Cancer Biomarkers 33 (2022) 173–184
DOI 10.3233/CBM-210301
IOS Press

Review

Applications of artificial intelligence (AI) in ovarian cancer, pancreatic cancer, and image biomarker discovery

Dina Mikdadi[a], Kyle A. O'Connell[a,b], Philip J. Meacham[a], Madeleine A. Dugan[a], Michael O. Ojiere[a], Thaddeus B. Carlson[a] and Juergen A. Klenk[a,*]
[a]*Biomedical Data Science Lab, Deloitte Consulting LLP, Arlington, VA, USA*
[b]*Department of Biology, George Washington University, Washington, DC, USA*

Received 31 May 2021
Accepted 11 September 2021

Abstract.
BACKGROUND: Artificial intelligence (AI), including machine learning (ML) and deep learning, has the potential to revolutionize biomedical research. Defined as the ability to "mimic" human intelligence by machines executing trained algorithms, AI methods are deployed for biomarker discovery.
OBJECTIVE: We detail the advancements and challenges in the use of AI for biomarker discovery in ovarian and pancreatic cancer. We also provide an overview of associated regulatory and ethical considerations.
METHODS: We conducted a literature review using PubMed and Google Scholar to survey the published findings on the use of AI in ovarian cancer, pancreatic cancer, and cancer biomarkers.
RESULTS: Most AI models associated with ovarian and pancreatic cancer have yet to be applied in clinical settings, and imaging data in many studies are not publicly available. Low disease prevalence and asymptomatic disease limits data availability required for AI models. The FDA has yet to qualify imaging biomarkers as effective diagnostic tools for these cancers.
CONCLUSIONS: Challenges associated with data availability, quality, bias, as well as AI transparency and explainability, will likely persist. Explainable and trustworthy AI efforts will need to continue so that the research community can better understand and construct effective models for biomarker discovery in rare cancers.

Keywords: Artificial intelligence, bias, biomarkers, machine learning, rare cancer

1. Introduction

Artificial intelligence (AI) has the potential to revolutionize healthcare [1], and in fact, is already being taken from theoretical development to clinical application, particularly with imaging analysis [2,3]. As the global population ages, pressures will mount on healthcare systems and increase the burden on practitioners. New digital technologies (often AI enabled) have the potential to disrupt current practices, largely by enhancing, rather than replacing the abilities of practitioners [4,5]. AI is widely defined as a computer's ability to "mimic" human intelligence by executing code contained in various algorithms [6]. Machine learning (ML) is a subset of AI, where statistical methods are used to develop and refine algorithms. Deep learning, in turn, is a subset of ML based on layers of neural networks that permit a computer to train itself on a particular task. While AI has garnered excitement across life sciences and

*Corresponding author: Juergen A. Klenk, Principal, Federal Health, Deloitte Consulting LLP, 1919 N. Lynn Street Arlington, VA 22209, USA. Tel.: +1 571 858 1141; E-mail: jklenk@deloitte.com.

ISSN 1574-0153/$35.00

healthcare, core challenges pertaining to data availability, quality, model training, and bias persist. Addressing these issues, and other limitations, will be crucial to reap the benefits of such technology for healthcare advancement. One important application of AI will be in the field of cancer biomarker discovery.

In this review, we define AI as activity or code that encompasses both machine learning and deep learning through a variety of neural networks. We define data availability as relevant, diverse, AI-ready data that is accessible for researchers and bias refers to AI model bias that occurs when data used in the machine learning process is not adequately representative, therefore producing prejudiced outputs. We also highlight advancements and challenges in the use of AI for biomarker discovery in two rare, but very lethal (i.e. high case-fatality) cancers – ovarian and pancreatic. These 'silent killer' cancers are especially aggressive in part due to the lack of early symptoms and early detection. The successful application of AI technologies and ML methods will have a significant impact in reducing cancer-associated mortality and morbidity, specifically in ovarian and pancreatic cancers given the current difficulty in diagnosing these malignant tumors early. We conducted a literature review by searching both PubMed and Google Scholar to survey the published medical research on the use of AI in ovarian cancer, pancreatic cancer, and cancer biomarkers. Here we summarize an overview of the landscape, including the regulatory and ethical considerations, and we identify future directions for the application of AI in rare cancers and biomarker discovery.

2. Ovarian cancer

2.1. Background

Ovarian cancer is relatively rare, accounting for fewer than 4% of cancers among women worldwide [7]. However, it is a leading cause of cancer-attributable deaths and is the most fatal gynecological cancer [8,9] due to late stage diagnosis and a high (70%) rate of recurrence [10,11]. According to the International Federation of Obstetrics and Gynecology (FIGO) staging [12], 5-year survival rates range between 70% and 90% when disease is limited to the ovaries (stage I) or pelvis (stage II) [13,14] but dramatically decrease to less than 30% once the disease metastasizes (stage III or IV) [14,15]. Incidence rates of ovarian cancer are greatest in developed countries but vary by age and race [7,13]. The epidemiological diversity of ovarian cancer throughout the world is due in part to the multifactorial etiology of the disease [10] as well as differences in clinical management and disparities in access to diagnostic services [16]. The majority (80%) of ovarian tumors are benign [17], although differentiating benign tumors from malignant disease remains a clinical challenge. Among the different pathologies, ovarian epithelial cancer (OEC) accounts for nearly 90% of malignant ovarian tumors [12]. OEC is a heterogeneous disease of distinct histologic subtypes with varying etiologies, morphologies, clinical presentations, and prognoses [14,15]. The primary risk factor of poor clinical outcomes in OEC is late-stage detection, and currently there is no standard screening test. Unfortunately due to the asymptomatic nature of the disease, fewer than 25% of women with OEC are diagnosed early (i.e. stage I or II) when the disease can be easily managed [11]. Increasing the rate of early detection has been suggested to lower the mortality rate by as much as 30% [18]. Given the low prevalence of ovarian cancers, including OEC, epidemiological rules require that for a screening test to be effective it must have a high sensitivity ($> 75\%$) and specificity ($> 99\%$) [19]. The emerging use of AI in the discovery of biomarkers can provide significant clinical benefits for early detection, new treatments, and improved prognosis.

2.2. Ovarian cancer biomarkers overview

Biomarkers play a critical role in personalized medicine and are urgently needed for early detection of ovarian cancers, especially OEC, due to the lack of a standard screening evaluation [20] and the high rate of recurrence. Imaging technologies such as transvaginal ultrasonography (TVS), positron emission tomography/computed tomography with fluorodeoxyglucose (FDG-PET/CT), and magnetic resonance imaging (MRI) can be utilized for detecting early-stage OEC. However, on their own, these techniques have poor sensitivity and specificity [21–24] which leads to false positives [25]. Additionally, PET/CT and MRI are not widely used for detection due to the radiation exposure and the high cost, respectively [26,27]. In addition to imaging techniques, a wide range of biochemical markers have been evaluated for early detection, screening, treatment response, and prognosis [28–30]. These include protein tumor biomarkers such as serum cancer antigen 125 (CA125) [31–35] and human epididymis protein (HE4) [36–38], genetic markers such as germline mutations in *BRCA1*/*BRCA2* [39], and epi-

genetic biomarkers such as DNA methylation [30,40] and microRNA expression [41]. None of these markers provide sufficient sensitivity or specificity to detect early-stage OEC [30,41], and there is a lack of evidence showing statistically significant decreases in mortality rates when using these markers as screening tools [42,43]. Currently, no single biomarker meets the required threshold for both sensitivity and specificity to be effective in detecting ovarian cancer early. Multivariate assays that combine biomarkers and clinical factors are being developed and evaluated for diagnostic accuracy [28,44,45]. The Risk of Malignancy Index (RMI) enhances the robustness of using CA125 alone by factoring in ultrasound imaging and menopausal status for the prediction of ovarian cancer in women with a pelvic mass [46]. Similarly, the Risk of Ovarian Malignancy Algorithm (ROMA) combines HE4 and CA125 to predict the likelihood of OEC in women with a pelvic mass [47]. There are two FDA-approved multivariate biomarker assays, Ova1 [48,49] and Overa [50,51], with relatively high sensitivity (96% and 91%, respectively) but low specificity (54% and 69%, respectively) [49,52]. Notably, these are not preferred screening tests for early detection, but rather prediction algorithms to determine the probability of a malignant tumor and the need for referral to a gynecologic oncologist.

