A review of using data mining and machine learning for predicting drug loading modeling in solid lipid nanoparticles containing curcumin

R Widyati¹*, A Ashari², Afiahayati²

¹ Computer Science Department Universitas Negeri Jakarta Indonesia,
² Computer Science and Electronica Department Universitas Gadjah Mada Indonesia

*Corresponding author: ratna.widyati@unj.ac.id

Abstract. This article provides a comprehensive review of the use of data mining and machine learning to develop predictive models for drug loading in nanoparticles. Solid lipid nanoparticle technology is a promising new approach to lipophile drug delivery. Solid lipid nanoparticles (SLNs) are an important advance in this area. The bio-acceptable and biodegradable properties of SLN make it less toxic than polymer nanoparticles. This review article contains a series that applies computer-oriented processes and tools to extract information, analyze data and finally extract the correlation and meaning of the results obtained regarding solid lipid nanoparticles especially those containing curcumin. The purpose of this review is to describe the development of several research results that have been published over a period that is useful for new insights on drug loading modeling.

1. Introduction

One of the situations in the treatment of disease is the administration of the right concentration of an efficacious drug to the site of action in a controlled and continuous manner. Nanoparticles are an important particulate carrier system. Nanoparticles are solid colloidal particles ranging in size from 1 to 1000 nm and are composed of macromolecular material [1]. Nanoparticles can be either polymeric or lipidic (SLNs). Industry estimates suggest that approximately 40% of lipophilic drug candidates fail due to issues of solubility and formulation stability, leading to a significant failure of research activity in advanced lipophile delivery technologies. Solid lipid nanoparticle technology is a promising new approach to lipophile drug delivery. Solid Lipid Nanoparticles (SLN) are an important advance in this area. The bio-acceptable and biodegradable properties of SLN make it less toxic than polymer nanoparticles. Their small size extends the blood circulation time, is suitable for large-scale production and the absence of burst effects makes them attractive candidates to study.

Solid lipid nanoparticles, introduced in 1991, are colloid carriers for regulating drug delivery systems through the development of emulsions, liposomes, microparticle polymers, and nanoparticles, with an average particle size of between 50-1000 nm [2]. Solid lipid nanoparticles combine the advantages and disadvantages of other colloid carriers. Its advantages include allowing controlled drug release and drug targeting, high oral bioavailability, increased drug stability, enabling the combination of lipophilic and
hydrophilic drugs, absence of toxicity from carriers, avoiding the use of organic solvents, and easy large-scale production.

However, solid lipid nanoparticles also have disadvantages such as they can cause drug degradation if they are made using high pressure and a gelation phenomenon can occur which describes the dispersed viscosity of solid lipid nanoparticles from low viscosity to gel-like viscosity. Solid lipid nanoparticles are prepared by homogenizing the liquid secretions of lipids and emulsifiers. The lipids used herein are broadly used and all classes of emulsifiers have been used to stabilize the lipid disperse. Solid lipid nanoparticles are one of the new potential colloid carriers in the 100-150nm range system as an alternative material for polymers that are identical to oil-in-water emulsions for parenteral nutrition, but emulsion liquid lipids have been replaced by solid lipids. They have many advantages such as good biocompatibility, low toxicity, and better lipophilic drug delivered in solid form of lipid nanoparticles and the system is physically stable. Solid lipid nanoparticles promise a sustained release and drug targeting system for lipophilic CNS antitumor drugs.

Curcumin is the active ingredient obtained from turmeric. Turmeric has been reported to have many biological activities such as anti-cancer, anti-inflammatory, anti-microbial, and antioxidant properties. However, poor oral absorption due to low water solubility (11 ng / ml) and slow dissolution results in low systemic oral bioavailability resulting in limited clinical use. [3] Turmeric (Curcuma longa L.) is a unique plant that combines the properties of spices, dyes, cosmetic and medicinal properties that are useful for some diseases. It has been used as a spice, herbal medicine, coloring agent, and cosmetics since Vedic times. The health and nutritional significance of turmeric have been widely recognized since the discovery of the pharmaceutical properties of the naturally occurring phenolic compounds in it. It has been found that dried turmeric rhizome is a rich source of beneficial phenolic compounds known as curcuminoids. Curcuminoids are the most important component of turmeric, which refers to a phenolic group of compounds, chemically related to its main ingredient, curcumin. Extensive investigations over the last five decades have shown that curcumin reduces blood cholesterol, prevents LDL oxidation, inhibits platelet aggregation, suppresses thrombosis and myocardial infarction (MI), suppresses the symptoms associated with type II diabetes, rheumatoid arthritis, multiple sclerosis, and Alzheimer's disease.

