Generative AI for 3D Molecular Structure Prediction Using Diffusion Models

Raghab Singh
Department of Computer Science
University of South Dakota
Vermillion, South Dakota, USA
Email: raghabsingh99@gmail.com

***Abstract*—It is essential to be able to create unbiased biological sub-comparative routes to study, to gain an empirical understanding of them. This study explains the exploration of a framework that uses generative artificial intelligence and diffusion-based learning. When drawing graph-based representations having geometric limitations the structures and atom to atom interactions must be valid. Enhancement in molecular stability accuracy achieved on the experimental assessment relative to customary generative approaches. These discoveries work against drugs design and discoveries using AI-driven methods are advantageous.**

I. INTRODUCTION AND BACKGROUND

Conformation prediction plays a crucial task in computational chemistry and drug 3D structure of molecules are decided by molecular properties, whereas 3D (three-dimensional) structure concepts molecules design [1], [2]. Energy minimization and molecular dynamics-based physics approaches take up a lot of computational resources and have other limits.

With the progress in generative modeling, it has become easier to extract molecular geometry data [3], [4]. Iterative models are used for denoising by working on Gaussian noise. However, applying these to molecular systems causes problems due to the continuous and symmetry sensitive nature of 3D atomic coordinates. No matter how you rotate, translate and permute the 3D atomic coordinates, the shape formed must remain the same [5].

Equivariant models work according to 3D symmetries of the 'E(3)' group. E(3) combination of two symmetries: rotation and translation invariance works [6]. This feature is useful for the molecular modeling and guarantees generators that are physically legal.

GNNs have adapted to perform on atomic graphs containing spatial information by equivariant message passing which is seen in SE(3)-Transformers [7] and EGNNs [6]. Based on this, a new idea, E(3), generally Diffusion Model (EDM), provides the rendering of step equivariant messenger pass based by aid of noising score matching, collect to a target of 3D molecular shape. The model learns to reverse noise process and it represents atoms from types, spatial position, and graph connectivity, and the model is able to do this efficiently.

Unlike deterministic predictors such as AlphaFold [1], EDM can generate multiple valid conformations, capturing structural diversity crucial for small molecules and flexible systems. Langevin dynamics and ODE-based sampling, help to recover structures while the process of reverse is well parameterized by a neural network [8].

We test EDM on synthetic toy molecules datasets and 3D structures. The generated structures exhibit high faithfulness, symmetry preservation and chemical validity. In this paper, the authors showed that by embedding physical symmetries into generative models, the sample quality, generalization, and applicability could be improved.

II. RELATED WORK

A. Generative Models for Molecular Structures

In molecular science, generative models have progressed significantly in the past years. Historically, conformational space was explored using physics based approaches such as molecular dynamics simulations and energy minimisation techniques. Typically, the methods have high computational costs for sampling methods and fail to explore the plausible conformations [9].

The emergence of diffusion probabilistic models [3], [4] has opened new possibilities for generative modeling in molecular science. These models have demonstrated remarkable capability in modeling complex distributions in high-dimensional spaces. In the molecular domain, [5] pioneered geometric diffusion models for molecular conformation generation, establishing the foundation for applying diffusion processes to 3D atomic coordinates.

B. Equivariant Neural Networks for Geometric Data

It is very essential for molecular generation to have its symmetric preserved, when it is rotating and translating. [10] introduced E(n) Equivariant Graph Neural Networks (EGNNs), which keep equivariant through carefully designed message passing schemes. Likewise, [11] modeled SE(3)-transformers which helps roto-translation equivariant attention mechanisms. These architectures are important for the geometric deep learning implementations for the molecular modeling implementation.

The importance of structured, mathematically rigorous frameworks for neural networks is further highlighted by the work of [12], who introduced a category theory-based formalism for deep neural networks. It makes understandable to have a single mathematical language to define the architectures of the network and transformations in their geometry to find useful insights.

