

# **Р е ц е н з и я**

от

проф. д-р **Стефан Цаковски**

на дисертационен труд с автор **Людмила Христова Нанева**

на тема:

## **ХЕМОМЕТРИЧЕН ПОДХОД ЗА КЛАСИФИЦИРАНЕ НА ХРАНИТЕЛНИ ПРОТЕИНИ В КАТЕГОРИИ „АЛЕРГЕНИ“ И „НЕАЛЕРГЕНИ“**

За присъждане на образователната и научна степен „**Доктор**”  
по научна специалност „**Аналитична химия**”,  
професионално направление 4.2. „**Химични науки**”  
научна област „**Природни науки, математика и информатика**”

### **Общо представяне на материалите и биографични данни**

Представеният ми за рецензиране дисертационен труд **включва целия изискуем комплект материали** съгласно Правилника за условията и реда за придобиване на научни степени и заемане на академични длъжности в СУ “Св. Климент Охридски”. Людмила Христова Нанева е зачислена като докторант на самостоятелна подготовка към катедра “Аналитична химия”, Факултет по химия и фармация на 15.11.2017 г. и е отчислена на 09.11.2018 г. Прегледът на документите показва, че са изпълнени всички задължения по учебния план и **докторантът отговаря на административните изисквания** за присъждане на образователната и научна степен "Доктор".

### **Актуалност на тематиката**

Темата на дисертационното изследване е **изключително актуална**, тъй като решава проблеми в една важна и комплексна област каквато е имунологията – разработва хемометрични модели за разпознаване на алергенни протеини от хранителен произход. Този тип изследвания са **силно интердисциплинарни** като съчетават области на имунологията с биоинформатиката и имат сериозен потенциал за приложение, имайки предвид какъв бич са хранителните алергии за съвременното общество и в частност за подрастващите.

## Структура и изложение на дисертационния труд

Дисертационният труд е стандартно структуриран и съдържа **въведение, литературен обзор, цел и задачи на дисертационния труд, материали и методи, резултати, обсъждане, приноси и списък на цитираната литература, както и две приложения.**

В **литературния обзор** са разгледани механизма на алергията и някои от основните *in silico* подходи за разпознаване на алергени, хеометрични и биоинформатични подходи, които се използват при работата с биомакромолекули. Като основни многовариационни подходи са застъпени: множественият линеен регресионен анализ, анализ на главни компоненти, метод на частично най-малките квадрати, кластерен анализ.

Логично следва **целта на изследването**, а именно създаване на хеометричен модел за разпознаване на алергенни протеини с хранителен произход, базиращ се на свойствата на аминокиселинните остатъци, изграждащи протеините. **Задачите** са детайлно и коректно формулирани.

В главите **резултати и обсъждане** са представени два модела за разпознаване на алергени: базирани на PLS-DA (метод на частично най-малките квадрати с дискриминационен анализ). Моделите са разгледани детайлно с оглед броя на компонентите и важността на променливите, както и са надлежно валидирани чрез вътрешно кръстосано валидиране по групи и валидиране върху външна тестова група. Представен е изводеният на базата на първия модел мотив на алергенност и е анализирано присъствието/отсъствието му в алергени/неалергени. Допълнително е представен подход за разделяне на алергенни от неалергенни протеини чрез кластерен анализ.

### Анализ на приносите

В резултат на проведените изследвания са дефинирани приноси, които **отговарят на поставената цел и коректно обобщават получените резултати**. Те се свеждат до:

1. Извеждане на хеометрични модели за класифициране на протеини от хранителен произход като алергени/неалергени.
2. Дефиниране на мотив на алергенност, който при успешно валидиране може да бъде прилаган при разпознаване на алергенни хранителни протеини.

По отношение на дисертационния труд имам следните **коментари, въпроси и препоръки**:

1. Във въведението, при обосновка на актуалността на тематиката, са цитирани източници от 2003 г. Действително големият интерес към полето е продуцирал множество публикации и могат да се намерят далеч по-съвременни и актуални такива. При това източник 25 представя принципи за анализ на риска от храни, получени в резултат на съвременните биотехнологии и трудно може да се отнесе към приложението на биоинформатичните подходи.
2. Литературен обзор: (i) В т. 2.4. се въвежда понятието биоинформатика. Логично е да се въведе и онзи неин клон, който прилага изчислителни подходи за решаване на проблеми в имунологията, а именно имуноинформатика (изчислителна имунология). Още повече, това е полето на дисертационния труд; (ii) т. 2.5. ненужно се разпростира върху приложения на хемометрията в естествените науки и медицината, които са изключително многобройни и не конкретно по темата на дисертацията. Вместо това, тук е мястото да бъдат представени съществуващите модели в имуноинформатиката и в частност тези, за предсказване на алергенност/неалергенност. Опит за подобно кратко описание е направен в т. 2.2., далеч обаче преди да се въведат изчислителните подходи. Това, както и липсата на изводи от литературния обзор възпрепятства идентифицирането на постиженията/нишите в научната област и респективно, мястото на дисертационното изследване в нея.
3. Авторът представя два модела, базирани респективно на две групи данни (Приложение 1 и 2). Посочено е че „Целта на втората група данни е да се създаде подобна матрица, с цел прилагане на методологията и проверка на модела за неговата валидност“. За постигането на тази цел е необходимо структурните пространства на двете бази данни да имат минимално припокриване. В тази връзка моля докторантът да коментира какъв е процентът на припокриване между двете бази данни.
4. На какво се базира изборът на хидрофобността като свойство за разделянето на алергени от неалергени чрез кластерен анализ? Как са получени 98те дескриптора във входната матрица от обучаваща серия от 38 протеина при кластерния анализ?

5. Бих препоръчал на докторанта по-голяма прецизност при оформянето на дисертационния труд. Част от забелязаните неточности включва: (i) стр. 8 – използвано е понятието „алергичност“ вместо алергенност; (ii) стр. 20 - *in silico* експерименти са всички компютърно-подпомогнати модели, в този смисъл понятието включва голяма част от предходните изброени задачи на биоинформатиката; (iii) на стр. 34 е посочено, че валидирането в групи е илюстрирано на фиг.6. Цитираната фигура, обаче илюстрира LOO (leave one out) валидиране (iv) включените изображения в таблица 9 са неинформативни, обемни и затрудняват възприемането ѝ; (v) в списъка с публикации е сгрешена поредността на авторите в публикация 1: J. Chemometrics 2014; 28:282-286, DOI: 10.1002/cem.2597; (v) в заглавието на приложение 1 е записано „дипломна работа“ вместо „дисертация“ и т.н.

### **Наукометрични показатели**

Изследвания с участието на докторанта са публикувани в **две статии** в научни списания с импакт фактор, едното от които (Journal of Chemometrics) е добре познато и утвърдено в областта на дисертационния труд. Забелязани са и **пет цитата** на публикациите, които са добър атестат за тяхното приемане в научната общност. **Множеството участия в научни форуми и богата проектна дейност** без съмнение са допринесли значително за професионалното развитие на кандидата.

### **Заклучение**

Имайки предвид изискванията на ЗРАСРБ и Правилника за неговото прилагане в СУ „Св. Кл. Охридски“, както и обемът на представените изследвания и използваните подходи от докторанта и неговите наукометрични показатели, давам **положителна оценка** и препоръчвам на почитаемото Научно жури да присъди на **Людмила Христова Нанева** образователната и научна степен **“Доктор”**.

14.01.2019 г

Рецензент:

(проф. Ст. Цаковски)