

Tytuł: Projektowanie mikroemulsyjnych postaci leku w oparciu o techniki eksploracji danych

Autor: Jakub Szlęk

Promotor: Prof. dr hab. Renata Jachowicz

Promotor pomocniczy: Dr hab. Aleksander Mendyk

Streszczenie

Założenia badawcze pracy dotyczyły oceny możliwości zastosowania technik eksploracji danych w formułowaniu mikroemulsyjnych postaci leku. Obejmowały trzy główne etapy: eksplorację danych i utworzenie systemu wspomagania decyzji, napisanie kodu aplikacji komputerowej oraz walidację systemu w praktyce.

Badania zrealizowano z wykorzystaniem metod obliczeniowych i laboratoryjnych. W ramach części obliczeniowej utworzono bazę danych zawierającą liczbową reprezentację składu ilościowego i jakościowego mikroemulsji, przeprowadzono modelowanie heurystyczne, opracowano system wspomagania decyzji, a także napisano kod programu ME_DSS do jego obsługi. Natomiast w oparciu o analizę właściwości fizykochemicznych i badania stabilności sporządzonych mikroemulsji, w ramach części laboratoryjnej, oceniono efektywność predykcji ME_DSS. W założeniach uwzględniono opracowanie składu mikroemulsji do podania na skórę zawierającej ibuprofen jako modelową substancję leczniczą trudno rozpuszczalną w wodzie.

Na podstawie piśmiennictwa utworzono zbiór danych zawierający liczbową reprezentację składu ilościowego i jakościowego mikroemulsji, który zawierał około 305 000 rekordów. Każdy z nich zawierał informacje dotyczące procentowej zawartości składników mikroemulsji oraz ich parametry fizykochemiczne w tym liczbę HLB, gęstość fazy olejowej, siłę jonową fazy wodnej, oraz ponad 400 deskryptorów molekularnych. Na podstawie analizy wrażliwości zredukowano ilość zmiennych niezależnych do siedemnastu. Tak przetworzona baza danych stanowiła podstawę do użycia technik eksploracji danych, tj. sztucznych sieci neuronowych oraz algorytmów klasyfikacyjnych. Biorąc pod uwagę rodzaje technik eksploracji danych oraz zastosowaną metodykę łącznie wytrenowano około 60 tysięcy modeli, z których wyselekcjonowano optymalne do przewidywania obszaru występowania mikroemulsji. Do systemu wspomagania decyzji włączono 5 modeli sztucznych sieci neuronowych oraz model RandomForest. Napisano kod źródłowy aplikacji komputerowej ME_DSS, który umieszczono na serwerze sourceforege.net, jest ona dostępna na warunkach licencji Wolnego Oprogramowania.

Zastosowanie ME_DSS pozwoliło na predykcję obszarów występowania ponad 1000 układów fazowych. Na podstawie wielowartościowego rankingu wybrano 20 układów fazowych i w oparciu o porównanie predykcji ME_DSS z obszarem występowania mikroemulsji wyznaczonym eksperymentalnie dokonano oceny efektywności programu. Do dalszych badań wybrano 20 formułacji mikroemulsji i oceniono ich właściwości fizykochemicznych oraz stabilność. Podobną ocenę dokonano dla dwóch formułacji z modelową substancją leczniczą.

Summary

Aim of the dissertation was to evaluate the application of data mining techniques in the process of microemulsion formulation. Three main stages of the work were established: data mining and the development of the decision support system, writing a code of the computer application and validation of the system in practice.

Both computational and technological methods were used to carry out the research. As a part of computational procedure following objectives have been done: the database was created which contained a numeric representation of qualitative and quantitative composition of the microemulsion, the heuristics modeling was conducted, the decision support system was developed, as well as the ME_DSS program code was written. To assess the ME_DSS prediction efficiency analysis of physicochemical properties and stability studies were made as a part of technological process. The microemulsion formulation composition was for dermal administration of ibuprofen as a model sparingly soluble in water drug active pharmaceutical ingredient.

Based on the literature survey a database was created, which contained about 305 000 records, a numeric representation of the qualitative and quantitative composition of the microemulsion formulations. Each data record contained information about the percentage of the components of microemulsions and their physical and chemical parameters e.g. the HLB value, density of the oil phase, ionic strength of the aqueous phase, and more than 400 molecular descriptors. Based on the results of sensitivity analysis number of independent variables was reduced to seventeen. The database was further utilized during data mining process in which techniques as artificial neural networks and classification algorithms were used. Jointly about 60 000 models were trained and tested in order to select the optimal ones to predict the occurrence of microemulsion. The decision support system was combining the results of five artificial neural networks and the RandomForest model. Computer application integrating the support system and its source code was uploaded on the sourceforge.net server under the Open Source license.

ME_DSS application was used to predict microemulsion occurrence in over 1000 phase triangles. Based on multivariate ranking 20 phase triangles were chosen to assess in laboratory the efficiency of the ME_DSS. Twenty formulations were chosen to assess their physicochemical properties and stability. Similar assays were conducted for two microemulsions with drug substance