

РЕЦЕНЗИЯ

на дисертационен труд на тема

Изкуствен интелект в радиогеномиката: Иновации в подкрепа на диагностика, терапевтични решения и прогностика в онкологията

за придобиване на образователна и научна степен „доктор“,
в област на висше образование: 4. Природни науки, математика и информатика по
професионално направление: 4.6. Информатика и компютърни науки, докторска
програма „Компютърни науки - Машинно самообучение“

с автор Мария Димитрова Митева

Научен ръководител: проф. д-р Мария Михайлова Нишева-Павлова, Факултет по математика и информатика, Софийски университет „Св. Климент Охридски“

Рецензент: проф. д-р Евдокия Николаева Сотирова, Факултет по природни науки, Бургаски държавен университет „Проф. д-р Асен Златаров“

Със заповед № РД 38-233/07.05.2025 г. на Ректора на Софийски университет съм назначена за член на научно жури по процедурата за защита на дисертационен труд на тема „Изкуствен интелект в радиогеномиката: Иновации в подкрепа на диагностика, терапевтични решения и прогностика в онкологията“ по професионално направление (ПН) 4.6. Информатика и компютърни науки, докторска програма „Компютърни науки - Машинно самообучение“ с автор Мария Димитрова Митева. На първото заседание на научното жури, проведено на 13.05.2025 г. съм определена за рецензент.

Настоящата рецензия е изготвена в съответствие със Закона за развитие на академичния състав в Република България, Правилника за неговото прилагане, както и с Правилника за условията и реда за придобиване на научни степени и за заемане на академични длъжности в СУ „Св. Климент Охридски“ (СУ).

1. Описание на представените материали

За участие в процедурата Мария Димитрова Митева е представила следните материали: заявление за откриване на процедура за вътрешна защита; дисертационен труд на английски език; автореферат на английски език; автореферат на български език; декларация от докторанта за оригиналност на дисертационния труд; доклад за сходство; протокол и становище от ръководителя за оригиналността на дисертационния труд; справка за изпълнението на минималните национални изисквания; доклад от научния ръководител за готовност за защита; заповед за зачисляване в докторската програма; удостоверение за положени изпити по учебната програма; заповед за отчисляване с право на защита; протокол за откриване на процедура за публична защита; списък на научните публикации по темата на дисертационния труд; декларация за съавторство; дипломи за висше образование (за ОКС „бакалавър“ и ОКС „магистър“); автобиография на докторанта; научни приноси на докторанта.

Предоставеният комплект материали е в съответствие с Правилника за условията и реда за придобиване на научни степени и за заемане на научни длъжности в СУ „Св. Климент Охридски“. Документите са пълни и прецизно подредени.

2. Данни за кандидата: образование, професионален опит и квалификация

Съгласно представените документи Мария Димитрова Митева е родена на 24.03.1982 г. в гр. Сургут, Русия. През 2005 г. се дипломира като бакалавър по специалност „Физика и математика“, професионална квалификация „Учител по физика и математика“ във Физически факултет, СУ „Св. Климент Охридски“ (Диплома серия А-04СУ No B175638). През 2010 г. се дипломира като магистър по специалност „Приложна математика“, професионална квалификация „Математик“ във Факултет по математика и информатика (ФМИ), СУ (Диплома серия А-09СУ No M200217). Съгласно приложената автобиография, докторантката има и магистърска степен по статистика, придобита в периода 2007-2008 г. в Католически университет Айхщет-Инголщат, Германия.

Мария Митева е зачислена като редовен докторант към катедра „Компютърна информатика“ със заповед №РД 20-1258 от 01.07.2021 г. по ПН 4.6. Информатика и компютърни науки, докторска програма „Компютърни науки - Машинно самообучение“. Със заповед №РД 20-1293 от 12.07.2024 г. е отчислена с право на защита, считано от 01.07.2024 г.

Докторантката има над 15 години професионален опит в международни компании и аналитични звена, където последователно е заемала длъжности, изискващи високо ниво на техническа експертиза, стратегическо мислене и лидерски умения. Последователно е заемала както технически, така и ръководни длъжности, изграждайки профил на специалист, съчетаващ научна компетентност с приложна експертиза и управленски опит в сферата на изкуствения интелект. Към момента заема позиция Мениджър „Наука за данните“ в международната консултантска компания PricewaterhouseCoopers.