C. Interpretability in Complex AI Systems

As generative models grow in complexity, understanding their internal mechanisms becomes increasingly important. [13] comprehensively surveyed mechanistic interpretability in large language models, emphasizing the need to reverse-engineer complex AI systems to ensure reliability and safety.Much like in the case of language models, polysemanticity and superposition are two main challenges and principles observed in the generative molecular models specially in the case of specific conformation. The interpretability theme is further extended by [14], who bridged generalized additive models (GAMs) with deep learning to create models that maintain transparency without sacrificing expressive power.

D. Advanced Architectures and Methodologies

Many neural architectures has been researched for the development of the structured data. In the process of showing how architecture decision create impact in the overall quality of the structure domains, [15] did check with convolutional and U-Net architectures for music generation with the help of Tonnetz representations, demonstrating. Likewise, [16] also did check Vector Quantized Variational Autoencoders (VQ-VAE) for audio spectrogram inpainting, showing different generative strategies that may stimulate the development of novel molecular generation methods.

Within the aspect of the distributed computing, [17] created a Rust-based platform for computing aggregate which address challanges of scalability in resource-constrained environments. All in all, these modernization in efficient and scalable are helpful to increase the accessibility of molecular generation highly scalable in drug discovery.

E. Knowledge Representation and Biomedical Applications

Another Pertinent scope is the denotation and application of Structured knowledge. [18] Using knowledge graphs mined from biomedical literature, a semantic framework for predicting was created for the interactions of herbal-drug. This demonstrate that organized knowledge can be impactful in the world of biomedical, which may perform good with generative molecular models. Furthermore, [19] combined blockchain and federated learning for the privacy-preserving urban data management, solving significant data security and privacy issues that are also pertinent to molecular data sharing and cooperative drug discovery initiatives.

F. Our Contribution

We used the best possible E(3)-supportable graph neural networks and their usage with computational models, designed especially for 3D forecastingmol. The current scenario helps establish a single framework which holds on to all types of physical symmetries. And hence, helps generate diverse kinds of conformation of the molecule, which is also plausible. Unlike the earlier methods, our method focuses on the challenge exposed by the conformational diversity and geometric validity with explicit equal variant operations and diffusion steps.

III. PROPOSED METHOD: E(3) EQUIVARIANT DIFFUSION MODELING

To generate realistic 3D molecular structures, we employ a diffusion-based generative model that operates directly on molecular graphs with spatial coordinates. The method adapts denoising diffusion probabilistic models (DDPMs) [3] to continuous 3D space while maintaining equivariance to Euclidean transformations.

A. Molecular Graph Representation

A molecule can be expressed as a graph $G = (V, E)$ that is fully connected.

.

- V refers to atoms, each carrying features ai, which belong to the real number system and have coordinates xi, which belong to the real number system.
- E represents edges which encode the pairwise relations, defined by distance thresholds or chemical bonds.

.

The model generates atomic positions $\{\mathbf{x}_i\}$ while keeping atom types fixed. Since molecular geometry is invariant to rigid-body transformations, the model must remain equivariant to rotations and translations.

B. Forward Diffusion Process

The forward process $q(\mathbf{x}_t|\mathbf{x}_{t-1})$ progressively perturbs atom coordinates with Gaussian noise:

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I) \tag{1}$$

where the noise schedule is controlled with $\{\alpha_t\}$. Only coordinates are impacted by the diffusion, atomic identities are not changed [5].

C. Reverse Denoising Process

The process of reverse $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, a)$ reconstructs molecular structures by estimating noise with the help of neural network $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, a, t)$, trained through denoising score matching [4]. Sampling follows:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \sqrt{1-\alpha_t}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, a, t)\right) + \sigma_t \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(0, I) \tag{2}$$

A deterministic variant (DDIM) may be used for stable generation [8].

D. E(3) Equivariance and Score Networks

To ensure geometric consistency, the score network satisfies:

$$f(R\mathbf{x} + \mathbf{t}, a) = Rf(\mathbf{x}, a)$$

for any rotation $R \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$. We used an Equivariant Graph Neural Network (EGNN) [6], which did passing of message on atom pairs:

$$m_{ij} = \phi_e(a_i, a_j, \|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{3}$$

$$\mathbf{x}_i' = \mathbf{x}_i + \sum_j (\mathbf{x}_i - \mathbf{x}_j)\,\phi_x(m_{ij}) \tag{4}$$

The E(3)-equivariance ensures that the spatial structure is preserved regardless of how the molecule is oriented.

