Federated Learning for Materials Property Prediction

Sadman Sadeed Omee[1], Md. Hasibul Amin[2]

University of South Carolina,

Email: [1]somee@email.sc.edu, [2]ma77@email.sc.edu

Abstract: Federated learning (FL) has emerged as a promising paradigm for collaborative machine learning, particularly in privacy-sensitive domains. This work presents the first application of FL to materials property prediction, addressing critical challenges in this field, including data privacy, decentralization, and data scarcity. By leveraging advanced graph neural network architectures such as DeeperGATGNN, SchNet, and MPNN, we explore the potential of FL to enable distributed training across diverse datasets held by different institutions. Experiments were conducted on benchmark datasets to evaluate model performance, out-of-distribution (OOD) datasets to test generalization capabilities, and scalability setups to evaluate system robustness under varying client configurations. Results demonstrate that FL achieves competitive performance on benchmark datasets while preserving data privacy. However, OOD evaluations reveal a persistent generalization gap, highlighting the complexity of materials data. Scalability experiments indicate that FL can scale up to 200 clients, achieving optimal performance before encountering resource constraints. Despite the limitations, this study provides valuable insights into the application of FL for materials discovery, paving the way for future advancements in privacy-preserving collaborative materials science.

Keywords: Federated learning, materials property prediction, graph neural networks, server-client communication, scalability

1 INTRODUCTION

The prediction of materials properties plays a pivotal role in the discovery and design of novel materials with desired functionalities [1, 7-8]. By accurately predicting these properties, researchers can identify promising candidates without the need for exhaustive experimental trials, significantly accelerating the innovation cycle. This is particularly critical in domains such as energy storage, semiconductors, catalysis, and structural materials, where time and resources are often constrained. Recent advancements in machine learning (ML), particularly graph neural networks (GNNs), have demonstrated their potential as powerful tools for materials property prediction [1-6]. By encoding complex relationships within material structures, GNNs can learn to predict properties such as bandgap, formation energy, elasticity, or dielectric constant directly from atomic configurations. This ability to extract meaningful representations of materials structures positions GNNs as a cornerstone in modern computational materials science.

However, despite the promise of ML in this domain, its application faces significant challenges. One of the most pressing issues is the vast discrepancy between the potential materials space and the relatively small number of documented materials in existing databases [11]. Theoretical estimates suggest that the possible materials space is practically infinite, with trillions of hypothetical materials that could exhibit unique and desirable properties. In contrast, only a tiny fraction of this space has been explored, characterized, and documented in public and proprietary databases. This data scarcity

restricts the training of robust ML models, which rely on large and diverse datasets to generalize effectively. Furthermore, the available datasets are often heterogeneous, derived from different experimental or computational sources, and may lack uniformity in data quality and representation.

Compounding this challenge is the distributed and proprietary nature of materials data. High-quality datasets are scattered across institutions, industries, and research labs, often protected by competitive or intellectual property concerns. These barriers hinder collaborative data sharing, limiting the development of models trained on comprehensive datasets. Privacy concerns are particularly pronounced in this context, as many organizations are unwilling to share raw materials data due to the risk of revealing sensitive or commercially valuable information. Consequently, traditional centralized approaches, where data is aggregated into a single repository for training, become impractical and undesirable due to privacy constraints, logistical challenges, and regulatory requirements.

Federated Learning (FL) offers a transformative solution to these challenges [12-15]. FL is a decentralized training paradigm that enables multiple clients to collaboratively train a shared model without exchanging raw data. Instead, each client trains a local model on their private dataset and only shares the model updates (e.g., gradients or parameters) with a central server. This decentralized approach preserves data privacy and ensures compliance with data-sharing regulations while enabling collaborative training. FL is uniquely suited to address the distributed nature of materials science data, where organizations such as research labs, industries, and universities hold distinct datasets. By employing FL, these organizations can collaboratively build powerful ML models without compromising the confidentiality of their data.

Recent materials property prediction benchmarks have demonstrated that graph neural networks (GNNs) are the state-of-the-art materials property predictors [7-8, 11]. For this reason, this study explores the application of FL in materials property prediction, focusing on the integration of advanced GNN architectures—DeeperGATGNN [3], SchNet [4], and MPNN [5]—into the FL framework. These architectures are particularly suited for modeling atomic and molecular systems due to their ability to capture intricate spatial, chemical, and electronic interactions. By leveraging FL, this work aims to assess the trade-offs between preserving data privacy, achieving high predictive performance, and ensuring scalability across a growing number of clients. The combination of FL with state-of-the-art GNNs is expected to unlock new opportunities for collaborative materials discovery.

The scope of this work includes experiments on multiple benchmark datasets that represent a range of material properties, including bandgap, formation energy, and mechanical properties. Additionally, the study investigates out-of-distribution (OOD) datasets to evaluate the generalization capabilities of FL models. OOD generalization is critical in materials science, as models are often required to predict properties for materials that are fundamentally different from those seen during training [22]. Furthermore, scalability experiments are conducted to assess the performance of FL systems under varying numbers of clients, from small-scale collaborations to large, distributed networks, measuring the impact on communication overhead, model convergence, and memory efficiency.

Preliminary results demonstrate that FL can achieve performance competitive with centralized training approaches while addressing privacy concerns. However, scalability experiments reveal challenges such as increased communication overhead and memory consumption, particularly when scaling to thousands of clients. These findings highlight the need

for further optimization in communication-efficient aggregation algorithms, client selection strategies, and resource allocation mechanisms.

This report aims to contribute to the growing body of knowledge on FL by demonstrating its applicability to materials property prediction. It identifies areas where FL excels and highlights opportunities for improvement. By addressing the unique challenges of this field, FL has the potential to revolutionize collaborative efforts in materials discovery, overcoming barriers posed by data silos and privacy concerns. This work also lays the groundwork for future studies to refine and expand the capabilities of FL in scientific and industrial applications, enabling breakthroughs in the discovery and design of advanced materials.

2 THE PROBLEM

2.1. INPUT-OUTPUT

The input and output for this problem are explained below:

- **Input:** The structure of a material, which provides crucial information about its atomic arrangement and elemental makeup. The structure consists of a *lattice*, a three-dimensional array of points representing the periodic arrangement of atoms, ions, or molecules. This lattice defines the geometry and symmetry of the crystal, with a *unit cell* as its smallest repeating unit. The lattice parameters (edge lengths and angles) describe the crystal's structure, forming patterns like cubic, hexagonal, or tetragonal arrangements. The structures are read, processed and encoded using the ASE [16] and PyMatGen [17] libraries.
- **Output:** The predicted properties of the material, which are essential for determining its suitability for various applications. Important materials properties include formation energy, band-gap, bulk and shear modulus, superconductivity, and thermal conductivity, etc.

Figure 1 shows explain the input and output of the materials property prediction problem.

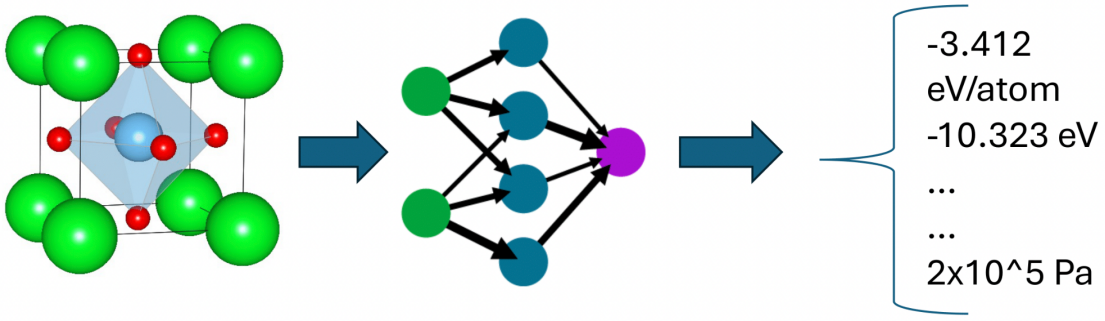


Figure 1: Input (materials structure, left) and output (materials properties, right) of the materials property prediction problem. ML models (middle) are generally employed to solve this problem.

