AWARD NUMBER: W81XWH-21-1-0725

TITLE: Analysis of Tissue Architecture to Identify Lethal Prostate Cancer in the Veteran Population

PRINCIPAL INVESTIGATOR: Dr. Beatrice Knudsen

CONTRACTING ORGANIZATION: The University of Utah

REPORT DATE: OCTOBER 2022

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
OCTOBER 2022	ANNUAL	1SEPT2021 - 31AUG2022

4. TITLE AND SUBTITLE

Analysis of Tissue Architecture to Identify Lethal Prostate Cancer in the Veteran Population

5a. CONTRACT NUMBER
W81XWH-21-1-0725

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Beatrice Knudsen

E-Mail:

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The University of Utah

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT:

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. **ABSTRACT: The project in this grant is aimed to improve the precision medicine treatment of men with high-grade prostate cancer who are diagnosed within the VA healthcare system. Recent clinical trials demonstrate that multi-modal upfront treatment with combination of two – three cancer drugs improve overall survival and disease-free survival of men with aggressive prostate cancer. However, clinical parameters and genomic test results do not reliably identify men who are at greatest risk of dying from lethal prostate cancer. The Gleason grade of the cancer is a good predictor of disease course overall, but does not predict who is at risk for metastatic progression within the high-grade cancer group. We propose that in addition to the cancer grade, computer can detect salient features of the cancer that are associated with risk of metastasis. Using artificial intelligence frameworks, we propose to train models on high-grade cancer regions that predict metastatic risk. We identified approximately 12,000 men at the VA which we propose to study. Of those, we will enroll at least 600 for the initial development of algorithms that is funded by this grant. During the first year of funding, we identified the study cohorts, established the basic enrolment system using STARLIMS and generated 1337 digital slides from 279 cases. We also consolidated multiple A.I. algorithms that can be used to identify regions of high-grade cancer and are testing which of them works best with VA cases. To perform above tasks, we completed all IRB and regulatory requirements and recruited all the necessary expertise for the project. The generous funding by the Department of Defense will help us with the development of an affordable and easily deployable software tool that has a chance to improve the care of veterans diagnosed with aggressive prostate cancer.**

15. SUBJECT TERMS NONE LISTED

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	31	**19b. TELEPHONE NUMBER** *(include area code)*

TABLE OF CONTENTS

1. **INTRODUCTION:** *Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.*

   This population sciences award project is designed to develop machine learning algorithms that predict the risk of metastases in men diagnosed with prostate cancer. The purpose of the research is to extract information from pathology slides about the aggressive nature of the cancer and to use the information for treatment decisions. The scope of the multi-disciplinary and multi-site research involves the development, validation of testing of machine learning models for the diverse population of men who receive medical care in the VA system.

2. **KEYWORDS:** *Provide a brief list of keywords (limit to 20 words).*

   Prostate cancer, needle biopsy, machine learning, pathology slides, algorithm, image analysis, veteran administration

3. **ACCOMPLISHMENTS:** *The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction.*

   What were the major goals of the project?
   List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

   1. Major task 1: **Cohort assembly (SOW completion at the end of year 1)**
   Subtask 1: Regulatory IRB, and data sharing agreements for image analysis: Team A (Knudsen, Lynch) DA with VINCI for pathology annotations, computer generated Gleason scores and Metastatic Risk scores.

   Subtask 2: Enrolment of participants into cohort
   - Team A (Lynch & DuVall): NLP for selecting 2000 cases for cohort. Project database.
   - Team B (Garraway): patient selection, data quality assessment for enrolment
   - Team A (Brintz): patient assignment to study group

   Subtask 3: Requesting cases from VA hospitals for the population-based cohort
   Team B (Garraway): requesting cases and follow-up until receipt of 900 – 1300 cases, accessioning cases into tracking system, Q/C of H&E, slide repair and new slide generation, sending slides to team A. Team A (Knudsen): monitoring slide receivals

   Subtask 4: S*lide scanning of 600 cases, pathology annotation of 100 cases and Gleason scoring of 600 cases*
   - Team A (Knudsen, Post): Slide scanning and access to digital slides
   - Team A (Knudsen, Post and pathologists): Slide annotations using QuIP software
   - Team A (Whitaker, Knudsen): computer- assisted Gleason scores

Milestone 1. Accessioning Cohort (n = 900 – 1300) – cases in each group depend on availability of cases in path archive: each group (M0- NMP no metastasis at diagnosis or follow-up; M0-MP future metastatic progression, M1 metastasis at diagnosis), will consist of 300 - 450 cases. Of those, 200 per group will be enrolled into the final study.

What was accomplished under these goals?
For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

A. Major task: Cohort Assembly
a. Major task 1.1: *Regulatory IRB, and data sharing agreements for image analysis*
1. Major activities: Execution of IRB and DSA/DUA
2. Specific objectives: HRPO approval
3. Significant result or key outcomes (major findings, developments, conclusions):

- The IRB protocol 00144620 for the study has been approved by a committee that does reciprocal approvals between the Salt Lake City VA and the University of Utah on 11/22/2021. The title of the IRB is: Pathology Image Analysis of Veterans Diagnosed with prostate Cancer. The status of the protocol according to the IRB review is "EXEMPT"
- A Data Use Agreement between the VA and the U of Utah was executed at the Greater Los Angeles VA in February 2022. This agreement allows deidentified image tiles to be transferred between the VA and the U of Utah.

4. Other achievements: none

b. Major task 1.2: *Enrolment of participants into cohort*
1. Major activities: Identify enrolment criteria for M0, M1 and M0-P cohorts shown in Table 1, write the code to search the Sequel database for case identification, refine the search function,
2. Specific objectives: identify 600 cases for the study, 200 cases in the M1, M0-P and M0 groups
3. Significant result or key outcomes (major findings, developments, conclusions): 3 lists of cases totaling ~1,200 in the M1 and M0-P groups with surgical pathology numbers delivered to GLA-VA, Tampa and SLC VA sites. The SLC site received a list of 465 cases with SP numbers.

   Enrolment and matching criteria for M0 group to obtain cases that qualify.
4. Other achievements: Regular meetings with team for searching VA databases and generating tables for case retrieval and statistical analysis