Мария Митева притежава впечатляваща допълнителна квалификация, удостоверена чрез международно признати сертификати в областта на изкуствения интелект, дълбокото обучение, обработката на медицински образи и програмирането с Python и PySpark. Завършените обучения от EdX, Coursera, DataCamp и Udemy, както и сертифицирани управленски програми от Allianz Group и Harvard Business Publishing, свидетелстват за широкоспектърна подготовка, стратегическо мислене и практически опит в прилагането на иновативни технологии на изкуствения интелект в медицинския и бизнес контекст.

3. Оценка на дисертационното изследване

Актуалност на тематиката

Дисертационният труд е посветен на изключително актуална и стратегически значима тема, съчетаваща изкуствен интелект, медицинска образна диагностика и геномика. Изследването е съсредоточено върху разработването на модели за машинно обучение за предсказване на метилирания статус на промотора на гена MGMT, който е ключов биомаркер за прогнозиране на терапевтичен отговор при пациенти с глиобластом. Това заболяване е с изключително ниска преживяемост и

навременното и прецизно определяне на генетичните характеристики на тумора е от критично значение за персонализираната медицина. Прилагането на дълбоки невронни мрежи, трансферно обучение и ансамблови модели за анализ на медицински изображения и геномни данни е свързано с редица предизвикателства, напр. хетерогенност на туморите, ограничено количество анотирани данни, трудности при интеграцията на мултимодални източници, необходимост от прозрачни и надеждни системи с изкуствен интелект и т.н.

Разработените в рамките на настоящото изследване модели имат потенциала да предложат по-бърза, по-безопасна и по-прецизна алтернатива на съществуващите диагностични методи. Считаю, че това допълнително подчертава както научната значимост, така и клиничната актуалност на темата.

Познаване на проблема

Представеният дисертационен труд свидетелства за задълбочено познаване на изследванията в областта на радиогеномиката, машинното обучение и приложението на изкуствения интелект в медицинската диагностика. Докторантката е направила обширен преглед на научната литература, и на водещи публикации в области като дълбокото обучение, трансферното обучение, ансамбловите модели и радиомичния анализ.

Особено внимание е отделено на предизвикателствата, свързани с хетерогенността на туморите, ограничеността на анотирани медицински данни и сложността на интеграцията между образна, геномна и клинична информация. Докторантката демонстрира способност да обобщи наличните подходи и решения и да идентифицира ясно научната празнина, която настоящото изследване се стреми да запълни, а именно липсата на достатъчно точни, обясними и надеждни модели за неинвазивно предсказване на MGMT метилация чрез ЯМР образи.

Критичният анализ на съществуващите модели и методи е коректно структуриран и подкрепен с примери от рецензирани източници, което показва много добро ориентиране в динамично развиваща се научна област и изградена способност за оценка на техните силни и слаби страни.

Методика на изследването

Считаю, че методиката на изследването е подходящо подбрана и обоснована спрямо поставените цели и задачи на дисертационния труд. Докторантката прилага комплексен подход, основан на методите на машинното и дълбокото обучение, с фокус върху предсказване на метилационния статус на MGMT промотора въз основа на ЯМР изображения и допълнителни клинични и геномни данни.

Изследването използва мултимодален подход, включващ интеграция на образни, клинични и радиомични данни. Приложени са утвърдени техники за предварителна обработка на изображения (нормализация, сегментация, извличане на характеристики), както и за управление на високодименционални и несбалансиранни данни. Използвани са дълбоки невронни мрежи (вкл. конволюционни невронни мрежи, CNN), ансамблови методи (stacking, hard voting), трансферно обучение и техники за валидация като кръстосана проверка и бутстрединг. Особено внимание заслужава подходът към моделната валидация, използващ външни и независими тестови данни, което повишава приложимостта и обобщаемостта на резултатите.

Различните стратегии за интеграция на мултимодални данни са описани ясно и са сравнени техните предимства и ограничения в контекста на поставения проблем.

Методологията е представена ясно и систематизирано, прилагайки добри практики от биомедицинската информатика и изкуствения интелект. Изборът на алгоритми, архитектури и процеси е аргументиран, а реализацията демонстрира високо ниво на техническа компетентност и изследователска задълбоченост.