2.2. WHY IS THIS PROBLEM INTERESTING?

- **Scientific Importance:** Accurate prediction of materials properties is critical for accelerating the discovery of new materials [7, 11]. This has a direct impact on advancing fields like renewable energy (e.g., high-

efficiency solar panels), electronics (e.g., low-resistance materials), and pharmaceuticals (e.g., drug discovery).

- **Practical Advantages of Federated Learning (FL):**
    - **Enhanced Model Performance:** FL enables the aggregation of models trained on distributed datasets across institutions. This approach effectively increases the diversity and size of the training data, leading to improved generalization and accuracy of predictions.
    - **Data Privacy Preservation:** Traditional data-sharing approaches are often impeded by concerns over intellectual property and confidentiality. FL allows organizations to collaborate without sharing raw data, maintaining privacy while benefiting from collective learning.
    - **Unexplored Opportunity:** While FL has been explored for molecular property prediction, it has not been widely applied to materials property prediction. This represents a novel application area with vast potential.

These factors underline the importance of solving this problem, not just as a technical challenge but also as a transformative approach to collaborative innovation in materials science. By leveraging FL, organizations and researchers can break through existing barriers and accelerate progress in materials discovery.

3 MOTIVATING SCENARIOS

The application of Federated Learning (FL) to materials property prediction presents transformative possibilities in both scientific research and industrial applications. Below are key motivating scenarios that highlight the importance and potential impact of this work:

I. Drug Discovery: In the pharmaceutical industry, the search for effective drugs requires screening a vast chemical space for compounds with desired properties. Companies rely on proprietary drug libraries, which they are reluctant to share due to intellectual property concerns. FL enables multiple companies to collaboratively train machine learning models, improving the prediction of drug efficacy without exposing proprietary data. This ensures faster and more accurate identification of promising candidates, accelerating the drug discovery process.

II. Renewable Energy: Energy companies face the challenge of designing high-efficiency solar panels and batteries. Materials with optimal electrical conductivity, thermal stability, or band-gap properties are critical for these advancements. Through FL, these companies can jointly develop predictive models for material discovery while keeping their experimental datasets confidential. This collaborative approach speeds up the innovation cycle for renewable energy technologies.

III. Vaccine Development: As seen in recent global health crises, vaccine development demands rapid identification of effective compounds and antigens. Pharmaceutical firms can use FL to predict vaccine efficacy for novel viruses. By leveraging shared computational insights without compromising the privacy of proprietary vaccine candidates, FL facilitates quicker and more effective vaccine design.

IV. Consumer Electronics: The development of high-performance consumer electronics, such as optical displays or advanced semiconductors, depends on discovering materials with properties like low refractive index or high electrical

conductivity. FL allows companies to jointly optimize material search algorithms, accelerating innovation while safeguarding competitive advantages and proprietary data about synthesized materials.

V. Finding Superconducting Materials: Research institutions and labs worldwide are in a race to discover superconducting materials that operate at room temperature, a breakthrough with revolutionary implications for energy transmission and storage. Using FL, these labs can pool their predictive models for superconducting properties without sharing sensitive experimental results, fostering global collaboration without compromising confidentiality.

VI. Economic and Time Efficiency: Traditional methods like density functional theory (DFT) are precise but computationally expensive and time intensive. Machine learning provides a promising alternative by trading off some accuracy for significant gains in speed. FL extends this capability by allowing organizations to aggregate models trained on distributed datasets, effectively enlarging the training dataset without sharing raw data. This reduces the dependency on DFT and expedites the material screening process, making the discovery of novel materials more accessible and efficient.

These scenarios illustrate how FL addresses both scientific challenges and industrial needs, fostering collaboration across sectors while maintaining data privacy and proprietary control. By leveraging FL, organizations can not only accelerate material discovery but also unlock transformative innovations across a range of fields.

4 RELATED WORKS

While federated learning (FL) has not been extensively applied to materials science, there have been notable applications in related domains such as molecular and protein property prediction. In these contexts, FL has shown promise in preserving data privacy while enabling collaborative model training across distributed datasets.

For instance, Zhu et al. [9] explored FL for molecular property prediction, focusing on the application of GNNs to learn molecular representations without sharing raw data, demonstrating the potential of FL to enhance performance while ensuring privacy. Similarly, Hausleitner et al. [10] applied FL to protein structure prediction, emphasizing its ability to integrate diverse data sources while maintaining confidentiality, which is crucial in fields with proprietary datasets.

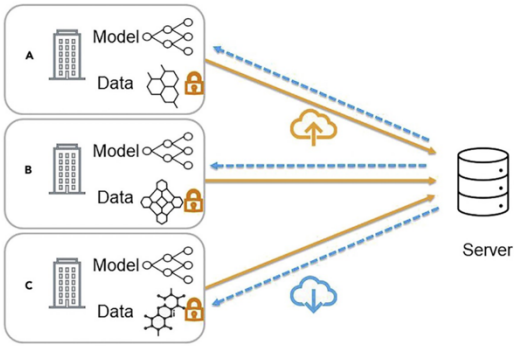
Our work draws inspiration from these approaches, particularly the FedChem [9] framework, which utilizes FL for molecular property prediction. FedChem integrates advanced GNN architectures into an FL setting to enable collaborative learning across distributed data sources. We have adapted this framework to materials property prediction, leveraging its privacy-preserving capabilities and scalability as a basis for our experiments. By extending FedChem's methodologies to materials science, we aim to address the unique challenges posed by this domain, such as the scarcity of high-quality data and the complexity of material properties.

Graph Neural Networks (GNNs) have become a cornerstone for materials property prediction due to their ability to effectively capture the structural and chemical properties of materials [7-8]. Various GNN architectures have been proposed to address different challenges in this domain. CGCNN [1] was among the first to encode material structures as graphs, where atoms are represented as nodes and bonds as edges, enabling predictions directly from crystal structures. MEGNet [2] extended this concept by integrating global state variables alongside atomic and bond features, enhancing its capacity to model diverse properties. DeeperGATGNN [3], with its deep graph attention mechanisms, further improved

prediction accuracy by emphasizing important atomic interactions. Similarly, SchNet [4] has been widely adopted for its ability to learn spatial and chemical representations of materials through continuous-filter convolutions, making it highly suitable for quantum mechanical property prediction. MPNNs [5] generalized the GNN framework by enabling flexible message-passing protocols that aggregate information from neighboring nodes, which is crucial for modeling complex material systems. ALIGNN [6] introduced a dual-graph representation by incorporating both atomic and bond graphs, capturing higher-order geometric features that are particularly relevant for material systems with intricate local environments. Together, these architectures represent a suite of powerful tools, each designed to tackle specific challenges in the prediction of materials properties, and provide a strong foundation for advancing computational materials science

5 SOLUTION

To address the challenges of privacy, data scarcity, and distributed datasets in materials science, this project implements a Federated Learning (FL) pipeline specifically designed for crystalline materials property prediction. While FL has been successfully applied to molecular property prediction, this pipeline extends its application to crystalline materials, which present unique complexities such as larger structural heterogeneity and a need for advanced data representations. Figure 2 shows a schematic for the FL pipeline and the general FL algorithm.



