**AI-Powered Multitarget Drug Discovery using Deep Learning**

**Prepared by:** Panuganti Jaswanth Sri

**Date:** 27-07-2025

**GitHub repository**: <https://github.com/jaswanthsri-4957/ai-drug-dectection>

**Project Overview**

This project focuses on building a deep learning model to predict bio activity of drug molecules against multiple protein targets. It uses SMILES-based fingerprint features and modern machine learning tools to create a robust, interpret able prediction system.

**Key Components**

**Dataset Used**

* **Source:** Kaggle — MultiTarget Bioactivity ChEMBL dataset
* <https://www.kaggle.com/datasets/xjoannax88/multitarget-bioactivity-chembl?utm_source=chatgpt.com>
* **Size:** 8,700 molecules
* **Targets:** 4 proteins (EGFR, HDAC2, PPARG, SRD5A2)
* **Features:** SMILES strings + molecular descriptors (LogP, TPSA, HBA, HBD)

**Data Cleaning & Preprocessing**

* Handled missing values using mean/median imputation
* Removed outliers using IQR method
* Removed invalid SMILES entries

**Feature Extraction**

* SMILES → 2048-bit molecular fingerprints using RDKit MorganGenerator
* Encoded protein labels using Label Encoder

**Model Training**

* Deep Neural Network (DNN): 512 → 256 → 128 → softmax
* Activation: ReLU + Dropout (0.3)
* Loss: Categorical Crossentropy
* Optimizer: Adam
* Epochs: 20, Batch Size: 64

**Evaluation**

* Accuracy on test set
* Classification report (Precision, Recall, F1-score)
* Confusion matrix
* Per-class accuracy

**Visual Analysis (EDA)**

* Protein target distribution bar plot
* Boxplots for descriptors by target
* Heatmap of descriptor correlation
* t-SNE projection of molecular clusters

**Novelty in This Project**

* **Multi target prediction model** instead of single-target classifiers
* Use of **RDKit MorganGenerator** (latest method)
* Applied **outlier removal** and deep EDA
* t-SNE fingerprint visualization
* Clean, end-to-end pipeline suitable for deployment or dashboard use

**Model Export**

* Final trained model saved as protein\_classifier\_model.h5

**Files Submitted**

* molecule\_classification\_dataset.csv (source data)
* Jupyter Notebook (.ipynb) with full pipeline
* Trained model file (.h5)
* Dashboard/visualization (optional Streamlit app)
* Code explanation video

**Reference list**

* Chakraborty, C., Bhattacharya, M., Lee, S.-S., Wen, Z.-H. and Lo, Y.-H. (2024). The changing scenario of drug discovery using artificial intelligence (AI) to deep learning (DL): Recent advancement, success stories, collaborations, and challenges. *Molecular Therapy — Nucleic Acids*, pp.102295–102295. doi:https://doi.org/10.1016/j.omtn.2024.102295.
* Gupta, U., Ayushman Pranav, Kohli, A., Ghosh, S. and Singh, D. (2024). The Contribution of Artificial Intelligence to Drug Discovery: Current Progress and Prospects for the Future. *Microorganisms for sustainability*, pp.1–23. doi:https://doi.org/10.1007/978-981-99-9621-6\_1.
* Hashemi, S., Parisa Vosough, Taghizadeh, S. and Amir Savardashtaki (2024). Therapeutic Peptide Development Revolutionized: Harnessing the Power of Artificial Intelligence for Drug Discovery. *Heliyon*, 10(22), pp.e40265–e40265. doi:https://doi.org/10.1016/j.heliyon.2024.e40265.
* Huang, D., Yang, M., Wen, X., Xia, S. and Yuan, B. (2024). AI-Driven Drug Discovery: Accelerating the Development of Novel Therapeutics in Biopharmaceuticals. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, [online] 3(3), pp.206–224. doi:https://doi.org/10.60087/jklst.vol3.n3.p.206-224.
* Mahilnan, V., Nandhini, Ms.B., Santhosh, N., MelkyAbishake, R., Ranganathan, S. and Maheshwaran, P. (2024). AI-Powered Drug Discovery: Accelerating Precision Medicine with Deep Learning. *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, [online] pp.1–7. doi:https://doi.org/10.1109/icses63760.2024.10910685.
* Nishan, H. (2025). AI-powered drug discovery for neglected diseases: accelerating public health solutions in the developing world. *Journal of Global Health*, 15. doi:https://doi.org/10.7189/jogh.15.03002.
* Odah, M. (2025). Artificial Intelligence Meets Drug Discovery: A Systematic Review on AI-Powered Target Identification and Molecular Design. [online] doi:https://doi.org/10.20944/preprints202503.0912.v1.
* Phuvamin Suriyaamporn, Boonnada Pamornpathomkul, Prasopchai Patrojanasophon, Tanasait Ngawhirunpat, Theerasak Rojanarata and Praneet Opanasopit (2024). The Artificial Intelligence-Powered New Era in Pharmaceutical Research and Development: A Review. *AAPS PharmSciTech*, [online] 25(6). doi:https://doi.org/10.1208/s12249-024-02901-y.
* Pun, F.W., Ozerov, I.V. and Zhavoronkov, A. (2023). AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*, 44(9), pp.561–572. doi:https://doi.org/10.1016/j.tips.2023.06.010.
* Tiwari, P., Pal, R., Chaudhary, M.J. and Nath, R. (2023). Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. *Drug Development Research*, 84(8). doi:https://doi.org/10.1002/ddr.22115.
* Verma, A. and Awasthi, A. (2024). Revolutionizing Drug Discovery: The Role of Artificial Intelligence and Machine Learning. *Current Pharmaceutical Design*, 30(11), pp.807–810. doi:https://doi.org/10.2174/0113816128298691240222054120.