International Journal of Multidisciplinary Science and Innovation (IJMSI)
Website: https://ijmsi.in/
ISSN: 3107-5754
Vol. 2, Issue 1, Jan-Jun, 2026
Page No.: 103-110

Research

Application of Deep Learning and Machine Learning Algorithms in Predicting Drug Stability and Bioavailability

Lubna Nousheen[1*], Mohammad Shamim Qureshi[2], Sameera Fatima[3]

[1,3]Department of Pharmaceutics, Anwarul Uloom College of Pharmacy, New Mallepally, Hyderabad, Telangana 500001, India.

[2]Department of Pharmacognosy, Anwarul Uloom College of Pharmacy, New Mallepally, Hyderabad, Telangana 500001, India.

Corresponding Author:
Dr. Lubna Nousheen

Email:
drlubnanousheen@gmail.com

DOI: *10.62896/ijmsi.2.1.14*

Conflict of interest: *NIL*

Article History
Received: 08/02/2026
Accepted: 10/03/2026
Published: 13/04/2026

Abstract:
Predicting the shelf life and oral absorption of drug candidates during the early phases of pharmaceutical development remains a formidable challenge, frequently contributing to expensive late-stage failures. In this investigation, deep learning (DL) and machine learning (ML) techniques were employed to forecast two pivotal pharmaceutical parameters chemical stability and oral bioavailability using a carefully assembled dataset of 450 small-molecule compounds with experimentally validated profiles. Five distinct computational architectures were constructed and systematically benchmarked: Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Artificial Neural Network (ANN), and Convolutional Neural Network (CNN). Molecular descriptors and extended-connectivity fingerprints served as input representations, and model performance was assessed through root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). For bioavailability estimation, the CNN architecture produced the strongest predictive capability ($R^2 = 0.91$, RMSE = 8.34%), whereas the GBM model outperformed all others in stability forecasting ($R^2 = 0.89$, RMSE = 4.21 months). Interpretability analysis using SHapley Additive exPlanations (SHAP) identified lipophilicity, molecular weight, hydrogen bond donor count, and topological polar surface area as the descriptors with the greatest influence on prediction outcomes. These results illustrate the capacity of DL and ML methodologies to meaningfully accelerate pharmaceutical screening by providing early-stage estimates of stability and bioavailability, thereby supporting more informed decision-making throughout the drug development pipeline.

Keywords: Deep learning; Machine learning; Drug stability; Bioavailability prediction; Molecular descriptors; Convolutional Neural Network; Gradient Boosting; SHAP; Pharmaceutical development; QSAR

1. Introduction

Bringing a new drug from initial synthesis to market approval typically spans a decade or longer, with estimated costs that can exceed several billion dollars [1]. Throughout this arduous trajectory, two physicochemical attributes exert a decisive influence on whether a candidate molecule progresses or fails: its chemical stability under storage conditions and its oral bioavailability. Stability, broadly defined as the capacity of an active pharmaceutical ingredient (API) to retain its potency, purity, and physical form over time, underpins the regulatory acceptability and commercial viability of any finished dosage form [2]. Oral bioavailability the proportion of an administered dose that enters the systemic circulation without chemical alteration governs the therapeutic window of orally delivered medicines and is, consequently, among the most scrutinized pharmacokinetic endpoints in drug evaluation [3].

Conventional experimental workflows for assessing these parameters are inherently time-consuming. Stability evaluation follows protocols laid down by the International Council for Harmonisation (ICH), which mandate accelerated and long-term studies extending over months to years under controlled temperature and humidity conditions [4]. Bioavailability determination relies on in vivo pharmacokinetic trials in animal models or human subjects, entailing substantial financial outlay, ethical oversight, and logistical coordination [5]. Such timelines sit uncomfortably alongside the rapid-throughput demands of contemporary drug discovery, in which thousands of candidate molecules may require concurrent profiling.

