ASPIRA WOMEN'S HEALTH®

Harnessing AI for the development of a blood based diagnostic test for Adnexal Mass Risk Assessment

Featuring: Manjusha Roy Choudhury

Date: 11/14/2024

Harnessing AI for the development of a blood based diagnostic test for Adnexal Mass Risk Assessment

- **Development of a deep neural network-based model for clinical management of patients with adnexal mass.**

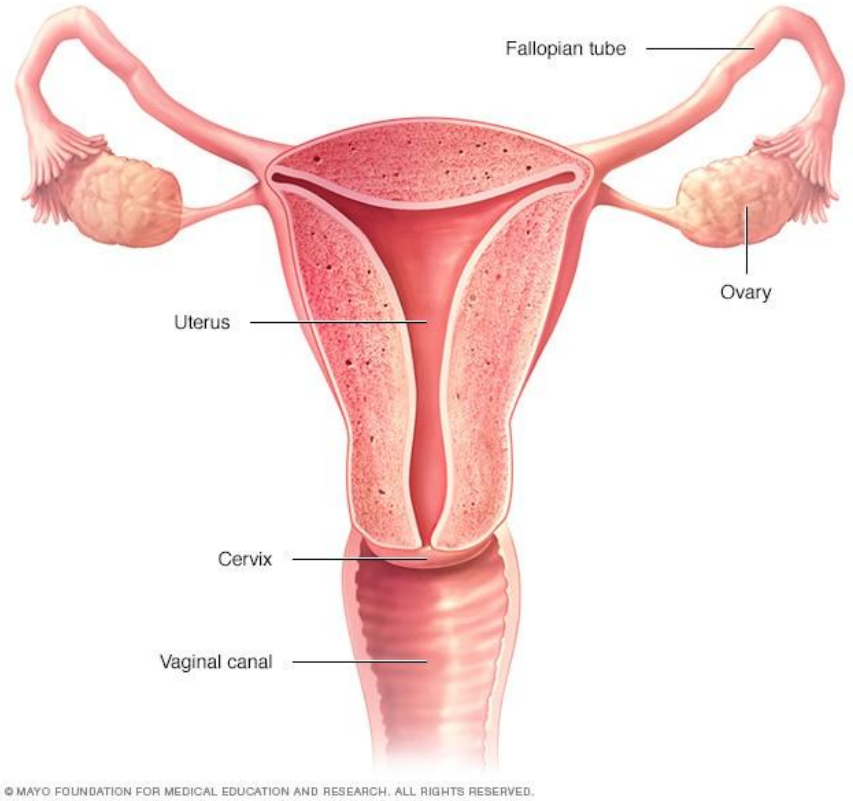
- Clinical Utility of the diagnostic test: Aid the physician in surgical consideration decision for adnexal mass risk.

- Understanding the reliability and accuracy of AI powered diagnostic.

Adnexal Mass

• What is an adnexal mass?

"A lump in tissue near the uterus, usually in the ovary or fallopian tube. Adnexal masses include ovarian cysts, ectopic (tubal) pregnancies, and benign (not cancer) or malignant (cancer) tumors."



• Prevalence of adnexal mass

5-10% of women in US will develop an adnexal mass at some point in their lifetime.

• Malignancy rate of adnexal mass

5-10% of patients scheduled for surgery are found to be malignant.

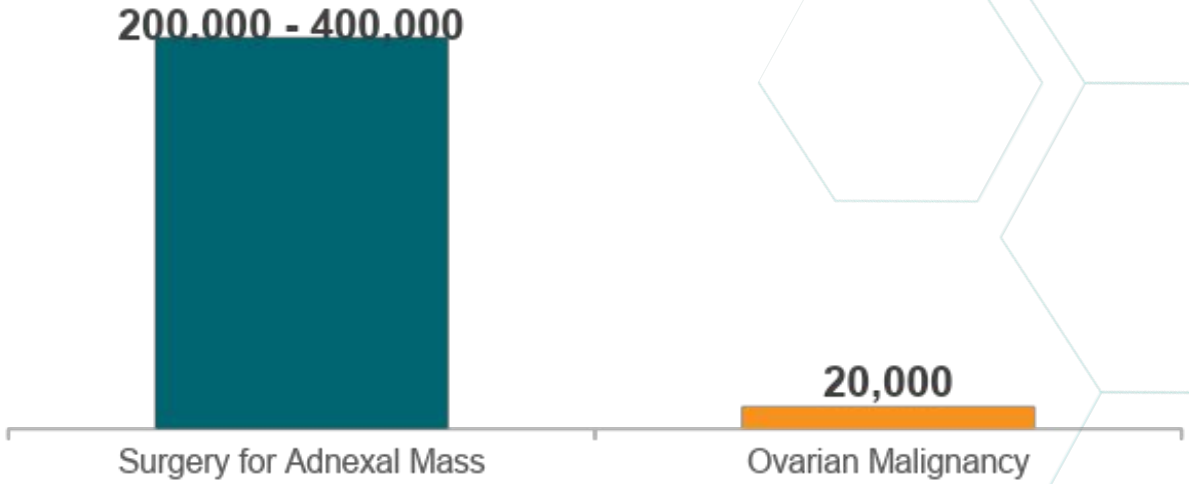
1. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/adnexal-mass 2. https://www.mayoclinic.org/diseases-conditions/adnexal-tumors/symptoms-causes/syc-20355053#dialogId40315602

US Annual Statistic for Adnexal Masses

Each Year:

