International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Machine Learning-Based Computational Intelligence Frameworks for Drug Discovery

Sadhana Singh, Vishal Bharati
Dept of Computer Science,
Suyash Institute of Information Technology, U.P., India
sadhanasingh9651@gmail.com, vishalbharati161715@gmail.com

Abstract: A simplified and application-oriented machine learning framework has been proposed for the estimation of enzyme inhibition parameters under both healthy and impaired kidney function conditions using two representative pharmaceutical compounds. The developed framework is implemented in the MATLAB environment, allowing flexible modification, rapid deployment, and efficient reuse for a wide range of drug discovery and pharmaceutical research applications. Due to its high computational performance, clinical significance, and user-friendly graphical user interface (GUI), the system functions not only as a reliable analytical platform for pharmaceutical studies but also as an interactive educational tool for understanding complex biomedical and biochemical processes. The integration of computational intelligence techniques enables dynamic visualization, predictive analysis, and improved conceptual interpretation of enzyme-related mechanisms. In addition, the findings of this research provide meaningful insights into enzyme inhibition behavior and support future investigations related to disease progression, drug response, and kidney-associated infections, thereby contributing toward the development of more effective therapeutic and diagnostic strategies in modern healthcare and precision medicine.

Keywords: Artificial intelligence, Drug discovery, Machine learning, Target validation.

1. Introduction:

Drugs constitute one of the most essential components of modern healthcare systems, playing a crucial role in the prevention, diagnosis, management, and treatment of a wide variety of acute and chronic diseases. Over the past several decades, pharmaceutical research has contributed significantly to improving global life expectancy and quality of life through the development of innovative therapeutic agents and treatment strategies. However, despite remarkable progress in medicinal chemistry, biotechnology, and clinical pharmacology, a large number of medical conditions still lack effective therapeutic solutions. Many complex diseases, including cancer, neurodegenerative disorders, rare genetic conditions, and emerging infectious diseases, continue to pose serious challenges to healthcare systems worldwide. The urgent need for rapid and efficient therapeutic development became especially evident during the coronavirus disease (COVID-19) pandemic, which exposed critical limitations in conventional drug discovery and development pipelines. The pandemic highlighted the importance of accelerated pharmaceutical innovation and demonstrated the necessity for advanced computational approaches capable of reducing the time required to identify and develop safe and effective drugs.

The process of drug discovery and development is inherently complex, multidisciplinary, and highly resource-intensive. It involves several interconnected stages, including target identification, hit discovery, lead optimization, preclinical testing, clinical trials, and regulatory approval. Each phase requires the integration of knowledge from multiple scientific disciplines such as molecular biology, chemistry, pharmacology, toxicology, bioinformatics, and clinical medicine. Traditionally, the development of a single successful therapeutic compound may require more than ten to fifteen years of continuous research and financial investments amounting to billions of dollars. Reports indicate that the average cost associated with bringing a new drug to market can range between 1.5 and 3 billion US dollars, depending on the complexity of the disease, the therapeutic approach, and the probability of clinical success. Despite these substantial investments, the overall success rate of drug candidates remains relatively low. A significant proportion of compounds fail during preclinical or clinical evaluation stages due to insufficient efficacy, unexpected toxicity, poor pharmacokinetic properties, or lack of therapeutic benefit. Studies have shown that only a small percentage of drug development programs ultimately receive regulatory approval, while therapies designed for rare diseases or orphan conditions exhibit even lower success rates. Consequently, the pharmaceutical industry continues to face increasing economic pressure associated with high research and development expenditures combined with declining productivity and limited approval rates.

To address these challenges, computational methods and computer-aided drug design (CADD) technologies were introduced to enhance the efficiency of pharmaceutical research processes. CADD approaches utilize computational modeling, molecular docking, virtual screening, simulation techniques, and predictive analytics to support the identification and optimization of potential drug candidates. These technologies have contributed to reducing certain experimental costs and improving decision-making during early-stage drug development. Nevertheless, despite the widespread adoption of computational tools, the

pharmaceutical sector continues to encounter significant limitations in productivity, scalability, and predictive reliability. Traditional computational models often depend heavily on handcrafted molecular descriptors, domain expertise, and simplified assumptions that may not fully capture the complexity of biological systems and molecular interactions. As a result, there remains a strong demand for more adaptive, intelligent, and data-driven approaches capable of transforming the conventional drug discovery paradigm.

