AIRA MATRIX | Artificial Intelligence Transformed Analyses

Publications & Posters

About Us

We deliver cutting-edge artificial intelligence solutions for healthcare and life sciences applications. Our innovative products and services drive efficiency, enhance diagnostic accuracy, and provide data-driven insights from digitally scanned images.

Drug Discovery & Preclinical Research

Our comprehensive AI-driven suite for drug discovery and development, streamlines pre-analytical quality control, facilitates tissue triage, and optimizes safety and efficacy assessments, accelerating breakthrough discoveries.

Healthcare Solutions

Our healthcare portfolio offers diagnostic, prognostic, and predictive tools that facilitate risk stratification and enable personalized therapy planning, thus improving treatment outcomes.

Write to solutions@airamatrix.com for more

AIRA MATRIX | Artificial Intelligence Transformed Analyses

Healthcare Solutions

Comparison of Pathologist and Artificial Intelligence–based Grading for Prediction of Metastatic Outcomes After Radical Prostatectomy

Lia D. Oliveira, Jiayun Lu, Eric Erak, Adrianna A. Mendes, Oluwademilade Dairo, Onur Ertunc, Ibrahim Kulac, Javier A. Baena-Del Valle, Tracy Jones, Jessica L. Hicks, Stephanie Glavaris, Gunes Guner, Igor D. Vidal, Bruce J. Trock, Uttara Joshi, Chaith Kondragunta, Saikiran Bonthu, Corinne Joshu, Nitin Singhal, Angelo M. De Marzo, Tamara L. Lotan

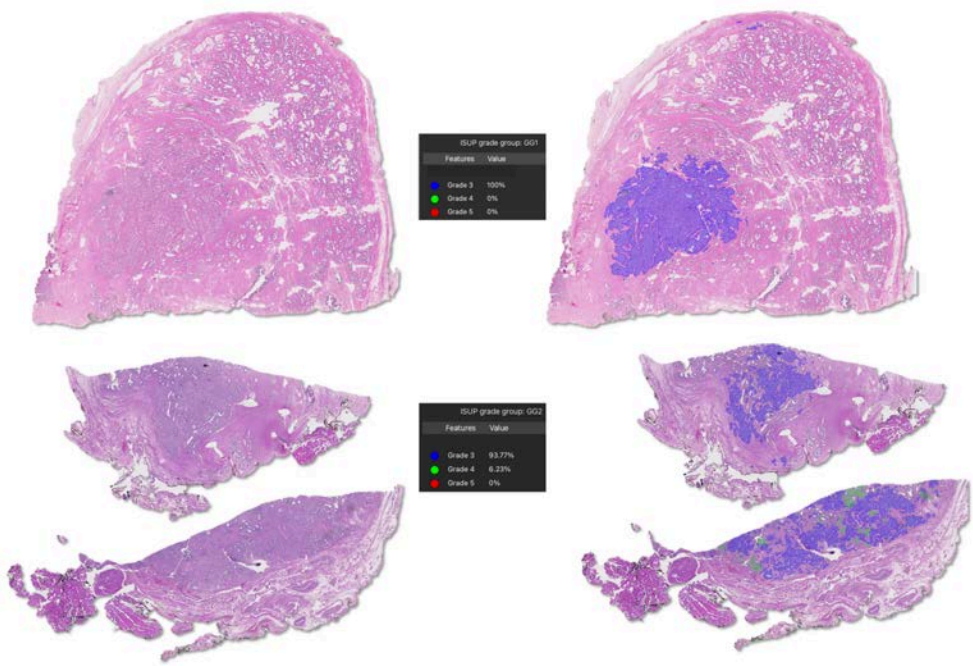
In collaboration with Johns Hopkins University School of Medicine

European Urology Oncology, Volume 0, Issue 0

Abstract

Gleason grade group (GG) is the most powerful prognostic variable in localized prostate cancer; however, interobserver variability remains a challenge. Artificial intelligence algorithms applied to histopathologic images standardize grading, but most have been tested only for agreement with pathologist GG, without assessment of performance with respect to oncologic outcomes. We compared deep learning-based and pathologist-based GGs for an association with metastatic outcomes in three surgical cohorts comprising 777 unique patients. A digitized whole slide image of the representative hematoxylin and eosin–stained slide of the dominant tumor nodule was assigned a GG by an artificial intelligence-based grading algorithm and was compared with the GG assigned by a contemporary pathologist or the original pathologist-assigned GG for the entire prostatectomy. Harrell's C-indices based on Cox models for time to metastasis were compared. In a combined analysis of all cohorts, the C-index for the artificial intelligence–assigned GG was 0.77 (95% confidence interval [CI]: 0.73–0.81), compared with 0.77 (95% CI: 0.73–0.81) for the pathologist-assigned GG. By comparison, the original pathologist-assigned GG for the entire case had a C-index of 0.78 (95% CI: 0.73–0.82).

Results



Evaluation of Artificial Intelligence-Based Gleason Grading Algorithms “in the Wild”

Khrystyna Faryna, Leslie Tessier, Juan Retamero, Saikiran Bonthu, Pranab Samanta, Nitin Singhal, Solene-Florence Kammerer-Jacquet, Camelia Radulescu, Vittorio Agosti, Alexandre Collin, Xavier Farre, Jacqueline Fontugne, Rainer Grobholz, Agnes Marije Hoogland, Katia Ramos Moreira Leite, Murat Oktay, Antonio Polonia, Paromita Roy, Paulo Guilherme Salles, Theodurus H. van der Kwast, Jolique van Ipenburg, Jeroen var der Laak, Geert Litjens

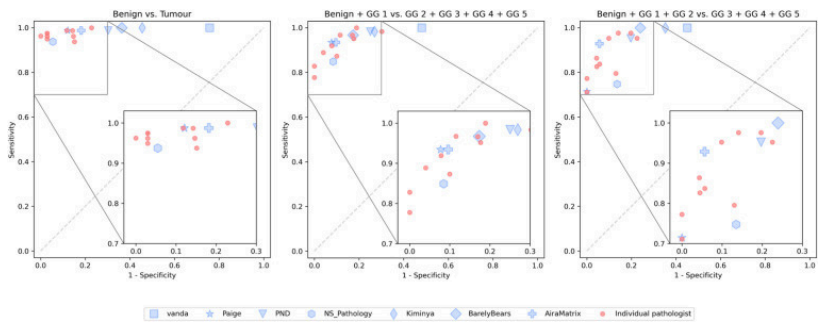
In collaboration with Radboud University Medical Center

***Modern Pathology*, Vol 37 (11) 100563**

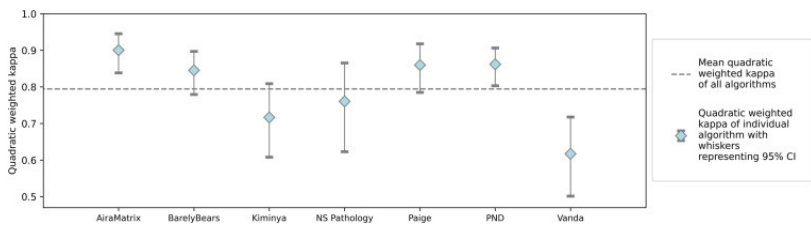
Abstract

The biopsy Gleason score is an important prognostic marker for prostate cancer patients. It is, however, subject to substantial variability among pathologists. Artificial intelligence (AI)–based algorithms employing deep learning have shown their ability to match pathologists’ performance in assigning Gleason scores, with the potential to enhance pathologists’ grading accuracy. The performance of Gleason AI algorithms in research is mostly reported on common benchmark data sets or within public challenges. In contrast, many commercial algorithms are evaluated in clinical studies, for which data are not publicly released. As commercial AI vendors typically do not publish performance on public benchmarks, comparison between research and commercial AI is difficult. The aims of this study are to evaluate and compare the performance of top-ranked public and commercial algorithms using real-world data. We curated a diverse data set of whole-slide prostate biopsy images through crowdsourcing containing images with a range of Gleason scores and from diverse sources. Predictions were obtained from 5 top-ranked public algorithms from the Prostate cANcer graDe Assessment (PANDA) challenge and 2 commercial Gleason grading algorithms. Additionally, 10 pathologists (A.C., C.R., J.v.I., K.R.M.L., P.R., P.G.S., R.G., S.F.K.J., T.v.d.K., X.F.) evaluated the data set in a reader study. Overall, the pairwise quadratic weighted kappa among pathologists ranged from 0.777 to 0.916. Both public and commercial algorithms showed high agreement with pathologists, with quadratic kappa ranging from 0.617 to 0.900. Commercial algorithms performed on par or outperformed top public algorithms.