Table 1. Summary data of patient from M0, M0-P and M1 subcohorts

	M0 NON-METATASTIC (N=9038)	M0 METASTATIC (N=1218)	M1 (N=871)
AGE AT DIAGNOSIS			
MEAN (SD)	66 (7.8)	65 (8.5)	70 (9.6)
MEADIAN (RANGE)	66 (34-95)	65 (42-93)	70 (43-99)
RACE			
WHITE	6116 (67.7%)	783 (64.3%)	551 (63.3%)
NON-WHITE	2922 (32.3%)	435 (35.7%)	320 (36.7%)
PSA PRIOR TO DIAGNOSIS			
</= 20	8220 (90.9%)	931 (76.4%)	302 (34.7%)
>20	550 (6.1%)	238 (19.5%)	499 (57.3%)
UNKNOWN	268 (3.0%)	49 (4.1%)	70 (8.0%)
CLINICAL T STAGE AT DIAGNOSIS			
T1/T2	8176 (90.5%)	917 (75.2%)	68 (7.8%)
T3/T4	450 (5.0%)	281 (23.2%)	744 (85.4%)
UNKNOWN	412 (4.5%)	20 (1.6%)	59 (6.8%)
CLINICAL N STAGE AT DIAGNOSIS			
N0	9038 (100%)	1210 (99.3%)	776 (89%)
N1	0 (0%)	8 (0.7%)	95 (11%)
NX	0 (0%)	0 (0%)	0 (0%)
METASTATIC SITES AT DIAGNOSIS			
BONE	NA	NA	525
LIVER	NA	NA	30
LUNG	NA	NA	37
SKIN	NA	NA	1
CNS	NA	NA	6
LYMPH NODES	NA	NA	95
PLEURA	NA	NA	0
PERITONEUM	NA	NA	3
GLEASON SUM AT DIAGNOSIS			
6	4276 (47.3%)	319 (26.2%)	45 (5.2%)
7	3326 (36.8%)	447 (36.7%)	148 (17.0%)
8	773 (8.6%)	201 (16.5%)	167 (19.2)
9	378 (4.2%)	209 (17.2%)	325 (37.3)
10	47 (0.52%)	24 (2.0%)	65 (7.5%)
UNKNOWN	238 (2.6%)	18 (1.4%)	121 (13.9%)
PRIMARY TUMOR TREATMENT (LOCALIZED)			
RP	972	301	NA
XRT +/- ADT	5103	900	NA
ADT	656	267	NA
WW/AS/UNKNOWN	2308	51	NA
VITAL STATUS			
ALIVE	3898 (43.1%)	499 (41.0%)	195 (22.4%)
DECEASED	5140 (56.9%)	719 (59.0%)	676 (77.6%)
MONTHS FOLLOW UP (DX TO LAST VA VISIT)			
MEAN (SD)	109.2 (57.3)	49.2 (58.5)	47.4 (45.01)
MEDIAN (RANGE)	104 (0-270)	116 (15-270)	32 (0-263)
LOCATION (1ST VA VISIT)			
WEST LOS ANGELES - 691	1637	301	212
SALT LAKE CITY – 660	1056	167	132
TAMPA – 673	2780	252	167
CLEVELAND - 541	2296	287	274
OTHER	1269	211	86

c. Major task 1.3: *Requesting cases from VA hospitals for the population-based cohort*

1. Major activities: Case lists deployed to VA sites
2. Specific objectives: working with pathology departments at respective VA sites to retrieve cases from pathology archive
3. Significant result or key outcomes (major findings, developments, conclusions): GLA-VA provides study coordinator and established a LIMS system for tracking. GLA-VA is entering 100 cases per week with data into the LIMS system. The cases are linked to medical records by a unique identifier.

   Slides are deidentified before scanning to increase data safety.
4. Other achievements: workflow for requesting, receiving, entering into LIMS and scanning slides has been established and validated. Dr. Andrew Borkowsky, a pathologist at the TAMPA VA received IRB approval to send cases to GLA-VA. Need to establish reimbursement strategy.

d. Major task 1.4: *Slide scanning, pathology annotations and development of automated Gleason scoring algorithm*

1. Major activities: At GLA-VA, at the time this report was written, 350 cases including 1,400 slides have been entered in the LIMS. 1000 of the slides have been scanned.
2. Specific objectives: Building pipeline for slide scanning, slide repository, pathology annotations and algorithm training
3. Significant result or key outcomes (major findings, developments, conclusions): GLA-VA purchased high throughput Aperio GT450 slide scanner, the slide scanner is connected to the VA network, GLA-VA hired a research assistant to scan slides and enter cases into the LIMS system. The VA STARLIMS was originally developed to track specimens from the Million Veteran Program by Dr. Saiju Pyarajan. Dr. Pyarajan has installed STARLIMS at GLA-VA for specimen acquisition and tracking for the VA MAPP biorepository.

   The pipeline is ready for SLC and Tampa sites to send cases to GLA-VA for entering into LIMS and scanning. The system is ready to upload a copy of deidentified slides to a folder in the high-performance computing center (CHPC) at University of Utah. The *Quantitative Imaging in Pathology "QuIP"* has been installed at CHPC and can be used by pathologists to open VA digital slides and provide annotations. The image analysis team in Utah at the Institute for Scientific Computing and Imaging (SCI) has access to the CHPC folder that contains annotated images for development of algorithms.
4. Other achievements: preliminary algorithms have been trained and are undergoing optimization

Milestone 1 completion (date for milestone 1 completion is the end of year 1):

- Accessioning Cohort of 900 – 1300 cases in each group: Cases have been identified. There are 1218 cases in cohort M0-P of which we will need 200 cases. There are 871 cases in cohort M1 of which we will need 200 cases. There are 9038 cases in cohort M0. From this cohort we will identify cases with PSA > 20 (550 cases in the parent cohort) and match M0 cases to M1 or M0-P cases by site and time period of biopsy.

Cases available for image analysis from GLA-VA:

Number of patients: 279

Number of scanned whole slide images available: 1337

Number of M0 non-metastatic (M0-NM) cases with Gleason pattern 4 or higher: 93

Number of M0-progressed cases: 65

Number of M1 cases: 41

Number of cases needing staging annotation: 69

- Retrieving 300 - 450 cases for Q/C and scanning. GLA-VA is scanning cases from M0-P and M1 cohorts. Tampa and SLC are resolving logistic issues of case retrieval and reimbursement

Milestone 2 completion (date for milestone 2 completion is 18 months after grant activation):

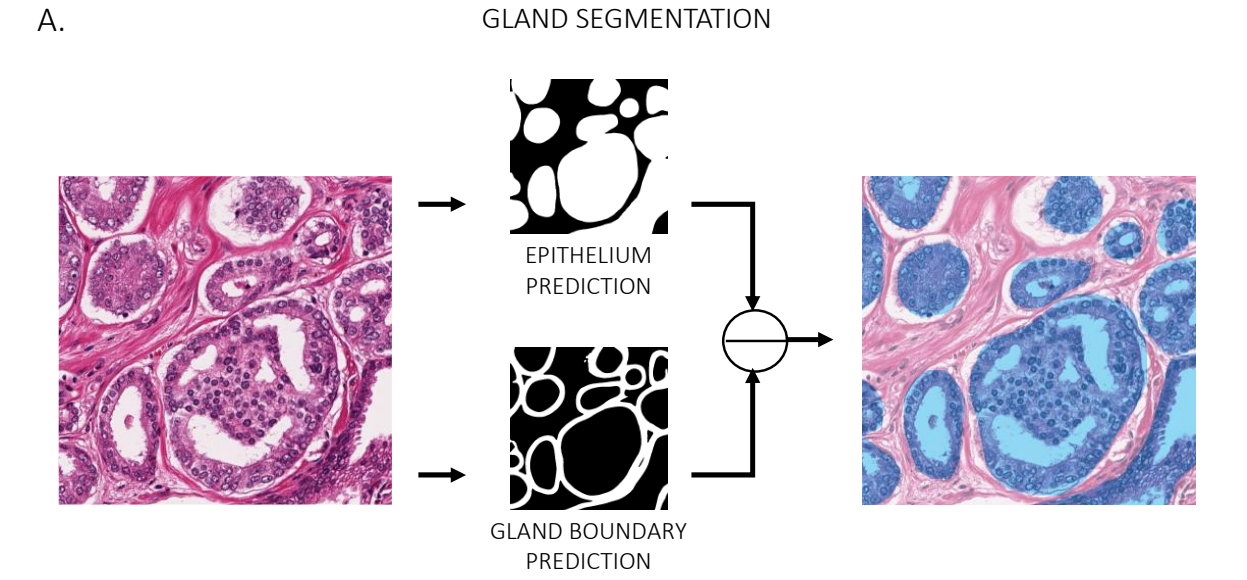
A.
GLAND SEGMENTATION
EPITHELIUM PREDICTION
GLAND BOUNDARY PREDICTION
B.
IDENTIFICATION OF CANCER GLANDS
ENCODER
FEATURE-WISE HISTOGRAMS
CLASSIFIER
Benign/cancer classification
C.
FEATURE-WISE HISTOGRAMS
Patch
Feature
1 2 ... n
1
2
128
Bin functions
D.
GRADING OF CANCEROUS GLANDS
CLASSIFIER
PROBABILITY SUM
Low/high grade classification