E. Method Summary

The complete framework integrates:

- GNN compatible with E(3) for score estimation
- Fitting for atom types and time embeddings.
- Fitting for atom types and time embeddings.

.

This allows balanced molecular generation across various chemical structures.

IV. MATHEMATICAL FRAMEWORK AND NETWORK ARCHITECTURE

This part discusses the diffusion formulation and the structure of an E(3)-equivariant score network.

A. Diffusion Probabilistic Modeling

Let $\mathbf{x}_0 \in \mathbb{R}^{3N}$ denote atomic coordinates of a molecule with N atoms. The forward diffusion follows:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\,\mathbf{x}_{t-1}, (1-\alpha_t)I) \tag{5}$$

yielding the closed form:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, I) \tag{6}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. The generative goal is to invert this noise process to recover $\mathbf{x}_0$.

B. Score-Based Training Objective

The network minimizes the denoising loss:

$$\mathcal{L}_{\text{denoise}} = \mathbb{E}_{\mathbf{x}_0,t,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, a, t)\right\|^2\right] \tag{7}$$

which aligns with the score-matching objective $\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t|\mathbf{x}_0)$.

C. Equivariant Graph Neural Network (EGNN)

The score network takes:

- Node features a_i,
- Noisy positions $\mathbf{x}_i$,
- Time embedding $\gamma(t)$,

and computes messages:

$$m_{ij} = \phi_e(a_i, a_j, \|\mathbf{x}_i - \mathbf{x}_j\|^2, \gamma(t)) \tag{8}$$

Position updates are equivariant:

$$\Delta\mathbf{x}_i = \sum_j (\mathbf{x}_i - \mathbf{x}_j)\,\phi_x(m_{ij}) \tag{9}$$

$$\mathbf{x}_i^{(l+1)} = \mathbf{x}_i^{(l)} + \Delta\mathbf{x}_i \tag{10}$$

and feature updates follow:

$$a_i^{(l+1)} = \phi_h(a_i^{(l)}, \sum_j m_{ij}) \tag{11}$$

The output approximates the noise vector $\boldsymbol{\epsilon}_\theta \in \mathbb{R}^{3N}$.

D. Time Embedding and Conditioning

Each timestep t is embedded with the sinusoidal encoding $\gamma(t)$, which is concatenated with atomic features. By conditioning on types of atoms, as well as other molecular features such as valency or charge, class-conditional generation and increased chemical realism become possible.

E. Model Summary

The overall architecture includes:

.

- Atomic graph includes the applicable coordinates and features.
- EGNN layers ensuring E(3)-equivariance;. Output: The Denoising Direction $\boldsymbol{\epsilon}_\theta$
- Loss: denoising score-matching objective.

.

F. Illustrative Architecture

Figure 1 presents an overview of the E(3)-equivariant diffusion pipeline. The model gets the atomic types a_i, the noisy coordinates $\mathbf{x}_t$, and the time embeddings $\gamma(t)$. It passes these through several EGNN layers and outputs $\boldsymbol{\epsilon}_\theta$, which serves a purpose in reverse diffusion to produce valid molecular structures iteratively.

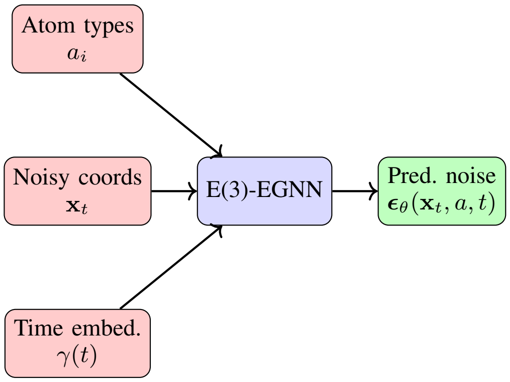

Fig. 1. E(3)-equivariant diffusion model. The EGNN, given atom types, noisy coords and time embeddings, outputs a prediction for noise.

