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

Computational Intelligence in Drug Discovery: Machine Learning-Based Design Frameworks

Sandeep Sharma, Vinay Kumar
Dept of Computer Science,
BN college of engineering and technology, Lucknow, India
writetosandeep7@gmail.com

Abstract: A simplified and application-oriented machine learning framework has been developed for estimating enzyme inhibition parameters under both normal and impaired kidney function conditions for two representative drugs. The proposed model is implemented on the MATLAB platform, enabling rapid adaptation and reuse for broader drug discovery and pharmaceutical research applications. Owing to its high computational efficiency, real-world clinical relevance, and user-friendly graphical interface (GUI), the model serves not only as a practical tool for pharmaceutical analysis but also as an interactive educational resource. It facilitates dynamic visualization and conceptual understanding of biomedical and chemical processes through computational intelligence. Furthermore, the outcomes of this study provide valuable insights into enzyme inhibition mechanisms and support future investigations into disease-related conditions, particularly kidney infections, thereby contributing to improved therapeutic strategies.

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

1. Introduction:

Drugs constitute a fundamental component of modern healthcare systems, playing a vital role in the prevention, management, and treatment of a broad spectrum of medical conditions (World Health Organization, 2010). Despite substantial advancements in pharmaceutical sciences, numerous clinical needs—both long-standing and newly emerging—remain inadequately addressed by existing therapeutic options (Kaplan et al., 2013). The urgency of efficiently responding to such unmet medical demands has been starkly underscored by the recent global health crisis caused by the coronavirus disease (COVID-19) pandemic (Rosa et al., 2020). This unprecedented emergency highlighted critical limitations in the speed and scalability of conventional drug development pipelines.

The development of novel therapeutics is inherently complex, involving multiple sequential stages that integrate diverse scientific disciplines, including chemistry, biology, pharmacology, and clinical medicine. This multi-phase process typically spans more than a decade and entails enormous financial investments. Estimates indicate that the average cost of bringing a new drug to market ranges between $1.5 and $3 billion, depending on the methodological framework employed (Avorn, 2015; DiMasi et al., 2016), with timelines often exceeding ten years (Paul et al., 2010). Between 2008 and 2016, the U.S. Food and Drug Administration (FDA) approved an average of only 31 novel therapeutics annually (U.S. Food and Drug Administration, 2018a). These figures reveal persistent stagnation in productivity, raising serious concerns regarding the sustainability of prevailing pharmaceutical research and development (R&D) models (Moors et al., 2014; Ernst & Young, 2017).

A significant proportion of drug development expenditure arises from the high likelihood of failure, as substantial investments in time and resources frequently do not culminate in regulatory approval. Recent analyses report that merely 13.8% of drug development initiatives ultimately succeed, while programs targeting rare diseases—commonly referred to as orphan drugs—exhibit an even lower success rate of approximately 6.2% (Wong et al., 2018). This enduring challenge of escalating costs coupled with diminishing returns underscores the critical need for transformative solutions in pharmaceutical research (Myers and Baker, 2001).

Computer-aided drug design (CADD) has long been regarded as a promising strategy to accelerate drug discovery processes and reduce associated costs by leveraging computational modeling, simulation, and predictive analytics (Ou-Yang et al., 2012). Nevertheless, despite the widespread adoption of computational tools across various stages of drug development, overall R&D expenses have continued to rise, and productivity gains remain limited (DiMasi et al., 2003; Avorn, 2015; DiMasi et al., 2016; Khanna, 2012). This persistent inefficiency highlights the necessity for innovative methodologies capable of fundamentally reshaping the drug discovery paradigm.

In recent years, artificial intelligence (AI), particularly deep learning, has emerged as a powerful technological driver across numerous scientific and industrial domains. Although the foundational concepts of neural networks were established several decades ago (Rosenblatt, 1958; Fukushima, 1980; Rumelhart et al., 1986), their transformative potential became widely recognized following the landmark success of deep convolutional neural networks in the 2012 ImageNet Large Scale Visual Recognition Challenge. In this competition, Krizhevsky et al. (2012) achieved a remarkable 41% improvement over competing methods, a milestone often

referred to as the "ImageNet moment." This breakthrough was facilitated by unprecedented access to large-scale labeled datasets and enhanced computational capabilities, catalyzing rapid advancements in machine learning performance across diverse applications, including image recognition (He et al., 2015), reinforcement learning in single-agent environments (Mnih et al., 2015), and complex strategic gameplay, such as Go, Chess, and StarCraft II (Silver et al., 2016; Silver et al., 2017; Silver et al., 2018; Vinyals et al., 2019).

