Research Report – AI in Drug Discovery

Student Name: Himani Raval

Date: July 27, 2025

## Paper 1:

🔗 Link: https://www.nature.com/articles/s41598-018-22811-x

📊 Dataset: ChEMBL dataset of kinase inhibitors

🧠 Methodology: Multitask Deep Neural Networks (DNNs)

✨ Novelty: Simultaneously trained on multiple kinase targets, improving generalization.

📈 Accuracy: 82%

🧪 Evaluation Metrics: AUC, Precision, ROC Curve

---

## Paper 2:

🔗 Link: https://www.nature.com/articles/s42256-019-0107-1

📊 Dataset: ZINC15 compound library

🧠 Methodology: Generative TensorFlow models (GENTRL)

✨ Novelty: Used reinforcement learning for rapid molecule generation.

📈 Accuracy: High hit rate of generated molecules against DDR1

🧪 Evaluation Metrics: Hit rate, Binding Affinity

---

## Paper 3:

🔗 Link: https://arxiv.org/abs/1905.12286

📊 Dataset: BindingDB, ChEMBL

🧠 Methodology: Graph Convolutional Networks (GCNs)

✨ Novelty: Learning molecular graphs directly using message passing.

📈 Accuracy: 85.3%

🧪 Evaluation Metrics: ROC-AUC, RMSE

---

## Paper 4:

🔗 Link: https://pubs.acs.org/doi/10.1021/acscentsci.9b00636

📊 Dataset: PubChem bioactivity assays

🧠 Methodology: One-shot learning with attention mechanisms

✨ Novelty: Few-shot learning approach for unseen targets.

📈 Accuracy: 70-90% (target-dependent)

🧪 Evaluation Metrics: AUC, Precision, Recall

---

## Paper 5:

🔗 Link: https://www.frontiersin.org/articles/10.3389/fphar.2021.650110/full

📊 Dataset: DrugBank, SIDER

🧠 Methodology: Autoencoder + MLP

✨ Novelty: Multi-omics data fusion in latent space

📈 Accuracy: 87.4%

🧪 Evaluation Metrics: F1-score, Accuracy, AUC

---

## Paper 6:

🔗 Link: https://www.nature.com/articles/s41586-019-0912-z

📊 Dataset: ZINC database

🧠 Methodology: Generative Adversarial Networks (GANs) for molecule generation

✨ Novelty: GANs used for de novo molecular generation

📈 Accuracy: N/A (evaluated using drug-likeness score)

🧪 Evaluation Metrics: Drug-likeness, Synthetic Accessibility

---

## Paper 7:

🔗 Link: https://arxiv.org/abs/2102.09548

📊 Dataset: PDBbind

🧠 Methodology: 3D CNN + attention for protein-ligand binding

✨ Novelty: Spatial learning from protein-ligand complex structures

📈 Accuracy: 76%

🧪 Evaluation Metrics: Binding Affinity RMSE, Pearson Correlation

---

## Paper 8:

🔗 Link: https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00441-1

📊 Dataset: PubChem

🧠 Methodology: SMILES-based RNN with transfer learning

✨ Novelty: Transfer learning on SMILES sequences

📈 Accuracy: 89%

🧪 Evaluation Metrics: Log-loss, AUC

---

## Paper 9:

🔗 Link: https://www.sciencedirect.com/science/article/pii/S200103702030397X

📊 Dataset: ChEMBL, DrugBank

🧠 Methodology: Transformer model for DTI prediction

✨ Novelty: Used attention on drug-target sequences

📈 Accuracy: 86.5%

🧪 Evaluation Metrics: AUC, Precision, F1-score

---

## Paper 10:

🔗 Link: https://academic.oup.com/bioinformatics/article/35/14/2436/5280163

📊 Dataset: BindingDB

🧠 Methodology: DeepDTA (CNN on SMILES + protein sequence)

✨ Novelty: Joint learning on drug and protein sequences

📈 Accuracy: 88.2%

🧪 Evaluation Metrics: Concordance Index, MSE

---