Input: # clients L, # local updates T, # Comms round C.
Output: Global Model F^g

Server initialize a global model F^g ▷ Server init.
while Communication Round < C **do**
  Server broadcasts F^g to clients
  $F^l \leftarrow F^g$ ▷ Client init.
  for l : 1 to L in parallel **do** ▷ Client Update
    for t : 1 to K **do** ▷ Update F^l for K steps
      Sample a minibatch $\{g_i^l, y_i^l\}_{i=1}^{B} \sim X^l$
      Update local model F^l by gradient descent
    end for
    Client sends updated model F^l to Server
  end for
  Server gets $F^g \leftarrow \sum_{l=1}^{L} \frac{|X^l|}{|X|} F^l$ ▷ Server Update
end while

Figure 2: (Left) An example FL system framework containing three clients and one "Server". Each client has one local dataset in this illustration, and the "Server" is used to receive and send the model parameters. (Right) The basic flow algorithm of an FL system.

5.1. FEDERATED LEARNING PIPELINE

5.1.1. SERVER INITIALIZATION

- The central server is the orchestrator of the FL process. It initializes a *global model* based on advanced Graph Neural Networks (GNNs) such as:
  - **DeeperGATGNN**: Designed for capturing intricate and long-range interactions in crystal structures.
  - **SchNet**: Optimized for both molecular and materials systems, employing continuous-filter convolutions.
  - **MPNN**: Uses message-passing techniques to extract relevant relational data for property regression tasks.
- The server does not have access to any client data. Instead, it focuses on managing communication and aggregation, ensuring the global model evolves through collaboration while preserving privacy.

5.1.2. CLIENT-SIDE TRAINING

- Each participating client, such as a research institution or a laboratory, retains its **local dataset**, which contains information about crystalline materials. These datasets are encoded as graphs where:
    - *Nodes* represent atoms.
    - *Edges* represent atomic interactions or spatial relationships (e.g., bonds or geometric proximity).
- Clients train the global model locally on their data using gradient descent. This training produces locally optimized model parameters that reflect the unique data distribution of each client.
- The data across clients is typically non-IID (non-independent and identically distributed)**,** as clients may specialize in different material types (e.g., semiconductors, alloys, or dielectric materials).
- For our experiments, we have both IID and out-of-distribution (OOD) datasets to test on.

5.1.3. MODEL UPDATES

- After completing local training, clients send model updates (e.g., weights, gradients) to the central server. This process ensures:
    - **Data Privacy**: No raw data is shared, protecting intellectual property and experimental findings.
    - **Efficient Communication**: Only the model parameters or gradients are transferred, reducing bandwidth requirements.

5.1.4. SERVER AGGREGATION

- The server aggregates the updates received from all clients using the *FedAvg* scheme. This aggregation computes a weighted average of the local updates, where the weights are proportional to the size of each client's dataset.
- This step allows the global model to benefit from the knowledge contained in all local datasets, improving its ability to generalize across diverse materials.

5.1.5. MODEL REDISTRIBUTION

- The server distributes the updated global model back to all clients. Each client incorporates the global model into its local training process, enabling further refinement based on its specific data.
- This iterative process of training, updating, aggregating, and redistributing continues over multiple communication rounds. Each round enhances the global model's accuracy and robustness.

5.1.6. FINAL GLOBAL MODEL

- After the specified number of communication rounds, the server produces a final global model. This model encapsulates the collective knowledge of all clients and is designed to perform well on a wide range of crystalline materials, even those outside individual datasets.

5.2. BENEFITS OF THE FEDERATED LEARNING PIPELINE

1. **Scalable Collaboration**:
    - The pipeline supports a flexible number of clients, ranging from a few institutions to hundreds, enabling broad participation in the materials science community.
2. **Enhanced Model Accuracy**:
    - By aggregating insights from diverse datasets, the global model achieves improved accuracy compared to models trained on isolated datasets.

3. **Data Privacy and Security**:
    - Institutions retain control over their data, addressing intellectual property concerns while benefiting from collaborative learning.
4. **Domain-Specific Adaptation**:
    - The pipeline extends FL's applicability from molecular to crystalline systems, incorporating tailored graph representations and GNN architectures.

6 EXPERIMENTAL SETUP

6.1. MODEL ARCHITECTURES

6.1.1. DEEPERGATGNN

DeeperGATGNN [3] is an advanced graph attention-based neural network designed to address critical limitations in conventional graph neural networks, particularly the over-smoothing issue that arises when stacking multiple layers. It extends the foundational Graph Attention Neural Network (GATGNN) architecture by integrating augmented graph attention (AGAT) layers, which employ multi-head attention mechanisms to enhance the learning of localized features within a crystal graph. These AGAT layers incorporate both node-level and edge-level features, allowing the model to capture complex atomic interactions effectively. Additionally, the architecture includes differentiable group normalization (DGN) [24], which stabilizes training in deeper networks and prevents the degradation of learned features as the network depth increases.

The model employs skip connections, inspired by ResNet [25] architectures, to retain critical feature information and ensure robust gradient flow through the network. DeeperGATGNN also incorporates a global attention layer, which aggregates localized features from AGAT layers into a global representation, crucial for predicting material properties such as formation energy and band gap. Its design supports scalability, allowing the model to operate with more than 30 graph convolution layers without significant performance degradation. This capability makes DeeperGATGNN highly suitable for extracting intricate relationships in large, complex crystal structures, achieving state-of-the-art performance in multiple materials science tasks. Figure 3 shows the DeeperGATGNN architecture.

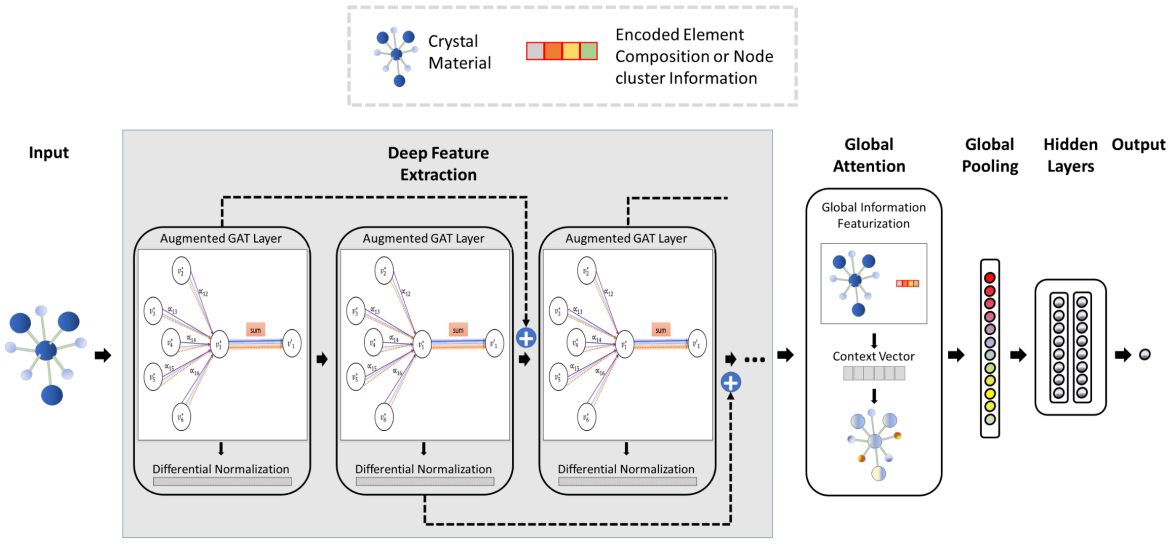


Figure 3: The DeeperGATGNN model architecture.