1.2-1.5M

Adnexal Masses Diagnosed

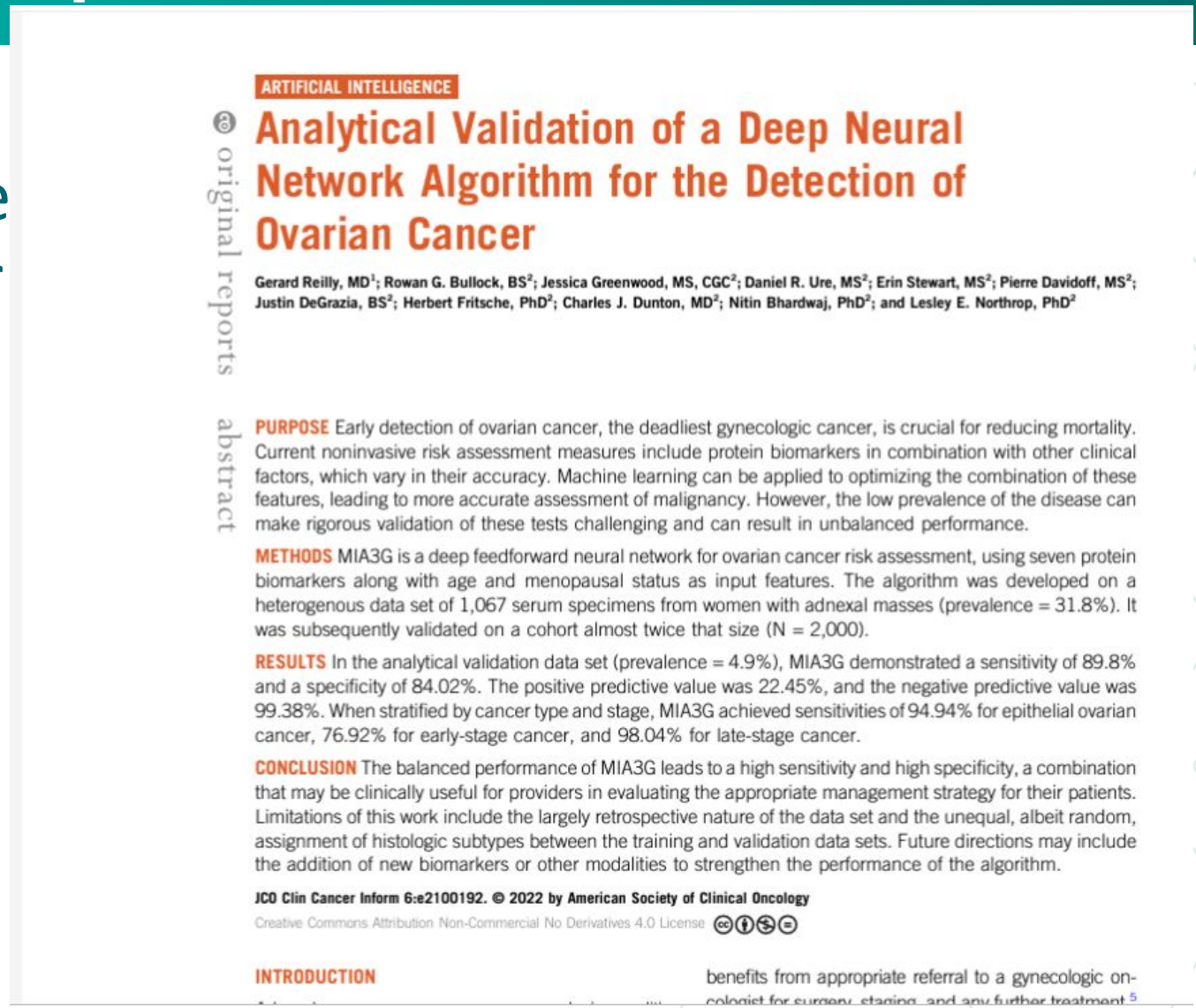


In the United States, there are approximately 9.1 surgeries per malignancy compared to the European International Ovarian Tumor Analysis center trials, with only 2.3 (oncology centers) and 5.9 (other centers) reported surgeries per malignancy, suggesting that there is room to improve our preoperative assessments.

1. Pavlik, E. J., Ueland, F. R., Miller, R. W., Ubellacker, J. M., DeSimone, C. P., Elder, J., ... & van Nagell Jr, J. R. (2013). Frequency and disposition of ovarian abnormalities followed with serial transvaginal ultrasonography. *Obstetrics & Gynecology*, *122*(2 PART 1), 210-217.
2. Ueland, F. R., & Fredericks, T. I. (2018). Ovarian masses: Surgery or surveillance. OBG Manag, 30(6), 17-26.
3. U.S. Census Bureau (2020). US Census Briefs, Age & Sex Composition: 2020, Table 2. Retrieved from https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-06.pdf.
4. Moore, R. G., McMeekin, D. S., Brown, A. K., DiSilvestro, P., Miller, M. C., Allard, W. J., ... & Skates, S. J. (2009). A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecologic oncology*, *112*(1), 40-46.
5. https://seer.cancer.gov/statfacts/html/ovary.html

Published OvaWatch Development

"Analytical Validation of a Dee Neural Network Algorithm for the Detection of Ovarian Cancer," was published in *JCO Clinical Cancer Informatics* in June 2022



ARTICIFIAL INTELLIGENCE

original reports abstract

Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer

Gerard Reilly, MD[1]; Rowan G. Bullock, BS[2]; Jessica Greenwood, MS, CGC[2]; Daniel R. Ure, MS[2]; Erin Stewart, MS[2]; Pierre Davidoff, MS[2]; Justin DeGrazia, BS[2]; Herbert Fritsche, PhD[2]; Charles J. Dunton, MD[2]; Nitin Bhardwaj, PhD[2]; and Lesley E. Northrop, PhD[2]

PURPOSE Early detection of ovarian cancer, the deadliest gynecologic cancer, is crucial for reducing mortality. Current noninvasive risk assessment measures include protein biomarkers in combination with other clinical factors, which vary in their accuracy. Machine learning can be applied to optimizing the combination of these features, leading to more accurate assessment of malignancy. However, the low prevalence of the disease can make rigorous validation of these tests challenging and can result in unbalanced performance.

METHODS MIA3G is a deep feedforward neural network for ovarian cancer risk assessment, using seven protein biomarkers along with age and menopausal status as input features. The algorithm was developed on a heterogenous data set of 1,067 serum specimens from women with adnexal masses (prevalence = 31.8%). It was subsequently validated on a cohort almost twice that size (N = 2,000).

RESULTS In the analytical validation data set (prevalence = 4.9%), MIA3G demonstrated a sensitivity of 89.8% and a specificity of 84.02%. The positive predictive value was 22.45%, and the negative predictive value was 99.38%. When stratified by cancer type and stage, MIA3G achieved sensitivities of 94.94% for epithelial ovarian cancer, 76.92% for early-stage cancer, and 98.04% for late-stage cancer.

CONCLUSION The balanced performance of MIA3G leads to a high sensitivity and high specificity, a combination that may be clinically useful for providers in evaluating the appropriate management strategy for their patients. Limitations of this work include the largely retrospective nature of the data set and the unequal, albeit random, assignment of histologic subtypes between the training and validation data sets. Future directions may include the addition of new biomarkers or other modalities to strengthen the performance of the algorithm.

JCO Clin Cancer Inform 6:e2100192.

INTRODUCTION

benefits from appropriate referral to a gynecologic on-

cologist for surgery, staging, and any further treatment.[5]

1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Feature Set

Biomarker Features

- **PreAlb** — Malnutrition, liver disease, acute inflammation
- **CA125** — Ovarian Cancer
- **HE4** — Ovarian & Endometrial Cancer
- **FSH** — Ovulation, Ovarian Reserve, Menopause etc.
- **B2M** — Cancers, kidney, autoimmune disorders
- **TRF** — Iron level, Liver Failure, Ovarian Cancer, COVID19
- **ApoA1** — Cancers, Cardiovascul-ar disease, stroke etc.