In recent years, artificial intelligence (AI) and machine learning (ML) technologies have emerged as powerful solutions capable of revolutionizing pharmaceutical research and drug discovery. Machine learning, a major branch of artificial intelligence, focuses on developing algorithms that can automatically learn patterns, relationships, and predictive models from large datasets without explicit programming. The rapid growth of biological, chemical, pharmacological, and clinical datasets, combined with substantial advancements in high-performance computing and graphical processing units (GPUs), has significantly accelerated the adoption of ML-based approaches in biomedical research. Although the fundamental concepts of artificial neural networks and deep learning were introduced decades ago, the true potential of deep learning became widely recognized after the remarkable success of deep convolutional neural networks in large-scale image recognition tasks, particularly during the ImageNet competition in 2012. This breakthrough demonstrated the ability of deep learning models to automatically extract highly informative features from raw data, thereby outperforming traditional machine learning methods across various domains.

The success of deep learning rapidly influenced the field of cheminformatics and pharmaceutical sciences. Machine learning techniques such as Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Neural Networks (GNNs) are now extensively employed in multiple stages of drug discovery and development. These techniques enable efficient virtual screening of massive chemical libraries, prediction of molecular properties, estimation of biological activity, identification of drug-target interactions, and optimization of lead compounds. ML-based models are also widely used for evaluating critical pharmacokinetic and pharmacodynamic parameters, including absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties, which play a vital role in determining drug safety and efficacy. Furthermore, machine learning has significantly improved Quantitative Structure–Activity Relationship (QSAR) modeling by enabling the automatic extraction of meaningful molecular representations directly from raw chemical structures, thereby reducing dependence on manually engineered descriptors and enhancing predictive performance.

Another major advantage of modern ML and deep learning approaches is their ability to integrate heterogeneous biomedical data sources, including genomics, proteomics, transcriptomics, metabolomics, electronic health records, and biomedical literature. The incorporation of natural language processing (NLP) techniques further supports automated knowledge extraction, biomarker identification, disease association analysis, and target discovery from scientific publications and clinical reports. In addition, the increasing availability of publicly accessible biochemical and structural databases, generated through high-throughput screening technologies and experimental advancements, has created unprecedented opportunities for data-driven drug discovery research. Landmark achievements such as AlphaFold and AlphaFold2 have further demonstrated the transformative capability of deep learning in accurately predicting protein structures with near-experimental precision, thereby opening new possibilities for understanding biological mechanisms and accelerating structure-based drug design.

Despite these significant advancements, several important challenges still limit the full-scale adoption of machine learning in pharmaceutical research. Issues related to data quality, dataset imbalance, model interpretability, reproducibility, overfitting, and limited generalization across diverse biological systems remain critical concerns. Many deep learning models operate as “black-box” systems, making it difficult for researchers and clinicians to interpret prediction outcomes and validate biological relevance. To overcome these limitations, recent studies have increasingly focused on explainable artificial intelligence (XAI), transfer learning, federated learning, reinforcement learning, and multi-task learning approaches aimed at improving transparency, robustness, and reliability. Overall, the integration of artificial intelligence, machine learning, bioinformatics, and computational intelligence is fundamentally reshaping the future of drug discovery and pharmaceutical development. These emerging technologies hold tremendous potential to accelerate therapeutic innovation, reduce research costs, improve prediction accuracy, and enable the development of safer, more efficient, and highly personalized treatment strategies for complex human diseases.

