

To:
Graduate Studies Committee
Faculty of Medicine and Organization of Healthcare
al-Farabi Kazakh National University
Almaty
Kazakhstan

Fachbereich Humanmedizin
Klinik für Innere Medizin mit
Schwerpunkt Pneumologie
Translationales Forschungslabor

Dr. Nurlan Dauletbayev, Leiter

Tel.: +49 (0) 6421-586-3304

Fax: +49 (0) 6421-587-6012

E-Mail:

nurlan.dauletbayev@uni-marburg.de

Anschrift: Herz-Lungenzentrum,
Büro -1/34320
Baldingerstraße
35043 Marburg

Web: www.uni-marburg.de

Marburg, den 23. November 2024

Betr.: Evaluation of PhD thesis of Ms. Gauhar Dunenova, shortened on December 27, 2024

The following is the evaluation of PhD thesis of the aforementioned doctoral candidate.

Summary and general comments:

The thesis presents results of a multidisciplinary investigation into several aspects related to breast cancer. The thesis has studied the epidemiology of breast cancer in Kazakhstan, and described the unmet need in the areas of digital pathology and telepathology. The thesis also presents the findings of a systematic review (first-authored by doctoral candidate; PMID: 39123488) of published literature on computational diagnostic algorithms for analysis of breast cancer immunohistochemistry. This review also utilized the Newcastle-Ottawa Scale to appraise the quality of clinical studies described in the included publications. The diagnostic algorithms themselves were not appraised. In particular, specific appraisal tools (such as QUADAS-2; PMID: 22007046) were not utilized to evaluate the selection publications, a limitation acknowledged by the authors in Discussion of the publication. The thesis next briefly describes the dataset of immunohistochemical images of HER2 staining. The dataset was manually compiled by the doctoral candidate from digital scans of archived pathology glass slides. In addition to the aforementioned investigations, this thesis also provides some suggestions on how to mitigate the unmet need in digital pathology and telepathology (e.g., capacity building in digital pathology, creation of image repositories).

Major achievements of the thesis:

Creation of a digital dataset of immunohistochemistry images of breast cancer biopsies (initially annotated by doctoral candidate) is the major achievement of this thesis.

Major shortcomings of the thesis and recommendations:

Research Objective no. 3 was achieved only in part.

1. The second part of the Objective (“... determine the conditions ...”) was not addressed in Chapter 3.3.3. The project should have utilized the frameworks recommended in implementation research. For example, the Consolidated Framework for Implementation Research (CFIR; PMID: 36309746 and <https://cfirguide.org/>) could have been employed. While some domains of CFIR were partly addressed in the thesis, the thesis investigations did not fully adhere to this framework. In particular, the description of web portal “Qazhisto.com” only focuses on the characteristics of the personnel that operates this web portal and the technical specifications of “Qazhisto.com”. In contrast, the capacity of the end users (both in digital training and digital infrastructure available to them) was neither studied nor presented. **Recommendation:** The CFIR domains that were not studied in the thesis should be listed as limitations in Discussion.
2. Chapter 3.5 is named “Recommendations on the use of the algorithm and development prospects”. This Chapter, however, does not provide recommendations on the diagnostic algorithm. In fact, all recommendations of this Chapter are related to the use of the aforementioned web portal Qazhisto.com. **Recommendation:** This Chapter does not provide information that would distinctly differ from previous Chapters. Therefore, Chapter 3.5 should be combined with Chapter 3.3.3.
3. Furthermore, Chapter 3.5 should have presented quantitative metrics. It is presently difficult to evaluate whether the thesis recommendations are realistic. For example, Figure 27 suggests “Specialist Training” (the right part of this Figure, outlined in green). The number of required training sessions for existing personnel and/or training of new pathologists is neither stated nor justified. **Recommendation:** When Chapter 3.5 is combined with Chapter 3.3.3., doctoral candidate should add appropriate quantitative metrics, such as number of new pathologists required, how many training sessions are needed, etc.
4. The statement “It is necessary to ensure the availability of modern equipment, ...” (page 140) is not supported by cost estimates. There are specific frameworks for health technology assessment. None of these appear to be applied in this thesis. **Recommendation:** To update the statement of page 140 with quantitative metrics (e.g., costs of purchase of new servers).
5. This also concerns other suggestions of the thesis, for example, on page 153 (“To increase the number of pathologists ...”). This suggestion does not address (a) how many pathologists would be needed, (b) whether a reform of a residency program

would be required, (c) how the repository and platform would be utilized in the training (as a simulation center?), (d) how much this would cost, (e) how to reorganize the licensing exams. The thesis also does not address how to mitigate the acute and substantial shortage of pathologists in Kazakhstan (PMID: 38680227, reference 43 of the thesis). **Recommendation:** The recommendations should not stop at stating the obvious need for more pathologists. The thesis should offer a structured, well-described, and quantifiable path for implementation of the proposed changes.

