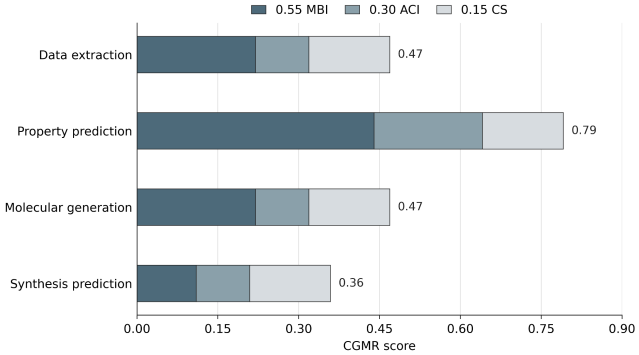

Figure 5. Weighted construction of the CGMR scores. The stacked bars show the contribution of the 0.55 MBI, 0.30 ACI, and 0.15 CS weights to each task score, with the final value reported at the end of each bar.

Figure 5 explains how Eq. (8) becomes the reported ranking. The modality term contributes the largest share for property prediction because four of five modalities are active. The architecture term also contributes strongly because prediction activates both encoder and encoder–decoder routes. For data extraction and molecular generation, the stacked contributions are numerically similar, but the later route assignment remains different because the active architecture family is different. This figure therefore reinforces a key finding: numerical score and architecture direction must be interpreted together.

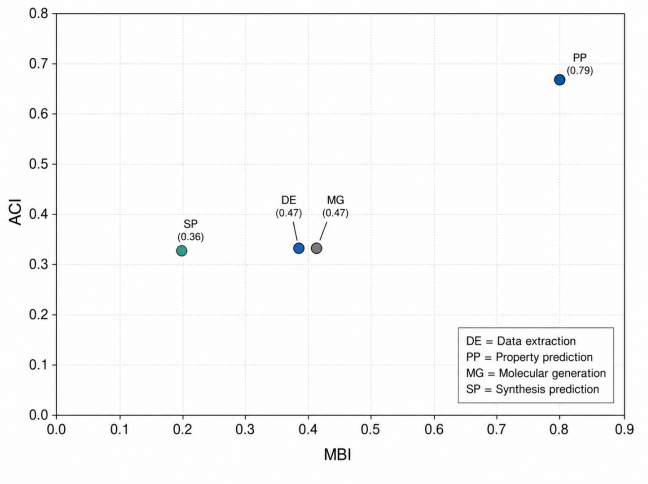

Figure 6. MBI–ACI routing map for the four CGMR task records. The scatter plot separates broad hybrid property prediction from narrower synthesis translation and from the two intermediate routes represented by data extraction and molecular generation.

Figure 6 places each task in the two-dimensional space formed by evidence breadth and architecture coupling. This representation is useful because it distinguishes different reasons for model-design burden. Property prediction is high in both dimensions, so it requires multimodal integration and hybrid route planning. Synthesis prediction is lower in modality breadth but still has a defined translation route. Data extraction and molecular generation sit in the intermediate region, showing comparable descriptor complexity but different scientific outputs. The scatter plot therefore provides a compact diagnostic map for early project planning.

Taken together, Table 3 and Figures 3–6 answer the numerical part of the research question. The compact table is not merely descriptive; once encoded, it produces a stable and interpretable ranking of route burden. The interpretation is materials-specific: high CGMR means that the workflow requires broader evidence integration and stronger architecture coordination, not simply that a larger model should be chosen.

3.3 Modality ablation

The modality-ablation analysis identifies which tasks are affected when each modality is removed. The results are reported in Table 4.

Table 4. Effects of removing each modality from the descriptor matrix.

Removed modality	Affected tasks	Interpretation
Text	Data extraction; property prediction; molecular generation; synthesis prediction	Text is the common scientific communication channel across all examined tasks, although its role changes by operation.
Image	Data extraction	Image evidence is specialized for document-level recognition where figures, plots, structures, or visual tables contain non-textual information.
Graph	Property prediction; molecular generation	Graph representation connects predictive and generative molecular tasks by encoding topology, connectivity, and local chemical environment.
Structural	Property prediction	Structural evidence is critical when spatial arrangement, periodicity, symmetry, or crystallographic environment controls material response.
Spectroscopic	Property prediction	Spectroscopic evidence contributes measurement-linked signatures that support property inference, uncertainty assessment, and experimental consistency.