Figure 1 ***Algorithm to identify regions of high-grade prostate cancer.*** (A) In the gland segmentation step, a Resnet-Unet architecture identifies the epithelium in the input tile, while another identical architecture segments the gland boundaries. A simple image subtraction, followed by a connected component analysis and a region-growing algorithm within the epithelium, reveals the individual glands. (B) The identification of cancer glands is the second step. Each identified gland is divided into 32x32 patches that a Resnet encoder processes separately to find features in the latent space. The feature-wise histogram function aggregates the information from the patches into histograms that a few fully connected layers use to provide the cancer/benign binary classification. (C) The grading of cancerous glands is performed by a classifier that processes single 64x64 patches from a gland, providing the probabilities that the patch belongs to a low or high grade cancer gland (p_l and p_h, respectively). Summing probabilities over all patches in a gland leads to the low/high grade classification for the whole gland. (D) The feature-wise histogram function takes one feature in the latent space as its input and constructs a histogram of the feature values along all patches in a gland. Each bin is represented by the same continuous function, shifted and centered at the bin's mean value.

- Algorithm for Gleason grading in VA system: We are meeting weekly with 2 computer science graduate students at U of Utah and CWRU to identify the most accurate algorithm for Gleason grading. The pipeline of the algorithm is illustrated in Figure 1

B. Major Task 2: Generating metastatic risk (MR) scores **(n = 400 cases, 200 M0 and 200 M1)**

1. Major activities: data procurement in major task 1
2. Specific objectives: training on high-grade cancer regions, cross-validation approach to determine the accuracy of the algorithm
3. Significant result or key outcomes (major findings, developments, conclusions): none
4. Other achievements: none

C. Major Task 3: Applying MR-algorithms from Major Task 2 to cases in the M0-MP cohort **(n = 400 cases, 200 M0 and 200 M0-P)**

5. Major activities: data procurement in major task 1
6. Specific objectives: testing performance of algorithm developed in Major Task 2 on cases in Major Task 3. If performance insufficient, we will develop a new algorithm using state-of-the-art approaches.
7. Significant result or key outcomes (major findings, developments, conclusions): developing frameworks for algorithms that are using cutting edge AI models: transformers, self-supervised and semi-supervised modules.
8. Other achievements: none

What opportunities for training and professional development has the project provided?

If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased

knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

The award supports two graduate students, one at U of Utah and the other at CWRU. Dr. Knudsen meets with the graduate students once a week to oversee the development, optimization and deployment of the algorithm for detection of high-grade prostate cancer.

Dr. Knudsen (Professor of Pathology) is a member of the thesis committee of Alessandro Ferrero, a graduate student in Computer Science at the U of Utah. I this role, Dr. Knudsen meets with Alessandro twice a week for one hour each, one of these meetings also includes, Dr. Whitaker, Professor of Computer Science at U of Utah. The goal is to educate a new generation of scientists who work in a multi-disciplinary team on development of pathology image analysis.

How were the results disseminated to communities of interest?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

Nothing to Report

What do you plan to do during the next reporting period to accomplish the goals?
If this is the final report, state "Nothing to Report."

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

In the next year, we plan to accomplish the following goals:

1. Identify and deploy the best performing algorithm for segmentation of high-grade prostate cancer in VA cases
2. Train and validate an A.I. model that outputs metastatic risk scores (MR-scores) for cases in the M0 + M1 cohort (400 cases) for Major Task 2
3. Train and validate an A.I. model that outputs MR-scores for cases in the M0 + M0-P cohort (400 cases) for Major Task 3
4. Determine whether the MR-score constitutes an independent variable for predicting metastasis in men diagnosed with high grade prostate cancer at the VA.

4. **IMPACT:** D*escribe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:*

What was the impact on the development of the principal discipline(s) of the project?

The project impacts on two disciplines: pathology and computer vision

The impact in pathology is that the proposed algorithm will allow urologists to have more data at hand to show to patients when they discuss treatment decisions. The model will allow to identify those patients who would benefit from aggressive treatment upfront and spare other patients with a low risk of metastasis the side effects of treatment. For patients diagnosed with high grade prostate cancer the tools to evaluate the risk of metastases are inaccurate and expensive. Pathology image analysis for MR-score can be made available to every man diagnosed with prostate cancer, since it only requires an image from the needle biopsy slide. The algorithm is a software that can potentially be more affordable than expensive genomic testing.

The impact of the project in the Computer Vision field involves novel methods and approaches that we developed as a part of the project. The main approach is a novel way to visualize the features that the computer learns to render the metastatic risk score. Knowing which pixel are read by the computer will allow us to understand better what the computer learns and will remove the black box of the A.I. algorithm. An algorithm that can be explained will have a greater chance of adaption and acceptance for clinical application.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

There is an impact on disciplines of urology and oncology in the more distant future. The MR-score may improve patient management decisions related to precision oncology. It could also enhance the point-of-care decision making process for patients traveling from far away to VA Hospitals for treatment.

What was the impact on technology transfer?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

Nothing to Report

What was the impact on society beyond science and technology?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

Nothing to report

5. **CHANGES/PROBLEMS:** *The PD/PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:*

   The overall scope of work in the grant application has not changed. However, there have been changes to the SOW in terms of the site where the tasks are accomplished because of unforeseeable changes in personnel. Key personnel in the grant left academia altogether, or moved to a different institution. There also have been changes in the allocation of the funding. Tasks and funding originally assigned to Case Western Reserve University (CWRU) site moved to the U of Utah site. Below is a list of specific changes:

   1. Dr. Sarah Markt, co-principle investigator and site PI at CWRU left CWRU and will no longer be able to serve as a site PI of the award.

   2. Dr. Anant Madabhushi left CWRU and was replaced by Dr. Satish Viswanath, who also functions as the site PI. The CWRU personnel consists of Dr. Satish Viswanath and one graduate student at CWRU.

   3. Dr. Brigid Wilson at Cleveland VA transferred statistical analysis tasks to Dr. Benjamin Brintz and Dr. Jian Ying at University of Utah.

   The budget was adjusted accordingly to above changes in personnel.

   Changes in approach and reasons for change
   Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.

   There are no changes in the approach

   Actual or anticipated problems or delays and actions or plans to resolve them
   Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

   1. **Delay retrieving cases from Pathology.** There is a delay retrieving glass slides from the pathology archive because of personnel issues in pathology. We are working with the pathology chair at the SLC VA to resolve personnel issues. We are also setting up the accounts to reimburse Tampa VA for slide retrieval.
   2. **Algorithm development.** A.I. algorithms require lots of data. This results in a problem because we may need to have images from all the cases before we can begin the training of the MR models. To overcome this problem, we are working on semi-supervised and self-supervised

methods and domain adaptation methods that vastly decrease the need for cases and labeled data. These methods will allow us to begin with algorithm testing as soon as we have 100 cases per group.