6.1.2. SCHNET

SchNet [4] is a neural network architecture explicitly developed for quantum chemistry and materials science, focusing on learning continuous representations of molecules and materials. Unlike conventional grid-based convolutional networks, SchNet employs continuous-filter convolutional layers that allow it to process data where atomic positions and spatial relationships are not confined to a grid. The architecture models atomic interactions by representing materials as a collection of atoms with associated feature vectors, which are updated iteratively through interaction blocks. Each interaction block incorporates a residual connection to enhance feature learning and ensure stability in training. The model captures rotational invariance in energy predictions and rotational equivariance in force predictions, aligning with fundamental quantum-mechanical principles.

SchNet uses a filter-generating neural network to produce continuous filters based on interatomic distances, enabling it to handle unevenly spaced atomic configurations. This approach ensures smooth energy predictions and physically consistent force fields, which are essential for accurate modeling of material behaviors. By leveraging its continuous-filter convolutional framework, SchNet excels in predicting total energy and interatomic forces for molecules and materials, demonstrating its versatility and robustness across a wide range of quantum-chemical benchmarks. Figure 4 shows the SchNet architecture.

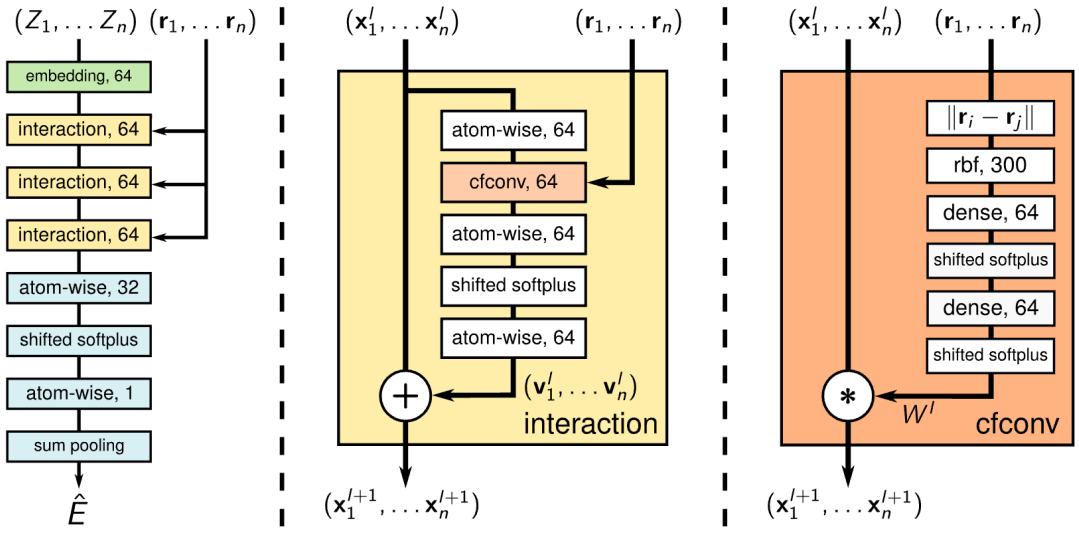


Figure 4: The SchNet model architecture.

6.1.3. MPNN

MPNN [5] offers a generalized framework for supervised learning on graph-structured data, including materials and molecules, by combining message-passing and readout phases. In the message-passing phase, information is exchanged between nodes based on their neighbors, with each node updating its representation using learned functions. This iterative process enables the model to capture localized atomic interactions within a crystal or molecular graph. The readout phase aggregates node-level features to produce a global representation for the entire graph, which is subsequently used for property prediction. The MPNN architecture is particularly well-suited for graph-based tasks because it naturally incorporates graph isomorphism invariance, ensuring consistent outputs regardless of node labeling.

MPNN extends traditional neural networks by allowing dynamic, learned representations of atomic interactions. For instance, edge features like bond strength or distances are used alongside node features, and both are updated iteratively. The flexibility in defining message functions and update mechanisms makes MPNNs highly adaptable for predicting material properties such as band gaps and thermal conductivity. Their scalability and capacity to learn complex dependencies in graph-structured data make them a cornerstone of modern machine learning applications in materials science and quantum chemistry. Figure 5 shows the MPNN architecture.

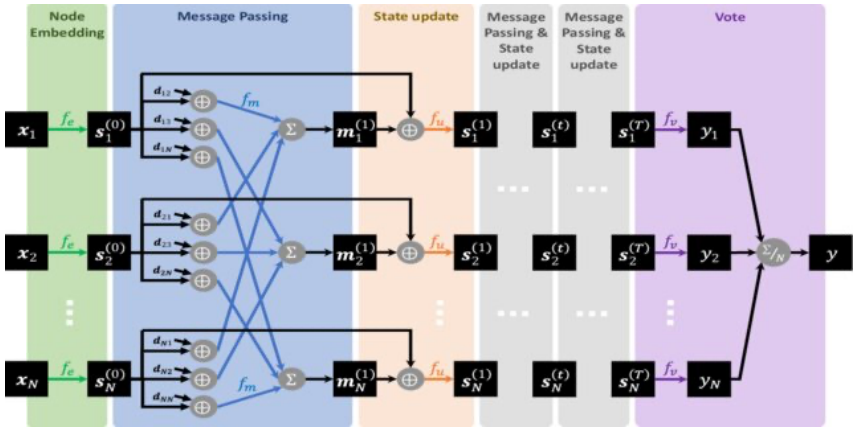


Figure 5: The MPNN model architecture.

6.2. DATASETS

To evaluate the effectiveness of federated learning (FL) for materials property prediction, experiments were conducted on diverse datasets representing a wide range of material properties. The datasets are categorized into three types:

6.2.1 BENCHMARK DATASETS

We use five benchmark datasets are widely used in the materials science community and serve as a reliable standard for comparing model performance We select the following datasets – formation energy [26], alloy surface [27], Pt-cluster [28], 2D materials [29], and band-gap [26]. The details are presented in Table 1.

Dataset	Property	# of Samples
Bulk Materials Formation energy	Formation Energy	36,839
Alloy Surface	Adsorption Energy	37,334
Pt-Cluster	Formation Energy	19,801
2D-Materials	Formation Energy	3,814
Band Gap	Band Gap	36,837

Table 1: Details of the five benchmark datasets used in this project.

6.2.2 OOD DATASETS

We use three different OOD datasets, each of which has 5 different folds/partitions, where the test set is systematically drawn from the training set. We utilize the OOD split method and datasets from [30]. The details are presented in Table 2.

Dataset	Property	# of Samples
Dielectric	Refractive Index	4764
Perovskites	Formation Energy	18,928
GVRH	Shear Modulus	10,987

Table 2: Details of the three OOD datasets used in this project.

6.2.3 SCALABILITY EXPERIMENT DATASETS

We use two datasets for this experiment and systematically increase the number of client nodes from 3 to 500 and observe the performance and memory consumption. The details are presented in Table 3.

Dataset	Property	# of Samples
Bulk Materials Formation energy	Formation Energy	36,839
Band Gap	Band Gap	36,837

Table 3: Details of the two datasets used for the scalability experiment in this project.