Results



Performance of pathologists and algorithms at clinically relevant decision thresholds



Individual Agreement (QWK) of algorithms with the majority vote of pathologists

Predicting Prostate Cancer Grade Reclassification on Active Surveillance Using a Deep Learning-Based Grading Algorithm

Chien-Kuang C Ding, Zhuo Tony Su, Erik Erak, Lia DePaula Oliveira, Daniela C Salles, Yuezhou Jing, Pranab Samanta, Saikiran Bonthu, Uttara Joshi, Chaith Kondragunta, Nitin Singhal, Angelo M De Marzo, Bruce J Trock, Christian P Pavlovich, Claire M de la Calle, Tamara L Lotan

In collaboration with Johns Hopkins University School of Medicine, University of California, and University of Washington

***JNCI: Journal of the National Cancer Institute*, 2024, djae139**

Abstract

Deep learning (DL)-based algorithms to determine prostate cancer (PCa) Grade Group (GG) on biopsy slides have not been validated by comparison to clinical outcomes. We used a DL-based algorithm, AIRAProstate, to re-grade initial prostate biopsies in two independent PCa active surveillance (AS) cohorts. In a cohort initially diagnosed with GG1 PCa using only systematic biopsies (n=138), upgrading of the initial biopsy to ≥GG2 by AIRAProstate was associated with rapid or extreme grade reclassification on AS (odds ratio 3.3, p=0.04), whereas upgrading of the initial biopsy by contemporary uropathologist reviews was not associated with this outcome. In a contemporary validation cohort that underwent prostate magnetic resonance imaging before initial biopsy (n=169), upgrading of the initial biopsy (all contemporary GG1 by uropathologist grading) by AIRAProstate was associated with grade reclassification on AS (hazard ratio 1.7, p=0.03). These results demonstrate the utility of a DL-based grading algorithm in PCa risk stratification for AS.

Results

Table 1: Univariable and multivariable logistic regression for rapid or extreme prostate cancer grade reclassification in the case-control active surveillance cohort (N = 138)

	Univariable regression			Multivariable regression[a]			Multivariable regression[b]		
Baseline characteristic	OR	95% CI	*P* value	OR	95% CI	*P* value	OR	95% CI	*P* value
Uropathologist consensus GG for initial biopsy							*(Excluded)*		
≤GG1	Reference			Reference			-		
≥GG2	1.69	0.51, 5.63	.39	2.90	0.69, 12.2	.15	-		
AIRAProstate GG for initial biopsy				*(Excluded)*					
≤GG1	Reference			-			Reference		
≥GG2	5.53	2.07, 14.8	**.001**	-			3.33	1.07, 10.3	**.04**
Year of biopsy (per one year increase)	1.21	1.10, 1.32	**<0.001**	1.26	1.13, 1.40	**<0.001**	1.22	1.09, 1.35	**<0.001**
Age at biopsy (per one year increase)	0.95	0.89, 1.02	.16	0.88	0.81, 0.96	**.01**	0.90	0.82, 0.98	**.01**
Race									
Other	Reference			Reference			Reference		
Black	6.19	0.70, 54.4	.10	5.69	0.53, 60.6	.15	5.01	0.46, 54.1	.18
PSA density (per 0.1 ng/mL^2 increase)	1.25	0.84, 1.86	.27	1.39	0.86, 2.23	.18	1.35	0.84, 2.17	.21
Number of positive biopsy cores (per 1 core increase)	1.16	0.87, 1.55	.32	0.90	0.60, 1.33	.59	0.92	0.61, 1.39	.70
Maximum percent core involvement (per 10% increase)	1.30	1.11, 1.52	**.001**	1.44	1.16, 1.78	**.001**	1.41	1.13, 1.76	**.002**
Underwent mpMRI before biopsy (vs. not)	1.16	0.07, 18.9	.92	0.13	0.004, 4.32	.25	0.17	0.01, 5.39	.31

Table 2: Univariable and multivariable Cox proportional-hazards regression for prostate cancer grade reclassification to Grade Group (GG) ≥2 in the contemporary active surveillance cohort (N = 169)

	Univariable regression			Multivariable regression[a]			Multivariable regression[b]		
Baseline characteristic	HR	95% CI	*P* value	HR	95% CI	*P* value	HR	95% CI	*P* value
AIRAProstate GG for initial biopsy							*(Excluded)*		
≤GG1	Reference			Reference			-		
≥GG2	2.04	1.28, 3.26	**.003**	1.71	1.05, 2.78	**.03**	-		
Year of biopsy (per one year increase)	1.52	1.31, 1.76	**<0.001**	1.53	1.32, 1.77	**<0.001**	1.54	1.32, 1.79	**<0.001**
Age at biopsy (per one year increase)	1.03	0.99, 1.07	.17	1.02	0.98, 1.07	.30	1.03	0.99, 1.08	.18
Race									
Other	Reference			Reference			Reference		
Black	0.88	0.35, 2.21	.79	0.67	0.26, 1.70	.40	0.62	0.24, 1.58	.32
PSA density (per 0.1 ng/mL^2 increase)	1.08	0.90, 1.30	.41	1.11	0.86, 1.43	.41	1.11	0.87, 1.42	.39
Number of positive biopsy cores (per 1 core increase)	1.20	1.01, 1.43	**.04**	1.03	0.85, 1.25	.76	1.08	0.89, 1.31	.43
Maximum percent core involvement (per 10% increase)	1.11	1.03, 1.21	**.01**	1.07	0.96, 1.19	.21	1.05	0.94, 1.16	.39
Highest PI-RADS score on pre-biopsy mpMRI									
≤3	Reference			Reference			Reference		
4	1.66	0.96, 2.89	.07	1.41	0.78, 2.52	.25	1.45	0.81, 2.58	.21
5	2.02	1.0003, 4.07	**.05**	1.66	0.78, 3.53	.19	1.63	0.77, 3.45	.20

Predicting Prostate Cancer Molecular Subtype with Deep Learning on Histopathologic Images

Eric Erak, Lia DePaula Oliveira, Adrianna A. Mendes, Oluwademilade Dairo, Onur Ertunc, Ibrahim Kulac, Javier A. Baena-Del Valle, Tracy Jones, Jessica L. Hicks, Stephanie Glavaris, Gunes Guner, Igor Damasceno Vidal, Mark Markowski, Claire de la Calle, Bruce J. Trock, Avaneesh Meena, Uttara Joshi, Chaith Kondragunta, Saikiran Bonthu, Nitin Singhal, Angelo M. De Marzo, Tamara L. Lotan

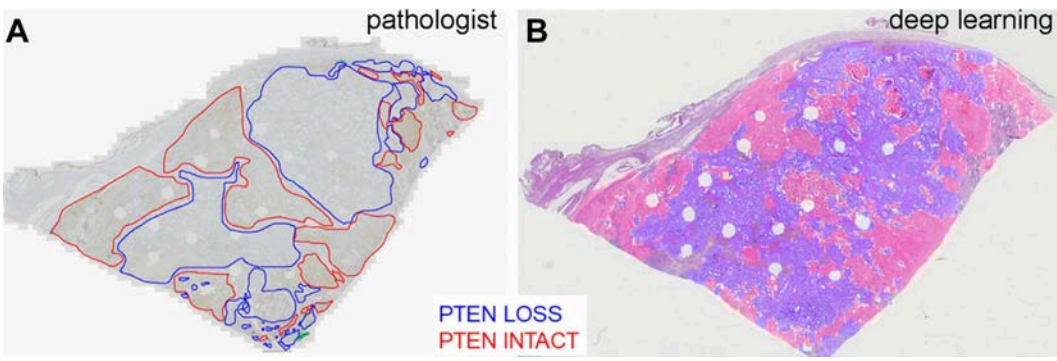
In collaboration with Johns Hopkins University School of Medicine, Suleyman Demirel University, Koç University School of Medicine, Fundacion Santa Fe de Bogota University Hospital, and University of Alabama School of Medicine