Changes that had a significant impact on expenditures
Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

Nothing to report

Significant changes in use or care of vertebrate animals

No vertebrate animals are used in the project.

Significant changes in use of biohazards and/or select agents

No biohazards or select agents are used in the project.

6. PRODUCTS: *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report.*

Publications, conference papers, and presentations
Report only the major publication(s) resulting from the work under this award.

Ferrero, A., Knudsen, B., Sirohi, D., Whitaker, R. (2022). A Pathologist-Informed Workflow for Classification of Prostate Glands in Histopathology. In: Huo, Y., Millis, B.A., Zhou, Y., Wang, X., Harrison, A.P., Xu, Z. (eds) Medical Optical Imaging and Virtual Microscopy Image Analysis. MOVI 2022. Lecture Notes in Computer Science, vol 13578. Springer, Cham. https://doi.org/10.1007/978-3-031-16961-8_6

Please see attached manuscript

Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

Books or other non-periodical, one-time publications. *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

Nothing to Report

Other publications, conference papers and presentations. *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.*

Dr. Knudsen included the results from this project in the following seminars, acknowledging support by DoD:

1. *Machine learning applied to pathology slides: prostate cancer aggression through the eye of a computer*. University of Arizona CBIO Seminar
2. *Unraveling Cell Differentiation and Chromosomal Instability through analysis of tissue architecture by machine learning and A.I.* University of California Irvine Cancer Center
3. *PathOMICS: next generation pathology for cancer research.* AACR, New Orleans 2022
4. *Machine Learning and AI Approaches for Quantitative Prostate Histopathology*. Coffee-Holden Prostate Cancer Academy, Los Angeles, CA
5. *PathOMICS: on the crossroad of quantitative pathology for cancer research*. University of Arizona Cancer Center, Tuscon AZ
6. *Challenges and Rewards of AI Software Applications in Pathology*. Park City, Utah
7. *Computational Pathology – update on opportunities for clinical implementation.* Research in progress, Department of Pathology, University of Utah
8. *Computational Pathology Solutions for Measuring Immune Cell Biomarkers*. PRIMP Department of Pathology, University of Utah
9. *Image analysis at the crossroads of digital and molecular pathology*. Pathology Informatics Summit, Pittsburg, PA
10. *Pathologist involvement in deep learning models to improve clinical adoption.* Pathology Informatics Summit, Pittsburg, PA

11. *Clinical Opportunities and Challenges of Digital and Computational Pathology.* 8th Digital Pathology and A.I. Congress: USA

Website(s) or other Internet site(s)
List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Nothing to Report

Technologies or techniques
Identify technologies or techniques that resulted from the research activities. Describe the technologies or techniques were shared.

See Figure 1 for a new modeling approach

Inventions, patent applications, and/or licenses
Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

Nothing to Report

Other Products
Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:

- *data or databases:* collection of pathology images with annotations, all cases entered into VA MAPP biorepository for use by VA investigators (VA cases) or general access under an material transfer agreement (Utah cases)
- *software:* software is deposited in GitHub
- *models: the HistoEm model that we developed in figure 1 in the MICCAI workshop proceedings (see attached manuscript)*

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?
Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change".

Name: *Beatrice Knudsen, M.D., Ph.D.*
Project Role: *PI*
Researcher Identifier (e.g. ORCID ID): 0000-0002-7589-7591
Nearest person month worked: 1
Contribution to Project: *Dr. Knudsen worked on IRB protocol approval and on a data sharing agreement with the VA. Dr. Knudsen meets twice a week with Alessandro Ferrero and Yufei Zhou graduate students to guide the development of the prostate cancer grading algorithm. Dr. Knudsen meets once a week with Dr. Ross Whitaker for his advice on machine learning models.*

Name: *Isla Garraway, M.D.-Ph.D.*
Project Role: *PI*
Nearest person months worked: *1*
Contribution to project: *Dr. Garraway oversees the identification and retrieval of cases at the VA. She meets regularly with Dr. Julie Lynch (co-I) to refine the NLP algorithm for detection of cases with metastatic prostate cancer. Dr. Garraway's team conducts a chart review to unequivocally assign cases to metastatic vs. non-metastatic groups. She completed the IRB and regulatory paperwork for the project at UCAL/Greater Los Angeles VA. In addition, Dr. Garraway joins bi-weekly conference calls with Dr. Markt and Dr. Knudsen to discuss progress and problems.*

Name: *Alessandro Ferrero, MS*
Project Role: *Graduate student in Computer Vision, U of Utah*
Nearest person months worked: *4*
Contribution to project: *Mr. Ferrero developed a new algorithm to find and grade prostate cancer. In contrast to other algorithms, this algorithm allows to "see" exactly what features the model learns in the image.*

Name: *Benjamin Brintz, Ph.D.*
Project Role: *Statistician*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *2*

Contribution to Project: *Dr. Brintz has WOC status at the VA and direct access to VA databases. He is working on a protocol to enroll cases into cohorts based on matching and enrolment criteria. He also conceives the statistical analysis methods for the project as it involves Bayesian statistics.*

Name: *Jian Ying, MS*
Project Role: *Statistician*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *2*

Contribution to Project: *Mr. Ying has WOC status at the VA and more than a decade of expertise working on analysis of VA data. He is also*

knowledgeable of methods involving probabilistic data from machine learning models. He works closely with Dr. Brintz.

Name: *Tao He*
Project Role: *Data manager, VA*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *1*

Contribution to Project: *Tao He has WOC status at the VA. He generates data views and data tables using data from VINCI within the VA system.*

Name: *Joanne Xin*
Project Role: *Project coordinator, VA*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *1*

Contribution to Project: *Ms. Xin is coordinating the case requests and receivals from the multiple VA sites involved with the project.*

Name: *Claudette Wong*
Project Role: *Histotechnologist, VA*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *2*

Contribution to Project: *Ms. Wong is generating H&E stained tissue sections on glass slides for cases that are sending blocks. She also repairs and restains slides that fail Q/C for scanning.*

Name: *Ananta Wadhwa*
Project Role: *Project manager*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*

Contribution to Project: *Ms. Wadhwa works directly with case identification, data procurement, organization, entering cases into STARLIMS and slide scanning.*

Name: *Yufei Zhou*
Project Role: *Graduate Student CWRU in Computer Vision, CWRU*
Researcher Identifier (e.g. ORCID ID):
Nearest person month worked: *3*

Contribution to Project: *Ms. Smith has performed work in the area of combined error-control and constrained coding.*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Nothing to Report

What other organizations were involved as partners?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.