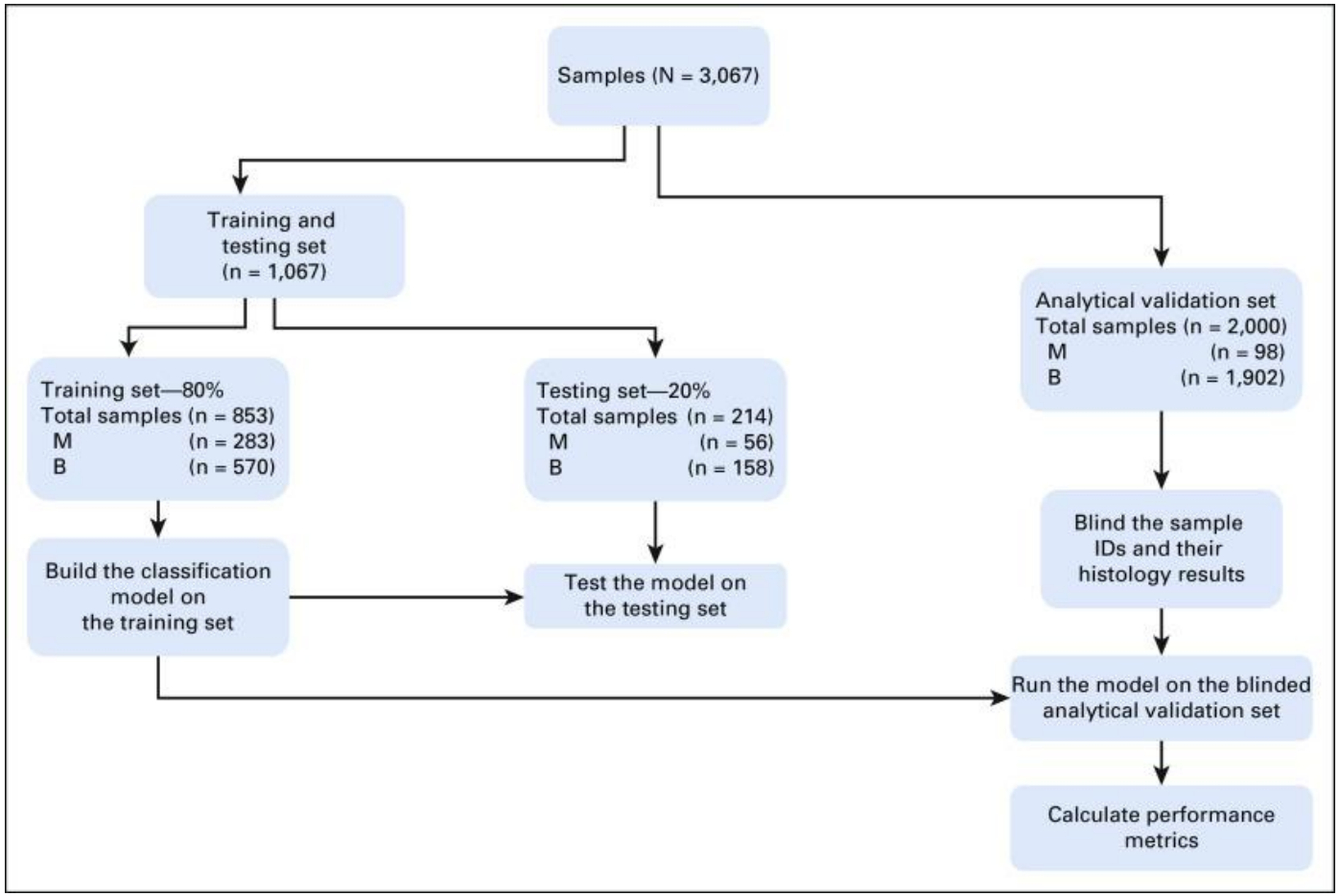
Demographic Features

- **Age**
- **Menopausal Status**

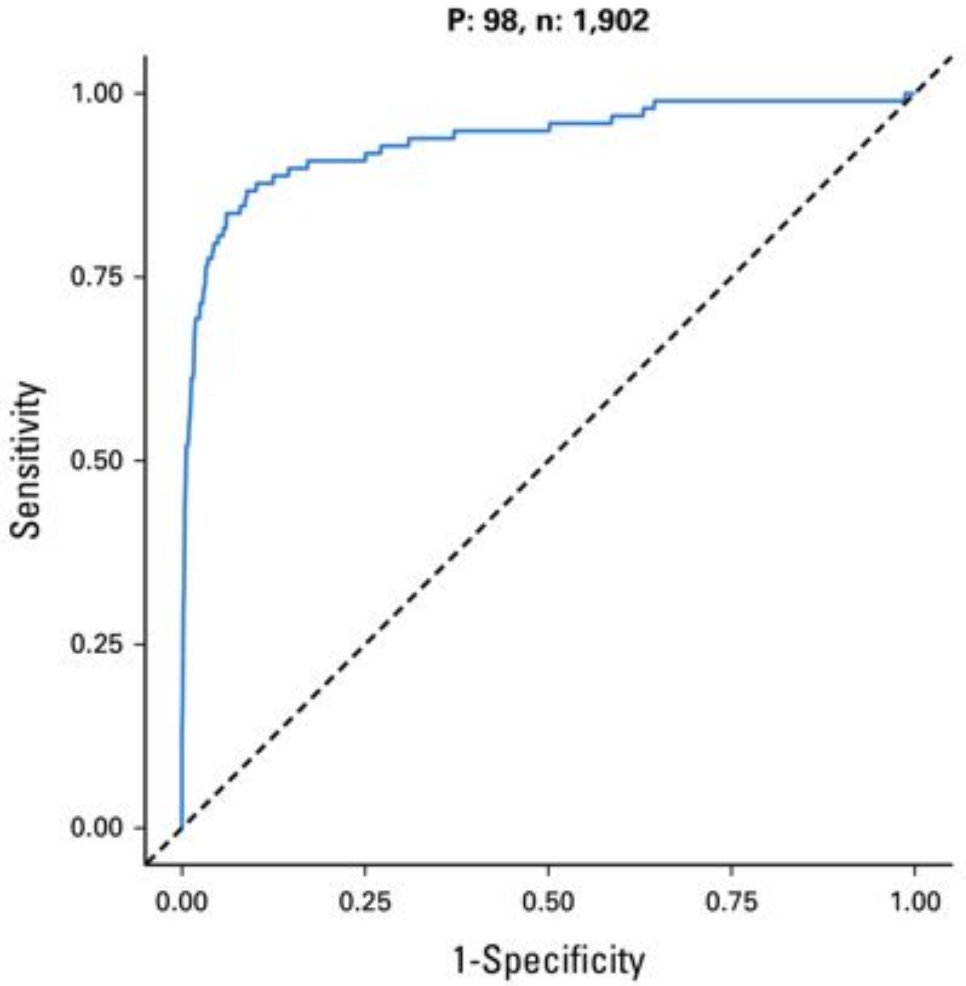
1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Data & OvaWatch Performance at a Glance