2. Methodology:

Angiotensin II (Ang II) is a potent vasoactive hormone that plays a central role in regulating arterial blood pressure by inducing vasoconstriction. Prolonged elevation of Ang II levels contributes to the development of hypertension, thereby increasing cardiac workload and predisposing individuals to a range of cardiovascular and renal complications. Angiotensin-converting enzyme (ACE) inhibitors constitute an important class of antihypertensive agents that suppress the enzymatic conversion of angiotensin I (Ang I) into Ang II. By reducing Ang II synthesis, these drugs effectively lower circulating hormone levels and promote blood pressure control.

The biochemical transformation of Ang I into Ang II represents a critical regulatory step within the renin–angiotensin system (RAS), a complex hormonal cascade responsible for maintaining vascular tone, electrolyte balance,

and fluid homeostasis (Fig. 1). Within this regulatory framework, Ang II exerts a negative feedback effect on renin secretion, thereby modulating the upstream conversion of angiotensinogen (AGT) into Ang I. In this study, we focus on a targeted subset of the RAS pathway comprising the principal hormones and enzymatic components that are directly influenced by ACE inhibition, as illustrated in Fig. 1.

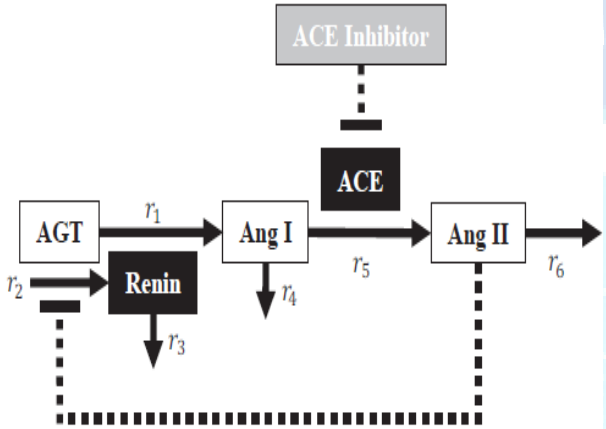

Fig 1: The reaction network for the renin-angiotensin system.

Angiotensin-converting enzyme (ACE) inhibitors are widely prescribed for the management of hypertension, congestive heart failure, and chronic kidney disease (CKD) (Balfour and Goa, 1991; Brown and Vaughan, 1998; Corbo et al., 2016). CKD represents a major complication associated with diabetes mellitus and is a significant contributor to morbidity and mortality in both type I and type II diabetic populations. As the leading cause of end-stage renal disease, uncontrolled CKD can ultimately progress to kidney failure, particularly in the presence of dysregulated Ang II levels. Consequently, delaying the progression of CKD prior to the onset of irreversible renal damage is of critical clinical importance, and ACE inhibitors have consistently demonstrated therapeutic efficacy in achieving this objective (Asher and Murray, 1991; Hoyer et al., 1993; Brown and Vaughan, 1998; Hsu et al., 2014; Yamout et al., 2014).

Given that both diabetes and CKD frequently coexist with hypertension, ACE inhibitors offer a highly effective pharmacological strategy for addressing multiple comorbid conditions simultaneously. Numerous ACE inhibitor formulations are currently available in clinical practice. In the present study, benazepril and cilazapril are selected due to their well-documented renoprotective properties and established use in the treatment of both hypertension and CKD (Hoyer et al., 1993; Niu et al., 2014). Moreover, extensive pharmacokinetic and pharmacodynamic investigations have been conducted to characterize the dose–response relationships and effects of these drugs on the renin–angiotensin system (RAS) in hypertensive patients with normal renal function (NRF) and impaired renal function (IRF) (Shionoiri et al., 1988, 1992; Kloke et al., 1996).

The modeling framework and parameter estimation methodologies proposed in this work are readily extensible to other ACE inhibitors, provided that sufficient experimental data are available to define their pharmacological profiles. It is noteworthy that both benazepril and cilazapril undergo extensive bioactivation to their respective diacid metabolites, which represent the pharmacologically active forms of these compounds—a characteristic shared by most ACE inhibitors (Hoyer et al., 1993; Toutain and Lefebvre, 2004; LeBlanc et al., 2006). Accordingly, all model parameters and computational analyses in this study are based exclusively on the active diacid derivatives of these drugs.