6.3. FL FRAMEWORK

- **Clients:** Default value 4 for benchmarking results, 5 for OOD results, and varying numbers (3, 4, 10... up to 500) for scalability testing.
- **Weight Update Scheme:** FedAvg.
- **Training Configuration:**
  - **Epochs:** 500
  - **Learning Rate:** 0.005
  - **Optimizer:** AdamW [31]
  - **Batch Size:** 64
  - **Graph convolution layers:** 2-5 (SchNet, MPNN), 20 (DeeperGATGNN)
  - **Latent embedding dimension:** 64
  - **Activation function:** SoftPlus [32] (DeeperGATGNN), ReLU [33] (SchNet, MPNN)
  - **Communication round:** 10

6.4. IMPLEMENTATION PLATFORMS:

6.4.1. FRAMEWORKS AND LIBRARIES:

Frameworks: PyTorch [18], PyTorch-Geometric [19], FedML [20] frameworks were utilized for implementing the project. Overall, the FedChem [9] repository was used as a base for implementing our project.

Libraries: PyMatGen [17], ASE [16], Scikit-learn [21], etc. libraries were used to process the materials dataset. Matplotlib [22], Seaborn [23], etc. libraries were used for the visualization.

6.4.2. HARDWARE:

The Hyperion cluster of the University of South Carolina were used to run the project. The code was run on the NVIDIA Tesla v100-32G GPUs.

7 RESULTS

We used Mean Absolute Error (MAE), which is a widely used metric in regression tasks to measure the average magnitude of errors between predicted values and the actual ground truth. MAE is used as the default performance metrics for materials property prediction tasks in most literature [1-8], so we also used it here. It is defined as the mean of the absolute differences between predicted values (y_{pred}) and true values (y_{true}), calculated as MAE = $1/n * \sum_{(i=1 \text{ to } n)} |y_{pred,i} - y_{true,i}|$, where n is the total number of data points.

7.1. RESULTS ON BENCHMARK MATERIALS DATASETS

The intuition behind this experiment is that FL can enhance results on benchmark datasets, even when they originate from a single distribution, due to its collaborative and decentralized nature. FL aggregates model updates from multiple clients, which introduces subtle variations in data preprocessing or structure, acting as implicit data augmentation and reducing overfitting. The frequent averaging of model parameters serves as a natural form of regularization, smoothing out irregular patterns and enhancing robustness. The results of baseline models, and the FL version of these models are shown in Table 4, where the baseline results were collected from [3].

Model	Band-gap	Bulk Materials Formation energy	2D Material	Alloy Surface	Pt-Cluster
DeeperGATGNN	Baseline: 0.24570 eV	Baseline: 0.02955 eV/atom	Baseline: 0.17185 eV/atom	Baseline: 0.04086 eV/atom	Baseline: 0.13210 eV/atom
	FL (4 clients): **0.2291 eV**	FL (4 clients): 0.07724 eV/atom	FL (4 clients): 0.2385 eV/atom	FL (4 clients): **0.03642 eV/atom**	FL (4 clients): 0.1468 eV/atom
SchNet	Baseline: 0.28168 eV	Baseline: 0.05 eV/atom	Baseline: 0.214 eV/atom	Baseline: 0.063 eV/atom	Baseline: 0.151 eV/atom
	FL (4 clients): 0.3005 eV	FL (4 clients): 0.108 eV/atom	FL (4 clients): 0.306 eV/atom	FL (4 clients): 0.09 eV/atom	FL (4 clients): 0.206 eV/atom
MPNN	Baseline: 0.26485 eV	Baseline: 0.046 eV/atom	Baseline: 0.204 eV/atom	Baseline: 0.058 eV/atom	Baseline: 0.182 eV/atom
	FL (4 clients): 0.2892 eV	FL (4 clients): 0.9293 eV/atom	FL (4 clients): 0.2974 eV/atom	FL (4 clients): 0.8805 eV/atom	FL (4 clients): 0.2655 eV/atom

Table 4: Performance comparison of baseline and federated version of DeeperGATGNN, SchNet, and MPNN for different benchmark datasets. The baseline results outperformed the baseline results in most cases. Cases where FL version improves benchmark results are marked in bold letters. This shows the promise of FL method for model training for materials property prediction. Further fine-tuning of the models and hyperparameter training might also improve the results.

Performance Summary:

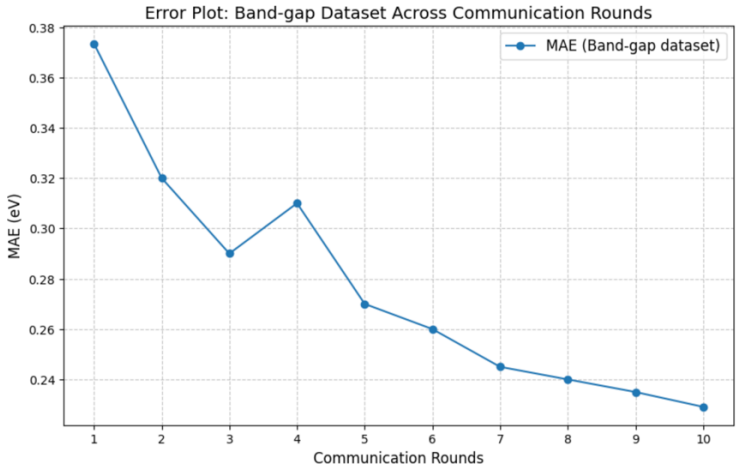
- The DeeperGATGNN model exhibited improved performance in the FL setup for the Band Gap dataset (achieving 0.2291 eV) and Alloy Surface dataset (0.03642 eV/atom), outperforming the centralized baseline results (0.24570 eV and 0.04086 eV/atom, respectively). This indicates that the collaborative nature of FL, even with distributed data, has potential to match or exceed centralized training under optimized conditions.
- SchNet and MPNN models underperformed compared to their respective baselines across most datasets in the FL setup, suggesting a need for further hyperparameter tuning and optimization.

Observations:

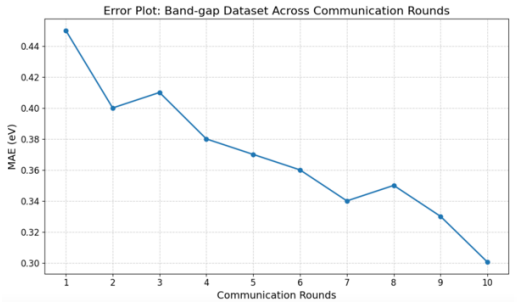
- The gap between FL and centralized baseline performance varied significantly across datasets, highlighting that FL may be more effective on some datasets than others, depending on the distribution and quality of the data.
- DeeperGATGNN consistently outperformed the other models in the FL framework, making it the most promising candidate for this approach.
- Fine-tuning FL hyperparameters such as client selection, aggregation strategies, and communication rounds could further reduce the MAE for all models.

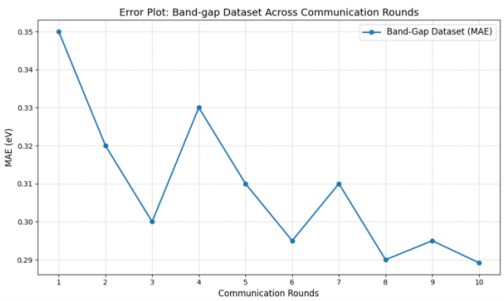
MAE vs. Communication Rounds:

The MAE vs. Communication Round plots for the band-gap dataset of DeeperGATGNN, SchNet, and MPNN are shown in Figure 6. illustrate the progression of the global model's performance across 10 communication rounds in the federated learning (FL) setup. At the outset, the MAE decreases significantly, showcasing that the training process was effectively working, and the model was learning from the aggregated updates across clients. This initial rapid error reduction highlights the capability of FL to leverage distributed data effectively, even in the presence of data heterogeneity.