![Figure 1. Structure of curcumin](#)

In line with these findings, curcumin is one of the natural compounds that has been extensively investigated for its antiviral effects [5]. Curcumin, a natural polyphenol compound extracted from the root of the Curcuma longa (Zingiberaceae family) rhizome plant, exhibits a wide variety of therapeutic properties including antioxidant, anti-microbial, anti-proliferative, anti-inflammatory, neuroprotective, and cardioprotective properties. Curcumin, the yellow pigment of turmeric is widely used in traditional Indian herbal medicine to cure many ailments associated with infection and inflammation for decades [8]. It has been reported that curcumin works against a broad spectrum of viruses including HIV, HSV-2, HPV virus, influenza virus, Zika virus, hepatitis virus, and adenovirus [6,7]. Recent studies have shown that the same SARS CoV, SARS-COV2 also attacks human host cells targeting the Angiotensin-Converting Enzyme 2 (ACE2) membrane receptor, where the coronavirus enters.

This review article consists of three parts, namely an introduction that contains the advantages and disadvantages of Solid Lipid Nanoparticles, the methodology section describes the data mining and machine learning techniques used in the reading source, and the final part of the conclusions from the reading sources that have been described.
2. Methodology
The research conducted is one of the efforts to utilize machine learning and artificial intelligence tools in the modeling and optimization of important nanocarriers such as solid lipid nanoparticles. Using the Gaussian Process, a supervised machine learning tool rarely used by pharmaceutical scientists but very accurate and sensitive, less percentage bias is obtained between the actual and predicted values of the docking binding energy compared to previously adopted methods such as ANN[9]. Solid lipid nanoparticles (SLN) are a nanoparticulate drug delivery system that is considered very attractive as a drug carrier especially for lipophilic drugs [10]. SLNs can protect these drugs and control their release. Moreover, they are used as an innovative colloid drug carrier for topical application, mainly because of their interaction with the stratum corneum (SC) and other skin layers [11]. Thus, modeling drug loading in this important nanoparticulate matrix is guaranteed to save the effort and time researchers and formulators spend in wet laboratory experiments and provide them with preliminary and accurate estimates of the fate of the drugs they are investigating in selected carriers.

Previously, it has been demonstrated that integrating multiple computational and chemoinformatic tools such as: data mining, molecular dynamics simulation, and experimental docking together with mathematical modeling and/or machine learning methods enables successful predictions of drug carrier interactions and consequently the good selection of drug carrier pairs [12, 13]. Machine learning methods cover two main categories; supervised and unsupervised [14,15], where supervised methods outperformed their counterparts in their ability to relate input variables to responses (outcome). Unsupervised methods such as Main Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) are mainly used to extract patterns or classify or explore certain features in data when supervised machine learning tools include Artificial Neural Networks (ANN), Partial Least Squares and the Gaussian Process (GP) is very efficient in developing correlations and estimating predictions [14]. The current study extends the hypothesis about successful virtual simulations and modeling of drug loading in solid lipid nanoparticles to other biochemical tools such as the Gaussian Process and molecular linking mechanisms as well as the assessment of different functions using the Molecular Operating Environment (MOE).