.

3. Machine Learning

Machine learning attempts to learn patterns directly from data without explicit functional pre-specification for use in prediction, decision making, or other out comes of interest (Mitchell, 1997; Murphy, 2012). These methodologies are often classified into several paradigms: supervised learning, unsupervised learning, and reinforcement learning. However, these paradigms are not mutually exclusive, and there are many connections between them.

In supervised learning we are interested in fitting a function f : X →Y using a data set, D, of n labelled observations

$$D = \{(x_i, y_i), i = 1 \ldots n\}$$

where $x_i \in X$ and $y_i \in Y$. Typical applications include regression and classification tasks.

In unsupervised learning, we do not have access to labels and thus our data set, D, consists of only observations of the source domain X, reducing to

$$D = \{x_i, i = 1 \ldots n\}$$

In this paradigm, the goal is to find some notion of internal structure or common featurisation. Clustering (Lloyd, 1982), anomaly detection (Hodge and Austin, 2004), and dimensionality reduction techniques (Maaten et al., 2007) are common unsupervised methodologies.

Finally, reinforcement learning aims to learn an optimal policy for an agent in an environment, given some notion of reward. While many formulations exist, a basic and natural one is

$$D = \{(si, ai, ri), i = 1 \ldots n\}$$

where $s_i \in S$ is the state of the system of environment, ai 2 A is the action taken by the agent, and ri 2 R is the reward given for taking action ai in state si. For the work presented in this thesis we mostly are concerned with the supervised and unsupervised paradigms, with a particular emphasis on deep learning-based models and their applications to drug discovery. In the remainder of this section, we will introduce two broad categories of machine learning algorithms that are applicable within any of the paradigms outlined above. The first, convolutional neural networks (CNNs), led to the “ImageNet moment” discussed previously.

4. Result and Discussion:

In the experimental data used to fit the parameters, the dose for benazepril was 5 mg and that for cilazapril was 1.25 mg.

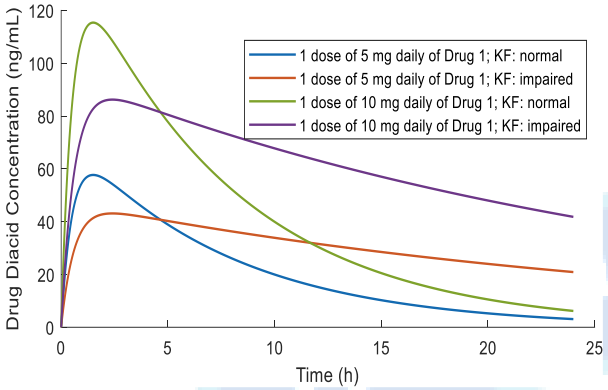

Fig 2: Benazepril validation results: diacid form of benazepril concentration versus time after a single dose for 5 mg and 10 mg doses for NRF and IRF.

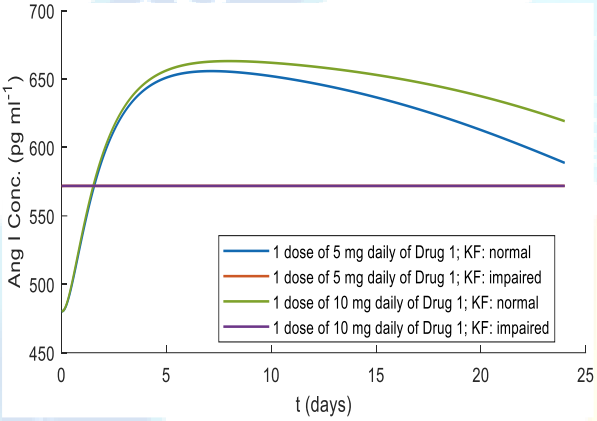

Fig 3: Benazepril validation results for doses of benazepril for NRF: Ang I concentration