(a)





(b) (c)

Figure 6: MAEs across communication rounds for (a) DeeperGATGNN, (b) SchNet, and (c) MPNN for the Band gap dataset. The error decreases on average with each communication round for each model, suggesting that more communication rounds might be needed for fine-tuning the results in the future.

As the training progresses, the rate of error reduction slows, indicating the model approaching convergence. Occasional spikes in the error are visible, which can be attributed to client-specific data variability or inconsistencies in model aggregation. Nevertheless, the general trend shows that the errors were steadily going down, affirming the training's efficacy.

Although the reduction in MAE is evident over 10 communication rounds, the plots suggest that additional communication rounds might lead to even better results, as the error does not appear to have completely plateaued. With more rounds, the model could potentially converge to lower error values, enhancing its predictive performance. This limitation to 10 rounds reflects computational constraints in the current setup, and future work could involve extending communication rounds using more powerful computational resources to fully explore FL's potential for these datasets.

Overall, the federated learning (FL) setup effectively preserved data privacy by ensuring that raw data remained localized on client devices, with only model updates being exchanged. This decentralized approach aligns with privacy and regulatory requirements, particularly in domains where data sensitivity or proprietary concerns are significant.

7.2. RESULTS ON OOD DATASETS

As FL has demonstrated in the past that it can improve OOD results, we construct a OOD experiment for that. We borrow the OOD data splitting mechanism and datasets from [30], where the test set is systematically drawn from a separate distribution than the training set. The mechanism and other OOD details can be found in [30]. We utilize three datasets – dielectric, GVRH (elasticity – shear modulus), and perovskites dataset, each having five folds or partitions – each for one different client to use. The results are shown in Table 5, where the ID results are found from [8].

Dataset	DeeperGATGNN	SchNet	MPNN
Dielectric	ID: 0.3355 (unitless)	ID: 0.3277 (unitless)	ID: 0.4682 (unitless)
	OOD: 0.5079 (unitless)	OOD: 0.6621 (unitless)	OOD: 0.587 (unitless)
GVRH	ID: 0.0903 log10(GPa)	ID: 0.0796 log10(GPa)	ID: 0.1206 log10(GPa)
	OOD: 0.1319 log10(GPa)	OOD: 0.1882 log10(GPa)	OOD: 0.1756 log10(GPa)
Perovskites	ID: 0.0288 eV/unit cell	ID: 0.0342 eV/unit cell	ID: 0.0621 eV/unit cell
	OOD: 0.0671 eV/unit cell	OOD: 0.0906 eV/unit cell	OOD: 0.1205 eV/unit cell

Table 5: Performance comparison of ID and OOD data of DeeperGATGNN, SchNet, and MPNN for different datasets. Although FL has shown good promise in improving OOD results in other tasks in the past, it did not however improve performances of OOD materials dataset results, demonstrating the complexity of material data and hardness of various distribution material data to cover.

Performance Summary:
FL consistently showed higher errors for out-of-distribution (OOD) datasets compared to in-distribution (ID) datasets across all evaluated models and datasets.

- In the Dielectric dataset**,** DeeperGATGNN achieved an OOD MAE of 0.5079, which is significantly higher than its ID MAE of 0.3355.
- Similarly, in the Perovskites dataset**,** SchNet exhibited a sharp increase in error, with an OOD MAE of 0.0906 eV/unit cell, compared to its ID MAE of 0.0342 eV/unit cell.
- MPNN also demonstrated a similar trend, with errors in OOD evaluations consistently exceeding those in ID evaluations for all datasets.

Observations:

Generalization Gap:

- The OOD errors were consistently higher than the ID errors for all three datasets. For example, in the Perovskites dataset, SchNet showed an OOD error of 0.0906 eV/unit cell, nearly triple its ID error of 0.0342 eV/unit cell.
- This highlights the difficulty FL frameworks face in capturing the subtility of out-of-distribution material properties.
- Although FL has shown good promise in improving OOD results in other tasks in the past, it did not however improve performances of OOD materials dataset results, demonstrating the complexity of material data and

hardness of various distribution material data to cover. A previous study however revealed an already existing generalization gap of GNNs for OOD property prediction, FL also however could not improve it.

Performance Across Models:

- DeeperGATGNN demonstrated relatively smaller gaps between ID and OOD results compared o SchNet and MPNN, potentially due to its ability to capture structural relationships more effectively.
- SchNet and MPNN underperformed in OOD scenarios, particularly in datasets like Perovskites, where the difference in material space between ID and OOD splits is substantial.

Model Robustness:

- While FL ensures privacy preservation and collaboration, it does not inherently improve robustness to OOD data. This limitation is apparent from the sharp increases in OOD errors across all datasets.

Potential Improvements:

- Incorporating domain-specific techniques such as data augmentation, invariant representation learning, or physics-guided embeddings could help bridge the generalization gap. These methods have shown promise in addressing OOD challenges in other applications of FL.

Fold-wise analysis:

Results of each fold for the perovskites dataset is shown in Figure 7. This figure shows that fold number 2 and 5 are particularly harder to achieve good results on, possibly because of more diverse distribution of each data point in those folds. Fine-tuning the models for both these folds separately might improve the overall OOD performances of these models.

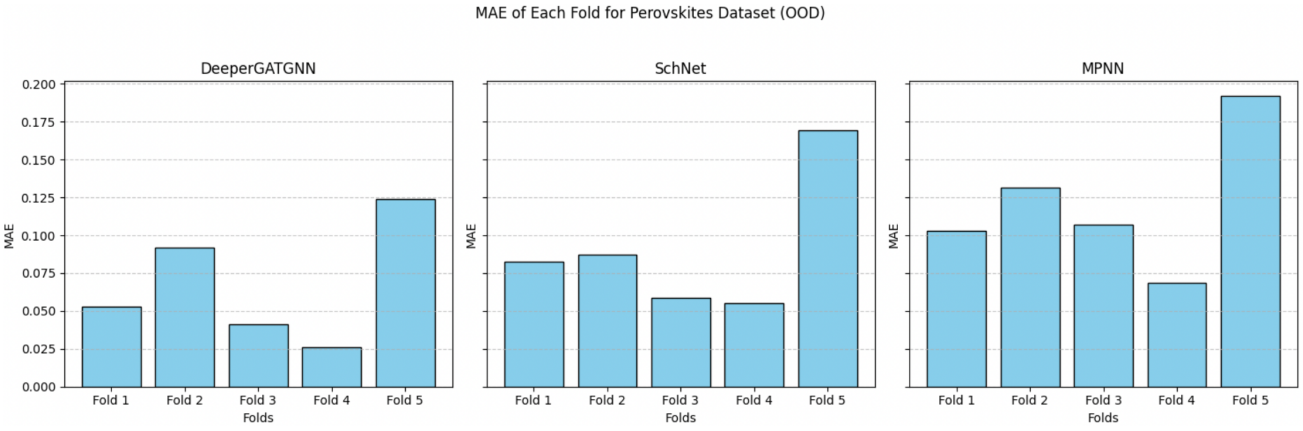


Figure 7: Fold-wise error of federated DeeperGATGNN, SchNet, and MPNN for the perovskites OOD dataset. It shows that the deterioration of results is mostly achieved because of subpar performances in 1-2 folds, indicating more fine-tuning is needed on those OOD folds.