The emergence of artificial intelligence (AI) particularly its sub-disciplines of machine learning and deep learning has begun to reshape the pharmaceutical landscape. ML algorithms discover predictive relationships within historical datasets and have already demonstrated value in quantitative structure–activity relationship (QSAR) modeling, toxicity estimation, and drug–target interaction screening [6,7]. Deep learning extends these capabilities through multi-layered neural architectures that can capture intricate, nonlinear dependencies in high-dimensional molecular representations, often surpassing classical ML techniques on tasks involving complex chemical spaces [8,9].

Within the specific domain of pharmacokinetics, several groups have applied ML methods to predict oral bioavailability. Wei et al. generated over one thousand two-dimensional molecular descriptors and trained a Random Forest classifier, achieving an accuracy of approximately 79% on a curated compound dataset [10]. Ng and Lu subsequently explored graph neural networks (GNNs) with transfer learning for bioavailability classification and reported improved performance relative to descriptor-based approaches [11]. On the stability front, Han et al. employed Random Forest to forecast the physical stability of solid dispersions over six-month horizons with notable success [12], while Ajdarić et al. used deep learning to predict the degradation kinetics of esomeprazole formulations [13]. More recently, Bayesian inference coupled with the Arrhenius equation has been proposed as a means of narrowing confidence intervals in early-stage stability predictions from accelerated test data [14].

Despite these promising developments, a systematic side-by-side comparison of multiple ML and DL architectures for the concurrent prediction of both drug stability and oral bioavailability using a unified dataset and consistent evaluation criteria remains underexplored. Moreover, there is growing recognition that prediction accuracy alone is insufficient; models must also furnish interpretable insights into the molecular determinants that drive favorable or unfavorable outcomes [15,16]. The SHapley Additive exPlanations (SHAP) framework, grounded in cooperative game theory, addresses this need by quantifying each feature's marginal contribution to a given prediction [17].

The present work aims to fill this gap by constructing, optimizing, and comparing five computational models RF, SVM, GBM, ANN, and CNN for predicting both chemical stability (expressed as shelf life in months) and oral bioavailability (expressed as a percentage) of drug molecules from molecular descriptors and fingerprints. Additionally, SHAP-based interpretability analysis is performed to elucidate which structural and physicochemical attributes most strongly govern prediction performance.

2. Methods

2.1 Dataset Assembly

A dataset comprising 450 small-molecule drug compounds was compiled from publicly accessible databases DrugBank, PubChem, and ChEMBL supplemented by experimentally validated values extracted from peer-reviewed literature and regulatory submissions filed with the United States Food and Drug Administration (FDA). For every compound, two target endpoints were recorded: (i) oral bioavailability, defined as the percentage of the administered dose reaching the systemic circulation; and (ii) chemical stability, expressed as shelf life in months under ICH long-term storage conditions

(25°C, 60% relative humidity) [4]. Only compounds with complete, experimentally determined values for both endpoints were retained.

2.2 Molecular Representation and Feature Engineering

Molecular descriptors were calculated using the RDKit open-source cheminformatics toolkit (version 2023.09) [18]. A total of 206 two-dimensional descriptors were generated per compound, covering physicochemical properties (molecular weight, LogP, topological polar surface area), constitutional descriptors (atom counts, bond counts, ring counts), and topological connectivity indices. In parallel, Extended-Connectivity Fingerprints of radius 2 (ECFP4, 1024-bit vectors) were computed as structural representations for the deep learning models, consistent with widely adopted practices in chemoinformatics research [19].

2.3 Data Preprocessing

Missing values, which accounted for less than 3% of the dataset, were addressed using k-nearest neighbors imputation (k = 5). Features with near-zero variance (standard deviation below 0.01) were excluded, and pairwise Pearson correlation analysis was used to remove highly redundant features ($|r| > 0.95$). The surviving descriptor set was standardized via z-score normalization. The dataset was then partitioned through stratified random sampling into training (70%, n = 315), validation (15%, n = 68), and test (15%, n = 67) subsets to ensure balanced representation of both target variables across splits.