Workflow for development of OvaWatch

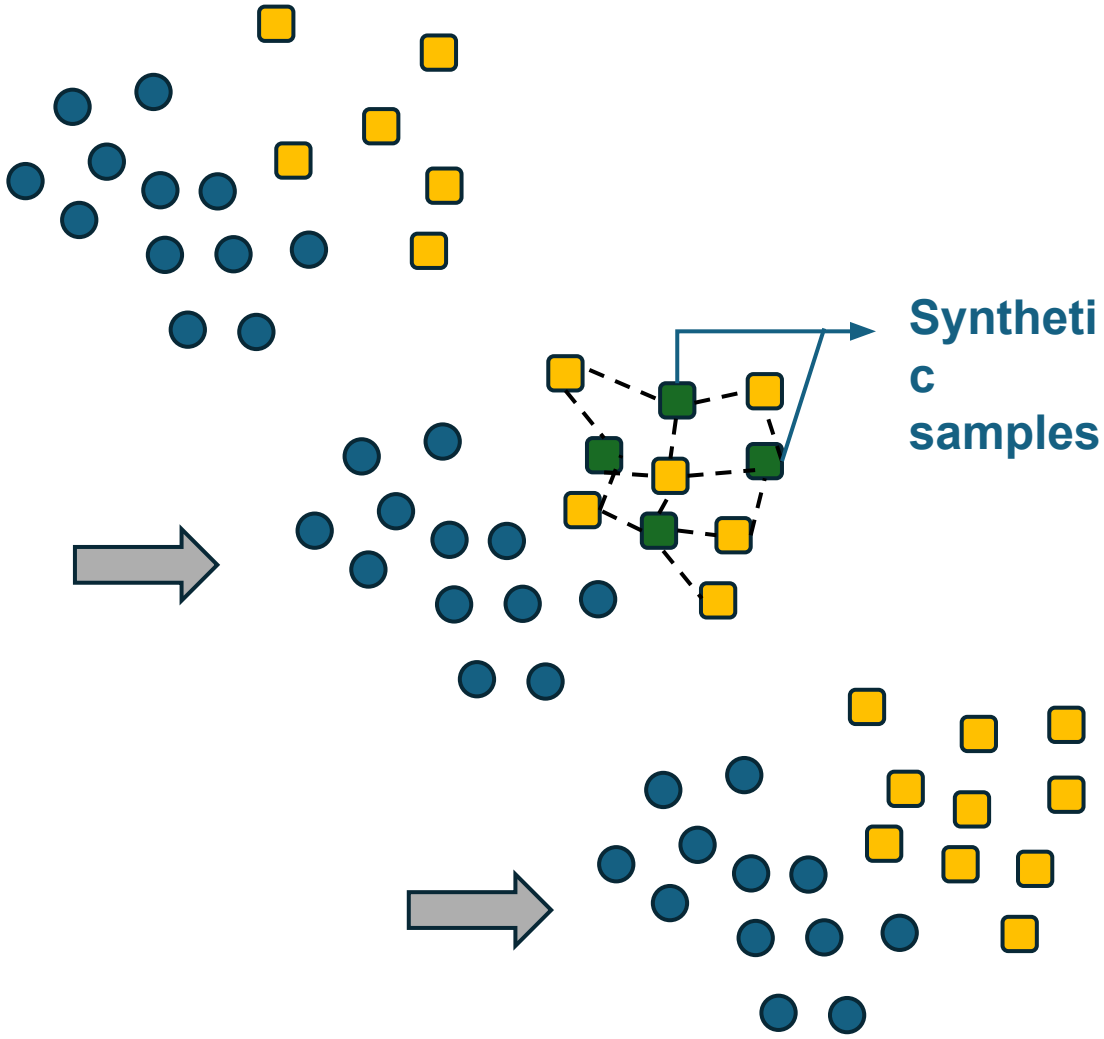


OvaWatch Performance – ROC Curve



1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Data Preparation – Data Augmentation

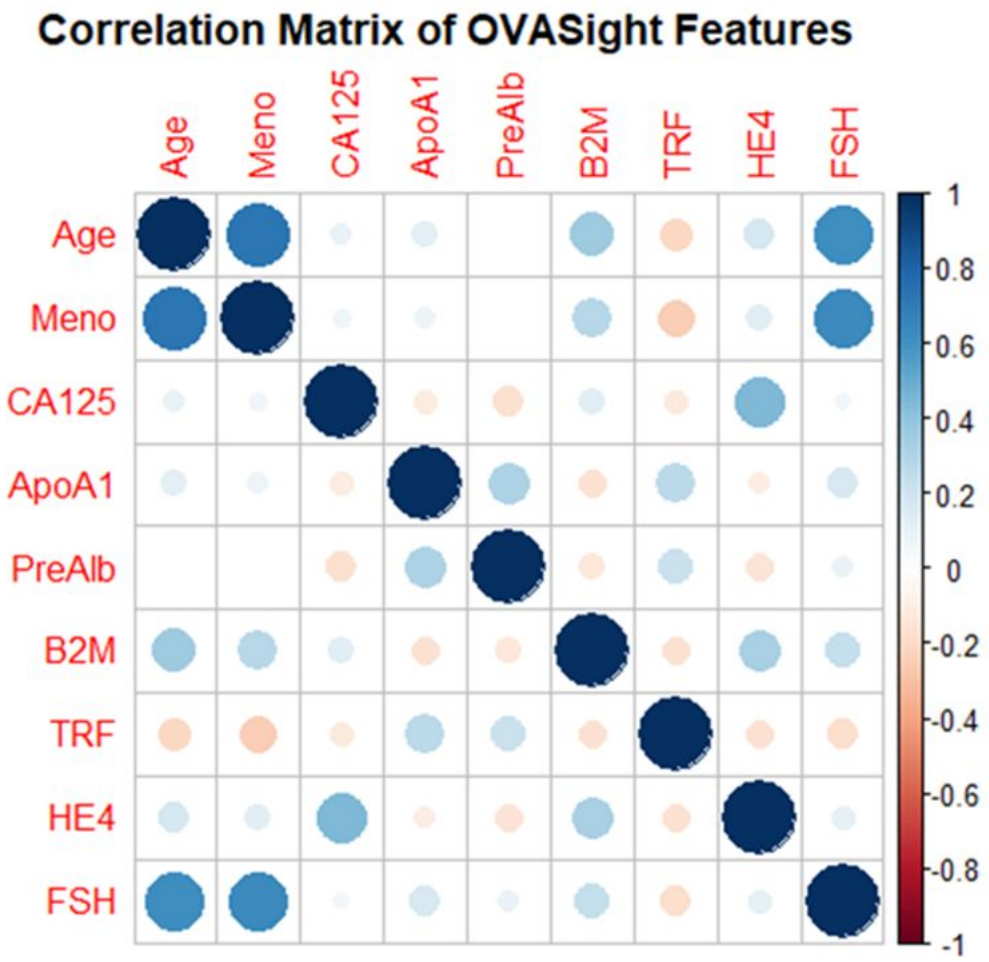


SMOTE/Borderline SMOTE

- **Minority Class** largely **underrepresented** and is class of interest, difficult for algorithm to learn decision boundary.
- SMOTE first selects a minority class instance a at random and finds **its k nearest minority class neighbors**.
- Synthetic instances are created by choosing one k nearest neighbors and finding a value between the two in the shared feature space.
- **Borderline SMOTE** focuses on the **instances of the minority class that are misclassified**, oversampling more of the difficult instances of classification.

1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Feature Correlation



- Correlation observed : **Age and Menopausal Status** . Removing menopausal status led to a mean of 3.8% (90.3% to 86.5%) decrease in sensitivity in the test data division.

- Correlation observed : **FSH with both age Menopausal status** . Removing FSH from the algorithm led to a 5.5% decrease in specificity (86.98% to 81.44%).

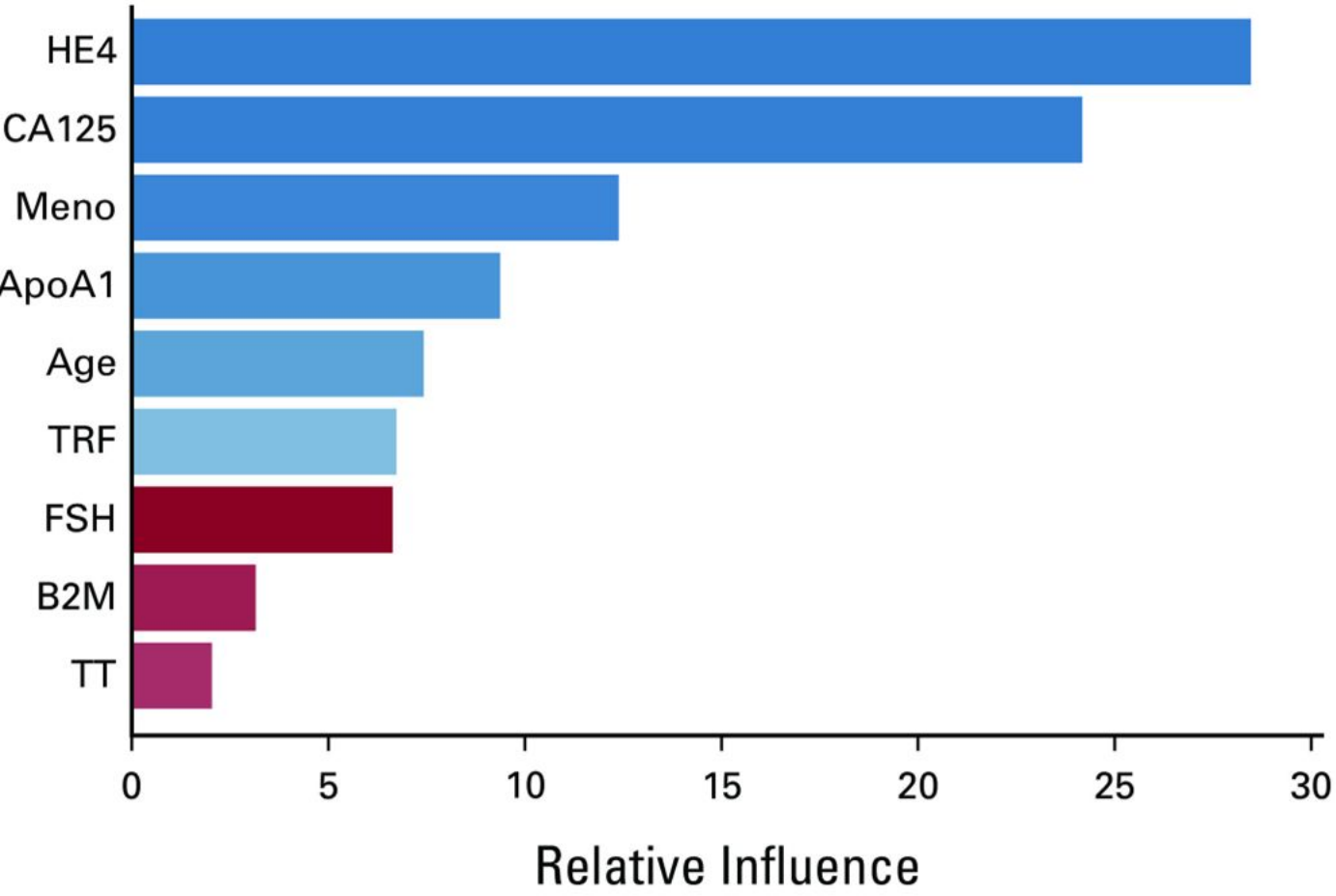
- There were no other correlations in the data that were either ≥ 0.5 or ≤ -0.5.