***Modern Pathology*, 2023, Vol. 36(10) 100247**

Abstract

Microscopic examination of prostate cancer has failed to reveal a reproducible association between molecular and morphologic features. However, deep-learning algorithms trained on hematoxylin and eosin (H&E)-stained whole slide images (WSI) may outperform the human eye and help to screen for clinically-relevant genomic alterations. We created deep-learning algorithms to identify prostate tumors with underlying ETS-related gene (ERG) fusions or PTEN deletions using the following 4 stages: (1) automated tumor identification, (2) feature representation learning, (3) classification, and (4) explainability map generation. A novel transformer-based hierarchical architecture was trained on a single representative WSI of the dominant tumor nodule from a radical prostatectomy (RP) cohort with known ERG/PTEN status (n = 224 and n = 205, respectively). Two distinct vision transformer-based networks were used for feature extraction, and a distinct transformer-based model was used for classification. The ERG algorithm performance was validated across 3 RP cohorts, including 64 WSI from the pretraining cohort (AUC, 0.91) and 248 and 375 WSI from 2 independent RP cohorts (AUC, 0.86 and 0.89, respectively). In addition, we tested the ERG algorithm performance in 2 needle biopsy cohorts comprised of 179 and 148 WSI (AUC, 0.78 and 0.80, respectively). Focusing on cases with homogeneous (clonal) PTEN status, PTEN algorithm performance was assessed using 50 WSI reserved from the pretraining cohort (AUC, 0.81), 201 and 337 WSI from 2 independent RP cohorts (AUC, 0.72 and 0.80, respectively), and 151 WSI from a needle biopsy cohort (AUC, 0.75). For explainability, the PTEN algorithm was also applied to 19 WSI with heterogeneous (subclonal) PTEN loss, where the percentage tumor area with predicted PTEN loss correlated with that based on immunohistochemistry (r = 0.58, P = .0097). These deep-learning algorithms to predict ERG/PTEN status prove that H&E images can be used to screen for underlying genomic alterations in prostate cancer.

Results



Correlation between pathologist annotations and deep-learning algorithm predictions for radical prostatectomy tumor regions with PTEN loss.

A Deep Learning System for Prostate Cancer Diagnosis and Grading in Whole Slide Images of Core Needle Biopsies

Nitin Singhal, Shailesh Soni, Saikiran Bonthu, Nilanjan Chattopadhyay, Pranab Samanta, Uttara Joshi, Amit Jojera, Taher Chharchhodawala, Ankur Agarwal, Mahesh Desai, and Arvind Ganpule

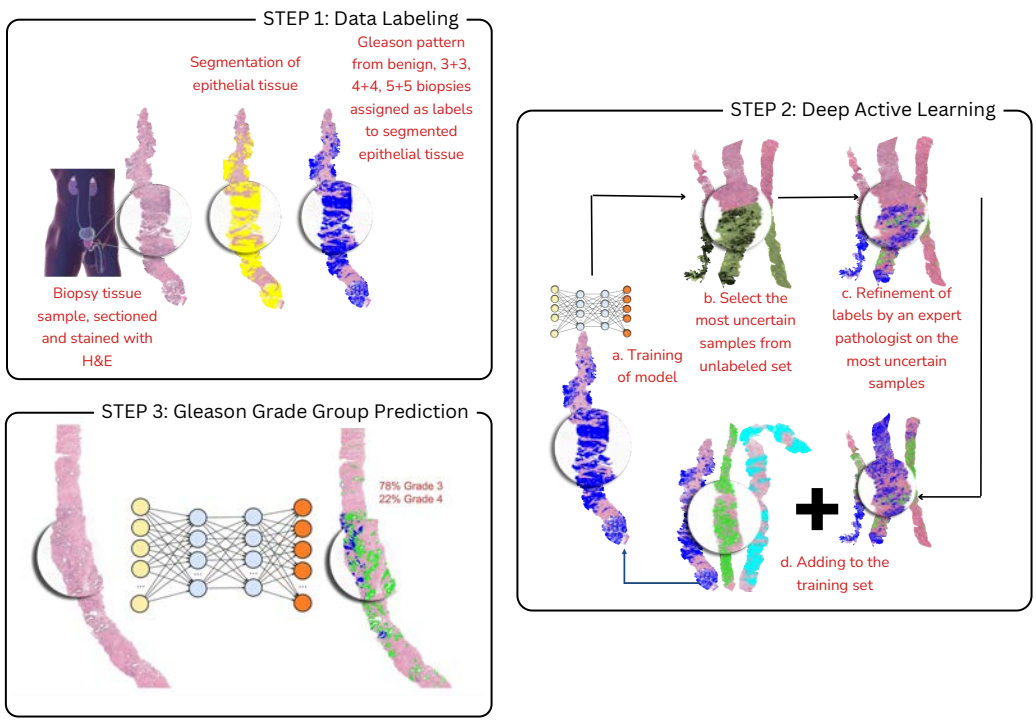
In collaboration with Muljibhai Patel Urological Hospital, Nadiad, India; Grant Medical Foundation, Ruby Hall Clinic, Pune, India; and Department of Lab Medicine, Manipal Hospital Jaipur, India

***Sci Rep* 12, 3383 (2022)**

Abstract

Gleason grading, a risk stratification method for prostate cancer, is subjective and dependent on experience and expertise of the reporting pathologist. Deep learning (DL) systems have shown promise in enhancing the objectivity and efciency of Gleason grading. However, DL networks exhibit domain shift and reduced performance on Whole Slide Images (WSI) from a source other than training data. We propose a DL approach for segmenting and grading epithelial tissue using a novel training methodology that learns domain agnostic features. In this retrospective study, we analyzed WSI from three cohorts of prostate cancer patients. 3741 core needle biopsies (CNBs) received from two centers were used for training. The κquad (quadratic-weighted kappa) and AUC were measured for grade group comparison and core-level detection accuracy, respectively. Accuracy of 89.4% and κquad of 0.92 on the internal test set of 425 CNB WSI and accuracy of 85.3% and κquad of 0.96 on an external set of 1201 images, was observed. The system showed an accuracy of 83.1% and κquad of 0.93 on 1303 WSI from the third institution (blind evaluation). Our DL system, used as an assistive tool for CNB review, can potentially improve the consistency and accuracy of grading, resulting in better patient outcomes.

Results



Active Learning-based semi-supervised deep learning system

USCAP| 2024

Comparison of Whole Slide Image-Based Deep Learning Algorithms and Genomic Classifiers for Assessing the Risk of Prostate Cancer Metastasis in Surgically Treated Patients

Presented

Genomic classifiers improve post-radical prostatectomy (RP) risk stratification compared to conventional clinical-pathologic parameters, but are tissue-destructive and expensive. In contrast, artificial intelligence algorithms utilizing diagnostic hematoxylin and eosin (H&E)-stained slides for risk stratification conserve tissue and could be made widely available at point-of-care. We compared the predictive output of a deep learning (DL)-based algorithm applied to H&E-stained whole slide images (WSI) of prostate tumors to commercial genomic classifiers such as Decipher and Prolaris in three diverse RP cohorts with follow-up for metastasis.

USCAP | 2024

Comparison of Pathologist and Deep Learning-Based Prostate Cancer Grading for Prediction of Metastatic Outcomes in Primary Prostate Cancer

Presented

Gleason grading is the most potent prognostic variable in primary prostate cancer, however inter-observer variability remains a major issue, particularly where subspeciality-trained pathologists are not available. Artificial intelligence algorithms for prostate cancer grading may improve health care equity by ensuring widespread access to standardized, high quality grading, however most algorithms have not been tested for performance with respect to oncologic outcomes. Here, we compared deep learning-based Gleason grading for prediction of metastatic outcome in three large radical prostatectomy cohorts.