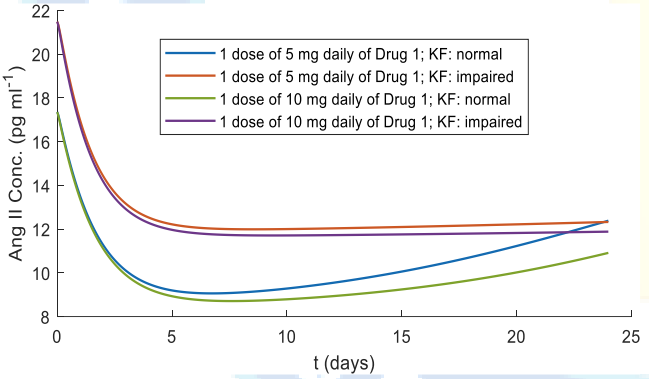

Fig 4: Benazepril validation results for doses of benazepril for Ang II concentration as functions of time

The following parameters for the PD model were estimated for four cases (benazepril and cilazapril for both normal function (NRF) and for impaired function (IRF): Vmax/KM, kR, kf, kAII, and f. The values reported for the fitted parameters were the median of the best-fit parameter sets along with the minima and maxima of the ranges of parameters (Table 5 for benazepril and Table 6 for cilazapril). The best-fit parameter sets were those with WSSR within 1% of that for the single best set (74, 93, 7, and 93 out of 101 multistart parameter sets for benazepril NRF, benazepril IRF, cilazapril NRF, and cilazapril IRF, respectively). The 95% prediction confidence intervals were determined using kernel density estimation with the best-fit parameter sets. The simulation results for the fitted parameters for all four cases are shown for output variables CAI, CAII, and PRA, respectively. Data obtained from (Shionoiri et al., 1992, 1988) are also shown in the figures.

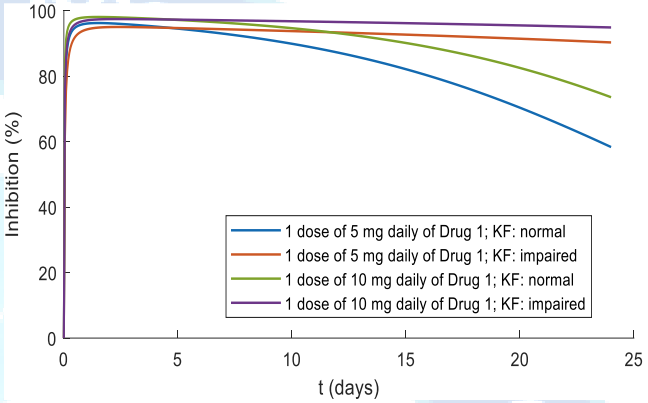

Fig 5: Benazepril validation results for doses of benazepril for Ang II concentration as functions of time

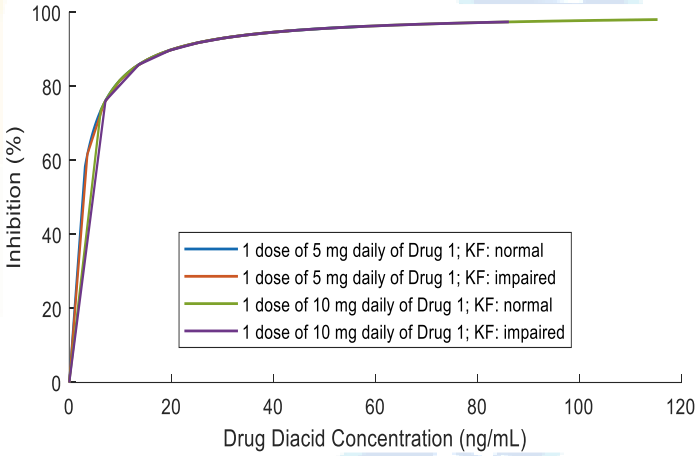

Fig 6: Validation results for doses of benazepril for inhibition as a function of drug diacid concentration