3. Results and Discussion
Hypotheses using several chemo/bioinformatics tools and statistical, computational methods can study and then predict the behavior of several drugs in nanoparticulate lipid models and polymer systems. Thus, two different matrices consisting of tripalmitin, a core component of solid lipid nanoparticles (SLN), and PLGA were first modeled using molecular dynamics simulations, and then drug interactions with this system were studied by calculating the binding free energy using molecular docking techniques. Therefore, this binding energy correlates with a load of this drug in the nanoparticles obtained experimentally from the available literature. The relationships obtained were verified experimentally in the laboratory using curcumin as a model drug. The artificial neural network is then used to determine the effect of the drug, molecular descriptors of binding energy and hence on drug loading. The results showed that the soft computing method used can provide the benefit of an accurate method for in silico prediction of drug load in tripalmitin and PLGA-based nanoparticulate systems. These results have prospects for application to other nanomedicine carrier systems, and this integrated statistical and chemo / bioinformatics approach offers a new toolbox for formulation science by proposing what is presented as a computer-assisted drug formulation design (CADFD) [11]. Figure 2. Venn diagram comparing the keywords used by curcumin-focused publications and those used by publications mentioning Curcuma longa or turmeric but not curcumin. Keywords from the former were more clinically relevant, whereas those from the latter were more related to plant science studies; commonly used keywords are presented in the middle of the diagram [23].
Table 1. The solid lipid nanoparticles of curcumin and its implications in various therapeutic applications

| References | Size (nm) ± SD | Applications against disease | Mechanism of actions | Advantage |
|------------|---------------|------------------------------|----------------------|-----------|
| [4]        | 194 ± 2.89    | Breast cancer                | Induces apoptosis in cancerous cells | Enhanced drug uptake due to improved drug delivery and retention Suitable for systemic administration |
| [16]       | 153.3 ± 0.2   | Pathology of the central nervous system Breast cancer | Involves a deficit in antioxidant defences | |
| [17]       | 30nm          | Breast cancer                | Regulate cell cycle, apoptosis and survival, proliferation, invasion, and metastasis. Brain uptake by the paracellular pathway through the opening of the tight junctions in brain microvasculature, passive diffusion, and endocytosis. Active targeting with receptors (apolipoprotein E) [204,206,210] | Increased cellular uptake when cells were treated with Cur-SLNs. High entrapment efficiency for hydrophobic drugs, biocompatibility, high physical stability and drug protection, controlled release, ease of formation methods (that can be easily scaled up and do not require organic solvents thus avoiding (neuro toxicity) |
| [18]       | 50-300nm      | Brain disease and Alzheimer’s disease | | |
| [19]       | 50–1000nm     | Cerebral ischemia, Colitis, Allergy, Breast cancer | Prolonged circulation of blood, Increased anti-inflammatory effects, Improved brain delivery | |
| [20]       | 166.5 ± 5.4   | Enhance the anti-tumour effect | High pressure homogenization | |
| [21]       | 190.4 ±10.6   | Asthma                       | Enhanced bioavailability Increased tissue concentrations Suppressed airway hyperresponsiveness and inflammatory cell infiltration | Increases the efficiency of curcumin uptake by the lungs. |

Data mining uses scientific literature databases using PubMed, Scopus, and Web of Science. The Gaussian process in predicting binding energy is carried out to calculate the important descriptors of the investigated drug, modeling using Gaussian Processes (GPs) and validating the predictive strength of the developed GP [9]. Data mining uses scientific database literature databases namely PubMed, Scopus, and Web of Science. Molecular Descriptors and Artificial Neural Networks in predicting energy binding to compute important explanations of the investigated drug, modeling using Artificial Neural networks (ANNs), validating the predictive power of developing ANNs [11].

4. Conclusion
This review is one of the efforts aimed at harnessing machine learning and artificial intelligence tools in the modeling and optimization of important nanocarriers such as solid lipid nanoparticles. Using the Gaussian Process, a supervised machine learning tool rarely used by pharmaceutical scientists but very accurate and sensitive, less percentage bias is obtained between the actual and predicted values of the docking binding energy compared to previously adopted methods such as ANN.

By integrating the use of several chemo/bioinformatics and statistical tools, the efficiency of loading drugs in a particular carrier can be predicted computationally. So it can be concluded that the relationship between the docking binding energy (ΔG) and both trapping percentage efficiency (% EE) and the
combined mass for different drugs is loaded on SLN and PLGA based on tripalmitine nanoparticles using molecular dynamics techniques and molecular docking. Such a cause-and-effect relationship would be enormously useful in the science of drug delivery and would save researchers a lot of experimental time in the optimal selection of the drug/carrier partner. The artificial neural network (ANN) can also capture strong relationships between specific drug molecular descriptors and their energy-anchored bonds in carrier systems with high predictive ability.