USCAP| 2024

Predicting Prostate Cancer Progression with Deep Learning on Low/Intermediate Risk Needle Core Biopsies

Presented

Risk Stratification of prostate cancer based on Gleason Grade Group (GG) has been the standard of care to guide disease management. Deep Learning (DL) algorithms trained on whole slide images (WSI) to provide GG have been largely validated by comparison to pathologist-determined GG. However, few studies have assessed the performance of DL grading of low/intermediate risk needle biopsies for predicting relevant clinical end-points.

USCAP | 2024

Evaluating Leading Commercial and Academic AI-Based Gleason Grading Algorithms “In the Wild”

Presented

Public and commercial Gleason grading algorithms based on artificial intelligence (AI) have demonstrated the ability to improve grading accuracy. The evaluation of commercial algorithms in histopathology happens within clinical trials, while academic algorithms are evaluated on benchmark datasets. Commercial algorithms rarely participate in challenges or report performance on public benchmarks, while the data from clinical trials is not available for the academic community to evaluate algorithms on. This research aims to assess and compare the performance of leading public and commercial Gleason grading algorithms using a crowdsourced real-world dataset.

Pathology Visions| 2023

AIRAQc: Deep Learning for Artefact Identification and Quantification in Digital Pathology

The quality of a whole slide image (WSI) of a histopathology slide may be impacted by artefacts introduced by fixation, processing, sectioning, and/or scanning. Artefacts in WSIs can render portions or the entirety of histopathology slides unsuitable for various types of analysis, and lead laboratory staff to create additional slides or scans that are less impacted. Quality control in preanalytical workflows to detect these artefacts is time and resource-consuming when performed manually and could lead to increased costs and slowed production timelines. In this study, we present a deep learning system trained to identify and quantify frequently occurring artefacts in WSI of H&E-stained tissues.

ECDP | 2023

Deep Learning-Based Identification of Lymph Node Metastasis in Prostate Cancer

Pelvic lymph node metastasis in prostate carcinoma is associated with poor prognosis among surgical candidates. These patients may benefit from adjuvant or salvage radiation and hormonal therapy. The mainstay for definitive diagnosis of lymph node metastasis is microscopic screening of all H&E slides containing dissected lymph nodes by surgical pathologists, often in cases containing only very minute foci of cancer cells that can be readily missed. To avoid false negatives, therefore, surgical pathologists tediously perform this lymph node screening task at medium to higher power in a time-consuming process. We present a deep-learning-based tool to assist pathologists in identifying micro-metastatic foci on hematoxylin and eosin (H&E) stained digitized images.

ECDP| 2023

Assessing Risk of Prostate Cancer Metastasis by Deep Learning in Surgically-Treated Patients

Prostate cancer is among the most prevalent types of cancer in males. More than 90 percent of patients without metastasis are expected to live at least five years, whereas patients with metastatic prostate cancer have a much poorer prognosis. Localized prostate cancer is commonly treated by radical prostatectomy (RP). Within 10 years, up to one-third of men who have undergone RP for clinically organ-confined prostate cancer develop a biochemical recurrence (BCR) and can ultimately develop metastases associated with a high mortality rate. Determining the risk of metastasis in prostate cancer patients after radical prostatectomy is challenging, and there are no effective risk prediction techniques. Current techniques largely rely on pathology data, such as primary and secondary Gleason scores, reporting of which is highly subjective and subject to inter- and intra-observer variation. Currently, there are no effective AI-based techniques for predicting the likelihood of metastasis in patients following radical prostatectomy.

USCAP | 2023

Predicting Prostate Cancer Molecular Subtype with Artificial Intelligence

Unlike other genitourinary cancers, such as renal cell carcinoma, visual microscopic examination of prostate cancer has failed to reveal a reproducible association between tumor molecular subtype and morphologic features. However, deep learning-based algorithms trained on whole slide images (WSI) from large cohorts with known molecular classification may outperform the human eye and provide a cost-effective and rapid method to identify cases with clinically relevant genomic alterations. As proof of principle, we described two such algorithms to identify prostate tumors with underlying ERG fusions and/or PTEN deletion.

AUA | 2022

A Novel Approach for 3D Reconstruction of the Prostate Gland that Allows Tumour Location, Volume Estimation, and Gleason Characterization

The ability to precisely define the tumour location on MRI may contribute to the development of enhanced prostate cancer interpretation tools and Deep Learning (DL) models. Radical prostatectomy specimens provide a unique opportunity to compare diagnostic mpMRI images with whole-mount histopathology images of the resected tissue, allowing accurate quantification of tumour location and volume. The extent of tumour can be immediately mapped to their matching mpMRI via registration of digital histopathology slides, resulting in accurate radiological tumour localization, including labels for tumour foci undetected by conventional mpMRI interpretation.

ECDP| 2022

Deep Learning for Sub-Classification of Gleason Pattern 4 in Prostate Cancer

Several studies have provided compelling evidence on the relevance of reporting the percentage of Gleason pattern 4 in a tissue biopsy with a Gleason score of 7. Its significance in risk stratification and therapy planning lets doctors to evaluate if a patient is suitable for active surveillance (AS) or definitive therapy, which is why the ISUP has stated that the proportion of Gleason pattern 4 for all tissue samples must be reported in 2019. Furthermore, some studies have suggested that not only the fraction of Gleason pattern 4 present in a sample, but also the histologic pattern of Gleason pattern 4 is crucial to record, as it can represent severity and influence prognosis. We created a computational technique that uses Deep Learning to identify and quantify Cribriform and Glomeruloid patterns in whole mount images (WMI) of radical prostatectomy (RP) specimens.

Pathology Visions | 2021

A Deep Learning Method for Tumor Region Identification and Tumor Proportion Score Estimation for PD-L1 Expression in Non-Small Cell Lung Carcinoma

The PD-L1 (Programmed Cell Death Ligand 1) protein expression has emerged as a critical biomarker for selecting individuals with advanced lung cancer who are most likely to respond to immune checkpoint inhibitor therapy. The inherent heterogeneous expression of PD-L1 and the availability of multiple PD-L1 assays, detection systems, platforms, and cut-offs have created challenges in ensuring reliable and reproducible reporting which continues to be based on subjective visual assessment by pathologists. Using Deep Learning, we propose a computational technique for recognizing tumour cells in whole slide images (WSI) of immunohistochemically (IHC) stained sections from Non-Small Cell Lung Carcinoma (NSCLC) and accurately computing the Tumour Proportion Score (TPS) for PD-L1 expression.

Pathology Visions| 2020

Deep Learning-Based Segmentation & Grading of Pancreatic Intraepithelial Neoplasia

To develop a computer-assisted algorithm for automated detection and grading of Pancreatic intraepithelial Neoplasia (PanIN) lesions into low and high grade, in histopathological sections of pancreas. Pancreatic Intraepithelial Neoplasia are possible microscopic epithelial precursor lesions of pancreatic ductal adenocarcinoma (PDAC). Existing imaging modalities cannot accurately identify PANIn lesions pre-operatively; they can be identified only on histopathological examination. We propose a deep learning-based method for the detection and grading of PanIN lesions in histopathological sections of pancreas.

USCAP | 2025

Deep learning-based image analysis model for detecting unlearned findings in rat liver

We previously developed supervised models capable of accurately detecting learned abnormalities in Whole Slide Images (WSIs) of the liver, kidney, testis, epididymis, and brain in Sprague Dawley (SD) rats. However, the models developed based on supervised approaches were unable to identify unlearned findings with sufficient accuracy. We developed the models using unsupervised development approaches into the current image analysis framework for rat liver and assessed its performance in identifying various toxicity findings in the rat liver.

USCAP | 2025

Deep Learning-Based Image Analysis Model for Detecting Unlearned Findings in Early Toxicity Screening Studies

Creating deep learning algorithms for identifying drug-induced changes in histology slides frequently involves training on prospectively identified histology findings. We previously created supervised learning models for image processing that accurately detect prospectively identified observations in rodent liver, kidney, and heart. While these supervised models are effective at detecting prospectively identified histology findings, our use-case in nonclinical toxicology study assessment requires sensitive detection of all potential histology findings. This study integrated an unsupervised learning methodology into the existing image analysis framework and evaluated its precision in detecting out of sample toxicity findings.