2.4 Machine Learning Algorithms

Three classical ML algorithms were implemented using the scikit-learn library (version 1.3) within a Python 3.11 environment:

Random Forest (RF): An ensemble of 500 bootstrap-aggregated decision trees with a maximum depth of 20 and a minimum of 5 samples per leaf. Feature importance was extracted via the mean decrease in impurity criterion [20].

Support Vector Machine (SVM): A regression variant using a radial basis function (RBF) kernel. The regularization constant C and kernel width γ were optimized through 5-fold cross-validated grid search across the ranges {0.1, 1, 10, 100} and {0.001, 0.01, 0.1, 1}, respectively [21].

Gradient Boosting Machine (GBM): Implemented via XGBoost (version 1.7). Hyperparameters including learning rate (0.01–0.3), tree depth (3–10), number of boosting rounds (100–1000), and column sampling ratio (0.6–1.0) were tuned through Bayesian optimization over 100 iterations [22].

2.5 Deep Learning Architectures

Two neural network models were developed using TensorFlow 2.14 with the Keras API:

Artificial Neural Network (ANN): A fully connected feedforward architecture comprising four hidden layers (256, 128, 64, and 32 neurons), each with rectified linear unit (ReLU) activations and batch normalization. Dropout regularization (rate = 0.3) was incorporated after each hidden layer. Training was conducted for up to 200 epochs with early stopping (patience = 20 epochs) using the Adam optimizer at an initial learning rate of 0.001 [8].

Convolutional Neural Network (CNN): A one-dimensional CNN designed to operate on the 1024-bit ECFP4 fingerprint vectors. The architecture comprised three convolutional blocks (64, 128, and 256 filters; kernel size = 3; ReLU activation; max pooling), followed by two dense layers (128 and 64 neurons) with dropout (rate = 0.4). Training parameters mirrored those of the ANN [9,23].

2.6 Performance Evaluation

All models were assessed on the withheld test set using three regression metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Generalizability was evaluated through 5-fold cross-validation on the training partition. Statistical differences between model performances were tested using paired t-tests with Bonferroni correction for multiple comparisons.

2.7 Interpretability via SHAP Analysis

To identify the molecular attributes most strongly associated with prediction outcomes, SHAP values were computed for each model [17]. SHAP assigns every feature an importance score reflecting its marginal contribution to the model output relative to the baseline prediction. Global feature importance rankings were derived by averaging the absolute SHAP values across all test set instances, following the practical framework described by Ponce-Bobadilla et al. [24]. This analysis enables chemically meaningful interpretation of the models' internal decision logic.

3. Results

3.1 Bioavailability Prediction Performance

The predictive accuracy of all five models for oral bioavailability is presented in Table 1. Among the architectures evaluated, the CNN yielded the strongest performance, achieving an R^2 of 0.91, an RMSE of 8.34%, and an MAE of 6.12%. The GBM emerged as the top-performing classical ML method (R^2 = 0.87, RMSE = 10.56%), while the ANN delivered closely comparable results (R^2 = 0.88, RMSE = 9.87%). Both RF and SVM exhibited moderate but meaningful predictive capacity.

Table 1. Performance metrics for oral bioavailability prediction across five computational models.

Model	R^2	RMSE (%)	MAE (%)	CV R^2 (±SD)
Random Forest	0.82	13.21	9.87	0.80 ± 0.03
SVM (RBF)	0.79	14.58	11.34	0.77 ± 0.04
GBM (XGBoost)	0.87	10.56	7.89	0.85 ± 0.02
ANN	0.88	9.87	7.23	0.86 ± 0.03
CNN	0.91	8.34	6.12	0.89 ± 0.02

3.2 Stability Prediction Performance

Table 2 reports the corresponding results for chemical stability prediction. In contrast to the bioavailability task, the GBM model attained the highest accuracy (R^2 = 0.89, RMSE = 4.21 months), surpassing both deep learning models. The CNN ranked second (R^2 = 0.87, RMSE = 4.89 months), followed by the ANN (R^2 = 0.85, RMSE = 5.56 months). The RF and SVM models produced lower but still appreciable predictive performance.