- All Features Retained.

1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Features and Algorithm Selection

Feature Importance



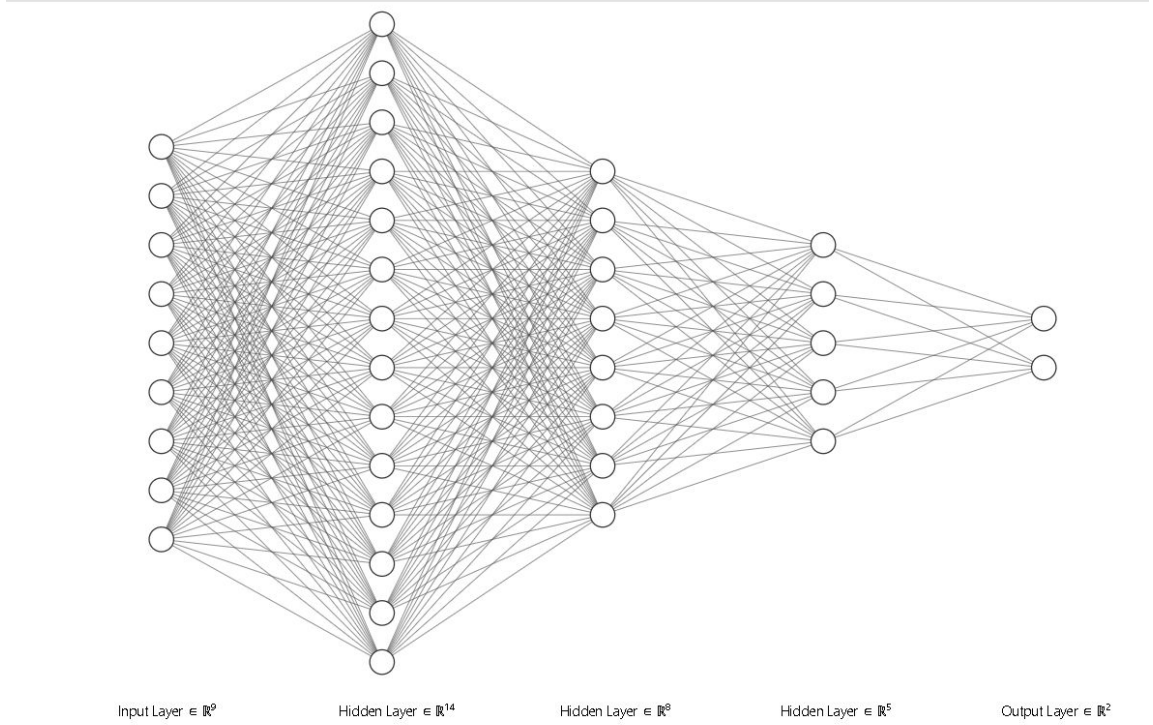
Performance comparison with other algorithms

TABLE A2. Performance of Other Methods in Comparison With Neural Networks, Which Demonstrated Highest Sensitivity and NPV

Model	Sens	Spec	PPV	NPV
C5.0	82.65	91.06	32.27	99.03
Naive Bayesian classifier	72.45	88.49	24.48	98.42
Boosted logistic regression	86.73	81.13	19.14	99.16
SVM with linear kernel	83.67	82.54	19.81	98.99
Boosted smoothing spline	79.59	86.54	23.35	98.80
Generalized linear model	83.67	83.39	20.60	99.00
Self-organizing maps	77.17	80.54	16.10	98.65
Heteroscedastic discriminatory analysis	59.18	98.26	63.74	97.90
Neural network	89.80	84.02	22.45	99.38

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity; SVM, support vector machine.

Deep Neural Network Architecture



*Not the actual architecture for the deep neural network.

- Neural network has multiple hidden layers, each with their own activation weighted nodes and activation function.
- **Regularized** using node dropout.
- The **final layer of the neural network has two nodes** and uses the ***softmax*** function to assign a binary classification: low or elevated risk of malignancy.

1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

OvaWatch Performance – Test Dataset

TABLE A1. Performance of MIA3G in the Test Data Set

Group	Malig	Benign	TP	TN	FP	FN	Sens (%)	Spec (%)	PPV (%)	NPV (%)
All	56	158	51	139	19	5	91.07	87.97	72.86	96.53
Premenopausal	18	87	16	83	4	2	88.89	95.40	80.00	97.65
Postmenopausal	38	71	35	56	15	3	92.11	78.87	70.00	94.92
EOC	45	—	42	—	—	3	93.33	—	—	—
Non-EOC	5	—	5	—	—	0	100.00	—	—	—
Stage I	15	—	12	—	—	3	80.00	—	—	—
Stage II	5	—	5	—	—	0	100.00	—	—	—
Stage III	24	—	24	—	—	0	100.00	—	—	—
Stage IV	4	—	4	—	—	0	100.00	—	—	—
Early stage (I and II)	20	—	17	—	—	3	85.00	—	—	—
Late stage (III and IV)	28	—	28	—	—	0	100.00	—	—	—
Not staged	2	—	2	—	—	0	100.00	—	—	—
Not primary to the ovary	6	—	4	—	—	2	66.67	—	—	—
LMP	—	6	—	3	3	—	—	50.00	—	—
Other benigns	—	152	—	136	16	—	—	89.47	—	—

NOTE. The number of cases or metrics not applicable for that category are displayed by —.

Abbreviations: EOC, epithelial ovarian cancer; FN, false negative; FP, false positive; LMP, low malignant potential/borderline tumor; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity; TN, true negative; TP, true positive.

1. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P, DeGrazia J, Fritsche H, Dunton CJ, Bhardwaj N, Northrop LE. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022 Jun;6:e2100192. doi: 10.1200/CCI.21.00192. PMID: 35671415; PMCID: PMC9225600.

Harnessing AI for the development of a blood based diagnostic test for Adnexal Mass Risk Assessment

- Development of a deep neural network-based model for clinical management of patients with adnexal mass.

- **Clinical Utility of the diagnostic test: Aid the physician in surgical consideration decision for adnexal mass risk.**

- Understanding the reliability and accuracy of AI powered diagnostic.