Research Objective no. 4

This Research Objective was also achieved only in part.

1. The unnecessary brevity of methodological description pertaining to the proposed diagnostic algorithm makes an objective evaluation very difficult. This evaluation acknowledges the fact that this thesis was prepared for a PhD degree in Public Health (that is, not a PhD degree in informatics or programming). However, this evaluation was guided by the wording of Research Objective 4, which was to develop an algorithm for computational diagnosis, implying a clinical application. Moreover, Chapter 3.5 states the goal of publishing this algorithm. This justifies an application of standard criteria to evaluate diagnostic algorithms. **Recommendation:** The methodology should be updated with sufficient detailization to ensure technical robustness and soundness of the proposed algorithm, as well as the reproducibility of thesis findings. It is obvious that this Public Health thesis won't achieve the depth of a standard informatics thesis. As an alternative, doctoral candidate can use a Bachelor of Science thesis (example: https://www.uni-bamberg.de/fileadmin/xai/material/231002_Bachelor-thesis_Hempel-Tom.pdf).
2. The image pre-processing should also be updated with pertinent details. **Recommendation:** To use this short but informative description (PMID: 36879615) as an example.
3. Similar criticism can be applied to description of actual results (pages 136 – 139). The diagnostic algorithm is described as a convolutional neural network (CNN). However, it is not clear whether doctoral candidate and her collaborators developed this CNN using existing Python libraries or completely anew. Reference 253 (cited in the legend to Figure 25) cites an image from Internet, not an actual publication. **Recommendation:** To update either the methodological description or Results with pertinent details of CNN. The aforementioned Bachelor thesis provides a good example of a sufficiently detailed description.
4. The training of diagnostic algorithm is described very superficially. Furthermore, important training parameters (i.e., number of “epochs”) or model fine-tuning were not provided in the submitted version of the thesis. It is also not clear whether the training was done using whole-slide images or smaller fragments of the images (called “patches”). Moreover, it is not clear whether the metrics presented on Figure 26 describe the model performance during training or testing. **Recommendation:** The aforementioned Bachelor thesis should be used as a good example of clear and detailed description of both methodology and experimental findings with a proposed algorithm. All statements in this comment should be addressed in detail.
5. The thesis provides no comparisons between the diagnostic performances of the proposal algorithms and the previously published ones, when both are applied to this dataset. For example, thesis states (page 138) that “The cumulative accuracy metric was 0.75, which indicates that the model correctly classifies 75% of all samples”. This is a rather low diagnostic performance for a classification algorithm. There are classification algorithms with a much higher model performance (e.g., PMID: 39004603). Furthermore, there are informatics approaches to estimate the range of model performance metrics. Without a comparison with the established algorithms, both in average metrics and the range, it is hard to judge whether the proposed algorithm offers a better or worse diagnostic performance. **Recommendation:** To remove all statements about the “diagnostic” nature of this algorithm from the thesis. The thesis should emphasize the pilot nature of its study, for example, with the following statement: “The presented thesis has tested, in a pilot study, a principal ability to apply deep learning algorithms (on an example of CNN) to the dataset compiled as part of thesis investigations”.
6. Another thesis statement (page 151) raises criticism: “The overall accuracy of the model is 75%, which indicates an average accuracy rank of the classification level, compared with the results of the systematic review”. As mentioned above, the systematic review first-authored by doctoral candidate had not applied QUADAS-2 (or a similar tool) to appraise the included diagnostic algorithms. It is possible that the “average accuracy rank” of diagnostic algorithms would have been much higher if QUADAS-2 tool had excluded the diagnostic algorithms with a substantial risk of bias. **Recommendation:** This comparison is not valid. The statement in question should be removed.
7. If the “diagnostic” nature of the algorithm is stated, then the algorithm should have been tested on an external dataset. This is called “external validation”, and this is an important part of the validation of diagnostic performances of computational algorithms (e.g., PMID: 33564405). Doctoral candidate described several open-source datasets in her systematic review. **Recommendation:** To remove all statements about the “diagnostic” nature of this algorithm and emphasize the pilot nature of this investigation. The lack of external validation should be stated as a limitation in Discussion (Conclusion).
8. The class imbalance of the dataset and interpretability of the proposed diagnostic algorithm were not addresses comprehensively. **Recommendation:** This should be stated as a limitation in Discussion (Conclusion).
9. Discussion (Conclusion) is written as rephrasing of the main findings. If international standards for PhD theses are to be followed, then Discussion (Conclusion) should put the thesis findings in a larger context. **Recommendation:** To update Discussion (Conclusion) with the aforementioned limitations. In addition, Discussion (Conclusion) could provide expanded statements on how to address the shortage of pathologists (with quantifiable plans), improve the digitization of pathology, and mitigate insufficient digital literacy of end users (again, with quantifiable plans).