2.3. Application of AI in ovarian cancer (diagnosis)

As medical research is beginning to focus on the clinical application of AI methods in oncology, more studies are needed to develop diagnostic tools for the early detection of ovarian cancer. Two-dimensional light scattering technology was employed by Chen and Zhang [53] for the early detection of single ovarian cancer cells. Results of 10-fold cross-validation by support vector machine algorithms show high sensitivity (95.9%) and moderately high specificity (87.5%) in detecting malignant ovarian cells. Computer-aided diagnosis (CADx) can be utilized to improve diagnostic accuracy of histologic subtypes of ovarian cancer (serous, mucous, endometrioid, and clear cell carcinomas). Using deep convolutional neural networks (DCNN) on 85 tissue specimens (24 serous carcinoma, 22 mucinous carcinoma, 21 endometrioid, and 18 clear cell carcinoma) from patients at Xinjiang Medical University between 2003 and 2016, Wu et al. [54] leveraged cytological images to automatically classify ovarian cancer subtypes with 72.8% accuracy. This increased to 78.2% accuracy only after image augmentation, which shows the correlation between model performance and the quantity and quality of the images for training DCNN. The application of AI also appears promising in diagnostic prediction of OEC prior to intervention with predictive algorithms benefiting personalized treatment options [55]. Machine learning models, compared to conventional regression-based analyses, may yield superior results in predicting clinical factors associated with OEC [55,56]. In 2019, Kawakami et al. [57] randomly assigned patients with OEC ($n = 334$) and those with benign ovarian tumors ($n = 101$) into a training group and a test group to establish a specific predictive framework for pretreatment of OEC patients. Machine learning classifiers, including random forest (RF), obtained diagnostic and prognostic information from 32 biomarkers and clinical factors commonly used in pretreatment peripheral blood tests. This method showed a statistically significant ability to discriminate OEC from benign ovarian tumors (accuracy $= 92.4\%$; AUC with RF $= 0.968$) with lower confidence at predicting clinical stage of OEC (accuracy $= 69.0\%$; AUC with RF $= 0.760$). These classifiers also underperformed in predicting histologic types of EOC (range of AUC: 0.597–0.785); however, this is likely due to the level of serum biomarkers not distinguishing the characteristics of these different tumor types.

2.4. Application of AI in ovarian cancer (prognosis)

Current literature evaluating the use of medical imaging data suggests that employing deep learning methods can improve the prediction of ovarian cancer patient prognosis. Enshaei et al. [58] developed an artificial neural network (ANN) algorithm using clinical and survival data on 668 OEC cases over a 10-year period to predict the overall five-year survival rate of OEC patients (accuracy $= 93\%$; AUC $= 0.74$). This AI model was also able to adequately predict surgical outcomes of complete, optimal, or suboptimal cytoreduction among the cases (accuracy $= 77.7\%$; AUC $= 0.73$). Wang et al. [59] developed a novel approach by combining a deep learning feature with conventional Cox proportional hazard regression (DL-CPH) to extract prognostic data from 8,917 CT images from 245 patients with high-grade serous ovarian cancer (HGSOC) across two different hospitals (feature-learning cohort, $n = 102$; primary cohort, $n = 49$; two independent validation cohorts, $n = 49$ and $n = 45$). To ensure minimal tumor selection bias influencing the robustness of the deep learning features, Wang et al. estimated the intraclass correlation coefficient (ICCC) using data from 40 patients corresponding to two radiologists selected at ran-

dom. All deep learning features were consistent (range of ICCC = 0.83–0.98) between the two radiologists. The DL-CPH model successfully identified two patient groups at high-risk ($p = 0.004$, AUC = 0.77) and low-risk ($p = 0.016$, AUC = 0.83) of recurrence at three years. If validated in future studies, this approach would allow for the prediction of HGSOC recurrence from CT images without the need for follow-up. Lu et al. [60] utilized machine learning models with 657 quantitative descriptors from preoperative CT images of 364 OEC patients to establish and validate a novel mathematical description of tumor phenotype and prognosis. This non-invasive measurement of the primary ovarian tumor consistently identified patients with median overall survival under 2 years and is significantly associated with progression-free survival ($p < 0.01$).

2.5. Future directions of AI for the early detection and prognosis of ovarian cancer

Conventional statistical methods are limited in their ability to analyze large, complex medical data. AI predictive algorithms seem to improve ovarian cancer diagnostic and prognostic accuracy prior to intervention [61,62], while outperforming most existing conventional methods [59,63], and performing near the same level as some gynecologic oncologists [64,65]. However, the AI algorithm that yields the greatest predictive power for a given set of variables is not yet understood. Future studies looking to improve diagnostic and prognostic accuracy in ovarian cancer need to ensure proper validation of the models to estimate unbiased generalization performance. It is not simply enough to select the approach with the strongest performance on trained data, but it also needs to perform well on data not yet seen by the model. More studies are needed, across different populations, that report on this generalization performance. One of the primary challenges of applying AI methods, especially neural networks, in ovarian cancer is the need for data collection on sufficiently large samples ($n > 1{,}000$) [66] to let the machines learn. Future studies will need to determine ways to increase sample size, possibly from large cohorts or by combining multi-site data, given the prevalence of ovarian cancer is low. One way to overcome the difficulty of increasing sample size in clinical studies is to employ novel technology such as generative adversarial networks [67] to augment existing data. Future studies should apply this in an ovarian cancer population comparable to a previous application in a breast cancer setting by Guan et al. [68] where synthetic data were generated using mammographic images from a digital mammography database. With continued improvements in AI, along with the use of big data and increased efficiency in computational resources, there is great potential for earlier detection of ovarian cancer and improved prognosis.

3. Pancreatic cancer

3.1. Background

Pancreatic ductal adenocarcinoma (PDAC) is the third leading cancer killer in the United States, [69] and ranks seventh globally [70]. The five-year survival for all diagnosed patients is below 10% and is only 3% for metastatic disease [71,72]. This high rate of mortality is in part due to chemotherapy resistance and a lack of targeted treatments [69]. This cancer is often diagnosed at a late stage when resection is not possible [73], and at the time of diagnosis 50% of patients have signs of metastatic disease [74]. Identification of tumors less than 2 cm via CT scan greatly improves the probability of survival [73,75]. However, invasive removal of non-cancerous lesions can increase the risk of morbidity and mortality for healthy patients. To date, few biomarkers have been identified and evaluated for PDAC, further hindering treatment [72,76]. Recently, several studies have successfully applied AI models to the detection and classification of pancreatic cancer from CT images [77].