The review article above shows that data mining can help determine formulation parameters for further research. The computerized formulation design approach will gain more basis because of the versatility of the computational tools used and the type of information obtained, which is predictive or explanatory [22]. Thus, the formulation design can be achieved.

Acknowledgment
Thank you to the Saudi Fund for Development (SFD) through the Universitas Negeri Jakarta for providing a doctoral scholarship to the author.

References
[1] Rajpoot K 2019 *Curr Pharm* 25(37) 3943
[2] Öztürk A A, Gündüz A B, Ozisik O 2018 *Comb Chem High Throughput Screen* 21 (9)693
[3] Akbar A., Kuanar A, Joshi R. K, Sandeep IS 2016 *Front. Plant Sci* 7 1507
[4] Shome S, Talukdar AD, Choudhury MD, Bhattacharya MK, Upadhyaya H 2016 *J Pharm Pharmacol* 68(12) 1481
[5] Manoharan Y, Haridas V, Vasanthakumar K C, Muthu S, Thavoorullah FF, Shetty P 2020 *J Clin Biochem* 35 373.
[6] Das S, Sarmah S, Lyndem S, Roy AS 2020 *J Biomol Struct Dyn.* 13 1
[7] Praditya D, Kirchhoff L, Bru¨ning J, Rachmawati H, Steinmann J, Steinmann E 2019 *Front Microbiol* 10:912
[8] Pang XF, Zhang LH, Bai F, Wang NP, Garner RE, McKallip RJ, Zhao Q 2015 *Drug Des Devel Ther.* 9 6043
[9] Hathout RM and Metwally AA 2016 *Eur. J. Pharm. Biopharm.* 108 262
[10] Rigon RB, Fachinetti N, Severino P, Santana MH and Chorilli M 2016 *Molecules* 21(1) 116
[11] Metwally AA, Hathout RM 2015 *Mol. Pharm* 12 2800
[12] Metwally AA, Hathout RM 2015 *Chem. Eng. Res. Des.* 104 453
[13] Gad HA, El-Ahmady SH, bou-Shoer MI, Al-Azizi MM 2013 *Phytochem. Anal.* 24 1
[14] Hathout RM 2014 *Pharm Dev. Technol.* 19 598
[15] Fagir W, Hathout RM, Sammour OA, ElShafeey AH 2015 *Nanomedicine (Lond)* 10 3373
[16] Santonocito D, Sarpietro MG, Carbone C, Panico A, Campisi A, Siciliano G, Castelli F, Pugli Ca 2020 *Molecules.* 25(13) 2991
[17] Wang W, Chen T, Xu H, Ren B, Cheng X,Qi R, Liu H, Wang Y, Yan L, Chen S, Yang Q, Chen C, 2018 *Molecules* 23(7) 1578
[18] Salehi B, Calina D, Docea AO, Koirala N, Aryal S, Lombardo D, Pasqua L, Taheri Y, Castillo C MS, Martorell M, Martins N, Iriti M, Suleria HAR, Rad JS 2020 *J Clin Med* 9(2):430
[19] Ghalandarlaki N, Alizadeh AM, Esfahani SA, 2014*Biomed Res Int.* 2014 394
[20] Chen Y, Lu Y, Lee RJ, Xiang G 2020 *Int J Nanomedicine* 15 3099
[21] Wang W, Zhu R, Xie Q, Li A, Xiao Y, Li K, Liu H, Cui D, Chen Y, Wang S 2012 *Int J Nanomedicine* 7 3667.
[22] Hathout RM and Metwally AA 2018 *Open J. Chem.* 4(1): 001
[23] Yeung AWK,Horbańczuk M, Tzvetkov NT, Mocan A, Carradori S, Maggi F, Marchewka J, Sut S, Acqua S D, Gan RY, Tancheva LP, Polgar T, Neagoe IB, Pirgozliev V, Šmejkal K, Atanasov AG 2019 *Molecules* 24(7)1393