Provide the following information for each partnership:
Organization Name: UCLA, Case Western Reserve University
Location of Organization: US
Partner's contribution to the project (identify one or more)
- *Collaboration (e.g., partner's staff work with project staff on the project);*

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: *For collaborative awards, independent reports are required from BOTH the Initiating Principal Investigator (PI) and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to* ***https://ebrap.org/eBRAP/public/index.htm*** *for each unique award.*

Not applicable

QUAD CHARTS: *If applicable, the Quad Chart (available on https://www.usamraa.army.mil/Pages/Resources.aspx) should be updated and submitted with attachments.*

Not applicable

9. APPENDICES: *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*

A Pathologist-Informed Workflow for Classification of Prostate Glands in Histopathology

Alessandro Ferrero[1,*], Beatrice Knudsen[2,†], Deepika Sirohi[2,‡], and Ross Whitaker[1,**]

[1] Scientific Computing and Imaging Institute, University of Utah
72 S Central Campus Drive, Room 3750, Salt Lake City, UT 84112
[2] University of Utah
201 President Circle, Salt Lake City, UT 84112
*alessandro.ferrero@utah.edu,†Beatrice.Knudsen@path.utah.edu,
‡Deepika.Sirohi@hsc.utah.edu,**whitaker@cs.utah.edu

Abstract. Pathologists diagnose and grade prostate cancer by examining tissue from needle biopsies on glass slides. The cancer's severity and risk of metastasis are determined by the Gleason grade, a score based on the organization and morphology of prostate cancer glands. For diagnostic work-up, pathologists first locate glands in the whole biopsy core, and—if they detect cancer—they assign a Gleason grade. This time-consuming process is subject to errors and significant inter-observer variability, despite strict diagnostic criteria. This paper proposes an automated workflow that follows pathologists' *modus operandi*, isolating and classifying multi-scale patches of individual glands in whole slide images (WSI) of biopsy tissues using distinct steps: (1) two fully convolutional networks segment epithelium versus stroma and gland boundaries, respectively; (2) a classifier network separates benign from cancer glands at high magnification; and (3) an additional classifier predicts the grade of each cancer gland at low magnification. Altogether, this process provides a gland-specific approach for prostate cancer grading that we compare against other machine-learning-based grading methods.

Keywords: Prostate cancer · microscopy imaging · segmentation · classification.

1 Introduction

Prostate cancer is the second most common cause of cancer death in men over 65 in the Unites States. A reliable diagnosis of prostate cancer can only be accomplished via a prostate needle biopsy. Pathologists examine the extracted tissue samples through a microscope and assign Gleason grades to cancerous regions as an indicator of cancer severity.

The main rationale for Gleason grading is to predict the risk of cancer progression and metastasis that informs treatment decisions. The Gleason grading

system encompasses four grades: Gleason grades 2 and 3 are considered low grade and almost never lead to metastatic progression, while Gleason grades 4 and 5 are high grade and carry a risk of metastatic spread. Within the normal prostate tissue, cells organize in tube-like structures called glands. Pathologists use several morphological features to distinguish between cancerous and benign glands. Non-cancerous (i.e., benign) glands consist of basal and luminal cell layers that make up the wall of the tube. The inside of the tube is referred to as the lumen. In contrast, cancerous glands typically loose the basal cell layer, while cancer cell nuclei enlarge and display prominent nucleoli. In addition, the cancerous gland's luminal edge is straight compared to the undulated edge of benign glands.

While cancer diagnosis relies on the cells' organization and appearance, the Gleason grading scheme uses the growth pattern and structural complexity of glands to score the disease's severity. Cancerous glands with a single lumen are classified as low-grade cancer; glands within glands with multiple lumina, or glands that have lost the ability to form a lumen, are high-grade glands.

Building on deep learning successes in image classification and segmentation, researchers all over the world have turned to neural networks to develop Gleason grading algorithms. Fully convolutional networks [16], in particular, have proven useful in a variety of medical image analysis settings. Typically, a network is trained to recognize and classify structures of interest in the input image, *producing pixel-wise probability maps*, one per class, with each pixel assigned to the class with the highest probability. For instance, Silva-Rodríguez et al. [21] show that segmenting tumor areas through a neural network achieves better results than traditional algorithms, such as [8] and [2]. The U-net ([20]) is a special form of convolutional-neural-network architecture designed for image segmentation, and many Gleason grading methods such as [3, 19, 13, 18] rely on variants of the U-net to process patches of a whole slide image (WSI) in order to produce a pixel-wise Gleason grade classification. Avinash et al. [15] designed their *Carcino-net's* architecture to include a pyramid pooling module [11] that employs different size convolutional kernels. They show their algorithm's high accuracy on low resolution images that include large cancer areas. However, this approach does not explicitly account for the gland-level patterns that define the pathology, and, as we will show in later sections, Carcino-net may arrive at incorrect conclusions on the gland's grade, even after summarizing pixel-level classification results. Other studies employ region-based convolutional neural networks (RCNNs) [9, 10] to first identify bounding boxes around areas of interest in a prostate biopsy, and then segment and classify the epithelium within the boxes. The method in [14] demonstrates that these RCNNs can identify gland clusters, but struggles when glands with different grades are packed within cancer regions.

This paper aims to accurately reproduce the pathologist's grading process, breaking the gland classification problem into three sequential tasks: the segmentation of single glands, the identification of malignant glands based on cellular structure, and the classification of glands into low- and high-grade cancer based on the complexity of glandular morphology. In particular, the cancer identification step employs a novel set-based neural network that processes large collec-

tions of image patches, summarizing the information into histograms, to distinguish between benign and cancer glands. This *Histogram-based (HB)* workflow provides high gland segmentation accuracy with limited training data. It also allows clinicians and engineers to examine the results at every step of the analysis process. A self-supervised strategy [1, 5, 22] utilizes nuclear-staining properties to allow better generalization.

2 Data

The *training dataset*, described in [14], encompasses more than $40,000$ glands, roughly equally distributed among the three classes of interest. The $2,200$ tiles, of 1200×1200 pixels, that contain the glands were extracted from whole slide images (WSI) at magnification 20X, with a pixel size of $0.5\mu m \times 0.5\mu m$. Several pathologists hand-annotated polygons and assigned a label to the gland outline, marking benign glands, and low-grade (GG3) or high-grade (GG4, GG5) cancer glands. The stroma (ST) between glands is considered background. Through the same process, $10,000$ additional glands from the same forty-one patients were gathered and labeled to form the *Internal test set* (537 tiles in total). Annotations mostly corresponded to glands, but clusters of small glands were often included in one outline, making an accurate segmentation difficult to learn. Therefore, 6100 polygons from the training set were later refined to separate all glands.

Fully testing the performance of ML-based histology-analysis algorithms requires generalization to data from *unseen patients*. As such, we created the *External test set* by selecting 14800 glands from WSIs in The Cancer Genome Atlas Program (TCGA). Two pathologists labeled the images from eighteen patients, initially at low resolution to identify regions of different Gleason grades. The 546 tiles extracted from these polygons were annotated a second time at high resolution, balancing the amount of tiles coming from each class.

For data augmentation, we used random rotation, flipping, and additive noise. As in [17], color augmentation is performed through histogram matching using color palettes from TCGA and PANDA [4] datasets as target color ranges.

3 Methods

3.1 *HB*-workflow

The paper's workflow consists of three sequential stages: gland segmentation, cancer gland detection, and cancer grading.

