SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE

Department of AI and Machine Learning

SYMBIOSIS Golden Jubilee — Celebrating 50 Years of Excellence

PROJECT TITLE: Comparing VAE and GAN Models for Molecular SMILES Generation and Property Prediction with GNNs

INTRODUCTION

The efficient design of novel molecules is vital for progress in pharmaceuticals, materials science, and chemical engineering. Traditional experimental methods are costly and time-consuming, prompting the adoption of computational approaches. This study evaluates Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) for generating molecular structures as SMILES strings and integrates them with Graph Neural Networks (GNNs) for property prediction. By leveraging the strengths of generative and predictive models, this framework accelerates molecular discovery, offering a computationally efficient alternative to traditional techniques.

OBJECTIVES & AIMS:

- To develop an efficient and scalable framework for molecular discovery using deep learning techniques, addressing challenges in molecular design and accelerating innovation.
- To compare VAEs and GANs for generating molecular structures as SMILES strings.
- To leverage GNNs for accurate molecular property prediction using graph-based representations.

OUTCOME OF THE PROJECT

The project developed a framework using VAEs and GANs for generating novel molecular structures, and GNNs for predicting their properties. It produced valid, diverse molecules with accurate predictions of quantum chemical properties like dipole moment and LogP, enabling advancements in molecular design and drug discovery.

METHODOLOGY

Overview
This chapter describes the methodologies for comparing VAEs and GANs in generating molecular SMILES strings and predicting properties with GNNs.

Data Acquisition and Preprocessing
The QM9 dataset was selected, preprocessed, and split to ensure balanced property distribution across training, validation, and test sets.

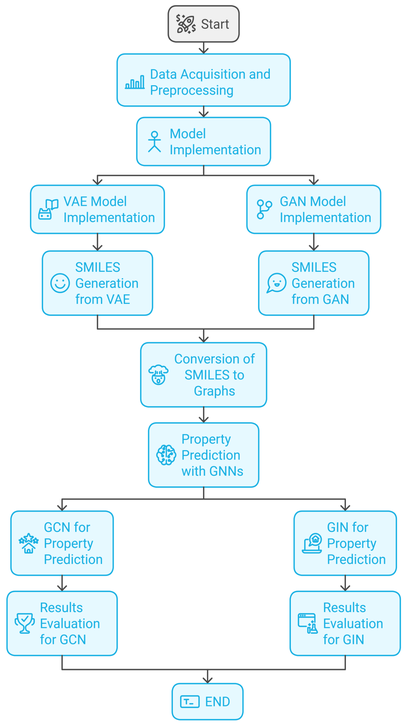
Variational Autoencoder (VAE)
The VAE model uses a GRU-based encoder-decoder architecture with a 64-dimensional latent space, optimized with reconstruction and KL divergence losses.

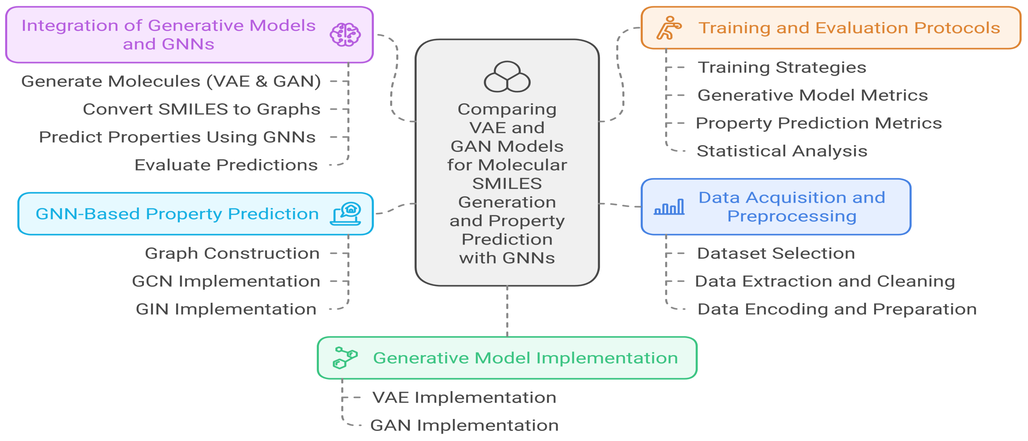
Generative Adversarial Network (GAN)
The GAN model includes an LSTM-based generator and bidirectional LSTM discriminator, trained adversarially with binary cross-entropy loss and policy gradient optimization.