Table 4 adds a second layer of interpretation beyond the score ranking. It shows that modalities differ in reach and specificity. Text is globally active, graph is a bridge between prediction and generation, while image, structural, and spectroscopic data are task-selective. This table is important because it prevents an overly simple conclusion that more modalities always improve a foundation model. The relevant question is whether the removed modality changes the scientific content needed by the task.

Figure 7 visually confirms the ablation logic in Table 4. The dense text column indicates that language is a shared carrier of materials information, but the sparse columns for image, structure, and spectra show that specialist modalities

should be deployed selectively. The figure is particularly useful for avoiding over-engineered multimodal models. If a modality does not alter the active task row, including it may increase preprocessing cost and alignment difficulty without improving the scientific route.

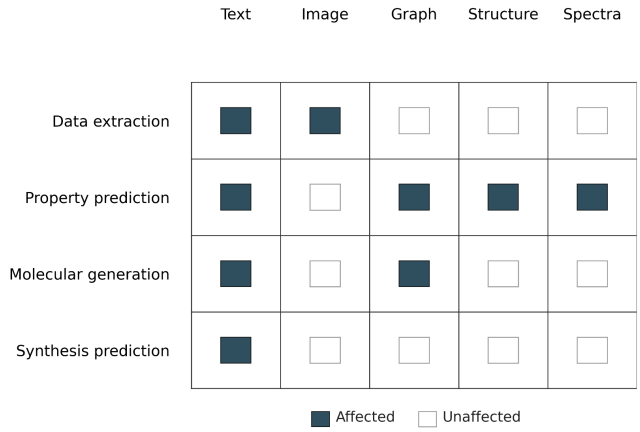

Figure 7. Task-level modality-ablation map. Filled cells identify task–modality pairs affected by removing an evidence channel, while open cells indicate unaffected pairs. The map confirms that text is globally active, whereas image, structure, and spectra have task-specific roles.

Figure 8 converts the ablation map into a count-level summary. Text affects all four tasks, graph affects two, and the remaining modalities affect one task each. This count should not be interpreted as a measure of scientific importance in general; it is a measure of importance within the defined descriptor matrix. For example, spectra affect only property prediction here, but within a spectroscopy-rich property study that single affected task may be decisive. The figure therefore supports task-specific modality prioritization rather than universal modality ranking.

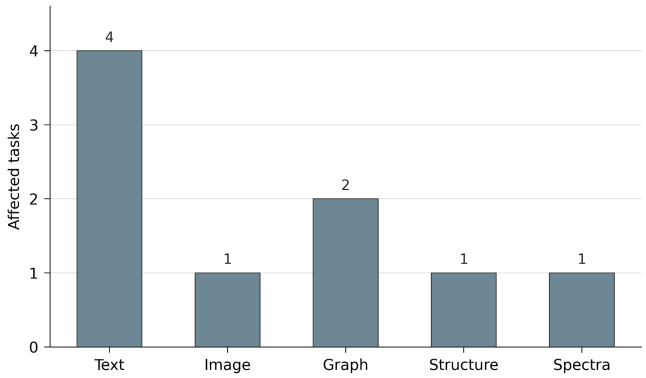

Figure 8. Number of CGMR task records affected by removal of each modality. Text has the largest impact because it participates in all four tasks; graph affects two tasks; image, structure, and spectra each affect one task in the present descriptor matrix.

The ablation results have practical consequences for foundation-model design. In a closed-loop materials-discovery workflow, text encoders may act as shared infrastructure because they support extraction, prediction context, design prompts, and synthesis language. Graph modules may be shared between property prediction and molecular generation, particularly in inverse design. Structural and spectroscopic encoders should be introduced when the prediction target depends on geometry, periodicity, or experimental response. Image encoders should be introduced when document figures or visual experimental records contain information not recoverable from text alone.

3.4 Architecture-routing recommendations

The CGMR routing recommendations are reported in Table 5. Each recommendation follows from the combination of task operation, modality breadth, and architecture concordance.

Table 5 is the practical output of CGMR. It translates numerical descriptors into model-design actions. The table also shows why the method is not a generic ranking scheme: each task receives an architecture route linked to its scientific operation. Data extraction is recognition, property prediction is multimodal inference with possible structured output,

molecular generation is candidate production, and synthesis prediction is chemical translation. These route assignments answer how the same compact descriptor table can guide foundation-model selection.

Table 5. CGMR-derived architecture routes for materials-discovery tasks.