Newly Published OvaWatch Study

"Ovarian Cancer Surgical Consideration is Markedly Improved by the Neural Network Powered- MIA3G Multivariate Index Assay," was published in *Frontiers in Medicine* in early May 2024

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 02 May 2024
DOI 10.3389/fmed.2024.1374836

Check for updates

OPEN ACCESS

EDITED BY
Carmine Conte,
Agostino Gemelli University Polyclinic (IRCCS), Italy

REVIEWED BY
José Luis Sánchez Iglesias,
Gynecology Oncology, Spain
Kah Teik Chew,
Universiti Kebangsaan Malaysia, Malaysia

*CORRESPONDENCE
Ryan T. Phan
rphan@prevenXhealth.com

†PRESENT ADDRESS
Ryan T. Phan,
PrevenX Health, Palo Alto, CA, United States

RECEIVED 22 January 2024
ACCEPTED 11 April 2024
PUBLISHED 02 May 2024

CITATION
Roy Choudhury M, Pappas TC, Twiggs LB, Caoili E, Fritsche H and Phan RT (2024) Ovarian Cancer surgical consideration is markedly improved by the neural network powered-MIA3G multivariate index assay. *Front. Med.* 11:1374836. doi: 10.3389/fmed.2024.1374836

COPYRIGHT
© 2024 Roy Choudhury, Pappas, Twiggs, Caoili, Fritsche and Phan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Ovarian Cancer surgical consideration is markedly improved by the neural network powered-MIA3G multivariate index assay

Manjusha Roy Choudhury[1], Todd C. Pappas[1], Leo B. Twiggs[2], Emma Caoili[3], Herbert Fritsche[4] and Ryan T. Phan[1,2,4*†]

[1]Department of Research and Development, Aspira Women's Health, Austin, TX, United States, [2]Division of Clinical Operations and Medical Affairs, Aspira Women's Health, Austin, TX, United States, [3]Department of Regulatory Affairs and Quality Assurance, Aspira Women's Health, Shelton, CT, United States, [4]Aspira Labs, Aspira Women's Health, Austin, TX, United States

Background: Surgery remains the main treatment option for an adnexal mass suspicious of ovarian cancer. The malignancy rate is, however, only 10–15% in women undergoing surgery. This results in a high number of unnecessary surgeries. A surveillance-based approach is recommended to form the basis for surgical referrals. We have previously reported the clinical performance of MIA3G, a deep neural network-based algorithm, for assessing ovarian cancer risk. In this study, we show that MIA3G markedly improves the surgical selection for women presenting with adnexal masses.

Methods: MIA3G employs seven serum biomarkers, patient age, and menopausal status. Serum samples were collected from 785 women (IQR: 39–55 years) across 12 centers that presented with adnexal masses. MIA3G risk scores were calculated for all subjects in this cohort. Physicians had no access to the MIA3G risk score when deciding upon a surgical referral. The performance of MIA3G for surgery referral was compared to clinical and surgical outcomes. MIA3G was also tested in an independent cohort comprising 29 women across 14 study sites, in which the physicians had access to and utilized MIA3G prior to surgical consideration.

Results: When compared to the actual number of surgeries ($n = 207$), referrals based on the MIA3G score would have reduced surgeries by 62% ($n = 79$). The reduction was higher in premenopausal patients (77%) and in patients ≤55 years old (70%). In addition, a 431% improvement in malignancy prediction would have been observed if physicians had utilized MIA3G scores for surgery selection. The accuracy of MIA3G referral was 90.00% (CI 87.89–92.11), while only 9.18% accuracy was observed when the MIA3G score was not used. These results were corroborated in an independent multi-site study of 29 patients in which the physicians utilized MIA3G in surgical consideration. The surgery reduction was 87% in this cohort. Moreover, the accuracy and concordance of MIA3G in this independent cohort were each 96.55%.

Conclusion: These findings demonstrate that MIA3G markedly augments the physician's decisions for surgical intervention and improves malignancy prediction in women presenting with adnexal masses. MIA3G utilization as a clinical diagnostic tool might help reduce unnecessary surgeries.

01 frontiersin.org

1. Roy Choudhury M, Pappas TC, Twiggs LB, Caoili E, Fritsche H, Phan RT. Ovarian Cancer surgical consideration is markedly improved by the neural network powered-MIA3G multivariate index assay. Front Med. 2024;11:1374836. doi: 10.3389/fmed.2024.1374836.

The Potential of Utilizing OvaWatch for Surgery Reduction

Differential Surgery Reduction

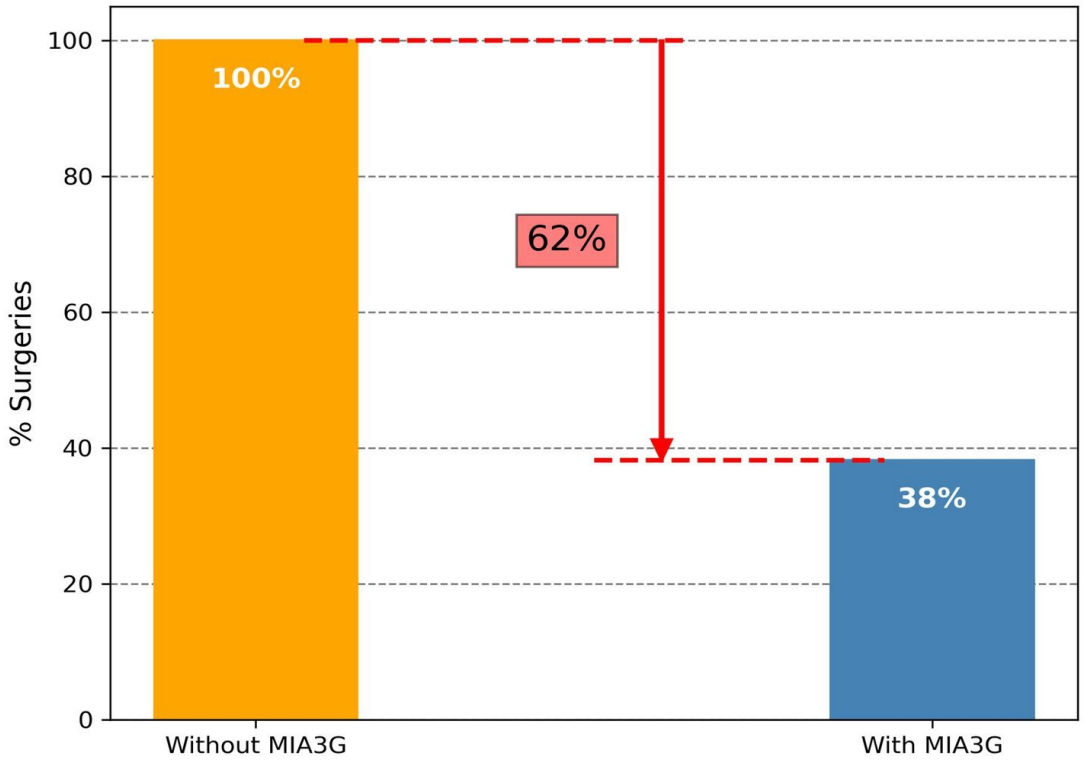


FIGURE 1
Differential surgery referrals without and with MIA3G stratification. The orange bar depicts surgery referrals without MIA3G (considered as 100%), and the blue bar depicts surgery referrals with MIA3G stratification (normalized to 100%).