7.3. RESULTS OF THE SCALABILITY ANALYSIS EXPERIMENTS

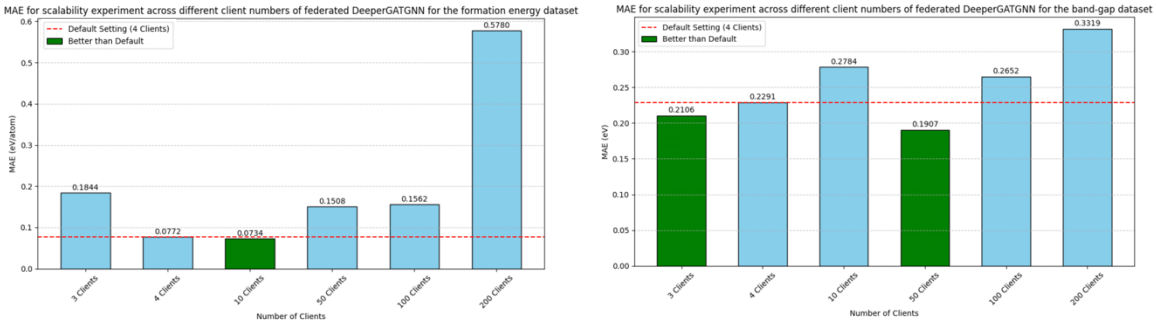
To understand how federated learning (FL) systems scale, we conducted experiments where the number of clients participating in training varied from 3 to 500. For these experiments, we focused solely on the federated DeeperGATGNN model, which showed the best performance among the three models evaluated (DeeperGATGNN, SchNet, and MPNN). The experiments utilized the Formation Energy and Band Gap datasets. Due to time constraints and resource limitations, other models were not included in the scalability tests. Metrics evaluated included model performance (Mean Absolute Error, MAE), convergence time, and memory requirements. The results for the formation energy dataset and the band-gap dataset are shown in Table 6 and 7, respectively. Figure 8 shows the side-by-side comparison for different client nodes.

# of Clients	Convergence Time (hr)	Formation Energy (MAE)
3	≈ 28	0.1844 eV/atom
4	≈ 35	0.0772 eV/atom
10	≈ 42	**0.0734 eV/atom**
50	≈ 66	0.1508 eV/atom
100	≈ 79	0.1562 eV/atom
200	≈ 151	0.5780 eV/atom
500	memory overload	-

Table 6: Convergence time and global-model MAE performance across different client configurations for the formation energy dataset of the federated DeeperGATGNN model. The client number changed improved the performance of the model in one case (marked in bold letter).

# of Clients	Convergence Time (hr)	Band Gap (MAE)
3	≈ 25	**0.2106 eV**
4	≈ 38	0.2291 eV
10	≈ 47	0.2784 eV
50	≈ 60	**0.1907 eV**
100	≈ 69	0.2652 eV
200	≈ 134	0.3319 eV
500	memory overload	-

Table 7: Convergence time and global-model MAE performance across different client configurations for the band-gap dataset of the federated DeeperGATGNN model. The client number changed improved the performance of the model in two cases (marked in bold letter).



(a) (b)

Figure 8: Comparison of MAEs for varying number of clients of the federated DeeperGATGNN model for (a) the formation energy dataset, and (b) the band-gap dataset. The dashed red horizontal line shows the default setting result (4 clients) used in benchmarking experiments. This shows that given more client nodes, that results can be improved as demonstrated for 10 clients for the formation energy dataset, and 50 clients for the band-gap dataset. But doing large scale experiments with more clients will also require more memory and high energy handling machines.

Performance Summary:

- **Formation Energy Dataset:**
  - The global model achieved its best performance with 10 clients, yielding a mean absolute error (MAE) of 0.0734 eV/atom.
  - Performance deteriorated as the client count increased beyond 50 clients, with a notable rise in MAE with communication overhead.
- **Band-Gap Dataset:**
  - Similar trends were observed, with the best results recorded at a moderate number of clients (10-20), after which performance degradation became apparent.
- **Convergence Time:**
  - Convergence time scaled linearly with the number of clients. While experiments with 100 clients took approximately 79 hours, the scalability experiment with 500 clients failed due to memory limitations, highlighting a significant constraint in the current FL framework.

Observations:

1. **Improvement with More Clients:**
   - Increasing the number of clients initially enhanced model performance as more data diversity and heterogeneity contributed to better generalization for the global model. This shows that given more client nodes, that results can be improved as demonstrated for 10 clients for the formation energy dataset, and 50 clients for the band-gap dataset. But doing large scale experiments with more clients will also require more memory and high energy handling machines.
   - This trend aligns with FL's objective of leveraging distributed datasets while preserving privacy.
2. **Performance Degradation with Too Many Clients:**
   - Beyond a certain threshold (e.g., 50 clients), the dataset size for each client became too small, leading to overfitting of local models. This overfitting caused inconsistencies in local updates and ultimately degraded the global model's performance.

    - The increased communication overhead further slowed down convergence, with diminishing returns in model improvement.
3. **Resource Constraints:**
    - Experiments with 500 clients encountered memory limitations, emphasizing the need for more efficient aggregation algorithms and better computational resource management in large-scale FL setups.

The results demonstrate that while increasing the number of clients can improve the model's performance up to a point, there exists a trade-off between data diversity and the size of local datasets. More clients provide greater data heterogeneity, which benefits global learning, but excessively small local datasets exacerbate overfitting and amplify the effects of non-IID data. This results in a performance plateau or even degradation. Addressing these issues will require exploring techniques such as:

- **Data Augmentation:** To enrich local datasets.
- **Client Clustering:** Grouping clients with similar data distributions to reduce inconsistencies in model updates.
- **Adaptive Aggregation Algorithms:** Dynamically adjusting weights based on the quality of local updates.

8 DISCUSSION

• First Work on Federated Learning for Materials Science
This study represents the pioneering work in applying federated learning (FL) to the domain of materials property prediction, specifically targeting challenges unique to this field. The adoption of FL marks a significant step forward in addressing data privacy and decentralization concerns, which are prevalent in materials science. By utilizing state-of-the-art models such as DeeperGATGNN, SchNet, and MPNN within an FL framework, this work establishes a foundation for collaborative machine learning efforts in materials discovery, enabling institutions to collectively benefit from data-driven insights without compromising data security.

• Addressing a Fundamental Scientific Challenge
Materials property prediction is a crucial problem in materials science, playing a pivotal role in accelerating the design and discovery of new materials for applications in energy, semiconductors, catalysis, and beyond. The colossal size of the theoretical materials space poses a unique challenge, as the number of documented materials in available datasets is vastly inadequate to represent this space comprehensively. This data scarcity, compounded by privacy concerns and scattered data ownership, necessitates innovative approaches like FL, which can facilitate collaborative modeling while respecting institutional data silos.

• Benchmark Results: Promising Yet Room for Improvement
The results from benchmark datasets demonstrate the potential of FL to achieve competitive performance compared to centralized models. However, the performance remains suboptimal in some cases, suggesting the need for further fine-tuning and hyperparameter optimization to extract the full potential of the models. The MAE vs. communication round plots clearly indicate that errors decrease on average with additional communication rounds, highlighting the potential benefits of extending the number of rounds. This provides a direction for future work, where more communication rounds could lead to even better results, given sufficient computational resources.