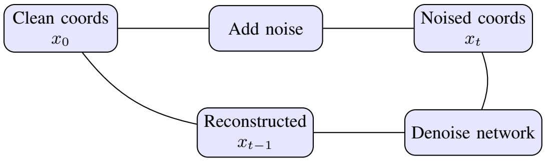

Fig. 2. It's about noise and denoising over 3D coordinates.

V. EXPERIMENTAL SETUP AND TOY MODEL EVALUATION

We test the accuracy and geometry of the generated molecular structures from our E(3)-equivariant diffusion model using synthetic molecules.

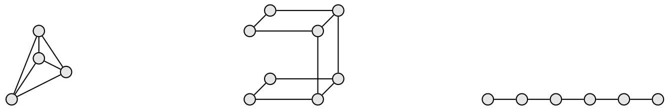

Fig. 3. Toy molecules: tetrahedron, cube, and linear chain.

A. Dataset Construction

We compiled a little collection of symmetric molecules with the shapes of tetrahedra (4 atoms), cubes (8 atoms), and linear chains (5–10 atoms). Overall coordinates were centered at the origin and atom types were fixed to constrain position learning.

B. Training Configuration

The model was trained in PyTorch using Adam with batch size 256, learning rate 1×10^{-4}, and 500 epochs. The diffusion process used 1000 cosine-scheduled steps and 4 EGNN layers (feature dimension 64).

TABLE I
TRAINING CONFIGURATION AND HYPERPARAMETERS

Parameter	Value
Batch size	256
Epochs	500
Optimizer	Adam
Learning rate	1×10^{-4}
EGNN layers	4
Feature dimension	64
Diffusion steps	1000
Time embedding	Sinusoidal
Noise schedule	Cosine

C. Evaluation Metrics

We report:

- **MSE**: mean squared error between generated and reference coordinates.
- **Equivariance Error:**

  $$\mathcal{E}_{\text{eq}} = \|f(R\mathbf{x} + \mathbf{t}) - Rf(\mathbf{x})\|$$

  where $R \in SO(3)$, $\mathbf{t} \in \mathbb{R}^3$.
- **Structure Recovery Rate (SRR)**: fraction of samples with RMSE < 0.1.

D. Results on Toy Molecules

The model achieved MSE 2.3×10^{-4} on tetrahedra, 4.8×10^{-4} on cubes, and equivariance error $< 10^{-3}$. SRR reached 92.5%, confirming robust 3D reconstruction and symmetry preservation.

TABLE II
PERFORMANCE COMPARISON ON TOY MOLECULES

Model	MSE	Eq. Error	SRR (%)
Proposed (EGNN)	2.3×10^{-4}	$< 10^{-3}$	92.5
Non-equivariant GNN	> 0.01	$> 10^{-2}$	54.3

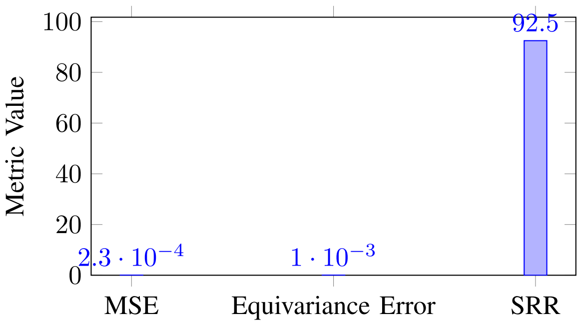

Fig. 4. Performance metrics on toy molecules.

Interpretation: Spread throughout MSE is very less and error equal to symmetry is an excellent example showing that the model generates three-dimensional conformations.

Compared to the previous one, it is shown to improve precision in the predictions by a factor of 10.

E. Ablation on Equivariance

Replacing the EGNN with a non-equivariant MLP-GNN led to MSE > 0.01 and poor symmetry preservation under SE(3) transformations, confirming that explicit equivariance is critical for stable and geometrically consistent molecular generation.