Graph Neural Network (GNN) for Property Prediction
Molecules were represented as graphs and processed by GCN and GIN models, which used convolutional layers and pooling to predict properties based on node and edge features.

Integration of Generative Models and GNNs
Generated molecules were converted to molecular graphs, and properties were predicted using pre-trained GCN and GIN models.

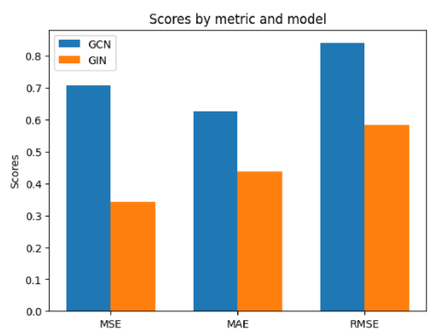
Training and Evaluation Protocols
Regularization, early stopping, and evaluation metrics like validity, uniqueness, novelty, MSE, and R^2 were used to assess generative and predictive model performance.

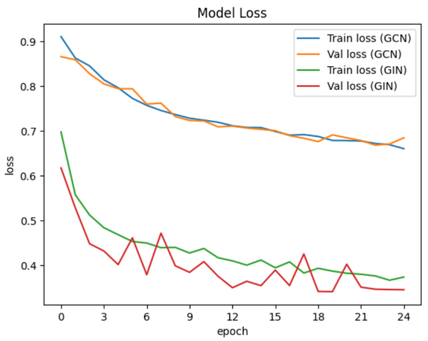


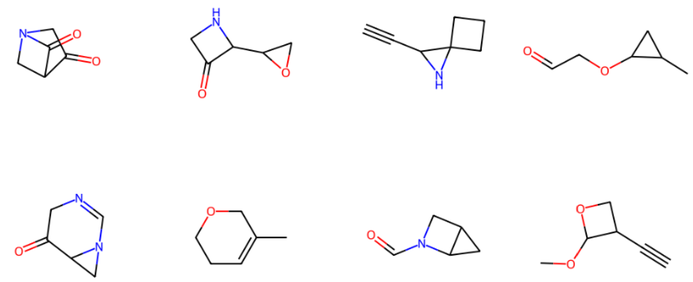


RESULTS:

Metric	VAE	GAN
Fréchet ChemNet Distance	47.773	118.981
Average Tanimoto Similarity	0.335	0.123
Internal Diversity	0.937	0.325
Validity Rate (%)	100.0	100.0
Uniqueness Rate (%)	99.678	75.327
Novelty Rate (%)	99.678	74.924
Average MolWt	126.964	108.390
Average LogP	0.412	0.088
Average NumHDonors	0.967	0.473
Average NumHAcceptors	2.260	2.061







Generated Molecular Structures

CONCLUSION AND FUTURE SCOPE:

- The VAE outperformed the GAN in generating unique and novel molecules, achieving a higher uniqueness (99.67%) and novelty (99.67%) rate. Its ability to closely reflect the training data distribution makes it a more reliable model for producing high-quality, chemically valid molecules.
- Future scope: Future work could focus on developing conditional GANs that generate molecules tailored to specific properties, enabling more targeted and efficient molecular design for applications like drug discovery and materials science.

GROUP MEMBERS: ARJUN TYAGI, KAUSTUBH RAYKAR, NACHIKET ROPIA

PROJECT MENTOR NAME: DR SHRUTI PATIL, DR SUNIL JAISWAL