Improvement in Malignancy Prediction Value

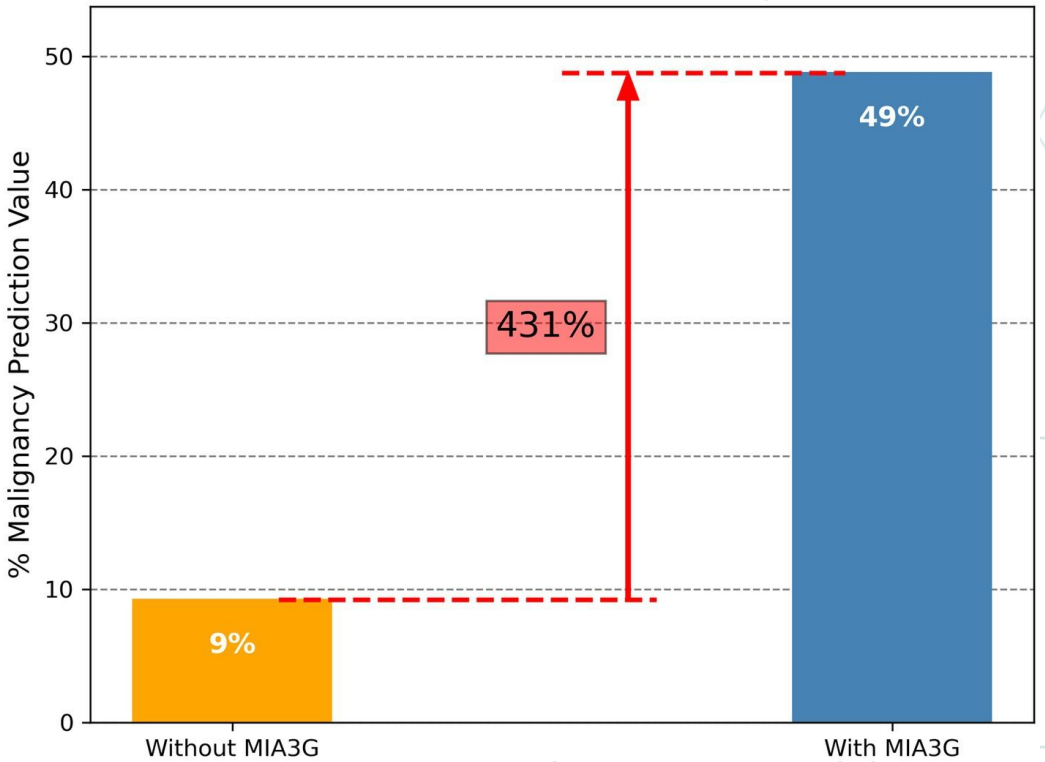
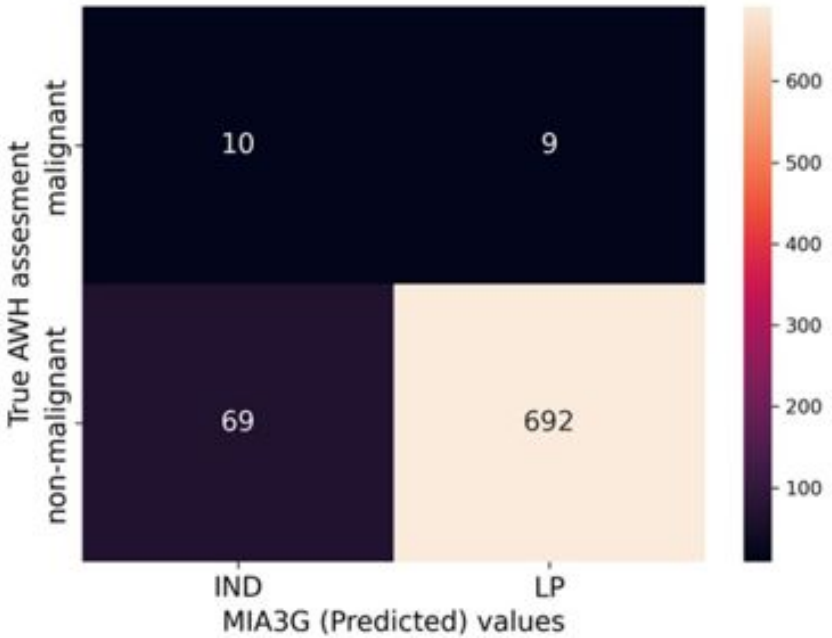


FIGURE 2
Malignancy prediction value without (orange bar) and with MIA3G
(blue bar).

1. Roy Choudhury M, Pappas TC, Twiggs LB, Caoili E, Fritsche H, Phan RT. Ovarian Cancer surgical consideration is markedly improved by the neural network powered-MIA3G multivariate index assay. Front Med. 2024;11:1374836. doi: 10.3389/fmed.2024.1374836.

Performance Analysis for OvaWatch

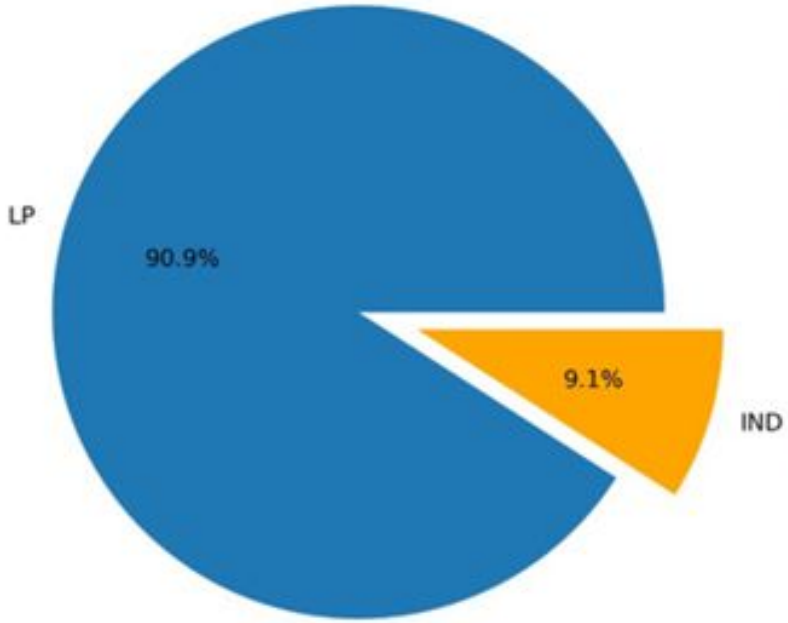
Confusion Matrix for total population



Accuracy of MIA3G stratification vs total patient clinical outcomes

	Count	Percent	95% CI
Total	780	90.00%	87.89 - 92.11

Concordance Analysis for non-malignant population



Concordance analysis for individial patient cohorts

	Total	Correctly Classified	Concordance	95% CI
All Patients	761	692	90.93%	88.89 – 92.97

FIGURE 3

(A) Confusion matrix generated for the total population with clinical outcome vs. MIA3G stratification **(B)** The pie chart depicts the concordance of MIA3G stratification for non-malignant cases

1. Roy Choudhury M, Pappas TC, Twiggs LB, Caoili E, Fritsche H, Phan RT. Ovarian Cancer surgical consideration is markedly improved by the neural network powered-MIA3G multivariate index assay. Front Med. 2024;11:1374836. doi: 10.3389/fmed.2024.1374836.