3.2. Pancreatic cancer biomarkers overview

PDAC is very rare [7], thus screening the general population is neither feasible nor advisable because the rate of false-positives would be high [78], potentially leading to unnecessary interventions [72]. Generally, the age-standardized incidence of PDAC is higher in higher-income countries [79], although prognosis does not differ between high, middle and low income countries [80]. Prevalence increases with age [81], and is correlated with comorbidities such as smoking, diabetes and obesity [82]. In particular, screening of populations identified as high risk for developing PDAC, such as family with an inherited risk, which accounts for about 10% of cases [79], people with pancreatic cystic lesions, and people older than 50 years who are newly diagnosed with type 2 diabetes [83,84] could help to identify precursor lesions while they are still treatable. Nonetheless, early stage tumors can be easily overlooked when using

CT and MRI, and it is possible that CNN models could help fill the gap as a 'second reader.' Further, some studies have suggested that CNN models have higher predictions when integrating image data with health, social media, or other data sources [84]. Although no studies have yet identified imaging biomarkers that are ready for clinical trial [72,76] this integrative approach will likely still greatly improve patient outcomes over current practices.

3.3. Application of AI in pancreatic cancer (diagnosis)

Most AI studies thus far have developed models focused around classification of cancerous lesions and healthy pancreas images using CT images, which are the standard diagnostic procedure for identifying PDAC [85]. Here we detail studies using AI for PDAC diagnosis, data for these studies is usually publically available unless specified herein. Chu et al., [86] used unsupervised clustering to extract 40 relevant features of pancreatic lesions from 190 cancerous and 190 healthy pancreas images. They classified cancerous and normal images using a random forest classification model that had 99.2% accuracy, 100% sensitivity, 98.5% specificity, and 99.9% AUC, correctly identifying all cases of PDAC. Kuwahara et al., [87] used 3,970 images from 50 patients to build a deep learning classification model (convolutional neural network) based on the original algorithm from ResNet50. Their aim was to diagnose intraductal papillary mucinous neoplasms (IPMNs) which are precursor lesions of PDAC. They evaluated their model using an AI prediction value defined as the predictive value of malignant probability averaged across all images for each patient. Their model achieved a mean AI value of 0.808 (probability between 0 and 1), 0.98 ($P < 0.001$) AUC, 95.7% sensitivity, 92.6% specificity, 94% accuracy, which was higher than the human diagnosis (source of statistic not defined). Sekaran et al., [88] used 19,000 publicly available images from 82 patients accessed from The Cancer Image Archive (TCIA). They developed a model that used lump feature detection, which allows for the extraction of a single feature from a noisy background, but they failed to specify how their model performed or make their model publicly available.

3.4. Application of AI in pancreatic cancer (prognosis)

Due to the dismal survival of PDAC patients in later stages, much focus in the field has been trained on building models that can detect cancer at earlier stages while the cancer is still treatable [83]. Thus, models built to identify precursor lesions can increase the likelihood of patient survival, but high-grade precursor lesions can be difficult to differentiate from low-grade lesions that never advance to carcinoma, leading to unnecessary interventions that increase patient morbidity and mortality [72]. As such, developing accurate detection models for high-grade precursor lesions as well as early tumors will significantly improve patient outcomes. To this end, Liu et al., [89] built a model focused on the detection of small tumors called Faster R-CNN that used VGG16. Their model was trained on 4000 images from 238 patients and validated on 1699 images from 100 patients, yielding a model with an AUC (trapezoidal rule) = 96% and 77% precision. Their model required only 0.2 seconds to process on CT image and highlight the advantages of this acceleration compared with clinicians. Likewise, Liu et al., [90] developed a CNN model modified from the Visual Geometry Group (VGG) to detect tumors less than 2 cm, of which 40% evade normal detection. They trained their model on images of 295 cancerous and 250 control patients from East Asian study participants. They validated their model on three datasets, including two East Asian datasets (75 cancerous and 64 controls, and 101 cancerous and 88 controls), and the TCIA dataset of North American samples (281 cancerous and 82 controls), demonstrating one of the first studies on PDAC imaging to include patient images from both East Asian and North American patient populations. Their model performed well when validated on the first East Asian dataset with 97% sensitivity, 100% specificity, 99% accuracy, and 99% AUC. On the second East Asian dataset, the model achieved 99% sensitivity, 99% specificity, 99% accuracy, and 100% AUC. On the North American validation dataset, the model performed less well, with 79% sensitivity, 98% specificity, 83% accuracy, and 92% AUC. Nonetheless, on the combined images of the two East Asian datasets the CNN yielded higher sensitivity than radiologists (98% vs. 93%), and perhaps more importantly, the model identified 11 of the 12 small tumors that were missed by the radiologists, while only missing 3/176 tumors (all less than 1.3 cm), yielding a small-tumor (< 2 cm) sensitivity value of 92.1% for the East Asian dataset and 63.1% for the North American dataset.

3.5. Future directions of AI for PDAC detection and prognosis

In the future, cancer screening may be done on whole regions of the body rather than organ by organ [77],

and a suite of models may be employed specific to each organ. To this aim, Wang et al., [59] developed a multi-organ segmentation model that would identify each organ of interest from abdominal CT images. Their model used statistical fusion of multiple layers and images to segregate organs from one with higher precision (based on Sørensen similarity coefficients and mean surface distances) than existing 2D and 3D batch-based methods. Zhu et al., [91] expanded on Wang et al. [59], to identify regions of interest for radiologists called a multiscale segmentation for classification model. The deep learning model iterates through three input volumes (training on each) of decreasing size to increase the probability of identifying small tumors. They compared their model to both the UNet and VNet algorithms, and trained and validated their model on 439 patients, with 136 cancerous and 303 control patients to achieve 94% sensitivity and 99% specificity. Chu et al., [77] leveraged the model developed by Zhu et al., [91] using CT images from 750 cancerous and 575 control patients. They first isolated the pancreas using multi-organ segmentation, achieving 87.8% accuracy, then classified PDAC cases using CT images from 156 PDAC and 300 control cases, yielding 94.1% sensitivity and 98.5% specificity. Not surprisingly, the model performed less well on tumors < 2 cm in diameter, but accuracy improved somewhat when informed by radiologist input regarding human readable features such as a dilated pancreatic duct. Future work will likely follow this system-wide approach, leveraging models trained on multi-organ-CT images to screen for various cancers at once in conjunction with practitioner input. As PDAC-specific models continue to improve, the early detection of tumors will lead to better patient outcomes and hopefully reduce the exceptionally high mortality rate.

4. Biomarkers and AI

4.1. Regulatory and ethical considerations

Global regulatory authorities continue to track AI technologies used for biomedical discovery and treatment. In Europe, high risk medical devices are regulated via the Conformité Européenne (European Conformity – "CE") mark that indicates that a device meets high safety, environmental, and health standards [92]. China's National Medical Product, similar in function to the United States' FDA, began tracking AI-based medical devices for the first time in 2018 prior to releasing publicly its *Technical Review Guidelines on AI-Assisted Software* in 2019. It has been argued that China's less restrictive data policies enable the nation to "liberate data for public health purposes," expediting their ability to discover new applications of AI/ML across sectors particularly in healthcare [93]. In the US, medical devices, including AI/ML based tools, are approved based on criteria addressing effectiveness and safety. The FDA has taken additional steps to mitigate bias with the release of their "AI/ML Software-as-a-Medical Device" action plan, which calls for greater transparency into the details of the datasets being used to train these AI/ML algorithms [94]. This proposed framework seems to promote detailed demographic breakdowns of datasets for public review. Encouraging researchers to obtain diverse datasets will build trust in medical devices and the algorithms they are built upon [2].