Table 2. Performance metrics for drug stability (shelf life) prediction across five computational models.

Model	R^2	RMSE (months)	MAE (months)	CV R^2 (±SD)
Random Forest	0.81	6.34	4.78	0.79 ± 0.03
SVM (RBF)	0.76	7.12	5.45	0.74 ± 0.05
GBM (XGBoost)	0.89	4.21	3.12	0.87 ± 0.02
ANN	0.85	5.56	4.11	0.83 ± 0.03
CNN	0.87	4.89	3.67	0.85 ± 0.02

3.3 Feature Importance Analysis

SHAP-based feature importance analysis identified a consistent set of molecular descriptors driving predictions across models. Table 3 lists the ten highest-ranking descriptors by mean absolute SHAP value, averaged across all five architectures. LogP emerged as the most influential descriptor for both prediction tasks. Molecular weight, hydrogen bond donor (HBD) count, topological polar surface area (TPSA), and rotatable bond count followed in importance for bioavailability prediction. For stability forecasting, the number of aromatic rings and the count of hydrolytically labile bonds (esters, amides) assumed heightened significance relative to the bioavailability task.

Table 3. Top ten molecular descriptors ranked by mean absolute SHAP value (averaged across all models).

Rank	Descriptor	Bioavailability SHAP	Stability SHAP
1	LogP (Lipophilicity)	0.142	0.138
2	Molecular Weight	0.128	0.121
3	HBD Count	0.115	0.098

4	TPSA	0.103	0.089
5	Rotatable Bond Count	0.091	0.082
6	HBA Count	0.078	0.075
7	Aromatic Ring Count	0.065	0.094
8	Fraction Csp3	0.058	0.067
9	Molar Refractivity	0.052	0.054
10	Labile Bond Count	0.041	0.088

3.4 Cross-Validation and Generalizability

Five-fold cross-validation yielded consistent R^2 values across folds for all models, with standard deviations ranging from 0.02 to 0.05 (Tables 1 and 2). These narrow confidence intervals indicate minimal overfitting and suggest that the models generalize satisfactorily to unseen data. Learning curve analysis performed on the CNN and GBM models showed performance convergence at approximately 300 training samples, indicating that the 450-compound dataset provided an adequate foundation for model construction.

4. Discussion

The outcomes of this investigation provide compelling evidence that both classical ML and modern DL architectures can generate practically useful predictions of drug stability and oral bioavailability from molecular representations. Several noteworthy patterns merit discussion.

The CNN's superiority in bioavailability prediction aligns with the capacity of convolutional layers to extract hierarchical patterns from binary fingerprint vectors, thereby capturing structural motifs such as aromatic scaffolds, hydrogen-bonding groups, and flexible chains that collectively govern intestinal permeability and first-pass hepatic metabolism [9,23]. This observation is consistent with prior work by Ng and Lu, who demonstrated that graph-based neural networks outperform conventional descriptor-driven models for bioavailability classification tasks [11].

Conversely, the GBM model's leading performance on the stability task is notable. Gradient boosting operates by iteratively correcting residual errors through sequential weak learners, an approach that appears well matched to the nature of chemical degradation data, where stability is governed by a relatively compact set of reactivity-related descriptors (labile bond counts, aromatic ring presence, oxidation susceptibility) acting in concert with storage-condition variables [12,22]. This finding echoes the broader machine learning literature, which consistently reports that gradient-boosted tree ensembles compete effectively with, and sometimes outperform, deep neural networks on tabular datasets of moderate size [25].

The feature importance rankings extracted via SHAP analysis carry clear pharmaceutical significance. The dominant role of LogP in both prediction tasks is well established: highly lipophilic molecules tend to exhibit poor aqueous solubility and incomplete dissolution, which limit absorption, while also being prone to oxidative and hydrolytic degradation [3,26]. Molecular weight influences membrane permeability in a manner consistent with the heuristics codified in Lipinski's Rule of Five [27], and larger molecules typically present more complex degradation pathways. The high ranking of TPSA is consistent with its recognized role as a surrogate for intestinal permeability compounds with excessive polar surface area traverse lipid bilayers less efficiently while hydrogen bond donor count affects both crystal packing energy (influencing physical stability) and passive transcellular transport [5,28].