Task	Recommended route	Design implication
Data extraction	Encoder-centered multimodal recognition	Combine text and image encoders with alignment layers for chemical entities, property values, table fields, figure labels, and document-level relations.
Property prediction	Hybrid encoder and encoder–decoder routing	Use modality-specific encoders for text, graph, structural, and spectroscopic inputs; add conditional output modules when explanation, prompt conditioning, or structured prediction is required.
Molecular generation	Decoder-centered conditional generation	Use chemical-string or graph decoders with validity control, property conditioning, novelty filtering, and post-generation screening.
Synthesis prediction	Encoder–decoder translation	Map targets, precursors, or reaction contexts into synthesis steps, routes, products, or retrosynthetic actions using sequence-to-sequence modeling.

Figure 9 condenses the complete routing decision into a design board. It shows not only which architecture is selected, but also what type of output the route should produce. This is valuable because materials-informatics workflows often involve multiple linked models. A literature-mining module may feed a property predictor; a property predictor may screen generated candidates; and a synthesis planner may translate selected candidates into feasible actions. The board therefore supports modular workflow design rather than a single monolithic foundation model.

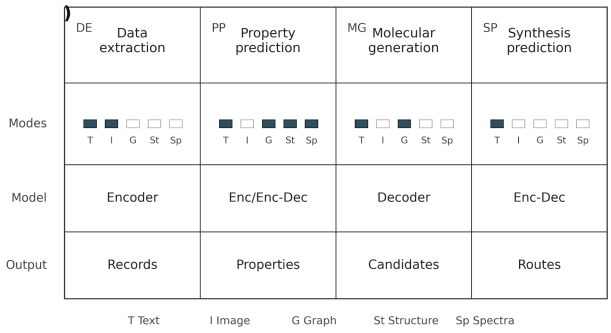

Figure 9. CGMR model-assignment board for the four materials-discovery tasks. The board condenses each task into active modes, selected model route, and output type, linking descriptor encoding to implementable foundation-model design.

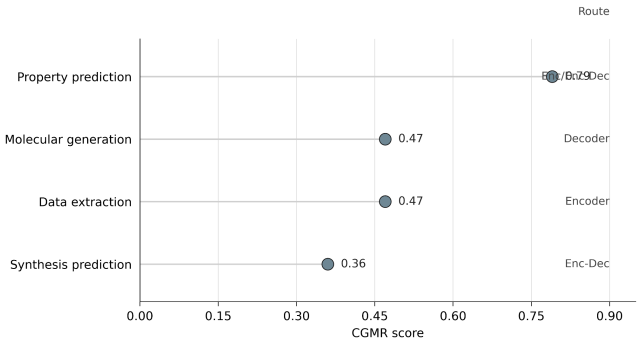

Figure 10. CGMR score–route panel. Each task is shown with its calculated score and its recommended route, clarifying that equal numerical scores can still correspond to different architecture directions.

Figure 10 makes the central interpretive point of the study explicit. Data extraction and molecular generation have the same CGMR score, yet they occupy different architecture routes. The score describes routing burden; the route describes computational direction. A correct interpretation requires both. This figure therefore guards against the common error

of treating a numerical score as a complete architecture decision.

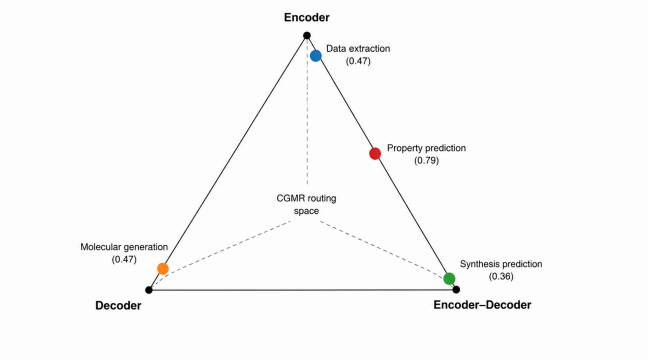

Figure 11. Triangular CGMR routing space. The vertices represent encoder, decoder, and encoder–decoder directions, while the task markers indicate the recommended model-route location and the corresponding CGMR score.

Figure 11 represents the architecture space as a two-dimensional shape. The triangular shape highlights that the four materials-discovery tasks do not fit on one line of complexity. Data extraction is closer to the encoder vertex, molecular generation is closer to the decoder vertex, synthesis prediction is closer to the encoder-decoder vertex, and property prediction occupies a hybrid position. The diagram is useful in planning multi-model systems because it shows how tasks may be distributed across various model routes within a single materials-discovery pipeline.

For data extraction, the recommended route is multimodal recognition around the encoder vertex. The goal is to detect entities and relations, so the architecture should not be generative. An example of such an architecture would include a domain language encoder and an image encoder, such that captions, tables, plots, molecule sketches, and contextual text could be combined. The route is particularly applicable to materials documents, since critical information tends to appear both in text and imagery.

