

## РЕЦЕНЗИЯ

на дисертационния труд за присъждане на образователната и научна степен „Доктор“

по докторска програма: “Автоматизирани системи за обработка на информация и управление”

**Автор на дисертационния труд:** Мая Петрова Тодорова

**Научен ръководител:** доц. д-р инж. Недялко Николов

**Тема на дисертационния труд:** „Изследване на методи на машинно обучение за анализ на онкологични заболявания“

**Рецензент:** доц. д-р инж. Милко Тодоров Маринов

Катедра „Компютърни системи и технологии“, Русенски университет

### 1. Актуалност на разработения в дисертационния труд проблем

В днешно време машинното обучение (което е част от цялата наука за изкуствения интелект) играе важна роля в много области на нашия живот – включително здравеопазването. Това може да включва различни, свързани със здравето аспекти като разработването на нови медицински процедури, справяне с данните за пациентите, прогнозиране на резултатите от лечението на заболявания и за идентифициране на вероятността за съществуването на някакво заболяване. В медицинската онкология методите на машинното обучение основно се прилагат в процеса на идентификация и диагностика на заболявания, като подпомагат медицинския екип при вземане на решения относно подобряване на начина на живот на пациента, предписание на терапия и медикаменти и ранна оценка на критичните състояния.

Разработеният в дисертационното изследване проблем е актуален, защото е посветен на автоматизиране процеса на обработка и анализ на голям обем от данни за пациенти с онкологични заболявания. Прилагането на предложения подход е особено полезен с увеличаване обема на данните, за да се генерират модели за предсказване на стадия на онкологично заболяване и класификация на онкологични лечебни заведения на база преживяемост на пациент и топографски код на заболяване. Това е модерно, актуално и нужно направление за научни изследвания.

## **2. Литературен преглед по дисертационния труд**

Списъкът на литературата съдържа 146 източници, от които 49 литературни източници на кирилица и 97 - на латиница. Анализът на литературните източници показва, че 21% от тях са издадени преди 2010 година, 24% - в периода 2010-2014, 30% - в периода 2015-2018, 25% - след 2019 година. Прави впечатление, че докторантката е използвала книги и учебни материали, които представляват класически литературни източници в предметната област на изследването.

Цитираната библиография е показател за равнището на труда, съответстващо на претендиращата степен. Рецензентът счита, че докторантката е запозната много добре с проблемите, чието решение търси в дисертационното изследване. Това ѝ е позволило да формулира правилно целта и задачите на изследването, отчитайки състоянието и актуалността на проблема.

## **3. Избрана методика на изследване**

Докторантката се е придържала към една класическа методика на дисертационно изследване. След направената класификация на видовете задачи при извличане на знания от големи обеми от данни, докторантката е направила анализ на над 30 научни публикации, повечето от които са обзори на съществуващи решения за прилагане методите на машинно обучение при решаване на класификационни и регресионни задачи в онкологията.

Въз основа на извършените теоретични изследвания докторантката реализира експериментални изследвания върху модели за предсказване стадия на онкологично заболяване чрез методите на машинното обучение и класификация на онкологични лечебни заведения. В рамките на предложената технология са проведени задълбочени експериментални изследвания и е извършена оценка на разработените подходи.

Реализираната от докторантката методика на изследване, предложените алгоритми, които в основата си са правилно обосновани и формулирани, позволяват да се решат успешно задачите на изследването и да се постигне поставената цел – да се изследва и оцени приложимостта и ефективността на методи на машинното обучение върху голям обем от данни за пациенти с онкологични заболявания. Тази цел е постигната чрез дефинирането на 4 основни задачи, които последователно са решени във втора, трета и четвърта глави.

#### **4. Характеристика на дисертационния труд**

Дисертационният труд е в обем от 144 печатни страници. Съдържа увод, четири глави, заключение, използвана литература и списък на публикациите, свързани с дисертацията, речник на използваните съкращения, речник на използваните термини, списък на фигурите и таблиците и едно приложение.

Структурата и елементите на дисертацията са правилно обосновани, а взаимовръзката им е логична и позволява да се получат достоверни резултати на изследваните параметри.

В дисертационното изследване докторантката използва съществуващ формален математически апарат, като показва добри умения при неговото цитиране, правилно прилагане за решаване на дефинираните в изследването задачи.

В първа глава е направен обзор на методите на машинно обучение и data mining за анализ на данни. Анализирани за над 30 научни публикации (предимно голяма част от тях обзори), в които се прилагат методите на машинно обучение в изследваната предметна област. В края на първа глава е дефинирана основната цел и задачите на дисертационното изследване.

Във втора глава са описани методите и алгоритмите на машинно обучение, използвани при провеждане на експерименталните изследвания. Дефинирани са метрики за оценка на качество и сравнение на класификационни и регресионни модели. В следващите глави тези метрики се използват за оценка на създадените от докторантката модели.

В глава трета е представено изследване на модели за предсказване стадия на онкологично заболяване. Докторантката е описала подробно и точно всички фази на проведените експериментални изследвания.

В четвърта глава са извършени експериментални изследвания с цел създаване на методика за класифициране на онкологични лечебни заведения.