In summary, this thesis represents research findings of both of theoretical and practical value. Its primary achievement is the compilation of a new digital dataset of immunohistochemistry staining of breast cancer biopsies.

Best regards,



Dr. Nurlan Dauletbayev

Head, Translational Research Lab / Department of Internal, Respiratory and Critical Care Medicine
Philipps-University of Marburg / Biomedical Research Centre (BMFZ), Office +1/61300
Hans-Meerwein-Str. 2, 35043 Marburg, Germany
Tel.: +49 (0) 6421-286-6041
Fax: +49 (0) 6421-286-5133
Email: nurlan.dauletbayev@uni-marburg.de

Adjunct Professor
Division of Respiratory Medicine / Montreal Children's Hospital
Department of Pediatrics, Faculty of Medicine
McGill University
1001 Decarie Boulevard
Montreal, Quebec, H4A 3J1, Canada
Email: nurlan.dauletbayev@mcgill.ca

Associate Investigator
Child Health and Human Development Program
Research Institute of McGill University Health Centre
1001 Decarie Boulevard
Montreal, Quebec, H4A 3J1, Canada

Ниже приводится оценка кандидатской диссертации вышеупомянутого докторанта.

Краткое содержание и общие замечания:

В диссертации представлены результаты междисциплинарного исследования нескольких аспектов, связанных с раком молочной железы. В диссертации изучалась эпидемиология рака молочной железы в Казахстане и описывались неудовлетворенные потребности в области цифровой патологии и телепатологии. В диссертации также представлены результаты систематического обзора (впервые подготовленного кандидатом медицинских наук; PMID: 39123488) опубликованной литературы по компьютерным диагностическим алгоритмам для иммуногистохимического анализа рака молочной железы. В этом обзоре также использовалась шкала Ньюкасла-Оттавы для оценки качества клинических исследований, описанных в включенных публикациях. Сами диагностические алгоритмы не оценивались. В частности, для оценки отобранных публикаций не использовались специальные инструменты оценки (такие как QUADAS-2; PMID: 22007046), что является ограничением, признанным авторами при обсуждении публикации. Далее в диссертации кратко описывается набор данных иммуногистохимических изображений окрашивания HER2. Набор данных был составлен докторантом вручную на основе цифровых сканирований архивных срезов патологоанатомических материалов. В дополнение к вышеупомянутым исследованиям, в этой диссертации также содержатся некоторые предложения о том, как уменьшить неудовлетворенную потребность в цифровой патологии и телепатологии (например, наращивание потенциала в области цифровой патологии, создание хранилищ изображений).

Основные достижения диссертации:

Создание цифрового набора данных иммуногистохимических изображений биоптатов рака молочной железы (первоначально с комментариями докторанта) является главным достижением этой диссертации.

Основные недостатки диссертации и рекомендации:

Цель исследования № 3 была достигнута лишь частично.

1. Вторая часть задачи (“... определить условия...”) не рассматривалась в главе 3.3.3. В проекте должны были использоваться рамки, рекомендованные в ходе исследования внедрения. Например, можно было бы использовать консолидированную структуру исследований по внедрению (CFIR; PMID: 36309746 и <https://cfirguide.org/>). Хотя некоторые области CFIR были частично рассмотрены в диссертации, исследования в диссертации не полностью соответствовали этой структуре. В частности, в описании веб-портала “Qazhisto.com” основное внимание уделяется только характеристикам персонала, который управляет этим веб-порталом, и техническим характеристикам “Qazhisto.com”. В отличие от этого, возможности конечных пользователей (как в области цифрового обучения, так и в доступной им цифровой инфраструктуре) не были изучены и не представлены. Рекомендация: Области CFIR, которые не были изучены в диссертации, должны быть указаны в качестве ограничений при обсуждении.