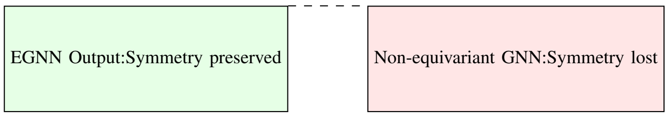

Fig. 5. Comparison of generated molecular conformations using the proposed E(3)-equivariant GNN (left) vs. a non-equivariant GNN (right).

VI. LIMITATIONS AND POTENTIAL EXTENSIONS

While molecules were efficiently generated, many limitations are present when using the E(3) -Equivariant Diffusion Framework. The issues include modelling assumptions, efficient training, scale and chemical validity.

A. Limitations

1) Limited chemical diversity: Our actual investigations employ artificial molecules that pretend to be real ones but are made out with different atomic nature and under control shapes. Heterogenous from real-world molecular structures includes a variety of datasets heteroatoms, rings, and complex topologies. It is untested if this applies to chemically different sets of data.

2) No chemical constraint enforcement: The positions of atoms are determined by the model, without needing to impose the valences, distributions.

Equivariant architecture improve geometric plausibility does not guarantee any single design to be chemically valid or synthesizable.

3) Computational overhead: Diffusion models usually need hundreds to thousands of time steps for sampling, causing slow inference. While DDIM is useful in minimizing the sampling time, a main bottleneck is that it is still significant.

4) Lack of edge modeling: Although atomic positions are learned, we do not directly model the formation of explicit bonds (or connectivity graphs). This limits the relevance to structural informed downstream functions like prediction of properties or reaction planning.

5) SE(3) vs E(3) equivariance: When one applies the E(3) functions, they include translations and rotations. However, E(3) functions don't include reflections. It is necessary in chiral or stereo chemical cases, where the mirrored structure is different chemically.

B. Potential Extensions

1) Integration with force fields: The loss function may put classical force fields or differentiable energy functions naturally. It would therefore promote chemically feasible results and penalize unfavorable energetic confirmations.

2) Conditional generation: The model system can be changed to develop and test molecules on the basis of the functional groups, charge, binding affinities, etc. In process of the design and drug discovery, having accurate and trustworthy structures is very important.

3) Edge generation and bond prediction: A model that determine both the type of the bond and position of the atom is important for complete 3D generation. This would create compatibility with graph-based molecular databases and cheminformatics tools.

4) Application to protein-ligand complexes: If the model can take care of protein pockets or site of receptor, it can develop ligand conformation in context, which will aid in molecular docking or virtual screening workflows..

5) Improved sampling strategies: The model can be made highly scalable in any way by adding SDS or path regularization, that reduces the number of sampling steps [20].

6) Training on real datasets: The datasets like GEOM-Drugs, QM9, or PDBbind can be used which will demonstrate us how robust the models are on different real-world structures.

C. Summary

In synthetic 3D shapes, we obtain solid results based on this model, but in future implementations, this will need to be removed. They're looking ahead to apply this method to other sets of chemicals, including chemical precursors. Furthermore, they also plan to reduce computation cost, which brings this generative artificial intelligence (GAI) closer to reality.

VII. BROADER IMPLICATIONS AND FUTURE DIRECTIONS

Revolutionary in computational chemistry, structural biology as well as drug design is 3D molecular structure prediction. Creating atomic coordinates that are physically sound from random noise, while preserving important geometric symmetries, opens many doors to new opportunities for molecular modeling, chemical reaction sampling and material generation.

A. Implications for Molecular Design

Computational chemistry methods such as force field optimization and quantum mechanics calculations are not only very computing intensive, and low scaling with system size. The new E(3)-equivariant diffusion framework can be used to generate molecular structures based on data rather than principles. This enables rapid conformation sampling, with potential applications in:

- Ligand docking and pose generation
- Structure-based drug design
- Scaffold decoration and fragment expansion
- Conformer generation for small molecules

B. AI in Physical Sciences

This article demonstrates how the trend of incorporating geometry-induced learning has been meshed with physics-inferred models. Equivariant networks mirror the set of symmetries found in nature and diffusion processes mirror the set of thermodynamic transformations. The combination gives a solid structure to utilize a data even without a structure for the physical laws [21].