Към всяка глава са направени изводи. В заключението на дисертационния труд са посочени значимите според докторантката изследвания и приноси на дисертационното изследване.

Дисертационният труд е структуриран правилно. Като цяло описанието е коректно и задълбочено.

## **5. Приноси на дисертационния труд**

Рецензентът счита, че основните приноси на дисертационния труд са с научно-приложен и приложен характер и могат да бъдат обобщени в следните групи:

- Чрез методите дърво на решения, опорни вектори и ансамблови алгоритми са създадени регресионни модели за предсказване стадия на онкологични заболявания за три независими съвкупности от данни;
- Създадените модели са оценени и сравнени в процес на обучение и процес на тестване;
- Създадена е методика за класифициране на онкологични лечебни заведения, базирайки се на проведени експериментални изследвания;
- Точността на резултатите, получени след проведените експериментални изследвания, са оценени чрез съответни метрики;
- Изследвано е влиянието на обема на данните върху точност, прецизност, пълнота и F-мярка на класификаторите.

## **6. Преценка на публикациите по дисертационния труд**

По темата на дисертацията са представени 6 научни публикации, които са отпечатани в пълен текст – 2 публикации са представени на международни конференции, индексирани в Scopus, 2 доклада са представени на международни конференции, чиито сборници са включени в Референтния списък на НАЦИД с научно рецензиране, 2 доклада са представени на международни конференции. 2 от публикациите са на английски език. 2 от публикациите са самостоятелни, а в 3 от останалите публикации докторантката е първи автор, което подчертава нейната водеща роля.

При така направените публикации докторантката покрива наукометричните критерии за придобиване на образователна и научна степен „Доктор“ в област „Технически науки“. Рецензентът счита, че публикационната дейност на докторантката е достатъчна и направените публикации съдържат основните приноси, за които се претендира в дисертацията.

## **7. Значимост на научно-приложните и приложните резултати от дисертационния труд**

Докторантката Мая Тодорова демонстрира самостоятелност и последователност при провеждане на изследванията си. Дисертационното изследване е базирано върху значителни по обем експериментални изследвания, резултатите от които могат да бъдат полезни за медицински специалисти при определяне стадия на онкологично заболяване и за пациенти при избор на лечебно заведение.

Рецензентът счита, че научно-приложните и приложни резултати от дисертационния труд са полезни, приложени са върху реални данни и са били представяни пред академичната аудитория у нас.

## **8. Автореферат**

Авторефератът правилно отразява съдържанието на дисертационния труд и приносите на дисертационното изследване.

## **9. Препоръки и критични бележки**

Не става ясно защо във втора глава са подбрани за анализ точно тези 4 метода на машинно обучение – kNN, SVM, NBC, DT.

В трета глава се използват 3 метода на машинно обучение – DT, опорни вектори и ансамблов метод. Защо ансамбловият метод не е включен в анализа във втора глава.

Тъй като става въпрос за анализ на голям обем данни, нека да бъде изяснено с по-голяма точност – какъв обем данни (в MB, GB, TB) са обработвани.

Добре би било някои от таблиците в експерименталните изследвания да се отделят като приложения. Това би осигурило по-стегнато изложение на материала.

Глава 3, стр. 55 – добре било да се покаже размерността на някои от атрибутите, за които тя не е ясна – хистология, лимфни възли, метастази.

На места има неправилно използване на понятието „информация“. Там се има предвид „данни“.

На стр. 7 – „Обемът на данните е около 1 000 000 ...“. Това 1000000 какво е?

Формулите в дисертацията са номерирани, но няма обръщение към индексите в самия текст.

Добре би било докторантката да помисли в близко бъдеще да модифицира някои от алгоритмите, за да подобри тяхната приложимост към изследваната предметна област. Голяма част от използваните алгоритми имат реализация на Python и са с отворен код. По този начин би могло да се създаде една уникална инструментална среда с конкретно приложение в областта на онкологията.

Посочените препоръки и критични бележки не намаляват достиженията на дисертационния труд. Рецензентът счита, че те би трябвало да се отчетат при бъдещата преподавателска и научна работа на докторантката.

Не познавам лично Мая Тодорова, но в качеството си на рецензент на дисертационния труд оценявам, че тя показва много добър потенциал за развитие, демонстрира умения за провеждане на самостоятелни изследвания в областта на приложение на методите на машинно обучение.

## **ЗАКЛЮЧЕНИЕ**

Оценявам положително научно-приложните и приложните приноси на дисертационния труд на инж. Мая Тодорова. Считаю, че изискванията на Закона за развитие на академичния състав и Правилника за неговото прилагане са изпълнени в представения дисертационен труд.

**Гореизложеното ми дава основание да предложа положителна комплексна оценка за представения дисертационен труд и да препоръчам на Научното жури да присъди на инж. Мая Петрова Тодорова образователната и научна степен „Доктор“ по докторска програма „Автоматизирани системи за обработка на информация и управление“ в професионално направление 5.3. Комуникационна и компютърна техника.**

20.09.2021 г.  
гр. Русе

Рецензент: .....

/доц. д-р Милко Маринов/