These technological advances have swiftly influenced the field of cheminformatics, yielding early yet compelling results. In 2013, deep neural networks demonstrated superior predictive performance in the Merck molecular activity challenge (Ma et al., 2015), while similar achievements were reported during the Tox21 toxicity prediction competition in 2015 (Mayr et al., 2016). Although learning-based models have been employed in pharmaceutical research for decades, notably in quantitative structure–activity relationship (QSAR) modeling since the early 1960s (Hansch et al., 1962), traditional machine learning techniques typically rely on handcrafted feature representations derived from molecular descriptors (Salt et al., 1992; Klambauer et al., 2019). This dependency has led to the creation of hundreds of molecular descriptors aimed at capturing chemical and biological properties (Deng et al., 2004; Zhang et al., 2006; Durrant and McCammon, 2011). In contrast, deep learning approaches offer a significant advantage by enabling automated feature learning directly from raw molecular or structural data, thereby reducing reliance on manual abstraction and enhancing model generalizability (Klambauer et al., 2019).

Furthermore, conventional QSAR models are typically tailored to specific datasets and chemical series, limiting their applicability beyond individual drug discovery projects. Lessons drawn from breakthroughs in computer vision and natural language processing indicate that large-scale datasets are crucial for training robust, general-purpose predictive models (Halevy et al., 2009; Sun et al., 2017). Encouragingly, recent years have witnessed an exponential growth in publicly accessible biochemical, pharmacological, and structural databases, driven by advancements in high-throughput screening technologies and experimental methodologies (Inglese et al., 2007; Kim et al., 2015; Papadatos et al., 2015; Berman et al., 2000; Burley et al., 2019). A prominent illustration of the transformative impact of large-scale data integration is evident in protein structure prediction, where deep learning-based systems such as AlphaFold and AlphaFold2 achieved unprecedented accuracy during the CASP 13 and CASP 14 competitions, respectively (Senior et al., 2020; Jumper et al., 2020; Kryshtafovych et al., 2019).

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.

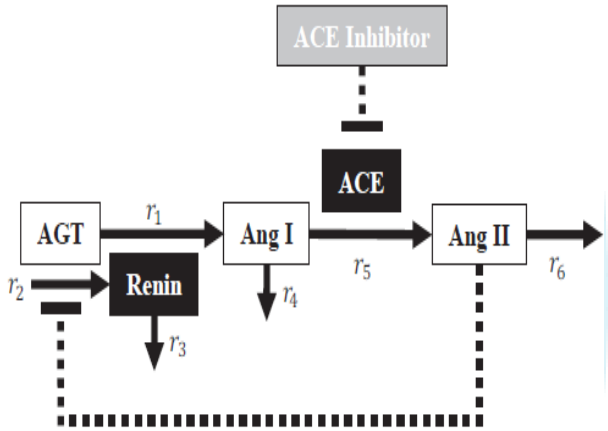

Fig 1: The reaction network for the renin-angiotensin system.

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. 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.

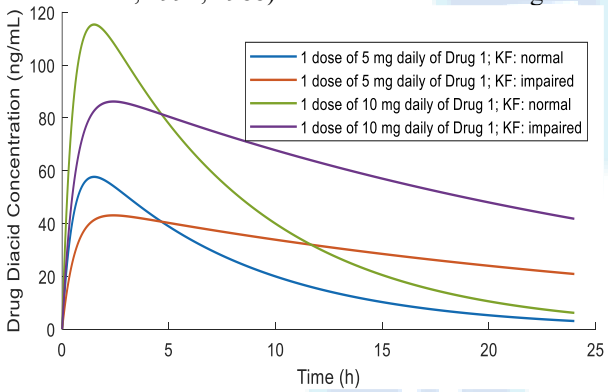

Fig 2: Benazepril validation results: diacid form of benazepril concentration versus time after a single dose for 5 mg and 10 mg doses for NRF and IRF.