2. Глава 3.5 называется “Рекомендации по использованию алгоритма и перспективам развития”. Однако в этой главе не приводятся рекомендации по диагностическому алгоритму. Фактически, все рекомендации этой главы относятся к использованию вышеупомянутого веб-портала Qazhisto.com. Рекомендация: Данная глава не содержит информации, которая бы существенно отличалась от предыдущих глав. Поэтому главу 3.5 следует объединить с главой 3.3.3.

3. Кроме того, в главе 3.5 должны были быть представлены количественные показатели. В настоящее время трудно оценить, насколько реалистичны рекомендации диссертации. Например, на рисунке 27 указано “Подготовка специалистов” (правая часть этого рисунка выделена зеленым цветом). Количество необходимых учебных занятий для существующего персонала и/или подготовки новых патологоанатомов не указано и не обосновано. Рекомендация: При объединении главы 3.5 с главой 3.3.3 докторанту следует добавить соответствующие количественные показатели, такие как количество требуемых новых патологоанатомов, количество необходимых учебных занятий и т.д.

4. Утверждение “Необходимо обеспечить доступность современного оборудования...” (стр. 140) не подтверждается оценкой затрат. Существуют конкретные рамки для оценки технологий здравоохранения. Ни одна из них, по-видимому, не применяется в данной работе. Рекомендация: Дополнить описание на странице 140 количественными показателями (например, затратами на покупку новых серверов).

5. Это также касается других предложений диссертации, например, на странице 153 (“Увеличить число патологоанатомов...”). Это предложение не касается (а) того, сколько потребуется патологоанатомов, (б) потребуется ли реформирование программы ординатуры, (в) как хранилище и платформа будут использоваться при обучении (в качестве симуляционного центра?), (г) сколько это будет стоить, (е) как реорганизовать экзамены на получение лицензии. В диссертации также не рассматривается вопрос о том, как смягчить острую и существенную нехватку патологоанатомов в Казахстане (PMID: 38680227, ссылка 43 диссертации). Рекомендация: Рекомендации не должны ограничиваться констатацией очевидной потребности в большем количестве патологоанатомов. В диссертации должен быть предложен структурированный, хорошо описанный и поддающийся количественной оценке путь внедрения предлагаемых изменений.

Цель исследования № 4

Эта исследовательская цель также была достигнута лишь частично.

1. Излишняя краткость методологического описания, относящегося к предлагаемому диагностическому алгоритму, очень затрудняет объективную оценку. Данная оценка подтверждает тот факт, что данная диссертация была подготовлена для получения степени доктора философии в области общественного здравоохранения (то есть не для получения степени доктора философии в области информатики или программирования). Однако эта оценка проводилась в соответствии с формулировкой цели исследования 4, которая заключалась в разработке алгоритма компьютерной диагностики, предполагающего клиническое применение. Более того, в главе 3.5 указывается цель публикации этого алгоритма. Это оправдывает применение стандартных критериев для оценки диагностических алгоритмов. Рекомендация: Методология должна быть обновлена с достаточной степенью детализации, чтобы обеспечить техническую надежность и обоснованность предлагаемого алгоритма, а также воспроизводимость результатов исследования. Очевидно, что эта диссертация по общественному здравоохранению не достигнет уровня стандартной диссертации по информатике. В качестве альтернативы кандидат в докторанты может использовать диссертацию на степень бакалавра наук (пример: https://www.uni-bamberg.de/fileadmin/xai/material/231002_Bachelor-thesis_Hempel-Tom.pdf).

2. Предварительная обработка изображений также должна быть дополнена соответствующими деталями. Рекомендация: Использовать это краткое, но информативное описание (PMID: 36879615) в качестве примера.

3. Аналогичная критика может быть применена к описанию фактических результатов (стр. 136 – 139). Диагностический алгоритм описывается как сверточная нейронная сеть (CNN). Однако неясно, разработали ли докторант и ее коллеги этот CNN с использованием существующих библиотек Python или полностью заново. Ссылка 253 (приведена в подписи к рисунку 25) содержит изображение из Интернета, а не фактическую публикацию. Рекомендация: Обновить либо методологическое описание, либо результаты с учетом соответствующих деталей CNN. Вышеупомянутая дипломная работа бакалавра является хорошим примером достаточно подробного описания.