Характеристика и оценка на дисертационния труд

Дисертационният труд е структуриран в 151 страници и съдържа резюме, въведение, седем глави, библиография от 158 литературни източници, списък на фигурите, списък на таблиците, списък с използвани съкращения и речник на термините. В глава *1. Увод* (5 страници) е представен изследователския проблем, неговата значимост и сложност, както и актуалността му в контекста на съвременната медицинска диагностика и изкуствения интелект. На стр. 9 е представена целта на работата, а именно „да се създадат усъвършенствани модели на машинно обучение, които интегрират множество източници на данни с цел предсказване на статуса на MGMT метилиране при пациенти с глиобластом, използвайки неинвазивни методи като ЯМР изображения в комбинация с геномни и клинични данни“. За постигането на целта са структурирани четири задачи. В глава *2. Теоретични основи и актуално състояние на изследванията по проблема* (42 страници) е представена теоретичната рамка на изкуствения интелект и машинното обучение с фокус върху дълбокото обучение в медицинската образна диагностика. Разгледани са конволюционни невронни мрежи (CNNs), капсулни мрежи (CapsNets), трансферно обучение, ансамблови методи и меопорни векторни машини (SVMs), както и предизвикателства при анализа на медицински данни. В глава *3. ЯМР и DICOM в контекста на анализ на медицински изображения с помощта на дълбочинно обучение* (23 страници) са представени физичните принципи на ЯМР, координатните системи в медицинската визуализация и DICOM стандарта, акцентирайки върху тяхната роля в анализа на образи с методи на дълбокото обучение. Четвъртата глава, озаглавена *4. Етика в медицинската образна диагностика в ерата на изкуствени интелект* (7 страници), анализира ключови етични въпроси като защита на личните данни, алгоритмична пристрастност, прозрачност на моделите и справедлив достъп до технологии. В глава *5. Спецификация на данни и дизайн на експеримента* (10 страници) са описани използваните набори от данни (UPenn-GBM и BraTS), методологията за тяхната предварителна обработка, както и дизайна на експериментите, включително интеграция на мултимодални данни и архитектури за обучение. В глава *6. Резултати и интерпретативен анализ* (24 страници) са представени аналитичните резултати от проведените експерименти, сравнение с други подходи, ефекта от прилаганите методи за предварителна обработка и използваните ансамблови техники. Глава *7. Приноси и перспективи* (5 страници) обобщава теоретичните и експериментални приноси на дисертационния труд, посочва възможни насоки за бъдещи изследвания, научните публикации и участия на докторантката, както и декларация за оригиналност на постигнатите резултати.

Работата би могла да се обогати с по-изчерпателно разглеждане на методологични допускания, например при избора и обосновката на конкретни ансамблови модели в т. 6.1.4, както и с по-задълбочен анализ на ограниченията, свързани с интеграцията на мултимодални данни в т. 5.3, което би допринесло за по-пълно очертаване на границите на приложимост на предложения подход в реални клинични условия.

Мога да обобщя, че дисертационният труд има ясна структура, висока степен на научна задълбоченост и иновативен интердисциплинарен подход, и считам че има съществен принос към съвременното развитие на радиогеномиката и персонализираната медицина.

4. Аprobация на резултатите

Съгласно представените документи Мария Димитрова Митева е публикувала резултатите от изследванията си в три научни публикации, индексирани в Scopus, в които е първи автор. И трите публикации са в съавторство с научния ръководител, като съгласно представената „Декларация-съавтори“, ролята на докторантката е водеща и приносът ѝ е по-съществен от приноса на другия съавтор. Публикации [1] и [2] по приложения „Списък публикации“ са в сборници от конференции с SJR: 0.61 и 0.51, съответно в квартали Q2 и Q3 за 2023 г., което свидетелства за високо качество на изследванията и международна видимост. Докторантката е участвала в 3 специализирани научни конференции, едната в чужбина, и в Пролетната научна сесия на ФМИ, СУ. Представянето на резултатите на тези форуми потвърждава валидността и актуалността на разработените методи. Аprobацията обосновава приноса на дисертационния труд в областта на радиогеномиката и приложението на изкуствен интелект в медицинската диагностика.

Удовлетворени са минималните национални изисквания и изискванията на СУ „Св. Климент Охридски“ за придобиване на образователна и научна степен „доктор“ - при изискуеми 30 точки по *Група от показатели Г*, Мария Димитрова Митева отчита 78 точки.

Представените от докторантката дисертационен труд и научни трудове към него не дублират резултати от предходни процедури за академична длъжност. Няма доказано по законоустановения ред плагиатство в представения дисертационен труд и научни трудове.