AIRA MATRIX | Artificial Intelligence Transformed Analyses

Drug Discovery & Development

Inter-Rater and Intra-Rater Agreement in Scoring Severity of Rodent Cardiomyopathy and Relation to Artificial Intelligence-Based Scoring

Thomas Steinbach, Debra Tokarz, Caroll Co, Shawn Harris, Sandra McBride, Keith Shockley, Avinash Lokhade, Gargi Srivastava, Rajesh Ugalmugle, Arshad Kazi, Emily Singletary, Mark Cesta, Heath Thomas, Vivian Chen, Kristen Hobbie, and Torrie Crabbs

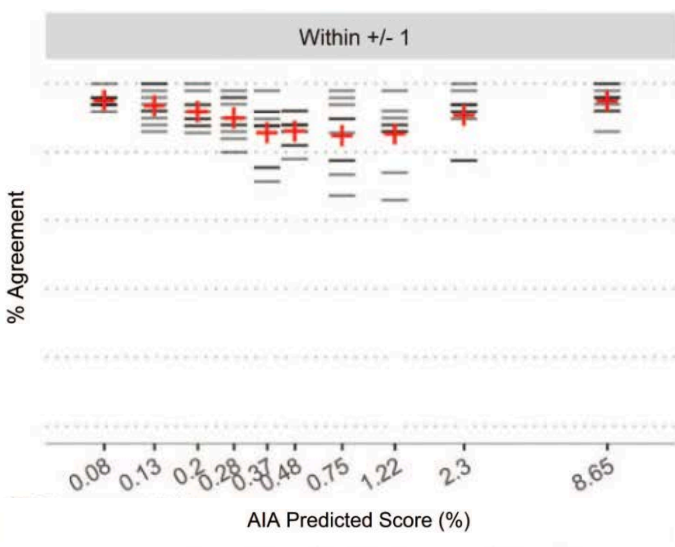
In collaboration with Experimental Pathology Laboratories Inc

***Toxicologic Pathology*, 2024; Vol. 0(0)**

Abstract

We previously developed a computer-assisted image analysis algorithm to detect and quantify the microscopic features of rodent progressive cardiomyopathy (PCM) in rat heart histologic sections and validated the results with a panel of five veterinary toxicologic pathologists using a multinominal logistics model. In this study, we assessed both the inter-rater and intra-rater agreement of the pathologists and compared pathologists ratings to the AI-predicted scores. Pathologists and the AI algorithm were presented with 500 slides of rodent heart. They qualified the amount of cardiomyopathy in each slide. A total of 200 of these slides were novel to this study, whereas 100 slides were intentionally selected for repetition from the previous study. After a washout period of more than six months, the repeated slides were examined to assess intra-rater agreement among pathologists. We found the intra-rater agreement to be substantial, with weighted Cohen's kappa values ranging from k = 0.64 to 0.80. Intra-rater variability is not a concern for the deterministic AI. The inter-rater agreements across pathologists was moderate (Cohen's kappa k = 0.56). These results demonstrate the utility of AI algorithms as a tool for pathologists to increase sensitivity and specificity for the histopathologic assessment of the heart in toxicology studies.

Results



Distribution of percent agreement across all pairs of raters.

Deep Learning-Based Spermatogenic Staging in Tissue Sections of Cynomolgus Macaque Testes

Lars Mecklenburg, C. Marc Luetjens, Annette Romeike, Rohit Garg, Pranab Samanta, Amogh Mohanty, Tijo Thomas, and Gerhard Weinbauer

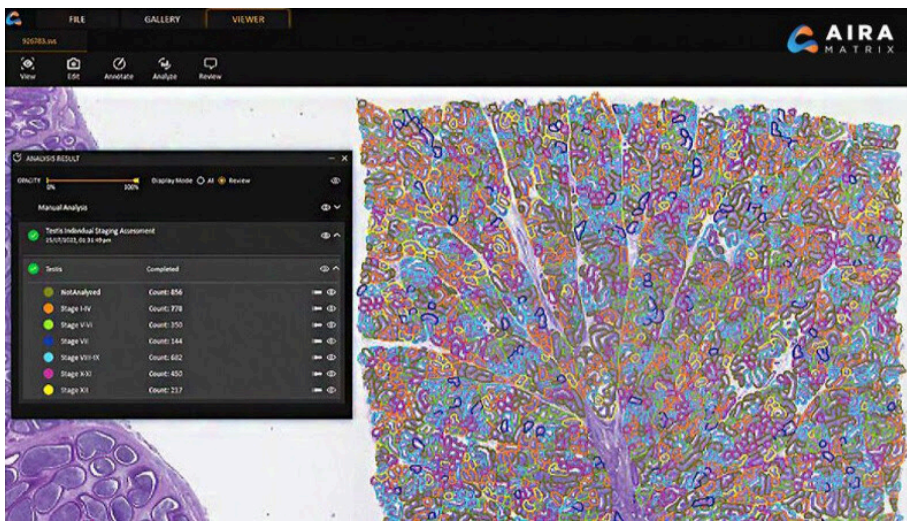
In collaboration with LabCorp Early Development Services GmbH

***Toxicologic Pathology*, 2024; Vol. 52(1) 4-12**

Abstract

The indirect assessment of adverse effects on fertility in cynomolgus monkeys requires that tissue sections of the testis be microscopically evaluated with awareness of the state of spermatogenesis that a particular cross-section of a seminiferous tubule is in. This difficult and subjective task could very much benefit from automation. Using digital whole slide images (WSIs) from tissue sections of the testis, we have developed a deep-learning model that can annotate the stage of each tubule with high sensitivity, precision, and accuracy. The model was validated on six WSI using a six-stage spermatogenic classification system. Whole slide images contained an average number of 4938 seminiferous tubule cross-sections. On average, 78% of these tubules were staged with 29% in stages I-IV, 12% in stage V-VI, 4% in stage VII, 19% in stage VIII-IX, 18% in stage X-XI, and 17% in stage XII. The deep learning model supports pathologists in conducting a stage-aware evaluation of the testis. It also allows the derivation of a stage-frequency map. The diagnostic value of this stage-frequency map. The diagnostic value of this stage-frequency map is still unclear, as further data on its variability and relevance need to be generated for testes with spermatogenic disturbances.

Results



Annotation of seminiferous tubule cross-sections in a WSI

Deep Learning-Based Image-Analysis Algorithm for Classification and Quantification of Multiple Histopathological Lesions in Rat Liver

Taishi Shimazaki, Ameya Deshpande, Anindya Hajra, Tijo Thomas, Kyotaka Muta, Naohito Yamada, Yuzo Yasui, and Toshiyuki Shoda

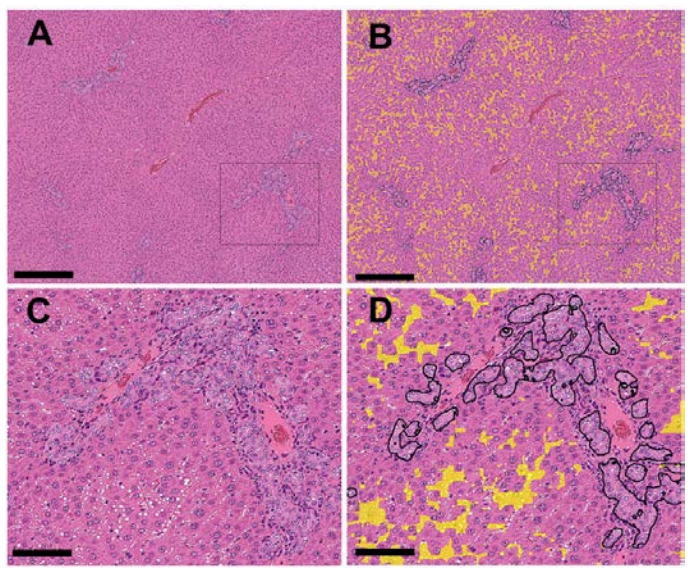
In collaboration with Japan Tobacco Inc.