4. Обучение диагностическому алгоритму описано очень поверхностно. Кроме того, в представленной версии диссертации не были указаны важные параметры обучения (например, количество “эпох”) или точная настройка модели. Также неясно, проводилось ли обучение с использованием изображений всего слайда или небольших фрагментов изображений (называемых “заплатками”). Кроме того, неясно, описывают ли показатели, представленные на рисунке 26, производительность модели во время обучения или тестирования. Рекомендация: Вышеупомянутую дипломную работу бакалавра следует использовать в качестве хорошего примера четкого и подробного описания как методологии, так и экспериментальных результатов с предлагаемым алгоритмом. Все утверждения в этом комментарии следует рассматривать подробно.

5. В диссертации не приводится сравнений между диагностическими характеристиками предлагаемых алгоритмов и ранее опубликованных, когда оба они применяются к данному набору данных. Например, в тезисе говорится (стр. 138), что “Совокупный показатель точности составил 0,75, что указывает на то, что модель правильно классифицирует 75% всех образцов”. Это довольно низкая диагностическая производительность для алгоритма классификации. Существуют алгоритмы классификации с гораздо более высокой производительностью модели (например, PMID: 39004603). Кроме того, существуют информационные подходы для оценки диапазона показателей эффективности модели. Без сравнения с установленными алгоритмами, как по средним показателям, так и по диапазону, трудно судить о том, обеспечивает ли предложенный алгоритм лучшую или худшую диагностическую производительность. Рекомендация: Убрать из диссертации все утверждения о “диагностическом” характере этого алгоритма. В диссертации следует подчеркнуть пилотный характер ее исследования, например, следующим утверждением: “В представленной диссертации в ходе пилотного исследования была протестирована принципиальная возможность применения алгоритмов глубокого обучения (на примере CNN) к набору данных, собранному в рамках исследований диссертации”.

6. Еще одно тезисное утверждение (стр. 151) вызывает критику: “Общая точность модели составляет 75%, что указывает на средний ранг точности уровня классификации по сравнению с результатами систематического обзора”. Как упоминалось выше, в систематическом обзоре, впервые подготовленном докторантом, не применялся QUADAS-2 (или аналогичный инструмент) для оценки включенных диагностических алгоритмов. Вполне возможно, что “средний уровень точности” диагностических алгоритмов был бы намного выше, если бы инструмент QUADAS-2 исключал

диагностические алгоритмы со значительным риском смещения. Рекомендация: Это сравнение некорректно. Данное утверждение следует удалить.

7. Если заявлен “диагностический” характер алгоритма, то алгоритм должен был быть протестирован на внешнем наборе данных. Это называется “внешняя проверка”, и это важная часть проверки диагностических характеристик вычислительных алгоритмов (например, PMID: 33564405). В своем систематическом обзоре кандидат в докторанты описала несколько наборов данных с открытым исходным кодом. Рекомендация: Убрать все заявления о “диагностическом” характере этого алгоритма и подчеркнуть пилотный характер этого исследования. Отсутствие внешней валидации должно быть указано в качестве ограничения при обсуждении (заключение).

8. Классовая несбалансированность набора данных и интерпретируемость предлагаемого диагностического алгоритма не были рассмотрены всесторонне. Рекомендация: Это должно быть указано в качестве ограничения при обсуждении (заключение).

9. Обсуждение (заключение) оформляется как перефразирование основных выводов. Если необходимо следовать международным стандартам для защиты докторских диссертаций, то обсуждение (заключение) должно рассматривать выводы диссертации в более широком контексте. Рекомендация: Дополнить обсуждение (заключение) вышеупомянутыми ограничениями. Кроме того, в ходе обсуждения (заключения) можно было бы сделать развернутые заявления о том, как решить проблему нехватки патологоанатомов (с помощью поддающихся количественной оценке планов), улучшить оцифровку патологии и снизить уровень недостаточной цифровой грамотности конечных пользователей (опять же, с помощью поддающихся количественной оценке планов).

В заключение, эта диссертация представляет собой результаты исследования, имеющие как теоретическую, так и практическую ценность. Ее основным достижением является составление нового цифрового набора данных иммуногистохимического окрашивания биоптатов при раке молочной железы.

С уважением,