For property prediction, the recommended route is hybrid. Since its score is 0.79, it should occupy a hybrid position in the architecture space of the triangle. Property prediction can be seen as the most challenging task of materials discovery because of the broad range of involved modalities and tight connection between modality and architecture components. A fusion route with modality-specific encoders will have to preserve scientific distinctions among graph, structural, spectral, and textual representations.

For molecular generation, the recommended route is conditional generation around the decoder vertex. The model will predict the output (a novel molecule), and thus, the route has to focus on validity, controllability, and screening of the candidates. The chemical string decoder may be fast, but the route will have to account for chemical validity of all generated strings. Alternatively, graph-based decoder may provide more flexibility in representing chemical structure, but it should also control the number of valence electrons and other chemical properties.

For synthesis prediction, the recommended route is encoder-decoder translation. The input is a chemical compound, procedure, or context. The output is an intermediate reaction step, final reaction outcome, list of precursors, or full chemical pathway. The difference from molecule generation is that the prediction must obey chemical reactions and conditions. The fact that its CGMR score is 0.36 indicates that the problem's descriptor is narrower, not easier. Reaction and route normalizations and condition controls are key priorities here.

3.5 Decision rules derived from the CGMR analysis

Table 6. Task-level decision rules derived from CGMR analysis.

Task	Primary CGMR finding	Model-design rule
Data extraction	Moderate score with encoder concordance	Use recognition-oriented encoders and align text with visual document evidence before extracting relations.
Property prediction	Highest score due to broad modality and architecture coupling	Build a fusion route with modality-specific encoders and preserve physical distinctions among graph, structure, spectra, and text.
Molecular generation	Moderate score with decoder concordance	Use controlled decoders and connect generation to validity, novelty, and downstream property screening.
Synthesis prediction	Lowest score but strong encoder–decoder concordance	Treat the problem as chemical translation and prioritize route feasibility and normalized reaction language.

The findings could be synthesized as a set of task-level rules. Table 6 outlines the rules based on the calculated scores, ablation patterns, and architectural routes.

Table 6 responds to the paper's research question. A concise descriptor table could indeed be used for quantitative routing in materials discovery. Each row in the table translates into a specific design rule. Of course, this score table should not be confused with actual performance. The purpose of CGMR is to provide the very first decision rule to help the researcher select the appropriate architecture family before any experimental validation or benchmark test.

3.6 Relevance to foundation-model design in materials science

The main benefit of using CGMR is interpretability. Each score can be mapped to either a modality or an architecture component. The interpretability will be beneficial for scientists working on designing foundation-models for materials discovery, since they will have a transparent way of justifying their choice of the route for further data collection, preprocessing, and training. The framework allows comparing various tasks in terms of their computational difficulty.

In this regard, property prediction requires the highest routing burden because of the broad evidence base and architecture requirements. Conversely, synthesis prediction has the lowest routing burden since its architecture has the narrowest route even though the underlying chemistry is complex. Data extraction and molecular generation examples show that the same descriptor complexity can imply quite different architectural requirements.

The framework is well-aligned with data-centric and knowledge-guided materials informatics. The qualities of descriptors, fidelity of their representations, and suitability to task requirements may sometimes be more significant for success than sheer model size. Large models have their benefits, but they are not always appropriate to a certain scientific problem because a model that does not properly account for the problem structure will simply fail. CGMR ensures that the very first architecture decision is explicitly defined, namely through choosing the scientific task and identifying relevant modalities.

The method could be applied to closed-loop material discovery. For example, an autonomous or semi-autonomous platform for materials design might include a document-mining module, a property-prediction module, a novel candidate generator, and a synthesis-planning module. Such a platform does not have to use a single architecture route. Instead, a modality-driven design can assign encoders for extraction, hybrid encoders for property learning, decoders for candidate generation, and encoder-decoders for synthesis prediction.

3.7 Limitations and future development

It is necessary to clarify several methodological limitations. First, the current matrix contains four task descriptions, so all scores should be understood as results for the described tasks rather than for all potential tasks of materials discovery. Second, each modality is treated as either present or absent, without measuring the volume of data, uncertainty, missingness, and other parameters. Third, the weightings in equation (8) were manually assigned. However, a more extensive study might learn these weightings empirically from the performance of multiple tasks.

The fourth limitation relates to architecture families, which were defined quite broadly. While CGMR uses only five categories, a more precise application might involve many types of encoders (e.g., transformers, equivariant neural nets), decoders (e.g., generators, seq2seq models), or encoder-decoders (e.g., transformers, autoregressive diffusion).