***Journal of Toxicologic Pathology*, 2022; 35: 135-147**

Abstract

Artificial intelligence (AI)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry. In this paper, we present an AI-based solution for preclinical toxicology studies. We trained a set of algorithms to learn and quantify multiple typical histopathological findings in whole slide images (WSIs) of the livers of young Sprague Dawley rats by using a U-Net-based deep learning network. The trained algorithms were validated using 255 liver WSIs to detect, classify, and quantify seven types of histopathological findings (including vacuolation, bile duct hyperplasia, and single-cell necrosis) in the liver. The algorithms showed consistently good performance in detecting abnormal areas. Approximately 75% of all specimens could be classified as true positive or true negative. In general, findings with clear boundaries with the surrounding normal structures, such as vacuolation and single-cell necrosis, were accurately detected with high statistical scores. The results of quantitative analyses and classification of the diagnosis based on the threshold values between “no findings” and “abnormal findings” correlated well with diagnoses made by professional pathologists. However, the scores for findings ambiguous boundaries, such as hepatocellular hypertrophy, were poor. These results suggest that deep learning-based algorithms can detect, classify, and quantify multiple findings simultaneously on rat liver WSIs. Thus, it can be a useful supportive tool for a histopathological evaluation, especially for primary screening in rat toxicity studies.

Results



Observation & Annotation by the Algorithm: A. & B. Drug-induced bile duct hyperplasia; C. & D. Bile duct hyperplasia and vacuolation of hepatocytes (Bar = 100 µm)

Deep Learning-Based Spermatogenic Staging Assessment for Hematoxylin and Eosin-Stained Sections of Rat Testes

Dianne M. Creasy, Satish T. Panchal, Rohit Garg, and Pranab Samanta

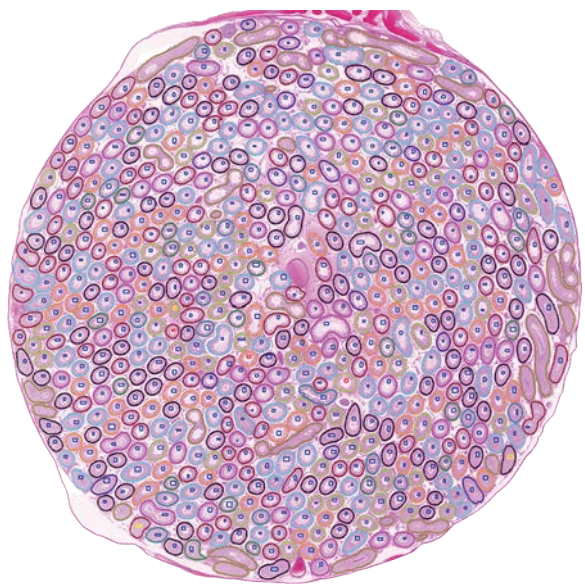
In collaboration with Dianne Creasy Consulting and Sun Pharma Advanced Research Co. Ltd.

***Toxicologic Pathology*, 2021, Vol. 49(4) 872-887**

Abstract

In preclinical toxicology studies, a “stage-aware” histopathological evaluation of testes is recognized as the most sensitive method to detect effects on spermatogenesis. A stage-aware evaluation requires the pathologist to be able to identify the different stages of the spermatogenic cycle. Classically, this evaluation has been performed using Periodic Acid-Schiff (PAS)-stained sections to visualize the morphology of the developing spermatid acrosome, but due to the complexity of the rat spermatogenic cycle and the subtlety of the criteria used to distinguish between the 14 stages of the cycle, staging of tubules is not only time consuming but also requires specialized training and practice to become competent. Using different criteria, based largely on the shape and movement of the elongating spermatids within the tubule and pooling some of the stages, it is possible to stage tubules using routine hematoxylin and eosin (H&E)-stained sections, thereby negating the need for a special PAS stain. These criteria have been used to develop an automated method to identify the stages of the rat spermatogenic cycle in digital images of H&E-stained Wistar rat testes. The algorithm identifies the spermatogenic stage of each tubule, thereby allowing the pathologist to quickly evaluate the testis in a stage-aware manner and rapidly calculate the stage frequencies.

Results



Tubules annotated by the algorithm and pathologist during validation of the algorithm.

Microscope-Based Automated Quantification of Liver Fibrosis in Mice Using a Deep Learning Algorithm

Yuval Ramot, Ameya Deshpande, Virginia Morello, Paolo Michieli, Tehila Shlomov, and Abraham Nyska

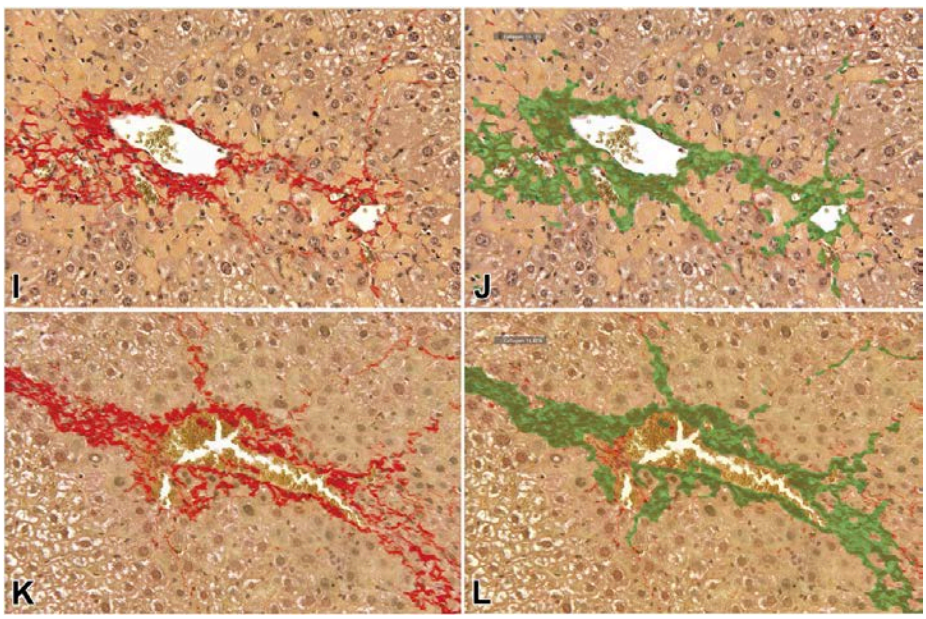
In collaboration with Hebrew University of Jerusalem, Hadassah Medical Center, AgomAb Therapeutics NV, Molecular Biotechnology Center, Tel Aviv University

***Toxicologic Pathology*, 2021, Vol. 49(5) 1126-1133**

Abstract

In preclinical studies that involve animal models for hepatic fibrosis, accurate quantification of the fibrosis is of utmost importance. The use of digital image analysis based on deep learning artificial intelligence (AI) algorithms can facilitate accurate evaluation of liver fibrosis in these models. In the present study, we compared the quantitative evaluation of collagen proportionate area in the carbon tetrachloride model of liver fibrosis in the mouse by a newly developed AI algorithm to the semiquantitative assessment of liver fibrosis performed by a board-certified toxicologic pathologist. We found an excellent correlation between the 2 methods of assessment, most evident in the higher magnification (×40) as compared to the lower magnification (×10). These findings strengthen the confidence of using digital tools in the toxicologic pathology field as an adjunct to an expert toxicologic pathologist.

Results



Quantitative assessment of liver fibrosis using image analysis.