Gland segmentation and processing. The close proximity of glands within the stroma presents a challenge in separating individual glands. To improve the segmentation, we propose a process that (1) performs epithelium (vs. stroma) segmentation, (2) finds the boundary of glands, (3) identifies the gland lumen

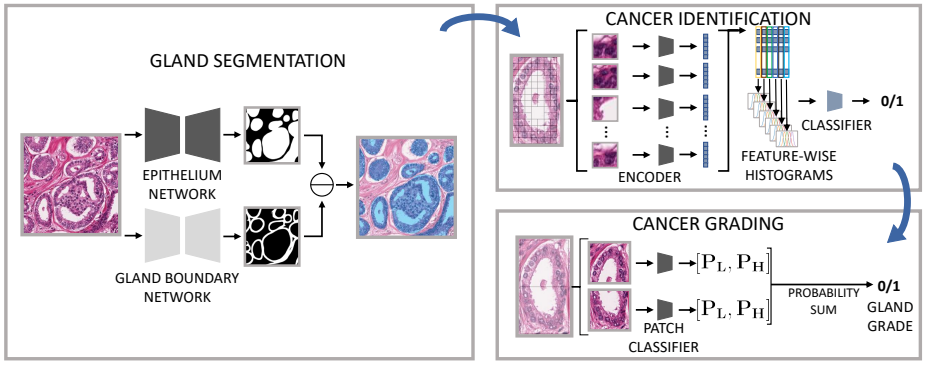

Fig. 1. *HB*-workflow. After segmentation, each gland is divided and processed in 32x32 pixel tiles by the cancer identification encoder. Features (n=128) from each gland are placed into 128 histogram bins, which are used to classify between benign and cancerous glands. Next, 64x64 cancer gland pixel tiles are used to distinguish low- and high-grade cancer tiles (P_L, P_H). If a gland spans over multiple tiles, the sum of prediction results over all the tiles, normalized to the gland area, determines the gland's predicted grade.

to form individual connected components, and (4) expands the connected components to the stromal boundaries to identify entire, distinct glands. Two identically structured U-net-like architectures perform the first two steps: one network recognizes the epithelium, and the second network finds boundaries around glands in input tiles. As shown by [6], short connections between layers increase the prediction accuracy when processing medical images; therefore we employ Resnet blocks [12] to capture relevant features. During the training phase, random 256×256 patches are extracted from the training tiles and the networks learn to minimize the cross-entropy loss between their predictions and the ground truth. The initial learning rate, set at 10^{-4}, decays every 10 epochs until the training stops at $2,000$ epochs. Pixel-wise subtraction of the epithelium and the gland boundary reveals the glands as individual, connected components. After removing small components, a region-growing algorithm ensures that gland instances include the entire epithelium.

Table 1. $mAP_{[0.5,0.9]}$ on the gland segmentation

Internal test set	External test set (TCGA)
0.67 ± 0.02	0.77 ± 0.03

Cancer gland identification. Once the individual glands have been segmented, the next stage evaluates the fine-scale cell structure to separate benign and cancerous glands. Since glands may vary widely in size and shape, we design a neural network that accepts sets of image patches from each gland, and outputs

a probability of cancer for the entire gland. To construct an appropriate set for analysis, each gland component is divided into 32×32 overlapping patches, each containing only a few luminal cells that span the thickness of the epithelium between the stroma and the lumen. The network is designed to extract useful cell features from each patch and properly aggregate that information, regardless of the gland size, helping the classifier output a probability of cancer.

Table 2. Results for CANCER IDENTIFICATION.

Internal test set

	Pixel-wise			**Gland-wise**		
	F1	**Sensitivity**	**Specificity**	**F1**	**Sensitivity**	**Specificity**
ST	0.95 ± 0.01	0.96 ± 0.01	0.92 ± 0.01	N/A	N/A	N/A
BN	0.89 ± 0.01	0.91 ± 0.02	0.97 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.02
CN	0.89 ± 0.01	0.86 ± 0.01	0.98 ± 0.01	0.96 ± 0.01	0.93 ± 0.02	0.99 ± 0.01

External test set

	Pixel-wise			**Gland-wise**		
	F1	**Sensitivity**	**Specificity**	**F1**	**Sensitivity**	**Specificity**
ST	0.97 ± 0.01	0.94 ± 0.01	0.98 ± 0.01	N/A	N/A	N/A
BN	0.74 ± 0.04	0.80 ± 0.06	0.93 ± 0.03	0.76 ± 0.04	0.81 ± 0.03	0.89 ± 0.06
CN	0.89 ± 0.01	0.88 ± 0.04	0.94 ± 0.01	0.96 ± 0.01	0.88 ± 0.04	0.83 ± 0.06

Inspired by the SetGAN discriminator's design in [7], the proposed NN architecture includes three modules: (1) an encoder that processes patches individually and generates a 128-dimensional feature vector; (2) a surrogate histogram function that summarizes each feature along all patches from a gland into a histogram with k bins (obtaining one histogram per feature), and (3) fully connected layers that use the resulting 128 histograms to output the gland classification. In this work, k is set to 5: empirically, fewer that 5 bins lead to an insufficiently descriptive latent representation, while more than 5 bins do not provide much additional information. This architecture has two advantages over a CNN: while the aggregating histogram function provides a rich representation of the whole gland, regardless of its size, the permutation invariance peculiar to the design (i.e., the patch order does not affect the classification) allows the network to study small cell groups, regardless of their location within the gland.

Cancer gland grading. The final step consists of classifying cancer glands into high or low grades. An analysis of the morphology of the entire gland, including its lumen, is necessary for this task. The above approach, developed for cancer detection, is not expected to work for cancer grading, since the unordered collection of small patches does not capture the complex morphology of high-grade cancer glands. Furthermore, Ma et al. [18] show that analyzing glands at a lower magnification yields better results when assigning Gleason grades. In order to format each malignant gland from the previous stage into a more appropriate

set of NN inputs, we reduce the magnification to 10X via a down sampling by a factor of 2 and then use 64×64 patches. A Resnet architecture learns to assign each cancer patch as high- and low-grade class. While most glands can be contained in a single patch, glands that span multiple patches are graded based on a majority vote.

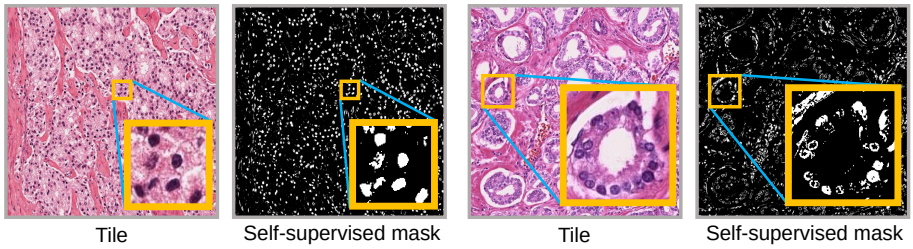

Fig. 2. Masks generated by thresholding the 10% darkest pixels.

3.2 Self-supervised strategy

To further improve generalization, we employ a self-supervised strategy throughout the entire pipeline. When prostate tissue is stained with hematoxylin and eosin (H&E), cell nuclei acquire a purple/blue hue, usually the darkest color in prostate WSIs. The self-supervised task labels are obtained by thresholding the 10% darkest training tile pixels (see Figure 2). The resulting mask highlights cell nuclei. All encoders train in association with a decoder that learns this coarse nuclear segmentation, encouraging encoders to learn nuclear features useful for segmenting and classifying glands. Altogether, our self-supervised approach is motivated by pathologist-defined diagnostic cues related to nuclear features.

4 Results

To quantitatively evaluate the model, the *HB*-workflow trained end-to-end ten times, using random validation sets of 2400 glands from the *External test set* and testing on the remaining samples. F1 scores, sensitivity and specificity are calculated per class (stroma - ST, benign - BN, low-grade - LG, high-grade - HG) to evaluate the pixel-wise and gland-wise performance. The *HB*-workflow's pixel-level data were generated using the gland-level classification labels. Gland-wise scores are obtained by majority vote of the corresponding pixels in the prediction mask, with the final score weight proportional to the gland's size.

Table 1 shows the mean average precision (mAP) between the manually segmented glands and the predicted glands. Although the *HB-workflow* tends to slightly oversegment glands (especially when the high-grade cancer consists of

Table 3. F1 scores on the *Internal test set*, similar to the training set.