Currently, the FDA has approved several molecular biomarkers for both ovarian and pancreatic cancers, including CA125, HE4, OVA1 test, ROMA test, and hCG for ovarian, and CEA and CA19-9 for pancreatic [95]. In addition to molecular markers, biomarkers identified from cystic fluid and pancreatic juices may be suited for clinical trials soon [76]. Encouragingly, the National Cancer Institute's Early Detection Research Network (EDRN) has identified promising directions for 300+ potential ovarian biomarkers and for 140+ pancreatic biomarkers [96]. One ongoing clinical trial related to AI discovery of novel biomarkers will analyze participant tissue and fluid samples with an AI platform to identify and validate biomarkers for use in early detection of several pancreatic diseases including cancer [97].

The FDA has yet to qualify imaging biomarkers for diagnosis or prognosis of ovarian or pancreatic cancers [98]. The biomarker validation process itself is rigorous, requiring thousands of samples to address potential variance within biomarker expression [99], and the high level of subjectivity inherent to imaging analysis may contribute to slower developments in the approval process for imaging biomarkers. Interpretation of images related to pancreatic cancer specifically presents challenges due to the difficulty in distinguishing conditions within images [100]. Magnetic resonance elastography (MRE) has shown promise in potentially serving as a valid image biomarker for pancreatic cancer, as this method has been used to diagnose lesions of other cancers including liver, breast, and kidney [101].

The use of AI in biomarker discovery is still relatively nascent, and currently FDA-cleared AI algorithms only exist for breast and lung cancer [102]. This is likely due to the much higher prevalence of breast

and lung cancer and subsequent data availability, allowing for robust image training and validation. For reference, TCIA has 32 collections related to lung cancer, 18 collections related to breast cancer, and only two collections related to either ovarian or pancreatic cancer [103]. In the Cancer Genome Atlas, there are 12,027 cases of lung cancer, 9,115 cases of breast cancer, and only 3,401 cases of ovarian cancer and 2,723 cases of pancreatic cancer [104]. However, pancreatic cancer is rapidly growing in prevalence and regular screening may be conducted for high risk patients [84], leading to additional data availability for AI models.

4.2. Bias in AI-driven biomarker discovery and implications for practice

Data availability and bias remains a concern across all cancer types and can render AI models ineffective regardless of application. Algorithms develop biases and produce prejudiced responses when the data that they are trained on are non-representative or incomplete. There are several ways in which bias can manifest in AI algorithms. For example, outputs can underestimate risk if a model is trained on a non-diverse dataset, or measurement bias existing in the data can lead to a discrepancy between what the algorithm should predict and what it actually predicts. Without correction, bias can inhibit models from making confident conclusions. These types of biases lead to inaccurate or unfair algorithms that can have unintentionally harmful consequences to underrepresented or unaccounted for populations.

Racial and ethnic minority groups may be more susceptible to pancreatic cancer due to associated comorbidities [105], but these groups are consistently underrepresented in clinical data [106]. The prevalence of pancreatic cancer is higher in men than women, naturally causing data to exhibit a gender skew [107]; this unequal representation in clinical training data could introduce bias into algorithms [108]. For example, a recent study trained a model to detect skin cancer with a dataset where 65% of the images were from Google Images, and only 5% of photographs were of dark-skinned individuals [109].

The lack of a diverse geographical sample also has significant implications for AI modelling. For example, ImageNet is a repository of millions of annotated images used for image classification, but $\sim$ 45% of data originates from the United States, while only $\sim$ 3% of images come from China or India [109]. A recent analysis also found that a majority of medical data used to train medical AI systems came from a small number of states in the United States, while a majority of states had no representation [110]. This has implications on model performance; Zech et al., [74] showed that their model performed significantly worse when deployed in differing locations from which the data were trained. Poorly trained models of this nature are not rare [111] and can pose significant risk to patients should their care be informed by such models.

4.3. Best practice in AI-driven biomarker discovery

Despite these challenges, US policy makers have prioritized AI, with Congress passing the National Artificial Intelligence Initiative Act, granting over $5B in funding towards AI research [112]. Likewise, EDRN has emphasized the importance of data science and AI to their research [113] which should provide more funding for AI-driven biomarker discovery research. Researchers continue to call for data collection reform to include geographically and racially diverse data [111] as well as rigorous methods testing to facilitate ethical AI [108]. Explainable and trustworthy AI campaigns attempt to rectify "black box" methodologies for algorithm development by constructing interfaces that allow humans to better understand and interrogate the AI model. Easier to understand models increase trust when the user better understands why a certain prediction was generated. However explainability can come at the expense of accuracy [114]. Calls for the "democratization of data" which ties heavily into explainable AI, makes data easier to access and understand to facilitate inclusion by those most susceptible to descrimination and bias [108]. These efforts outlined here can be implemented to help mitigate bias and facilitate reproducibility across the biomedical research enterprise.

5. Conclusion

The use of AI for biomedical research and biomarker discovery continues to hold great promise and will likely be the target of several research studies evaluating AI efficacy. This will be aided by the decreasing cost of compute resources, proliferation of various open source tools, and rapidly evolving biotechnology applications centered around imaging informatics such as pathology and radiology. Challenges associated with data availability, quality, bias, as well as AI transparency and explainability, will likely persist as the field expands further. The stakes are even higher for rare conditions such

as ovarian and pancreatic cancer, where the clinical application of AI is only just beginning. In these cancers, many of the models need to be validated in larger, clinical settings. More importantly, many studies use images that are not publicly available, limiting the pooling of resources that would build more representative and robust models. Larger and more diverse image databases for rare cancers combined across institutions (federated model) will both increase the probability of biomarker discovery and increase model generalizability across racially/ethnically diverse patient cohorts. Greater image availability will also facilitate model validation and reduce bias in cancer diagnosis and prognosis. Further, standardized reporting metrics will allow for quantitative comparisons of models across cohorts, and facilitate the evaluation of models for patient cohorts not used to train the model. Finally, AI based biomarkers will require explainable models. As funding for AI increases, regulatory agencies, research institutions, and other stakeholders need to be prepared to address these challenges in order to make a true impact on biomarker discovery.

References

[1] E.J. Topol, High-performance medicine: The convergence of human and artificial intelligence, *Nat. Med* **25** (2019), 44–56. doi: 10.1038/s41591-018-0300-7.

[2] E. Brodwin, 4 steps for AI developers to build trust in their clinical tools, in: *Promise Peril AI Transform. Health Care*, 2020, pp. 201–203. https://www.statnews.com/2021/01/13/4-steps-for-ai-developers-to-build-trust-in-their-clinical-tools/ (accessed May 25, 2021).

[3] A.L. Fogel and J.C. Kvedar, Artificial intelligence powers digital medicine, *Npj Digit. Med* **1** (2018), 1–4. doi: 10.1038/s41746-017-0012-2.

[4] D.F. Steiner, R. MacDonald, Y. Liu, P. Truszkowski, J.D. Hipp, C. Gammage, F. Thng, L. Peng and M.C. Stumpe, Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer, *Am. J. Surg. Pathol* **42** (2018), 1636–1646. doi: 10.1097/PAS.0000000000001151.

[5] P.A. Keane and E.J. Topol, AI-facilitated health care requires education of clinicians, *The Lancet* **397** (2021), 1254. doi: 10.1016/S0140-6736(21)00722-4.

[6] Artificial Intelligence – National Cancer Institute, 2020. https://www.cancer.gov/research/areas/diagnosis/artificial-intelligence (accessed May 24, 2021).

[7] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA. Cancer J. Clin* **68** (2018), 394–424. doi: 10.3322/caac.21492.

[8] S.B. Coburn, F. Bray, M.E. Sherman and B. Trabert, International patterns and trends in ovarian cancer incidence, overall and by histologic subtype, *Int. J. Cancer* **140** (2017), 2451–2460. doi: 10.1002/ijc.30676.