Using Artificial Intelligence to Detect, Classify, and Objectively Score Severity of Rodent Cardiomyopathy

Debra A. Tokarz, Thomas J. Steinbach, Avinash Lokhande, Gargi Srivastava, Rajesh Ugalmugle, Caroll A. Co, Keith R. Shockley, Emily Singletary, Mark F. Cesta, Heath C. Thomas, Vivian S. Chen, Kristen Hobbie, and Torrie A. Crabbs

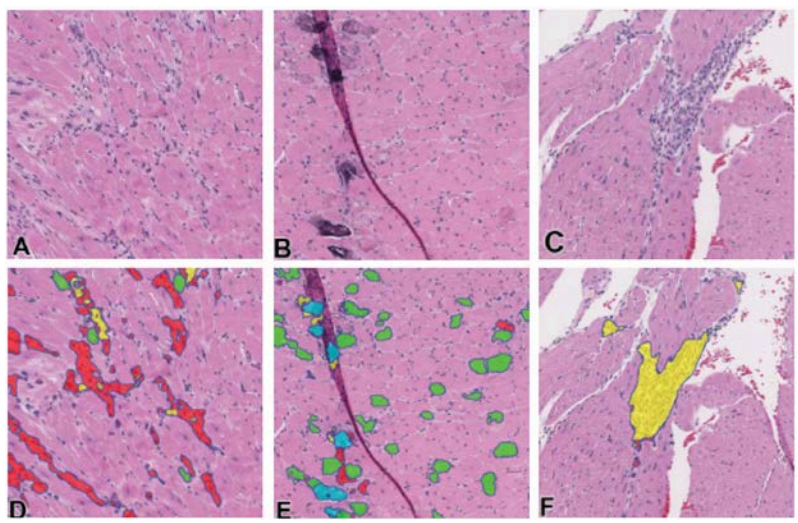
In collaboration with Experimental Pathology Laboratories Inc., Social and Scientific Systems, National Institute of Environmental Sciences, Aclairo Pharmaceuticals Development Group, Charles River Laboratories Inc., and Integrated Laboratory Systems, LLC

***Toxicologic Pathology*, 2021, Vol. 49(4) 888-896**

Abstract

Rodent progressive cardiomyopathy (PCM) encompasses a constellation of microscopic findings commonly seen as a spontaneous background change in rat and mouse hearts. Primary histologic features of PCM include varying degrees of cardiomyocyte degeneration/necrosis, mononuclear cell infiltration, and fibrosis. Mineralization can also occur. Cardiotoxicity may increase the incidence and severity of PCM, and toxicity-related morphologic changes can overlap with those of PCM. Consequently, sensitive and consistent detection and quantification of PCM features are needed to help differentiate spontaneous from test article-related findings. To address this, we developed a computer-assisted image analysis algorithm, facilitated by a fully convolutional network deep learning technique, to detect and quantify the microscopic features of PCM (degeneration/necrosis, fibrosis, mononuclear cell infiltration, mineralization) in rat heart histologic sections. The trained algorithm achieved high values for accuracy, intersection over union, and dice coefficient for each feature. Further, there was a strong positive correlation between the percentage area of the heart predicted to have PCM lesions by the algorithm and the median severity grade assigned by a panel of veterinary toxicologic pathologists following light microscopic evaluation. By providing objective and sensitive quantification of the microscopic features of PCM, deep learning algorithms could assist pathologists in discerning cardiotoxicity-associated changes.

Results



Sprague Dawley rat heart images before and after algorithm segmentation.

Utilization of a Deep Learning Algorithm for Microscope-Based Fatty Vacuole Quantification in a Fatty Liver Model in Mice

Yuval Ramot, Gil Zandani, Zecharia Madar, Sanket Deshmukh, and Abraham Nyska

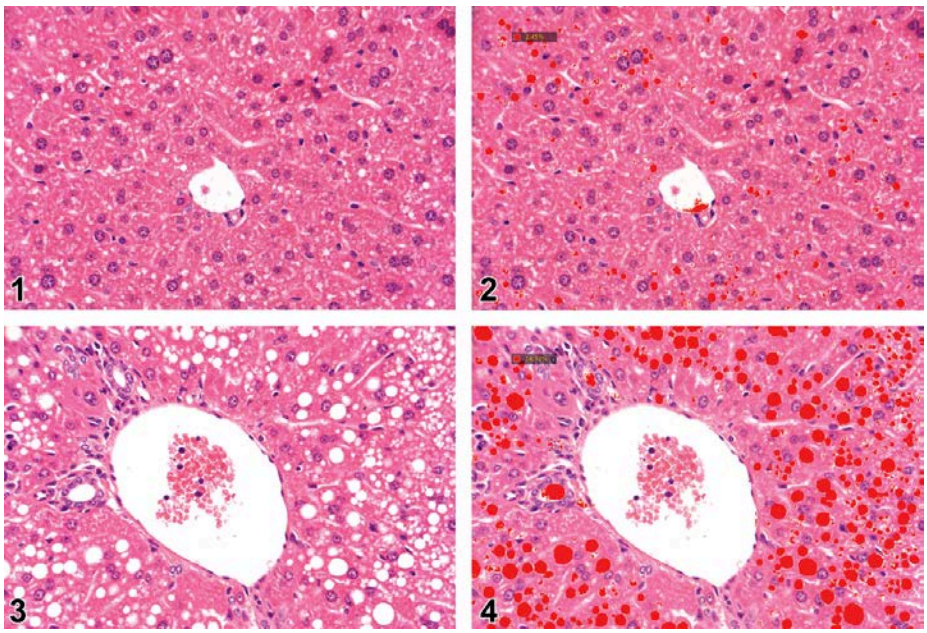
In collaboration with Hebrew University of Jerusalem and Tel Aviv University

***Toxicologic Pathology*, 2020, Vol. 48(5) 702-707**

Abstract

Quantification of fatty vacuoles in the liver, with differentiation from lumina of liver blood vessels and bile ducts, is an example where the traditional semiquantitative pathology assessment can be enhanced with artificial intelligence (AI) algorithms. Using glass slides of mice liver as a model for nonalcoholic fatty liver disease, a deep learning AI algorithm was developed. This algorithm uses a segmentation framework for vacuole quantification and can be deployed to analyze live histopathology fields during the microscope-based pathology assessment. We compared the manual semiquantitative microscope-based assessment with the quantitative output of the deep learning algorithm. The deep learning algorithm was able to recognize and quantify the percent of fatty vacuoles, exhibiting a strong and significant correlation ($r = 0.87$, $P < .001$) between the semiquantitative and quantitative assessment methods. The use of deep learning algorithms for difficult quantifications within the microscope-based pathology assessment can help improve outputs of toxicologic pathology workflows.

Results



Histology section of the perilobular region of the liver from a mouse fed normal diet (1 & 2) and high-fat diet (3 & 4) for 7 weeks. Hematoxylin and Eosin staining.

EUROTOX | 2024

Machine Learning for Rodent Liver Toxicity Prediction: Leveraging Drug Substructure Properties

To develop robust predictive models to assess liver toxicity in rodents (rats & mice) during the drug development process. This objective emphasizes the importance of accurate models for:
Safety Assessment: Early identification of potential liver toxicity risks
Regulatory Compliance: Meeting regulatory requirements for drug development
Cost and Time Savings: Avoiding unnecessary animal testing and delays in drug development

EUROTOX | 2024

In Vitro to In Vivo Extrapolation (IVIVE) for Predicting Drug-Induced Liver Toxicity

The objective of this study was to create a model for predicting drug-induced liver toxicity by identifying crucial genes linked to cell toxicity in in vitro settings and assessing their ability to predict toxicity in in vivo. The objective was to establish a connection between in vitro and in vivo data by employing quantitative in vitro to in vivo extrapolation (QIVIVE) techniques, which have the potential to reduce animal sacrifice in in vivo studies.

STP Annual Symposium | 2024

Deep Learning-based Method for Anatomical Subsite-wise Evaluation of Single Cell Necrosis and Vacuolation of Neuron/Nerve Fiber of CNS Toxicity in Rats

Assessment of changes within the central nervous system (CNS) including single cell necrosis and vacuolation of neuron/nerve fiber is a sensitive method to assess toxicity. A deep learning-based algorithm for analysis of single-cell necrosis and vacuolation is proposed for 7 levels of rat brain.

STP Annual Symposium | 2024

Deep Learning Solution for the Automated Assessment of the Rodent Thymus

Thymus, being a primary lymphoid organ, is a sensitive target following exposure to immunotoxins. Reduction in cortical lymphocytes is an important histopathological finding in compound-induced effects. Hence evaluating the cortico-medullary ratio, and assessing the cortical lymphocytes is important. Manual histopathological assessment of these features can be a time-consuming process with subjective outputs. We developed a deep-learning solution for the automated assessment of the rodent thymus. The solution separately identifies the cortex and medulla, computes the cortico-medullary ratio, and quantifies lymphocytes and apoptotic cells in each compartment.