	Pixel-wise F1 scores			**Gland-wise F1 scores**		
Class	U-net	Carcino-net	3-stage	U-net	Carcino-net	3-stage
ST	0.92 ± 0.01	$\mathbf{0.95 \pm 0.01}$	0.95 ± 0.01	N/A	N/A	N/A
BN	0.84 ± 0.02	$\mathbf{0.93 \pm 0.01}$	0.89 ± 0.01	0.95 ± 0.01	$\mathbf{0.97 \pm 0.01}$	0.94 ± 0.01
LG	0.52 ± 0.02	$\mathbf{0.79 \pm 0.01}$	0.73 ± 0.01	0.71 ± 0.01	$\mathbf{0.91 \pm 0.01}$	0.86 ± 0.02
HG	0.68 ± 0.02	$\mathbf{0.89 \pm 0.02}$	0.86 ± 0.01	0.80 ± 0.01	$\mathbf{0.94 \pm 0.01}$	0.93 ± 0.01
	Pixel-wise Sensitivity			**Gland-wise Sensitivity**		
Class	U-net	Carcino-net	3-stage	U-net	Carcino-net	3-stage
ST	0.93 ± 0.01	$\mathbf{0.94 \pm 0.01}$	0.96 ± 0.01	N/A	N/A	N/A
BN	0.80 ± 0.02	$\mathbf{0.94 \pm 0.01}$	0.91 ± 0.01	0.94 ± 0.01	$\mathbf{0.97 \pm 0.01}$	0.95 ± 0.02
LG	0.81 ± 0.02	$\mathbf{0.76 \pm 0.01}$	0.75 ± 0.03	0.95 ± 0.01	$\mathbf{0.86 \pm 0.01}$	0.84 ± 0.02
HG	0.53 ± 0.02	$\mathbf{0.91 \pm 0.01}$	0.81 ± 0.02	0.67 ± 0.01	$\mathbf{0.96 \pm 0.01}$	0.90 ± 0.01
	Pixel-wise Specificity			**Gland-wise Specificity**		
ST	0.89 ± 0.01	$\mathbf{0.93 \pm 0.01}$	0.92 ± 0.01	N/A	N/A	N/A
BN	0.97 ± 0.01	$\mathbf{0.97 \pm 0.01}$	0.95 ± 0.02	0.98 ± 0.01	$\mathbf{0.98 \pm 0.01}$	0.95 ± 0.02
LG	0.90 ± 0.01	$\mathbf{0.98 \pm 0.01}$	0.97 ± 0.01	0.86 ± 0.01	$\mathbf{0.99 \pm 0.01}$	0.98 ± 0.01
HG	0.98 ± 0.01	$\mathbf{0.96 \pm 0.01}$	0.98 ± 0.01	0.99 ± 0.01	$\mathbf{0.98 \pm 0.01}$	0.98 ± 0.01

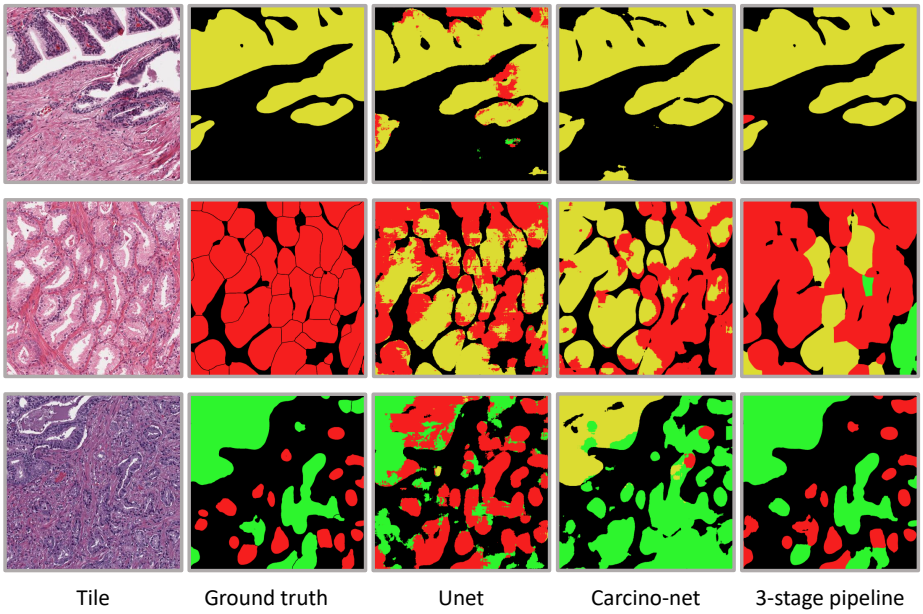

Fig. 3. Classification results. H&E stained image tiles with manual ground truth labels are compared to U-net and Carcino-net pixel-level classification results and gland level classification from the *HB*-workflow (yellow - BN, red - LG, green - HG).

free cells), the mAP values are high. Experiments showed that gland oversegmentation is less detrimental to the final workflow's output than undersegmentation, where the binary classifiers tend to give class probabilities closer to 0.5 when glands of different grades appear in the same segmented instance (i.e., the classifiers are less certain about their predictions and make more mistakes). The slight difference between the datasets' scores in Table 1 is mainly due to the gland conglomerates present in the *Internal test set*'s labels that decrease the mAP value. Despite this, the mAP values remain high.

Evaluations in Table 2 show that the cancer-identifying classifier distinguishes between benign and cancer glands with high accuracy, thanks to the rich representation provided by the learned histograms. Tables 3 and 4 compare U-net and Carcino-Net, trained as described in [15], to the *HB*-workflow.

Figure 3 confirms the high scores that the three architectures achieve in segmenting stroma. Panels displaying the U-net and Carcino-net pixel-level classification results demonstrate misclassified pixels within glands. These pixel labels are corrected by applying the majority voting scheme, resulting in a single label for each gland. The quantitative results in Tables 3 and 4 show that Carcino-Net performs better on the *Internal test set* (which is similar to the training data) than on the *External test set*, meaning that Carcino-Net is prone to overfitting. In contrast, the *HB*-workflow achieves higher F1 scores, pixel-wise and gland-wise, for all classes on the unseen data.

Table 4. Results on the *External test set (TCGA)*.