[9] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D.M. Parkin, M. Piñeros, A. Znaor and F. Bray, Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods, *Int. J. Cancer* **144** (2019), 1941–1953. doi: 10.1002/ijc.31937.

[10] Z. Momenimovahed, A. Tiznobaik, S. Taheri and H. Salehiniya, Ovarian cancer in the world: Epidemiology and risk factors, *Int. J. Womens Health* **11** (2019), 287–299. doi: 10.2147/IJWH.S197604.

[11] W.-L. Yang, Z. Lu and R.C. Bast, The role of biomarkers in the management of epithelial ovarian cancer, *Expert Rev. Mol. Diagn* **17** (2017), 577–591. doi: 10.1080/14737159.2017.1326820.

[12] B.M. Reid, J.B. Permuth and T.A. Sellers, Epidemiology of ovarian cancer: A review, *Cancer Biol. Med* **14** (2017), 9–32. doi: 10.20892/j.issn.2095-3941.2016.0084.

[13] L.A. Torre, B. Trabert, C.E. DeSantis, K.D. Miller, G. Samimi, C.D. Runowicz, M.M. Gaudet, A. Jemal and R.L. Siegel, Ovarian cancer statistics, 2018, *CA. Cancer J. Clin* **68** (2018), 284–296. doi: 10.3322/caac.21456.

[14] S. Lheureux, M. Braunstein and A.M. Oza, Epithelial ovarian cancer: Evolution of management in the era of precision medicine, *CA. Cancer J. Clin* **69** (2019), 280–304. doi: 10.3322/caac.21559.

[15] I. Romero, S. Leskela, B. Mies, A. Velasco and J. Palacious, Morphological and molecular heterogeneity of epithelial ovarian cancer: Therapeutic implications, *EJC Suppl* **15** (2020), 1–15.

[16] G. Chornokur, E.K. Amankwah, J.M. Schildkraut and C.M. Phelan, Global ovarian cancer health disparities, *Gynecol. Oncol* **129** (2013), 258–264. doi: 10.1016/j.ygyno.2012.12.016.

[17] S. Singh, D.V. Saxena, S. Khatri, S. Gupta, J. Garewal and K. Dubey, Histopathological evaluation of ovarian tumors, *Undefined* **2** (2016), 435–439.

[18] L.J. Havrilesky, G.D. Sanders, S. Kulasingam, J.P. Chino, A. Berchuck, J.R. Marks and E.R. Myers, Development of an ovarian cancer screening decision model that incorporates disease heterogeneity: Implications for potential mortality reduction, *Cancer* **117** (2011), 545–553. doi: 10.1002/cncr.25624.

[19] J.A. Rauh-Hain, T.C. Krivak, M.G. Del Carmen and A.B. Olawaiye, Ovarian cancer screening and early detection in the general population, *Rev. Obstet. Gynecol* **4** (2011), 15–21.

[20] US Preventive Services Task Force, D.C. Grossman, S.J. Curry, D.K. Owens, M.J. Barry, K.W. Davidson, C.A. Doubeni, J.W. Epling, A.R. Kemper, A.H. Krist, A.E. Kurth, C.S. Landefeld, C.M. Mangione, M.G. Phipps, M. Silverstein, M.A. Simon and C.-W. Tseng, Screening for ovarian cancer: US preventive services task force recommendation statement, *JAMA* **319** (2018), 588. doi: 10.1001/jama.2017.21926.

[21] S. Fenchel, D. Grab, K. Nuessle, J. Kotzerke, A. Rieber, R. Kreienberg, H.-J. Brambs and S.N. Reske, Asymptomatic adnexal masses: Correlation of FDG PET and histopathologic findings, *Radiology* **223** (2002), 780–788. doi: 10.1148/radiol.2233001850.

[22] S.A. Sohaib, T.D. Mills, A. Sahdev, J.A.W. Webb, P.O. VanTrappen, I.J. Jacobs and R.H. Reznek, The role of magnetic resonance imaging and ultrasound in patients with adnexal masses, *Clin. Radiol* **60** (2005), 340–348. doi: 10.1016/j.crad.2004.09.007.

[23] S. Risum, C. Hogdall, A. Loft, A. Berthelsen, E. Hogdall, L. Nedergaard, L. Lundvall and S. Engelholm, The diagnostic value of PET/CT for primary ovarian cancer – a prospective

study, *Gynecol. Oncol* **105** (2007), 145–149. doi: 10.1016/j.ygyno.2006.11.022.

[24] K.B. Mathieu, D.G. Bedi, S.L. Thrower, A. Qayyum and R.C. Bast, Screening for ovarian cancer: Imaging challenges and opportunities for improvement, *Ultrasound Obstet. Gynecol* **51** (2018), 293–303. doi: 10.1002/uog.17557.

[25] M.A. Rossing, K.G. Wicklund, K.L. Cushing-Haugen and N.S. Weiss, Predictive value of symptoms for early detection of ovarian cancer, *JNCI J. Natl. Cancer Inst* **102** (2010), 222–229. doi: 10.1093/jnci/djp500.

[26] B. Khiewvan, D.A. Torigian, S. Emamzadehfard, K. Paydary, A. Salavati, S. Houshmand, T.J. Werner and A. Alavi, An update on the role of PET/CT and PET/MRI in ovarian cancer, *Eur. J. Nucl. Med. Mol. Imaging* **44** (2017), 1079–1091. doi: 10.1007/s00259-017-3638-z.

[27] V.R. Iyer and S.I. Lee, MRI, CT, and PET/CT for ovarian cancer detection and adnexal lesion characterization, *Am. J. Roentgenol* **194** (2010), 311–321. doi: 10.2214/AJR.09.3522.

[28] M. Montagnana, M. Benati and E. Danese, Circulating biomarkers in epithelial ovarian cancer diagnosis: From present to future perspective, *Ann. Transl. Med* **5** (2017), 276–276. doi: 10.21037/atm.2017.05.13.

[29] R.C. Bast, Z. Lu, C.Y. Han, K.H. Lu, K.S. Anderson, C.W. Drescher and S.J. Skates, Biomarkers and strategies for early detection of ovarian cancer, *Cancer Epidemiol. Biomarkers Prev* **29** (2020), 2504–2512. doi: 10.1158/1055-9965.EPI-20-1057.

[30] R. Mari, E. Mamessier, E. Lambaudie, M. Provansal, D. Birnbaum, F. Bertucci and R. Sabatier, Liquid biopsies for ovarian carcinoma: How blood tests may improve the clinical management of a deadly disease, *Cancers* **11** (2019), 774. doi: 10.3390/cancers11060774.

[31] O. Dorigo and J.S. Berek, Personalizing CA125 levels for ovarian cancer screening, *Cancer Prev. Res. (Phila. Pa.)* **4** (2011), 1356–1359. doi: 10.1158/1940-6207.CAPR-11-0378.

[32] D.L. Meany, L.J. Sokoll and D.W. Chan, Early detection of cancer: Immunoassays for plasma tumor markers, *Expert Opin. Med. Diagn* **3** (2009), 597–605. doi: 10.1517/17530050903266830.

[33] P. Bottoni and R. Scatena, The Role of CA 125 as Tumor Marker: Biochemical and Clinical Aspects, in: R. Scatena (Ed.), Adv. Cancer Biomark, Springer Netherlands, Dordrecht, 2015, pp. 229–244. doi: 10.1007/978-94-017-7215-0_14.