STP Annual Symposium | 2024

AI-based Approach for Quantifying and Grading Bile Duct Hyperplasia in Mice

Bile duct hyperplasia in rodents can represent a direct or indirect effect of chemicals by cytotoxicity or interference with bile flow, respectively. It may also be observed as a spontaneous change in portal areas of older animals. We developed an AI model to for accurate severity assessment of bile duct hyperplasia in rodents.

STP Annual Symposium| 2024

Solving the Reference Pathologist Paradox in Machine Learning Development for Histopathology Scoring

Development of machine learning (ML) algorithms for scoring histology slides for nonclinical toxicology commonly involves training against example histopathology findings. This approach creates an ML performance boundary based on the list of diagnoses included in training and sensitivity of the pathologist who made the diagnoses. We hypothesized that a ML development strategy that does not require training against histopathology findings could increase algorithm sensitivity.

STP Annual Symposium | 2023

Deep Learning Method for Quantification of Mast Cells in Toluidine Blue Stained Tissue Sections of Mouse Skin

Mast cells play a significant role in skin immunity and the pathogenesis of multiple skin diseases, including atopic dermatitis, scleroderma, contact dermatitis, blistering cutaneous disorders, and chronic graft versus host disease. It is also responsible for physiological functions such as innate and adaptive responses, angiogenesis, regulation of vasodilation, vascular homeostasis, and detoxification of venom. A common characteristic across skin diseases is an increase in mast cells, which undergo degranulation in the affected skin. Accurate quantification of mast cells is an important step in assessing the efficacy of study models for skin diseases. We propose a deep learning (DL)-based method for quantification of mast cells in skin sections of mice stained with toluidine blue.

STP Annual Symposium | 2023

Deep Learning-Based Image Analysis Model for Evaluation of Testicular and Epididymal Lesions in Rats

Spermatogenic staging and assessment of testicular toxicities in rat tissue sections are time-consuming and requires the expertise of well-trained pathologists. Deep Learning (DL)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry. Here we present a DL-based solution for classifying 11 stage groups of spermatogenesis namely stages I, II-III, IV-VI, VII, VIII, IX, X, XI, XII-XIII, XIV, XIV and "Stage Not Analyzed" based on normal testicular tissue structure. In addition, the solution for identifying and quantifying testicular and epididymal lesions in rat toxicology studies is also shown.

STP Annual Symposium | 2022

Quantification of Fibrotic Changes in Hematoxylin and Eosin-Stained Wistar Rat Liver Sections using Deep Learning

Fibrosis is usually linked to chronic liver parenchymal injury and is considered irreversible. Perpetuation of the fibrotic reaction can lead to end-stage liver disease, cirrhosis, and hepatocellular carcinoma, with a worldwide increase in incidence. Disease models in rodents are utilized to assess efficacy of test compounds being investigated to treat liver fibrosis. Collagen specific histochemical stains such as Masson's Trichrome (MT) or Picrosirius Red (PSR) are commonly used for assessment of fibrosis in rodents. However, these staining techniques add to the cost and turnaround time. We provide a DL-based technique for quantification of fibrosis on H&E-stained liver sections of Wistar rats. A staining technique routinely used in animal studies.

Digital Pathology & AI Congress | 2022

Deep Learning-Based Method for Spermatogenic Staging and Assessing Testicular Toxicity Endpoints in Rats

Exposure to drugs and chemicals has significant adverse effects on the reproductive system. Histopathology evaluation of the testis is an important component when assessing drug safety and environmental toxicants. Understanding the cellular relationships occurring during the spermatogenic cycle helps recognize absent cells and detect subtle changes restricted to specific points in spermatogenesis. Earlier, we had developed a deep learning (DL)-based method to automate the staging which is easy to use, and provides results comparable to an expert on normal testes and to historical data. Taking this ahead, we propose a DL-based approach for the identification and quantification of the stage-specific and non-specific endpoints in rat testis.

STP Annual Symposium | 2022

Automated Identification and Quantification of Pancreatic Pathology in Rodents

Drug-induced pancreatic injury in preclinical toxicology studies is a serious liability in drug development. Pancreatic toxicity is generally characterized by dysregulation of lipid metabolism and edema in early reversible stages, followed by massive necrosis resulting in inflammation, with or without fibrosis at the advanced irreversible stages. Some patients with pancreatitis can also develop pancreatic cancer. Therefore, accurate identification and characterization of test compound induced pancreatic lesions in preclinical toxicity studies is important to understand the clinical translatability. Islet cell hyperplasia, acinar cell apoptosis, and atrophy are the commonly observed pathological lesions in pancreas in rodent studies. We present a DL-based method to quantify these histopathological changes in rodent pancreas.

STP Annual Symposium | 2022

Deep Learning Models for Quantifying Testicular Toxicity in Rats

Recognizing the 14 stages in the rat testis cycle is a difficult task. We have developed a deep learning (DL)based method to automate the staging which is easy to use, and provide results comparable to an expert on normal testes, and to historical data. Taking this ahead, we propose a DL-based method for identification and quantification of findings in rat testes. We have concentrated on a subset of findings from the INHAND nomenclature. We selected several Janssen toxicology studies on which early stages findings (spermatid retention after one month) and more chronic, degenerative findings were recorded. We have developed DL models to detect those finds and compared them to pathologists' results.

STP Annual Symposium | 2022

Deep Learning-Based Image Analysis Algorithm for Classification and Quantification of Multiple Histopathological Lesions of the Rat Liver and Kidney

Artificial Intelligence (AI)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry. In this study, we present a Deep Learning (DL)-based method for classification and quantification of multiple histopathological lesions in rodent liver and kidney. The trained algorithms were validated using 255 liver Whole Slide Image (WSIs) to detect, classify, and quantify the seven findings in the liver. A modified form of the U-Net DL Model was trained using data from WSIs of 92 liver sections and 90 kidney sections. The trained model was used for identifying and quantifying 7 types of histopathology findings in both liver (vacuolation, bile duct hyperplasia, single-cell necrosis, microgranuloma, EMH, and hypertrophy) and kidney (vacuolation, basophilia/degeneration/regeneration tubule, dilation, hyaline cast, mineralization, mononuclear cell infiltration, and cyst). The algorithm was validated by comparing the results with pathologists' findings on 255 liver sections and 285 kidney sections.

JSTP Annual Meeting | 2022

Deep Learning-Based Image Analysis Algorithm for Classification and Quantification of Multiple Histopathological Lesions of the Rat Liver

Artificial Intelligence (AI)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry. In this study, we present an AI-based solution for preclinical toxicology studies. We trained a set of algorithms to learn and quantify seven types of typical histopathological findings (including vacuolation, bile duct hyperplasia, and single cell necrosis) in whole slide images (WSIs) of the livers of young Sprague Dawley (SD) rats by using a U-Net-based deep learning network. The trained algorithms were validated using 255 liver WSIs to detect, classify. and quantify the seven findings in the liver.

STP Annual Symposium | 2021

Deep Learning-Based Method for Assessing the Phases of Estrus Cycle in H&E Stained Sections of Wistar Rat Uterus

Determining the phase of the estrus cycle accurately and consistently helps in understanding interference of chemical entities with the female reproductive function of animals during pre-clinical studies. The approach of ascribing phases to the animals based on the evaluation of a synchronous system that includes the vagina and uterus is widely accepted. In addition to our novel work on Deep Learning-based estrus staging method using vaginal histopathology, an effort has been made to identify and quantify the morphological changes in the uterus across different phases of the estrus cycle. A Deep Learning-based algorithm is proposed for an accurate identification of 4 phases, Proestrus, Estrus, Metestrus and Diestrus, of the estrus cycle using whole slide images (WSI) of conventional hematoxylin and eosin-stained sections of Wistar rat uterus.

EUROTOX Congress | 2021

Multi-Modal Approach for Prediction of Skin Toxicity in Small Molecules

Calculating precomputed descriptor values is a common method for identifying molecular data, however it is ineffective for large molecular sizes. To accurately characterize skin hazardous data, we devised a multimodal technique. We have also created a tool that reveals the class designation as well as the essential substructure found in skin hazardous compounds.

solutions@airamatrix.com

AIRA MATRIX

Artificial Intelligence
Transformed Analyses

AM-PP-PRECLIN-MAR-2025-V01