For future work, it is recommended to enlarge the descriptor universe for materials-discovery tasks. In addition to those studied here, the list might contain tasks for phase diagrams, defect prediction, microstructural segmentation, catalyst screening, polymer design, corrosion risk, processing optimization, and multi-fidelity simulation control. Also, it will be worthwhile to increase the modality taxonomy, which can cover, e.g., diffraction measurements, microscopy images, process logging, simulation results, and laboratory automation signals.

With additional effort, the framework could be developed by introducing data-quality penalties, uncertainty terms, cost-related indicators, and missingness markers. The method might also be improved by connecting it to active learning and changing routing decisions as new evidences arrive. The general principle is to develop CGMR until it becomes part of the closed-loop materials-discovery process.

Despite the mentioned limitations, the study showed that the compact materials-discovery table could be translated into a useful quantifiable decision tool. The key insight is that the table does not provide routing unless the scores are interpreted in combination with architecture direction. CGMR makes architecture choice interpretable and thus allows for better alignment between scientific goals and computational methods.

4 Conclusion

The paper introduced a new method of materials-discovery architecture selection based on the concepts of Modality Breadth Index, Architecture Coupling Index, and task–architecture Concordance Score. The method transforms a task-modality-architecture matrix into a vector representation, computes MBI and ACI, calculates the Concordance Score, and combines all three into a CGMR score. Four materials-discovery tasks were studied using the framework: data extraction,

property prediction, molecular generation, and synthesis prediction.

The research question posed is whether a compact task-modality-architecture table could be translated into an interpretable quantitative method for selecting foundation-model routes in materials discovery. The answer is positive, provided that one understands architecture direction. The framework provided quantitative estimates of routing difficulty, which helped understand why equally high scores may translate into different architecture directions. Thus, property prediction has the highest CGMR score of 0.79 and the most demanding architectural requirement, which is to use multimodal fusion architecture for preservation of modality distinction. Two other tasks, data extraction and molecular generation, have the same score of 0.47, but different architecture routes, i.e., encoder-centered recognition vs decoder-centered generation. Finally, synthesis prediction has the lowest score of 0.36 due to the narrower descriptor complexity, and thus, it requires encoder–decoder translation route.

In summary, the paper proposes that foundation-model architectures in materials discovery should be chosen task-specifically, i.e., based on the problem's scientific nature, evidences, and output type. CGMR helps in making an early choice between encoder, decoder, hybrid encoder, encoder-decoder, and multimodal fusion models.

References

[1] Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., and Kim, C. Machine learning in materials informatics: Recent applications and prospects. *npj Computational Materials* 3, 54 (2017).

[2] Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., and Walsh, A. Machine learning for molecular and materials science. *Nature* 559, 547–555 (2018).

[3] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* 1, 011002 (2013).

[4] Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source Python library for materials analysis. *Computational Materials Science* 68, 314–319 (2013).

[5] Dunn, A., Wang, Q., Ganose, A., Dopp, D., and Jain, A. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. *npj Computational Materials* 6, 138 (2020).

[6] Xie, T. and Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical Review Letters* 120, 145301 (2018).

[7] Sch"utt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and M"uller, K.-R. SchNet: A deep learning architecture for molecules and materials. *Journal of Chemical Physics* 148, 241722 (2018).

[8] Chen, C., Ye, W., Zuo, Y., Zheng, C., and Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials* 31, 3564–3572 (2019).

[9] Choudhary, K. and DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials* 7, 185 (2021).

[10] Chen, C. and Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* 2, 718–728 (2022).

[11] Swain, M. C. and Cole, J. M. ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature. *Journal of Chemical Information and Modeling* 56, 1894–1904 (2016).

[12] Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K. A., Ceder, G., and Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* 571, 95–98 (2019).

[13] Gupta, T., Zaki, M., Krishnan, N. M. A., and Mausam. MatSciBERT: A materials domain language model for text mining and information extraction. *npj Computational Materials* 8, 102 (2022).

[14] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* 28, 31–36 (1988).

[15] Krenn, M., Hase, F., Nigam, A., Friederich, P., and Aspuru-Guzik, A. SELFIES and the future of molecular string representations. *Patterns* 3, 100588 (2022).

[16] Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science* 4, 268–276 (2018).

[17] Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., and Lee, A. A. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science* 5, 1572–1583 (2019).

[18] Segler, M. H. S., Preuss, M., and Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555, 604–610 (2018).

[19] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., and others. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).