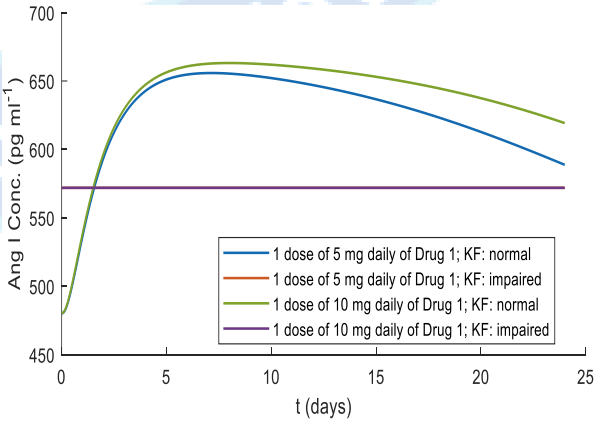

Fig 3: Benazepril validation results for doses of benazepril for NRF: Ang I concentration

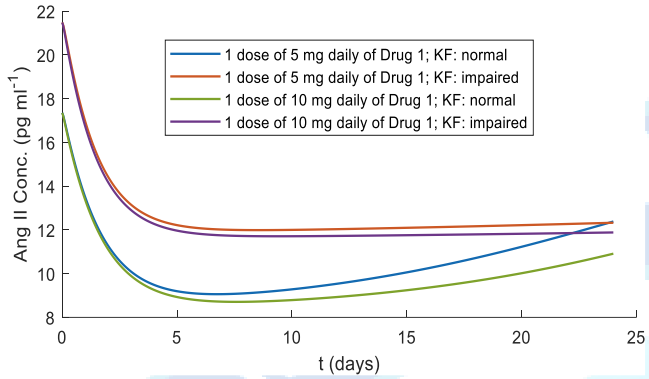

Fig 4: Benazepril validation results for doses of benazepril for Ang II concentration as functions of time

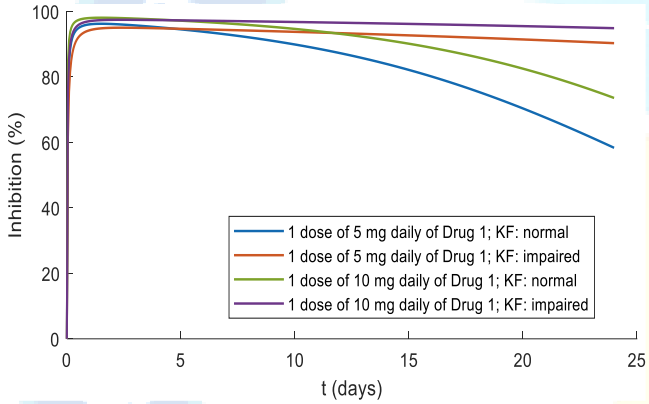

Fig 5: Benazepril validation results for doses of benazepril for Ang II concentration as functions of time

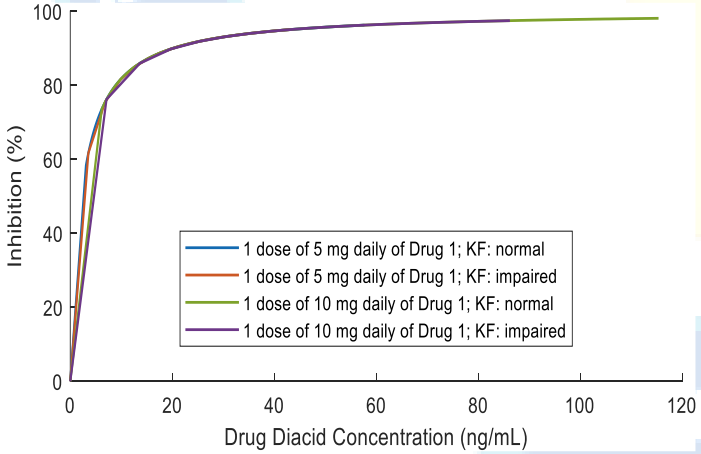

Fig 6: Validation results for doses of benazepril for inhibition as a function of drug diacid concentration

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.