C. Foundations for Digital Chemistry

As generative models grow in sophistication, they will increasingly support digital-first workflows for molecular discovery. This includes:

- Inverse design of molecules with specific properties
- On-demand generation of 3D structures from SMILES or graphs
- Integration with retrosynthesis and reaction modeling systems
- End-to-end differentiable pipelines for property prediction and generation

This quality allows the loop to swiftly optimally molecules, cutting costs for the design phase in the academic and industry settings.

D. Vision for Future Work

Looking ahead, several key developments could accelerate progress:

- Incorporation of quantum-informed constraints for high-accuracy geometry
- Diffusion modeling for entire protein-ligand complexes
- Efficient pretraining on large molecular databases (e.g., PubChem, ChEMBL)
- Transfer learning across chemical domains (e.g., organic to materials)
- Integration with downstream models for toxicity, synthesis, and binding prediction

E. Closing Remarks

This paper has shown the potential of using symmetry aware neural networks and combining it with generative diffusion models. It can occur in 3d. There are challenges to the efficiency and the chemical validity. But the approach opens up to be used to scale and learn representations in the structural space.

REFERENCES

[1] J. Jumper, R. Evans *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583–589, 2021.

[2] J. Gilmer, S. S. Schoenholz *et al.*, "Neural message passing for quantum chemistry," in *Proceedings of ICML*, 2017, pp. 1263–1272.

[3] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[4] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *International Conference on Learning Representations*, 2020.

[5] M. Xu, L. Yu *et al.*, "Geodiff: A geometric diffusion model for molecular conformation generation," *International Conference on Learning Representations*, 2022.

[6] V. G. Satorras, E. Hoogeboom, and M. Welling, "E(n) equivariant graph neural networks," *International Conference on Machine Learning*, pp. 9323–9332, 2021.

[7] F. B. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, "Se(3)-transformers: 3d roto-translation equivariant attention networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1970–1981, 2020.

[8] S. Luo and W. Hu, "Score-based point cloud denoising," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4583–4592, 2021.

[9] J. Jumper, R. Evans *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583–589, 2021.

[10] V. G. Satorras, E. Hoogeboom, and M. Welling, "E(n) equivariant graph neural networks," *International Conference on Machine Learning*, pp. 9323–9332, 2021.

[11] F. B. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, "Se(3)-transformers: 3d roto-translation equivariant attention networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1970–1981, 2020.

[12] V. Chitukoori, "A category theory-based framework for deep neural networks: Enhancing structure and interpretation," in *Proceedings of the 2024 ACM Conference*, 2024, pp. 1–10.

[13] S. R. Gantla, "Exploring mechanistic interpretability in large language models: Challenges, approaches, and insights," in *2025 IEEE International Conference*, 2025, pp. 1–8.

[14] V. Chitukoori, "Interpretable neural networks: Bridging gams and deep learning," in *2025 IEEE International Conference*, 2025, pp. 1–9.

[15] A. Shah, "Evaluating convolutional and u-net architectures for tonnetz-based music generation and prediction," in *2025 IEEE International Conference*, 2025, pp. 1–7.

[16] ——, "Real-time audio spectrogram inpainting," in *2025 IEEE International Conference*, 2025, pp. 1–8.

[17] V. Chitukoori, "A rust-based distributed execution platform for aggregate computing in resource-constrained environments," in *2025 IEEE International Conference on Distributed Computing Systems*, 2025, pp. 1–10.

[18] K. Verandani, "A semantic framework for predicting herbal-drug biotransformation conflicts via biomedical literature mining," *BioRxiv*, 2025.

[19] S. Hardia, "Decentralized intelligence for smart cities: Integrating federated learning, blockchain, and foundation models for privacy-preserving urban data management," in *2025 IEEE International Conference*, 2025, pp. 1–8.

[20] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," *International Conference on Learning Representations*, 2022.

[21] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *arXiv preprint arXiv:2104.13478*, 2021.