An important observation is the divergent importance of labile bond count and aromatic ring count across the two tasks. These descriptors ranked considerably higher for stability than for bioavailability, reflecting the well-known susceptibility of ester and amide linkages to hydrolytic cleavage under storage conditions, and the contribution of aromatic systems to photodegradation and oxidation pathways [4]. Such task-specific patterns underscore the value of SHAP analysis in moving beyond aggregate accuracy metrics toward mechanistically informative model interpretation [15,17].

Several limitations of the present study should be noted. The dataset of 450 compounds, though carefully curated, is modest relative to the chemical diversity of pharmaceutical space, and model performance may improve with larger, more heterogeneous collections. All molecular representations employed were two-dimensional; the inclusion of three-dimensional conformational descriptors could enhance prediction accuracy for endpoints influenced by stereochemistry and molecular shape [11,29]. Stability predictions were confined to chemical degradation under standardized ICH conditions and did not incorporate formulation-specific variables such as excipient compatibility, moisture uptake by packaging materials, or light exposure profiles.

Future research should explore graph neural network architectures, which operate directly on molecular graphs and have shown promise for pharmacokinetic property prediction [11,29]. Transfer learning from large pretrained chemical language models represents another avenue for improving accuracy in low-data regimes. The integration of three-dimensional descriptors and formulation parameters into multi-task learning frameworks may further extend the utility of these computational tools across a broader range of pharmaceutical development scenarios [6,30].

5. Conclusions

This study demonstrates that deep learning and machine learning algorithms can deliver robust, quantitative predictions of drug stability and oral bioavailability from molecular descriptors and fingerprints. Among the models evaluated, the CNN architecture achieved the strongest bioavailability prediction ($R^2 = 0.91$), while the GBM model excelled at stability forecasting ($R^2 = 0.89$). SHAP-based interpretability analysis identified lipophilicity, molecular weight, hydrogen bond donor count, and topological polar surface area as the molecular attributes most strongly influencing prediction accuracy for both endpoints. These computational tools hold considerable promise for accelerating early-stage pharmaceutical screening, enabling researchers to prioritize candidates with favorable stability and absorption characteristics before committing to resource-intensive experimental evaluation.

Acknowledgments

The authors gratefully acknowledge the computational infrastructure and institutional support provided by Anwarul Uloom College of Pharmacy, Hyderabad. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

[1] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spjuth, O. and Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), pp.463–477.

[2] Vora, L.K., Gholap, A.D., Jetha, K., Thakur, R.R.S., Solanki, H.K. and Chavda, V.P. (2023). Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics, 15(7), p.1916.

[3] Wang, J. and Hou, T. (2015). Advances in computationally modeling human oral bioavailability. Advanced Drug Delivery Reviews, 86, pp.11–16.

[4] ICH Expert Working Group (2003). Stability testing of new drug substances and products Q1A(R2). International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use.

[5] Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. and Shanley, M. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46(1–3), pp.3–26.

[6] Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.H.M. and Ahsan, M.J. (2022). Machine learning in drug discovery: a review. Artificial Intelligence Review, 55(3), pp.1947–1999.

[7] Lo, Y.C., Rensi, S.E., Torber, W. and Altman, R.B. (2018). Machine learning in chemoinformatics and drug discovery. Drug Discovery Today, 23(8), pp.1538–1546.

[8] Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. and Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), pp.1241–1250.

[9] Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gomez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A. and Adams, R.P. (2015). Convolutional networks on graphs for learning molecular fingerprints. Advances in Neural Information Processing Systems, 28, pp.2224–2232.

[10] Wei, M., Zhang, X., Pan, X., Wang, B., Ji, C., Qi, Y. and Zhang, J.Z.H. (2022). HobPre: accurate prediction of human oral bioavailability for small molecules. Journal of Cheminformatics, 14, p.1.