[34] G. Sölétormos, M.J. Duffy, S. Othman Abu Hassan, R.H.M. Verheijen, B. Tholander, R.C. Bast, K.N. Gaarenstroom, C.M. Sturgeon, J.M. Bonfrer, P.H. Petersen, H. Troonen, G. CarloTorre, J. Kanty Kulpa, M.K. Tuxen and R. Molina, Clinical use of cancer biomarkers in epithelial ovarian cancer: Updated guidelines from the european group on tumor markers, *Int. J. Gynecol. Cancer* **26** (2016), 43–51. doi: 10.1097/IGC.0000000000000586.

[35] M. Lycke, B. Kristjansdottir and K. Sundfeldt, A multicenter clinical trial validating the performance of HE4, CA125, risk of ovarian malignancy algorithm and risk of malignancy index, *Gynecol. Oncol* **151** (2018), 159–165. doi: 10.1016/j.ygyno.2018.08.025.

[36] M. Montagnana, E. Danese, S. Giudici, M. Franchi, G.C. Guidi, M. Plebani and G. Lippi, HE4 in ovarian cancer: From discovery to clinical application, *Adv. Clin. Chem* **55** (2011), 1–20.

[37] J. Wang, J. Gao, H. Yao, Z. Wu, M. Wang and J. Qi, Diagnostic accuracy of serum HE4, CA125 and ROMA in patients with ovarian cancer: A meta-analysis, *Tumor Biol* **35** (2014), 6127–6138. doi: 10.1007/s13277-014-1811-6.

[38] L. Zhang, Y. Chen and K. Wang, Comparison of CA125, HE4, and ROMA index for ovarian cancer diagnosis, *Curr. Probl. Cancer* **43** (2019), 135–144. doi: 10.1016/j.currproblcancer.2018.06.001.

[39] G.C. Jayson, E.C. Kohn, H.C. Kitchener and J.A. Ledermann, Ovarian cancer, *The Lancet* **384** (2014), 1376–1388. doi: 10.1016/S0140-6736(13)62146-7.

[40] T.E. Liggett, A. Melnikov, Q. Yi, C. Replogle, W. Hu, J. Rotmensch, A. Kamat, A.K. Sood and V. Levenson, Distinctive DNA methylation patterns of cell-free plasma DNA in women with malignant ovarian tumors, *Gynecol. Oncol* **120** (2011), 113–120. doi: 10.1016/j.ygyno.2010.09.019.

[41] A. Yokoi, J. Matsuzaki, Y. Yamamoto, Y. Yoneoka, K. Takahashi, H. Shimizu, T. Uehara, M. Ishikawa, S. Ikeda, T. Sonoda, J. Kawauchi, S. Takizawa, Y. Aoki, S. Niida, H. Sakamoto, K. Kato, T. Kato and T. Ochiya, Integrated extracellular microRNA profiling for ovarian cancer screening, *Nat. Commun* **9** (2018), 4319. doi: 10.1038/s41467-018-06434-4.

[42] S.S. Buys, Effect of screening on ovarian cancer mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial, *JAMA* **305** (2011), 2295. doi: 10.1001/jama.2011.766.

[43] P.F. Pinsky, K. Yu, B.S. Kramer, A. Black, S.S. Buys, E. Partridge, J. Gohagan, C.D. Berg and P.C. Prorok, Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up, *Gynecol. Oncol* **143** (2016), 270–275. doi: 10.1016/j.ygyno.2016.08.334.

[44] S. Kondalsamy-Chennakesavan, A. Hackethal, D. Bowtell and A. Obermair, Differentiating stage 1 epithelial ovarian cancer from benign ovarian tumours using a combination of tumour markers HE4, CA125, and CEA and patient's age, *Gynecol. Oncol* **129** (2013), 467–471. doi: 10.1016/j.ygyno.2013.03.001.

[45] T. Muinao, H.P. Deka Boruah and M. Pal, Multi-biomarker panel signature as the key to diagnosis of ovarian cancer, *Heliyon* **5** (2019), e02826. doi: 10.1016/j.heliyon.2019.e02826.

[46] R.G. Moore, D.S. McMeekin, A.K. Brown, P. DiSilvestro, M.C. Miller, W.J. Allard, W. Gajewski, R. Kurman, R.C. Bast and S.J. Skates, A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass, *Gynecol. Oncol* **112** (2009), 40–46. doi: 10.1016/j.ygyno.2008.08.031.

[47] R.G. Moore, M.C. Miller, P. Disilvestro, L.M. Landrum, W. Gajewski, J.J. Ball and S.J. Skates, Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass, *Obstet. Gynecol* **118** (2011), 280–288. doi: 10.1097/AOG.0b013e318224fce2.

[48] Z. Zhang and D.W. Chan, The road from discovery to clinical diagnostics: Lessons learned from the first FDA-cleared *in vitro* diagnostic multivariate index assay of proteomic biomarkers, *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol* **19** (2010), 2995–2999. doi: 10.1158/1055-9965.EPI-10-0580.

[49] F.R. Ueland, C.P. Desimone, L.G. Seamon, R.A. Miller, S. Goodrich, I. Podzielinski, L. Sokoll, A. Smith, J.R. van Nagell and Z. Zhang, Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors, *Obstet. Gynecol* **117** (2011), 1289–1297. doi: 10.1097/AOG.0b013e31821b5118.

[50] R.L. Coleman, T.J. Herzog, D.W. Chan, D.G. Munroe, T.C. Pappas, A. Smith, Z. Zhang and J. Wolf, Validation of a

second-generation multivariate index assay for malignancy risk of adnexal masses, *Am. J. Obstet. Gynecol* **215** (2016), 82.e1–82.e11. doi: 10.1016/j.ajog.2016.03.003.

[51] S. Kumari, Serum biomarker based algorithms in diagnosis of ovarian cancer: A review, *Indian J. Clin. Biochem* **33** (2018), 382–386. doi: 10.1007/s12291-018-0786-2.

[52] R.W. Miller, A. Smith, C.P. DeSimone, L. Seamon, S. Goodrich, I. Podzielinski, L. Sokoll, J.R. van Nagell, Z. Zhang and F.R. Ueland, Performance of the american college of obstetricians and gynecologists' ovarian tumor referral guidelines with a multivariate index assay, *Obstet. Gynecol* **117** (2011), 1298–1306. doi: 10.1097/AOG.0b013e31821b1d80.

[53] Q. Chen and J. Zhang, Classification and recognition of ovarian cells based on two-dimensional light scattering technology, *J. Med. Syst* **43** (2019), 127. doi: 10.1007/s10916-019-1211-y.

[54] M. Wu, C. Yan, H. Liu and Q. Liu, Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks, *Biosci. Rep* **38** (2018). doi: 10.1042/BSR20180289.

[55] J. Zhou, Z.Y. Zeng and L. Li, Progress of artificial intelligence in gynecological malignant tumors, *Cancer Manag. Res* **12** (2020), 12823–12840. doi: 10.2147/CMAR.S279990.

[56] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis and D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, *Comput. Struct. Biotechnol. J* **13** (2015), 8–17. doi: 10.1016/j.csbj.2014.11.005.

[57] E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, K. Koseki, Y. Iida, M. Saito, H. Komazaki, J.S. Shapiro, C. Goto, Y. Akiyama, R. Saito, M. Saito, H. Takano, K. Yamada and A. Okamoto, Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res* **25** (2019), 3006–3015. doi: 10.1158/1078-0432.CCR-18-3378.

[58] A. Enshaei, C.N. Robson and R.J. Edmondson, Artificial intelligence systems as prognostic and predictive tools in ovarian cancer, *Ann. Surg. Oncol* **22** (2015), 3970–3975. doi: 10.1245/s10434-015-4475-6.