5. Conclusion:

A simplified yet physiologically meaningful pharmacokinetic/pharmacodynamic (PK/PD) modeling framework has been developed to characterize angiotensin-converting enzyme (ACE) inhibition under conditions of normal and impaired renal function for two representative drugs. To promote accessibility and facilitate broader adoption, the model has been implemented as a MATLAB-based application, enabling rapid reuse by researchers, educators, and practitioners. Owing to its computational efficiency, clinical relevance, and intuitive graphical user interface (GUI), the platform serves as an effective educational tool, providing dynamic and interactive

visualizations to support learning in biomedical science and chemical engineering domains.

The predictive capability of the model offers valuable insights into the systemic effects of ACE inhibition and establishes a foundation for future investigations into disease-specific conditions, particularly chronic kidney disease (CKD). Furthermore, the predicted temporal profiles of circulating Ang I and Ang II concentrations may be integrated into clinically relevant multiscale frameworks to examine localized tissue-level responses, thereby enabling detailed exploration of microvascular complications associated with CKD, diabetes mellitus, and hypertension.

References:

[1] Collins FS, Varmus H. A new initiative on precision medicine. New England J Med 2015;372(9):793–5.

[2] Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486(7403):346–52.

[3] Romond EH, Perez EA, Bryant J, Suman VJ, Geyer Jr CE, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, et al. Trastuzumab plus adjuvant chemotherapy for operable her2-positive breast cancer. N Engl J Med 2005;353(16):1673–84.

[4] Blanco JL, Porto-Pazos AB, Pazos A, Fernandez-Lozano C. Prediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selection. Sci Rep 2018;8(1):1–11.

[5] Munteanu CR, Fernández-Blanco E, Seoane JA, Izquierdo-Novo P, Angel Rodriguez-Fernandez J, Maria Prieto-Gonzalez J, Rabunal JR, Pazos A. Drug discovery and design for complex diseases through qsar computational methods. Current Pharmaceutical Des 2010;16(24):2640–55.

[6] García I, Munteanu CR, Fall Y, Gómez G, Uriarte E, González-Díaz H. Qsar and complex network study of the chiral hmgr inhibitor structural diversity. Bioorganic Med Chem 2009;17(1):165–75.

[7] Liu Y, Tang S, Fernandez-Lozano C, Munteanu CR, Pazos A, Yu Y-Z, Tan Z, González-Díaz H. Experimental study and random forest prediction model of microbiome cell surface hydrophobicity. Expert Syst Appl 2017;72:306–16.

[8] Riera-Fernández P, Munteanu CR, Dorado J, Martin-Romalde R, Duardo- Sanchez A, Gonzalez-Diaz H. From chemical graphs in computer-aided drug design to general markov-galvez indices of drug-target, proteome, drugparasitic disease, technological, and social-legal networks. Current Computeraided Drug Des 2011;7(4):315–37.

[9] Shirvani P, Fassihi A. Molecular modelling study on pyrrolo [2, 3-b] pyridine derivatives as c-met kinase inhibitors, a combined approach using molecular docking, 3d-qsar modelling and molecular dynamics simulation. Mol Simul 2020:1–16.

[10] B. Suay-Garcia, J.I. Bueso-Bordils, A. Falcó, M.T. Pérez-Gracia, G. Antón-Fos, P. Alemán-López, Quantitative structure–activity relationship methods in the discovery and development of antibacterials, Wiley Interdisciplinary Reviews: Computational Molecular Science e1472.

[11] Fernandez-Lozano C, Gestal M, Munteanu CR, Dorado J, Pazos A. A methodology for the design of experiments in computational intelligence with multiple regression models. PeerJ 2016;4:e2721.

[12] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., Drugbank 5.0: a major update to the drugbank database for 2018, Nucleic acids research 46 (D1) (2018) D1074– D1082..

[13] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al. Pubchem 2019 update: improved access to chemical data. Nucleic Acids Res 2019;47(D1):D1102–9.

[14] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, et al. Chembl: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40(D1): D1100–7.