• Challenges in Generalization and OOD Performance
While FL has been shown in prior studies to enhance OOD performance, this was not observed in the context of materials data. The inability of FL to improve OOD results highlights the generalization gap in FL models applied to complex materials data. This underscores the inherent complexity and heterogeneity of materials datasets, which pose significant challenges for machine learning methods. Moreover, the poor overall MAE observed in some datasets was

primarily driven by poor results in 1-2 specific folds, suggesting that targeted fine-tuning and data-specific optimizations are necessary to improve performance on those folds.

- **Insights from Scalability Results**

Scalability experiments revealed both the promise and the limitations of FL in a distributed setup. The model showed consistent improvements in error reduction as the number of client nodes increased, achieving the best results at 10 and 50 clients for the formation energy dataset, and the band-gap dataset, respectively. However, beyond 200 clients, the model ran into memory constraints, highlighting the computational expense and resource intensity of scaling FL to large numbers of clients. These findings suggest that FL can achieve better results with more client nodes, but such configurations require significant advancements in memory and computational efficiency, as well as strategies to optimize communication overhead and energy usage.

- **Time Constraints and Future Directions**

Due to time constraints, this work could not explore all potential optimizations or perform exhaustive experiments across all possible configurations for fine-tuning. Despite this limitation, the project provided valuable insights into the applicability of FL to materials property prediction, the challenges of generalization, and the trade-offs in scalability. Future work could explore advanced aggregation methods, task-specific fine-tuning, different weight update scheme, alternative FL frameworks, and more scalability experiments to further improve the results. Another future work can be to provide different models to different clients and see collectively how they perform in a federated learning setting. This study serves as a stepping-stone for future research, aiming to make FL a robust and scalable solution for collaborative materials science efforts.

9 CONCLUSION

This work represents a significant step forward in applying federated learning (FL) to the field of materials property prediction, addressing critical challenges such as data privacy and decentralization. While benchmark results were promising and scalability experiments demonstrated the potential for improved performance with increasing client nodes, limitations such as the generalization gap for OOD datasets and resource constraints at larger scales highlight areas for future refinement. Despite the constraints, this study provided valuable insights into the complexities of applying FL to materials science and paves the way for further advancements in collaborative and privacy-preserving machine learning for scientific discovery

10 LINKS

- Project code link: https://github.com/csce585-mlsystems/FedMat/tree/main
- Project dataset link: https://github.com/csce585-mlsystems/FedMat/tree/main/data/datasets
- Project presentation link: https://github.com/csce585-mlsystems/FedMat/blob/main/Final_project_presentation_pdf/Final_presentation.pdf
- Project presentation video link: https://youtu.be/j_hZIPUTBZ0?si=u3-4A457fwbogc42

REFERENCES

[1] Xie, T. and Grossman, J.C., 2018. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, *120*(14), p.145301.

[2] Chen, C., Ye, W., Zuo, Y., Zheng, C. and Ong, S.P., 2019. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, *31*(9), pp.3564-3572.

[3] Omee, S.S., Louis, S.Y., Fu, N., Wei, L., Dey, S., Dong, R., Li, Q. and Hu, J., 2022. Scalable deeper graph neural networks for high-performance materials property prediction. *Patterns*, *3*(5).

[4] Schütt, K.T., Sauceda, H.E., Kindermans, P.J., Tkatchenko, A. and Müller, K.R., 2018. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, *148*(24).

[5] Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E., 2017, July. Neural message passing for quantum chemistry. In *International conference on machine learning* (pp. 1263-1272). PMLR.

[6] Choudhary, K. and DeCost, B., 2021. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, *7*(1), p.185.

[7] Fung, V., Zhang, J., Juarez, E. and Sumpter, B.G., 2021. Benchmarking graph neural networks for materials chemistry. *npj Computational Materials*, *7*(1), p.84.

[8] Dunn, A., Wang, Q., Ganose, A., Dopp, D. and Jain, A., 2020. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. *npj Computational Materials*, *6*(1), p.138.

[9] Zhu, W., Luo, J. and White, A.D., 2022. Federated learning of molecular properties with graph neural networks in a heterogeneous setting. *Patterns*, *3*(6).

[10] Hausleitner, C., Mueller, H., Holzinger, A. and Pfeifer, B., 2024. Collaborative weighting in federated graph neural networks for disease classification with the human-in-the-loop. *Scientific Reports*, *14*(1), p.21839.

[11] Omee, S.S., Fu, N., Dong, R., Hu, M. and Hu, J., 2024. Structure-based out-of-distribution (OOD) materials property prediction: a benchmark study. *npj Computational Materials*, *10*(1), p.144.

[12] McMahan, B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B.A., 2017, April. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

[13] Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R. and D'Oliveira, R.G., 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), pp.1-210.

[14] Li, Q., He, B. and Song, D., 2021. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10713-10722).

[15] Li, L., Fan, Y., Tse, M. and Lin, K.Y., 2020. A review of applications in federated learning. *Computers & Industrial Engineering*, *149*, p.106854.

[16] Larsen, A.H., Mortensen, J.J., Blomqvist, J., Castelli, I.E., Christensen, R., Dułak, M., Friis, J., Groves, M.N., Hammer, B., Hargus, C. and Hermes, E.D., 2017. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, *29*(27), p.273002

[17] 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., 2013. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, *68*, pp.314-319.

[18] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A., 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

[19] Fey, M. and Lenssen, J.E., 2019. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428*.

[20] He, C., Li, S., So, J., Zeng, X., Zhang, M., Wang, H., Wang, X., Vepakomma, P., Singh, A., Qiu, H. and Zhu, X., 2020. Fedml: A research library and benchmark for federated machine learning. *arXiv preprint arXiv:2007.13518*.

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp.2825-2830.

[22] Tosi, S., 2009. *Matplotlib for Python developers*. Packt Publishing Ltd.

[23] Waskom, M.L., 2021. Seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), p.3021.

[24] Zhou, K., Huang, X., Li, Y., Zha, D., Chen, R. and Hu, X., 2020. Towards deeper graph neural networks with differentiable group normalization. *Advances in neural information processing systems*, *33*, pp.4917-4928.

[25] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[26] 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., 2013. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, *1*(1).

[27] Mamun, O., Winther, K.T., Boes, J.R. and Bligaard, T., 2019. High-throughput calculations of catalytic properties of bimetallic alloy surfaces. *Scientific data*, *6*(1), p.76.

[28] Fung, V. and Jiang, D.E., 2017. Exploring structural diversity and fluxionality of pt n (n= 10–13) clusters from first-principles. *The Journal of Physical Chemistry C*, *121*(20), pp.10796-10802.

[29] Haastrup, S., Strange, M., Pandey, M., Deilmann, T., Schmidt, P.S., Hinsche, N.F., Gjerding, M.N., Torelli, D., Larsen, P.M., Riis-Jensen, A.C. and Gath, J., 2018. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials*, *5*(4), p.042002

[30] Fu, N., Omee, S.S. and Hu, J., 2024. Physical Encoding Improves OOD Performance in Deep Learning Materials Property Prediction. *arXiv preprint arXiv:2407.15214*.

[31] Loshchilov, I., 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

[32] Zheng, H., Yang, Z., Liu, W., Liang, J. and Li, Y., 2015, July. Improving deep neural networks using softplus units. In *2015 International joint conference on neural networks (IJCNN)* (pp. 1-4). IEEE.

[33] Agarap, A.F., 2018. Deep learning using rectified linear units. *arXiv preprint arXiv:1803.08375*.