[59] S. Wang, Z. Liu, Y. Rong, B. Zhou, Y. Bai, W. Wei, W. Wei, M. Wang, Y. Guo and J. Tian, Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer, *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol* **132** (2019), 171–177. doi: 10.1016/j.radonc.2018.10.019.

[60] H. Lu, M. Arshad, A. Thornton, G. Avesani, P. Cunnea, E. Curry, F. Kanavati, J. Liang, K. Nixon, S.T. Williams, M.A. Hassan, D.D.L. Bowtell, H. Gabra, C. Fotopoulou, A. Rockall and E.O. Aboagye, A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer, *Nat. Commun* **10** (2019), 764. doi: 10.1038/s41467-019-08718-9.

[61] A. Hosny, C. Parmar, J. Quackenbush, L.H. Schwartz and H.J.W.L. Aerts, Artificial intelligence in radiology, *Nat. Rev. Cancer* **18** (2018), 500–510. doi: 10.1038/s41568-018-0016-5.

[62] M.H. Hesamian, W. Jia, X. He and P. Kennedy, Deep learning techniques for medical image segmentation: Achievements and challenges, *J. Digit. Imaging* **32** (2019), 582–596. doi: 10.1007/s10278-019-00227-x.

[63] G. Bogani, D. Rossetti, A. Ditto, F. Martinelli, V. Chiappa, L. Mosca, U. Leone Roberti Maggiore, S. Ferla, D. Lorusso and F. Raspagliesi, Artificial intelligence weights the importance of factors predicting complete cytoreduction at secondary cytoreductive surgery for recurrent ovarian cancer, *J. Gynecol. Oncol* **29** (2018), e66. doi: 10.3802/jgo.2018.29.e66.

[64] V. Aramendía-Vidaurreta, R. Cabeza, A. Villanueva, J. Navallas and J.L. Alcázar, Ultrasound image discrimination between benign and malignant adnexal masses based on a neural network approach, *Ultrasound Med. Biol* **42** (2016), 742–752. doi: 10.1016/j.ultrasmedbio.2015.11.014.

[65] F. Sheikhzadeh, R.K. Ward, D. van Niekerk and M. Guillaud, Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks, *PLOS ONE* **13** (2018), e0190783. doi: 10.1371/journal.pone.0190783.

[66] K. Tanabe, M. Ikeda, M. Hayashi, K. Matsuo, M. Yasaka, H. Machida, M. Shida, T. Katahira, T. Imanishi, T. Hirasawa, K. Sato, H. Yoshida and M. Mikami, Comprehensive serum glycopeptide spectra analysis combined with artificial intelligence (CSGSA-AI) to diagnose early-stage ovarian cancer, *Cancers* **12** (2020), 2373. doi: 10.3390/cancers12092373.

[67] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative Adversarial Networks, in: *Proc. Neural Inf. Process. Syst. Conf.*, Montreal, QC, 2014, pp. 2672–2680.

[68] S. Guan, Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks, *J. Med. Imaging* **6** (2019), 1. doi: 10.1117/1.JMI.6.3.031411.

[69] T. Kamisawa, L.D. Wood, T. Itoi and K. Takaori, Pancreatic cancer, *The Lancet* **388** (2016), 73–85. doi: 10.1016/S0140-6736(16)00141-0.

[70] World Health Organization International Agency for Research on Cancer, Pancreas, 2020.

[71] R.L. Siegel, K.D. Miller and A. Jemal, Cancer statistics, 2019, *CA. Cancer J. Clin* **69** (2019), 7–34. doi: 10.3322/caac.21551.

[72] M.R. Young, N. Abrams, S. Ghosh, J.A.S. Rinaudo, G. Marquez and S. Srivastava, Prediagnostic image data, artificial intelligence, and pancreatic cancer: A tell-tale sign to early detection, *Pancreas* **49** (2020), 882–886. doi: 10.1097/MPA.0000000000001603.

[73] J. Hippisley-Cox and C. Coupland, Identifying patients with suspected pancreatic cancer in primary care: Derivation and validation of an algorithm, *Br. J. Gen. Pract. J. R. Coll. Gen. Pract* **62** (2012), e38–45. doi: 10.3399/bjgp12X616355.

[74] L.C. Chu, S. Park, S. Kawamoto, A.L. Yuille, R.H. Hruban and E.K. Fishman, Pancreatic cancer imaging: A new look at an old problem, *Curr. Probl. Diagn. Radiol* **50** (2021), 540–550. doi: 10.1067/j.cpradiol.2020.08.002.

[75] S. Gangi, J.G. Fletcher, M.A. Nathan, J.A. Christensen, W.S. Harmsen, B.S. Crownhart and S.T. Chari, Time interval between abnormalities seen on CT and the clinical diagnosis of pancreatic cancer: Retrospective review of CT scans obtained before diagnosis, *AJR Am. J. Roentgenol* **182** (2004), 897–903. doi: 10.2214/ajr.182.4.1820897.

[76] M.R. Young, P.D. Wagner, S. Ghosh, J.A. Rinaudo, S.G. Baker, K.S. Zaret, M. Goggins and S. Srivastava, Validation of biomarkers for early detection of pancreatic cancer: Summary of the alliance of pancreatic cancer consortia for biomarkers for early detection workshop, *Pancreas* **47** (2018), 135–141. doi: 10.1097/MPA.0000000000000973.

[77] L.C. Chu, S. Park, S. Kawamoto, Y. Wang, Y. Zhou, W. Shen, Z. Zhu, Y. Xia, L. Xie, F. Liu, Q. Yu, D.F. Fouladi, S. Shayesteh, E. Zinreich, J.S. Graves, K.M. Horton, A.L. Yuille, R.H. Hruban, K.W. Kinzler, B. Vogelstein and E.K.

Fishman, Application of deep learning to pancreatic cancer detection: Lessons learned from our initial experience, *J. Am. Coll. Radiol. JACR* **16** (2019), 1338–1342. doi: 10.1016/j.jacr.2019.05.034.

[78] B.J. Kenner, S.T. Chari, D.F. Cleeter and V.L.W. Go, Early detection of sporadic pancreatic cancer: Strategic map for innovation – a white paper, *Pancreas* **44** (2015), 686–692. doi: 10.1097/MPA.0000000000000369.

[79] C. Franck, C. Müller, R. Rosania, R.S. Croner, M. Pech and M. Venerito, Advanced pancreatic ductal adenocarcinoma: Moving forward, *Cancers* **12** (2020), 1955. doi: 10.3390/cancers12071955.

[80] A. McGuigan, P. Kelly, R.C. Turkington, C. Jones, H.G. Coleman and R.S. McCain, Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes, *World J. Gastroenterol* **24** (2018), 4846–4861. doi: 10.3748/wjg.v24.i43.4846.

[81] S. Midha, S. Chawla and P.K. Garg, Modifiable and non-modifiable risk factors for pancreatic cancer: A review, *Cancer Lett* **381** (2016), 269–277. doi: 10.1016/j.canlet.2016.07.022.

[82] GBD 2017 Pancreatic Cancer Collaborators, The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017, *Lancet Gastroenterol. Hepatol* **4** (2019), 934–947. doi: 10.1016/S2468-1253(19)30347-4.

[83] J. Klapman and M.P. Malafa, Early detection of pancreatic cancer: Why, who, and how to screen, *Cancer Control J. Moffitt Cancer Cent* **15** (2008), 280–287. doi: 10.1177/107327480801500402.

[84] S.P. Pereira, L. Oldfield, A. Ney, P.A. Hart, M.G. Keane, S.J. Pandol, D. Li, W. Greenhalf, C.Y. Jeon, E.J. Koay, C.V. Almario, C. Halloran, A.M. Lennon and E. Costello, Early detection of pancreatic cancer, *Lancet Gastroenterol. Hepatol* **5** (2020), 698–710. doi: 10.1016/S2468-1253(19)30416-9.