[11] Ng, S.S.S. and Lu, Y. (2023). Evaluating the use of graph neural networks and transfer learning for oral bioavailability prediction. Journal of Chemical Information and Modeling, 63(16), pp.5035–5044.

[12] Han, R., Xiong, H., Ye, Z., Yang, Y., Huang, T., Jing, Q., Lu, J., Pan, H., Ren, F. and Ouyang, D. (2019). Predicting physical stability of solid dispersions by machine learning techniques. Journal of Controlled Release, 311, pp.16–25.

[13] Ajdarić, J., Ibrić, S., Pavlović, A., Ignjatović, L. and Ivković, B. (2021). Prediction of drug stability using deep learning approach: case study of esomeprazole 40 mg freeze-dried powder for solution. Pharmaceutics, 13(6), p.829.

[14] Watanabe, S., Yonemochi, E. and Wakiyama, N. (2023). Algorithm for the early prediction of drug stability using Bayesian inference and multiple measurements: application for predicting the stability of silodosin tablets. International Journal of Pharmaceutics, 641, p.123089.

[15] Rodríguez-Pérez, R. and Bajorath, J. (2020). Interpretation of machine learning models using Shapley values: application to compound potency and multi-target activity predictions. Journal of Computer-Aided Molecular Design, 34, pp.1013–1026.

[16] Lavecchia, A. (2015). Machine-learning approaches in drug discovery: methods and applications. Drug Discovery Today, 20(3), pp.318–331.

[17] Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, pp.4765–4774.

[18] RDKit: Open-Source Cheminformatics Software. Available at: https://www.rdkit.org/ (Accessed: 2024).

[19] Rogers, D. and Hahn, M. (2010). Extended-connectivity fingerprints. Journal of Chemical Information and Modeling, 50(5), pp.742–754.

[20] Breiman, L. (2001). Random forests. Machine Learning, 45(1), pp.5–32.

[21] Burbidge, R., Trotter, M., Buxton, B. and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. Computers & Chemistry, 26(1), pp.5–14.

[22] Chen, T. and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794.

[23] Mayr, A., Klambauer, G., Unterthiner, T. and Hochreiter, S. (2016). DeepTox: toxicity prediction using deep learning. Frontiers in Environmental Science, 3, p.80.

[24] Ponce-Bobadilla, A.V., Schmitt, V., Maier, C.S., Mensing, S. and Stodtmann, S. (2024). Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clinical and Translational Science, 17, e70056.

[25] Grinsztajn, L., Oyallon, E. and Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems, 35, pp.507–520.

[26] Schneckener, S., Grimbs, S., Hey, J., Menz, S., Osmers, M., Schaper, S., Hillisch, A. and Göller, A.H. (2019). Prediction of oral bioavailability in rats: transferring insights from in vitro correlations to (deep) machine learning models using in silico model outputs and chemical structure parameters. Journal of Chemical Information and Modeling, 59(11), pp.4893–4905.

[27] Lipinski, C.A. (2004). Lead- and drug-like compounds: the rule-of-five revolution. Drug Discovery Today: Technologies, 1(4), pp.337–341.

[28] Egan, W.J., Merz, K.M. and Baldwin, J.J. (2000). Prediction of drug absorption using multivariate statistics. Journal of Medicinal Chemistry, 43(21), pp.3867–3877.

International Journal of Multidisciplinary Science and Innovation (IJMSI)
Website: https://ijmsi.in/
ISSN: 3107-5754
Vol. 2, Issue 1, Jan-Jun, 2026
Page No.: 103-110

[29] Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., Palmer, A., Settels, V., Jaakkola, T., Jensen, K. and Barzilay, R. (2019). Analyzing learned molecular representations for property prediction. Journal of Chemical Information and Modeling, 59(8), pp.3370–3388.

[30] Bao, Z., Bufton, J., Hickman, R.J., Aspuru-Guzik, A., Bannigan, P. and Allen, C. (2023). Revolutionizing drug formulation development: the increasing impact of machine learning. Advanced Drug Delivery Reviews, 202, p.115108.

*****