5. Приноси и значимост на разработката

Съгласна съм с предложеното в дисертационния труд (стр. 118-119) условно класифициране на приносите като теоретични, методологични и експериментални. В по-синтезиран вид представям основните постижения на Мария Митева по следния начин:

Теоретични и методологични приноси:

- Разработена е мултимодална методологична рамка за интеграция на ЯМР образи, геномни и клинични данни, с цел предсказване на MGMT метилация, която отразява актуалните тенденции в персонализираната медицина.
- Извършен е обоснован подбор и ранкиране на радиомични характеристики чрез техники за важност и селекция на признаци. Подчертана е значимостта

на характеристиката *FLAIR_NC_Morphologic_Eccentricity* за детекция на морфологична асиметрия и хетерогенност на тумора.

- Реализирана е ансамблова архитектура на база *stacking generalization*, която комбинира различни модалности и доказано повишава предиктивната точност спрямо единични модели.
- Подробно са изследвани ефекти от различни техники за предварителна обработка на ЯМР изображения (нормализация, *bias field correction*, регистрация), както и тяхното влияние върху качеството на радиомичните характеристики.

Експериментални приноси:

- Проведен е качествен анализ на 67 радиомични файла с различни конфигурации, като чрез оптимизиране на параметрите е подоброено улавянето на туморната хетерогенност и намален шумът.
- Доказано е, че характеристиките, извлечени от FLAIR изображенията, в частност от некротичните области, допринасят значимо за точността на предиктивните модели.
- Извършена е сравнение на ансамблови методи (вкл. *Hard Voting Classifier* с RF, XGBoost и SVM), които показват по-висока устойчивост и точност спрямо отделни класификатори.
- Постигнато е високо качество на моделиране чрез комбиниране на структурирани и неструктурирани данни – образни, геномни и клинични – като резултатите са валидирани върху независими набори от данни.
- Идентифицирани са най-подходящите стратегии за предварителна обработка и моделиране в контекста на радиогеномната класификация, което може да служи като референтен подход за бъдещи изследвания.

Предложената разработка има потенциал за трансляция в клиничната практика като инструмент за неинвазивна диагностика и подпомагане на терапевтичните решения. Тя отговаря на актуалните изисквания за развитие на прецизна медицина, като едновременно допринася за задълбочаване на познанията в областта на радиогеномиката и изкуствения интелект.

6. Качества на автореферата

Авторефератът е с обем 45 страници и представя по ясен и систематичен начин съдържанието на дисертационния труд. В него са отразени коректно целите на изследването, използваните методи, получените резултати и направените научни и приложни изводи. Текстът е добре структуриран, логически последователен и подходящо онагледен. Авторефератът е оформен в съответствие с изискванията на Правилника за условията и реда за придобиване на научни степени и заемане на академични длъжности в СУ „Св. Климент Охридски“.

7. Въпроси

Имам следните въпроси към докторантката:

В дисертационния труд е приложена мултимодална интеграция на ЯМР, радиомни и клинични данни чрез ансамблови модели.

1. Бихте ли обяснили какви бяха основните предизвикателства при синхронизирането на различните модалности и как се гарантира, че интеграцията не води до „information leakage“ или пренасищане на модела?

2. При мултимодална интеграция как се отразяват отделните данни в общия резултат на системата?

8. Заключение

Оценката ми за дисертационния труд, автореферата, научните публикации и получените резултати от работата на Мария Димитрова Митева е изцяло положителна. Получените оригинални резултати изцяло съответстват на поставената в дисертационния труд цел. Считаю, че докторантката има много добри познания в областта, в която работи и са получени оригинални научни и научно-приложни резултати и потвърждавам, че представеният дисертационен труд и научните публикации към него, както и качеството и оригиналността на представените в тях резултати и постижения, отговарят на изискванията на ЗРАСРБ, Правилника за приложението му и съответния Правилник на СУ „Св. Климент Охридски“ за придобиване от кандидата на образователната и научна степен „доктор“ в научната област 4. Природни науки, математика и информатика, професионално направление 4.6. Информатика и компютърни науки. Не е установено плагиатство в представените по конкурса научни трудове.

Въз основа на гореизложеното, **препоръчвам** на научното жури да присъди на Мария Димитрова Митева образователна и научна степен „доктор“ в научна област 4. Природни науки, математика и информатика, професионално направление 4.6. Информатика и компютърни науки, докторската програма „Компютърни науки - Машинно самообучение“.

06.06.2025 г.

Изготвил рецензията:
(проф. д-р Евдокия Николаева Сотирова)