[85] W. Zhao, L. Shen, B. Han, Y. Yang, K. Cheng, D.A. Toesca, A.C. Koong, D.T. Chang and L. Xing, Markerless pancreatic tumor target localization enabled by deep learning, *Int. J. Radiat. Oncol. Biol. Phys* **105** (2019), 432–439. doi: 10.1016/j.ijrobp.2019.05.071.

[86] L.C. Chu, S. Park, S. Kawamoto, D.F. Fouladi, S. Shayesteh, E.S. Zinreich, J.S. Graves, K.M. Horton, R.H. Hruban, A.L. Yuille, K.W. Kinzler, B. Vogelstein and E.K. Fishman, Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue, *AJR Am. J. Roentgenol* **213** (2019), 349–357. doi: 10.2214/AJR.18.20901.

[87] T. Kuwahara, K. Hara, N. Mizuno, N. Okuno, S. Matsumoto, M. Obata, Y. Kurita, H. Koda, K. Toriyama, S. Onishi, M. Ishihara, T. Tanaka, M. Tajika and Y. Niwa, Usefulness of deep learning analysis for the diagnosis of malignancy in intraductal papillary mucinous neoplasms of the pancreas, *Clin. Transl. Gastroenterol* **10** (2019). doi: 10.14309/ctg.0000000000000045.

[88] K. Sekaran, P. Chandana, N.M. Krishna and S. Kadry, Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer, *Multimed. Tools Appl* **79** (2020), 10233–10247. doi: 10.1007/s11042-019-7419-5.

[89] S.-L. Liu, S. Li, Y.-T. Guo, Y.-P. Zhou, Z.-D. Zhang, S. Li and Y. Lu, Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network, *Chin. Med. J. (Engl.)* **132** (2019), 2795–2803. doi: 10.1097/CM9.0000000000000544.

[90] K.-L. Liu, T. Wu, P.-T. Chen, Y.M. Tsai, H. Roth, M.-S. Wu, W.-C. Liao and W. Wang, Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: A retrospective study with cross-racial external validation, *Lancet Digit. Health* **2** (2020), e303–e313. doi: 10.1016/S2589-7500(20)30078-9.

[91] Z. Zhu, Y. Xia, L. Xie, E.K. Fishman and A.L. Yuille, Multi-scale Coarse-to-Fine Segmentation for Screening Pancreatic Ductal Adenocarcinoma, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap and A. Khan (Eds.), Med. Image Comput. Comput. Assist. Interv. – MICCAI 2019, Springer International Publishing, Cham, 2019, pp. 3–12. doi: 10.1007/978-3-030-32226-7_1.

[92] U.J. Muehlematter, P. Daniore and K.N. Vokinger, Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): A comparative analysis, *Lancet Digit. Health* **3** (2021), e195–e203. doi: 10.1016/S2589-7500(20)30292-2.

[93] L. Zhang, H. Wang, Q. Li, M.-H. Zhao and Q.-M. Zhan, Big data and medical research in China, *BMJ* (2018), j5910. doi: 10.1136/bmj.j5910.

[94] U.S. Food & Drug Administration, Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, U.S. Food & Drug Administration, 2021.

[95] A. Kirwan, M. Utratna, M.E. O'Dwyer, L. Joshi and M. Kilcoyne, Glycosylation-based serum biomarkers for cancer diagnostics and prognostics, *BioMed Res. Int* **2015** (2015), 1–16. doi: 10.1155/2015/490531.

[96] Biomarkers, *Early Detect. Res. Netw.* (n.d.). https://edrn.nci.nih.gov/data-and-resources/biomarkers (accessed May 8, 2021).

[97] Pancreatic Cancer Research Team, Project Survival-Prospective Biomarker Discovery to Transform Diagnosis and Treatment for Patients With Pancreatic Diseases and Cancer, clinicaltrials.gov, 2019. https://clinicaltrials.gov/ct2/show/NCT02781012 (accessed May 23, 2021).

[98] Center for Drug Evaluation and Research, List of Qualified Biomarkers, *FDA*. 2020. https://www.fda.gov/drugs/biomarker-qualification-program/list-qualified-biomarkers (accessed May 25, 2021).

[99] E.R. Sauter, Reliable biomarkers to identify new and recurrent cancer, *Eur. J. Breast Health* **13** (2017), 162–167. doi: 10.5152/ejbh.2017.3635.

[100] L. Zhang, S. Sanagapalli and A. Stoita, Challenges in diagnosis of pancreatic cancer, *World J. Gastroenterol* **24** (2018), 2047–2060. doi: 10.3748/wjg.v24.i19.2047.

[101] I. Dregely, D. Prezzi, C. Kelly-Morland, E. Roccia, R. Neji and V. Goh, Imaging biomarkers in oncology: Basics and application to MRI: MRI Biomarkers in Oncology, *J. Magn. Reson. Imaging* **48** (2018), 13–26. doi: 10.1002/jmri.26058.

[102] American College of Radiology Data Science Institute, FDA Cleared AI Algorithms, 2021. https://models.acrdsi.org/ (accessed May 5, 2021).

[103] The Cancer Imaging Archive, TCIA Collections, *Cancer Imaging Arch. TCIA*. (n.d.). https://www.cancerimagingarchive.net/collections/ (accessed May 16, 2021).

[104] National Cancer Institute, Genomic Data Commons Data Portal, *GDC Data Portal.* (n.d.). https://portal.gdc.cancer.gov/ (accessed May 16, 2021).

[105] American Cancer Society, Pancreatic Cancer Risk Factors, *Am. Cancer Soc.* 2020. https://www.cancer.org/cancer/pa

ncreatic-cancer/causes-risks-prevention/risk-factors.html (accessed May 15, 2021).
[106] M.A. Ma, D.E. Gutiérrez, J.M. Frausto and W.K. Al-Delaimy, Minority representation in clinical trials in the United States, *Mayo Clin. Proc* **96** (2021), 264–266. doi: 10.1016/j.mayocp.2020.10.027.
[107] American Cancer Society, Key Statistics for Pancreatic Cancer, *Am. Cancer Soc.* 2021. https://www.cancer.org/cancer/pancreatic-cancer/about/key-statistics.html (accessed May 25, 2021).
[108] S. Leavy, B. O'Sullivan and E. Siapera, Data, Power and Bias in Artificial Intelligence, *ArXiv200807341 Cs.* 2020. http://arxiv.org/abs/2008.07341 (accessed May 25, 2021).
[109] J. Zou and L. Schiebinger, AI can be sexist and racist – it's time to make it fair, *Nature* **559** (2018), 324–326. doi: 10.1038/d41586-018-05707-8.
[110] R. Robbins, Medical AI systems are disproportionately built with data from just three states, new research finds, in: Promise Peril AI Transform. Health Care, 2020, pp. 111–113.
[111] E. Wu, K. Wu, R. Daneshjou, D. Ouyang, D.E. Ho and J. Zou, How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals, *Nat. Med* **27** (2021), 582–584. doi: 10.1038/s41591-021-01312-x.
[112] E. Johnson, H.R. 6216 (116th): National Artificial Intelligence Initiative Act of 2020, 2020.
[113] S. Srivastava, Early Detection Research Network, 2021.
[114] P. Linardatos, V. Papastefanopoulos and S. Kotsiantis, Explainable AI: A review of machine learning interpretability methods, *Entropy* **23** (2020), 18. doi: 10.3390/e23010018.