	Pixel-wise F1 scores			**Gland-wise F1 scores**		
Class	U-net	Carcino-net	*HB workflow*	U-net	Carcino-net	*HB workflow*
ST	0.89 ± 0.01	0.91 ± 0.01	$\mathbf{0.97 \pm 0.01}$	N/A	N/A	N/A
BN	0.58 ± 0.02	0.72 ± 0.11	$\mathbf{0.74 \pm 0.04}$	0.69 ± 0.05	0.74 ± 0.06	$\mathbf{0.76 \pm 0.04}$
LG	0.58 ± 0.04	0.66 ± 0.03	$\mathbf{0.69 \pm 0.03}$	0.67 ± 0.07	0.68 ± 0.08	$\mathbf{0.71 \pm 0.04}$
HG	0.39 ± 0.14	0.55 ± 0.06	$\mathbf{0.69 \pm 0.10}$	0.54 ± 0.18	0.68 ± 0.09	$\mathbf{0.72 \pm 0.01}$
	Pixel-wise Sensitivity			**Gland-wise Sensitivity**		
Class	U-net	Carcino-net	*HB workflow*	U-net	Carcino-net	*HB workflow*
ST	0.85 ± 0.01	0.87 ± 0.01	$\mathbf{0.94 \pm 0.01}$	N/A	N/A	N/A
BN	0.52 ± 0.02	$\mathbf{0.87 \pm 0.01}$	0.80 ± 0.06	0.55 ± 0.01	$\mathbf{0.88 \pm 0.01}$	0.81 ± 0.03
LG	$\mathbf{0.83 \pm 0.03}$	0.62 ± 0.01	0.73 ± 0.02	$\mathbf{0.90 \pm 0.01}$	0.64 ± 0.01	0.73 ± 0.02
HG	0.36 ± 0.14	0.58 ± 0.01	$\mathbf{0.65 \pm 0.04}$	0.41 ± 0.01	0.63 ± 0.01	$\mathbf{0.66 \pm 0.01}$
	Pixel-wise Specificity			**Gland-wise Specificity**		
Class	U-net	Carcino-net	*HB workflow*	U-net	Carcino-net	*HB workflow*
ST	0.93 ± 0.01	0.93 ± 0.01	$\mathbf{0.98 \pm 0.01}$	N/A	N/A	N/A
BN	$\mathbf{0.98 \pm 0.02}$	0.89 ± 0.01	0.93 ± 0.03	$\mathbf{0.98 \pm 0.01}$	0.86 ± 0.01	0.89 ± 0.06
LG	0.80 ± 0.04	$\mathbf{0.93 \pm 0.01}$	0.90 ± 0.01	0.53 ± 0.01	$\mathbf{0.88 \pm 0.01}$	0.84 ± 0.04
HG	$\mathbf{0.94 \pm 0.05}$	0.89 ± 0.01	0.89 ± 0.01	$\mathbf{0.94 \pm 0.01}$	0.86 ± 0.01	0.86 ± 0.06

5 Conclusions

Pathologists diagnose and grade prostate cancer glands based on vastly different morphological criteria. The *HB*-workflow presented in this paper mimics the pathologist's workflow by separating gland segmentation, cancer detection and cancer grading into three separate stages. The division of tasks allows each neural network to focus on the relevant features for the task at hand. In particular, the histogram aggregation function provides a permutation invariant way to process sets of small gland patches, allowing the cancer identification network to focus on the cell morphology. The *HB*-workflow shows higher quantitative and qualitative results per class than other state-of-the-art methods. Future work includes the training of this pipeline on larger, multi-cohort datasets, and its use for identifying high grade cancer regions to cost-effectively predict cancer stage and prognosis.

Acnowledgments. We acknowledge the generous support from the Department of Defense Prostate Cancer Program Population Science Award W81XWH-21-1-0725-. We also acknowledge that we received the training data from Cedars-Sinai Hospital in Los Angeles and we thank Dr. Akadiusz Gertych for his work on establishing the tiles. The results presented here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

References

1. Abbasi Koohpayegani, S., Tejankar, A., Pirsiavash, H.: Compress: Self-supervised learning by compressing representations. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 12980–12992. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/975a1c8b9aee1c48d32e13ec30be7905-Paper.pdf
2. Avenel, C., Tolf, A., Dragomir, A., Carlbom, I.B.: Glandular segmentation of prostate cancer: An illustration of how the choice of histopathological stain is one key to success for computational pathology. Frontiers in Bioengineering and Biotechnology **7** (2019). https://doi.org/10.3389/fbioe.2019.00125, https://www.frontiersin.org/article/10.3389/fbioe.2019.00125
3. Bulten, W., Bandi, P., Hoven, J., van de Loo, R., Lotz, J., Weiss, N., van der Laak, J., Ginneken, B., Hulsbergen-van de Kaa, C., Litjens, G.: Epithelium segmentation using deep learning in h&e-stained prostate specimens with immunohistochemistry as reference standard. Scientific Reports **9**, 864 (01 2019). https://doi.org/10.1038/s41598-018-37257-4
4. Bulten, W., Litjens, G., Pinckaers, H., Ström, P., Eklund, M., Kartasalo, K., Demkin, M., Dane, S.: The panda challenge: Prostate cancer grade assessment using the gleason grading system (Mar 2020). https://doi.org/10.5281/zenodo.3715938, https://doi.org/10.5281/zenodo.3715938
5. Doersch, C., Zisserman, A.: Multi-task self-supervised visual learning. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2070–2079 (2017). https://doi.org/10.1109/ICCV.2017.226

6. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C.: The importance of skip connections in biomedical image segmentation (08 2016). https://doi.org/10.1007/978-3-319-46976-8_19
7. Ferrero, A., Elhabian, S., Whitaker, R.: Setgans: Enforcing distributional accuracy in generative adversarial networks (06 2019)
8. Gavrilovic, M., Azar, J.C., Lindblad, J., Wählby, C., Bengtsson, E., Busch, C., Carlbom, I.B.: Blind color decomposition of histological images. IEEE Transactions on Medical Imaging **32**(6), 983–994 (2013). https://doi.org/10.1109/TMI.2013.2239655
9. Girshick, R.: Fast r-cnn. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1440–1448 (2015). https://doi.org/10.1109/ICCV.2015.169
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988 (2017). https://doi.org/10.1109/ICCV.2017.322
11. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(9), 1904–1916 (2015). https://doi.org/10.1109/TPAMI.2015.2389824
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
13. Li, J., Sarma, K., Ho, K.C., Gertych, A., Knudsen, B., Arnold, C.: A multi-scale u-net for semantic segmentation of histological images from radical prostatectomies. AMIA ... Annual Symposium proceedings. AMIA Symposium **2017**, 1140–1148 (04 2018)
14. Li, W., Li, J., Sarma, K.V., Ho, K.C., Shen, S., Knudsen, B.S., Gertych, A., Arnold, C.W.: Path r-cnn for prostate cancer diagnosis and gleason grading of histological images. IEEE Transactions on Medical Imaging **38**(4), 945–954 (2019). https://doi.org/10.1109/TMI.2018.2875868
15. Lokhande, A., Bonthu, S., Singhal, N.: Carcino-net: A deep learning framework for automated gleason grading of prostate biopsies. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). pp. 1380–1383 (2020). https://doi.org/10.1109/EMBC44109.2020.9176235
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
17. Ma, J.: Histogram matching augmentation for domain adaptation with application to multi-centre, multi-vendor and multi-disease cardiac image segmentation (2020)
18. Ma, Z., Li, J., Salemi, H., Arnold, C., Knudsen, B., Gertych, A., Ing, N.: Semantic segmentation for prostate cancer grading by convolutional neural networks. p. 46 (03 2018). https://doi.org/10.1117/12.2293000
19. Nagpal, K., Foote, D., Liu, Y., Chen, P.H., Wulczyn, E., Tan, F., Olson, N., Smith, J., Mohtashamian, A., Wren, J., Corrado, G., MacDonald, R., Peng, L., Amin, M., Evans, A., Sangoi, A., Mermel, C., Hipp, J., Stumpe, M.: Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. npj Digital Medicine **2**, 48 (06 2019). https://doi.org/10.1038/s41746-019-0112-2
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)

21. Silva-Rodríguez, J., Payá-Bosch, E., García Pardo, J.G., Colomer, A., Naranjo, V.: Prostate Gland Segmentation in Histology Images via Residual and Multi-resolution U-NET, pp. 1–8 (10 2020). https://doi.org/10.1007/978-3-030-62362-3_1
22. Zheng, X., Wang, Y., Wang, G., Liu, J.: Fast and robust segmentation of white blood cell images by self-supervised learning. Micron **107** (02 2018). https://doi.org/10.1016/j.micron.2018.01.010