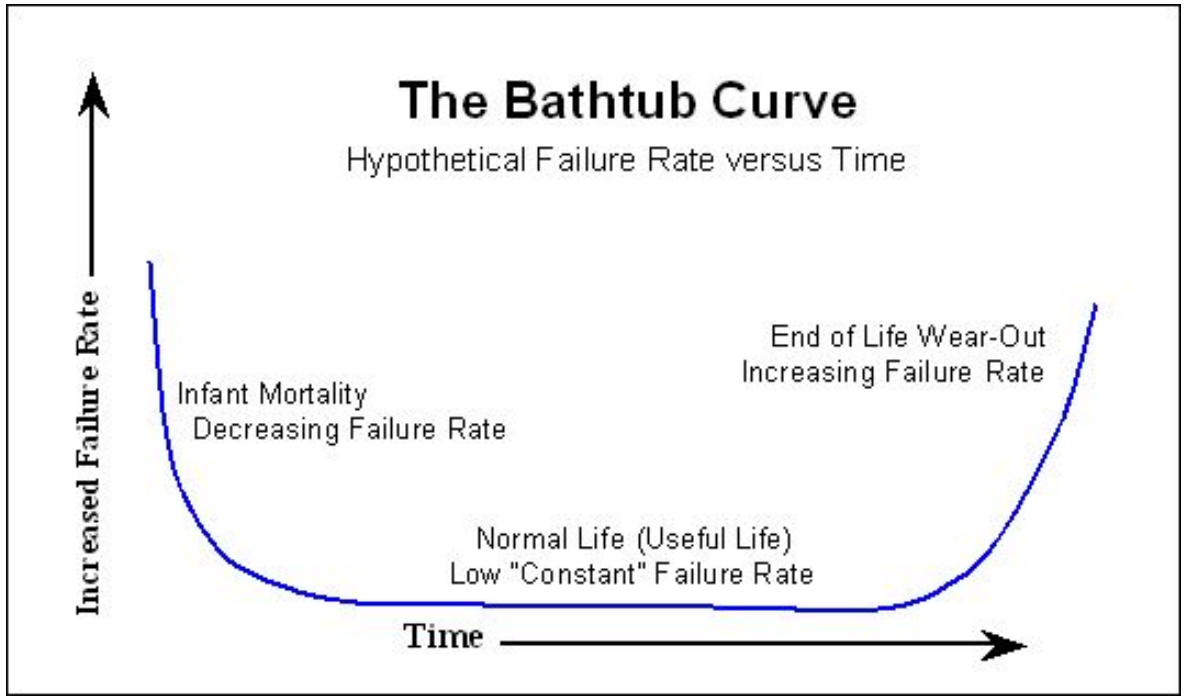
Harnessing AI for the development of a blood based diagnostic test for Adnexal Mass Risk Assessment

- Development of a deep neural network-based model for clinical management of patients with adnexal mass.
- Clinical Utility of the diagnostic test: Aid the physician in surgical consideration decision for adnexal mass risk.
- **Understanding the reliability and accuracy of AI powered diagnostic.**

AI Reliability

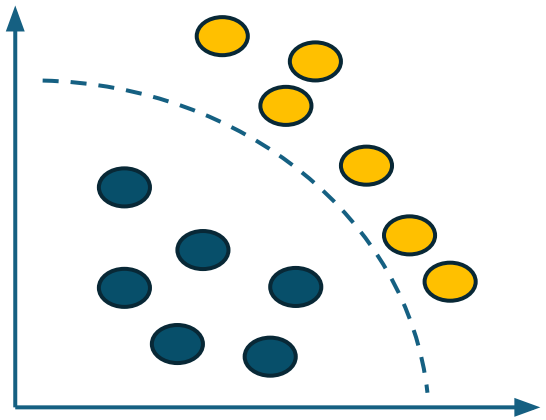
General Definition of Reliability?

Reliability is defined as the probability that a product, system, or service will perform its **intended** function adequately for a specified period of time or will operate in a defined environment without failure. (Source: ASQ)



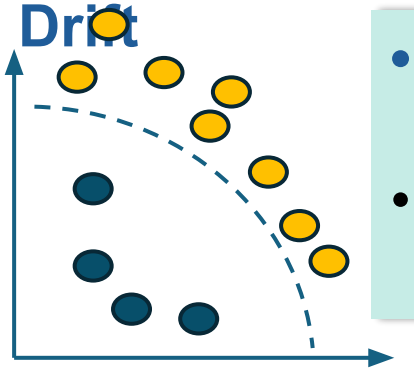
1. https://upkeep.com/learning/bathtub-curve/.
2. asq.org

Components of AI Reliability - Drift



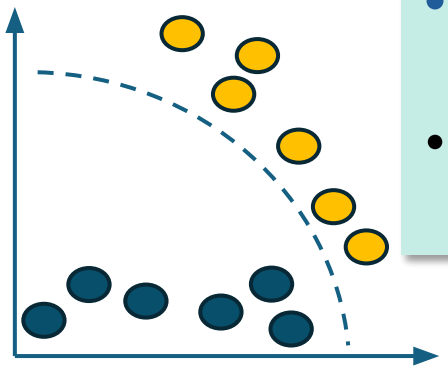
Training Data

Data Drift / Label Drift



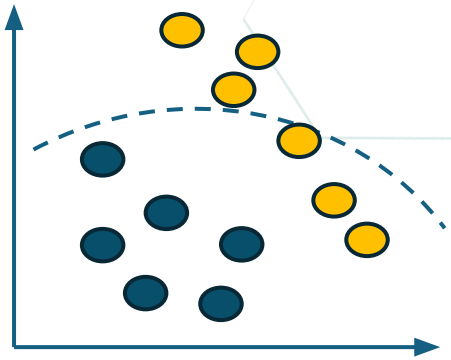
- **P(Y)** changes
- Output data shifts.

Feature Drift



- **P(X)** changes
- Input data shifts.

Concept Drift



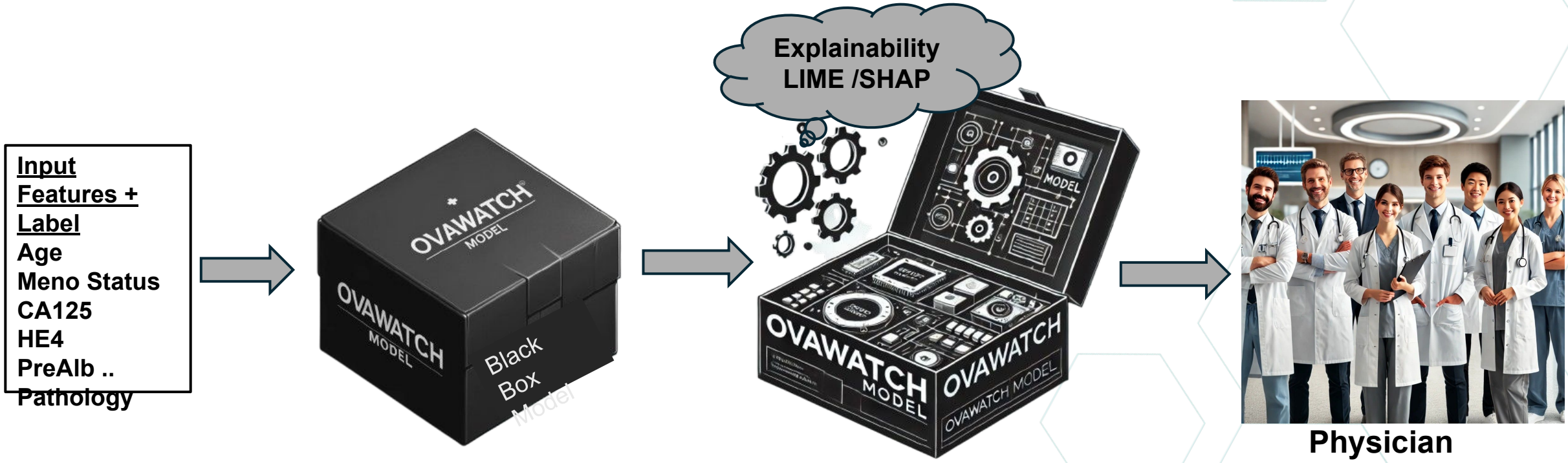
- **P(Y|X)** changes
- Relationship between input and output change but input does not change.

1. https://medium.com/@ajayverma23/understanding-and-detecting-different-types-of-drift-20d720c4727a 2
2. https://www.youtube.com/watch?app=desktop&v=uOG685WFO00

Components of AI Reliability - Explainability

AI Explainability

The methods and techniques used to make **the decision-making processes of artificial intelligence models understandable** and transparent to humans.

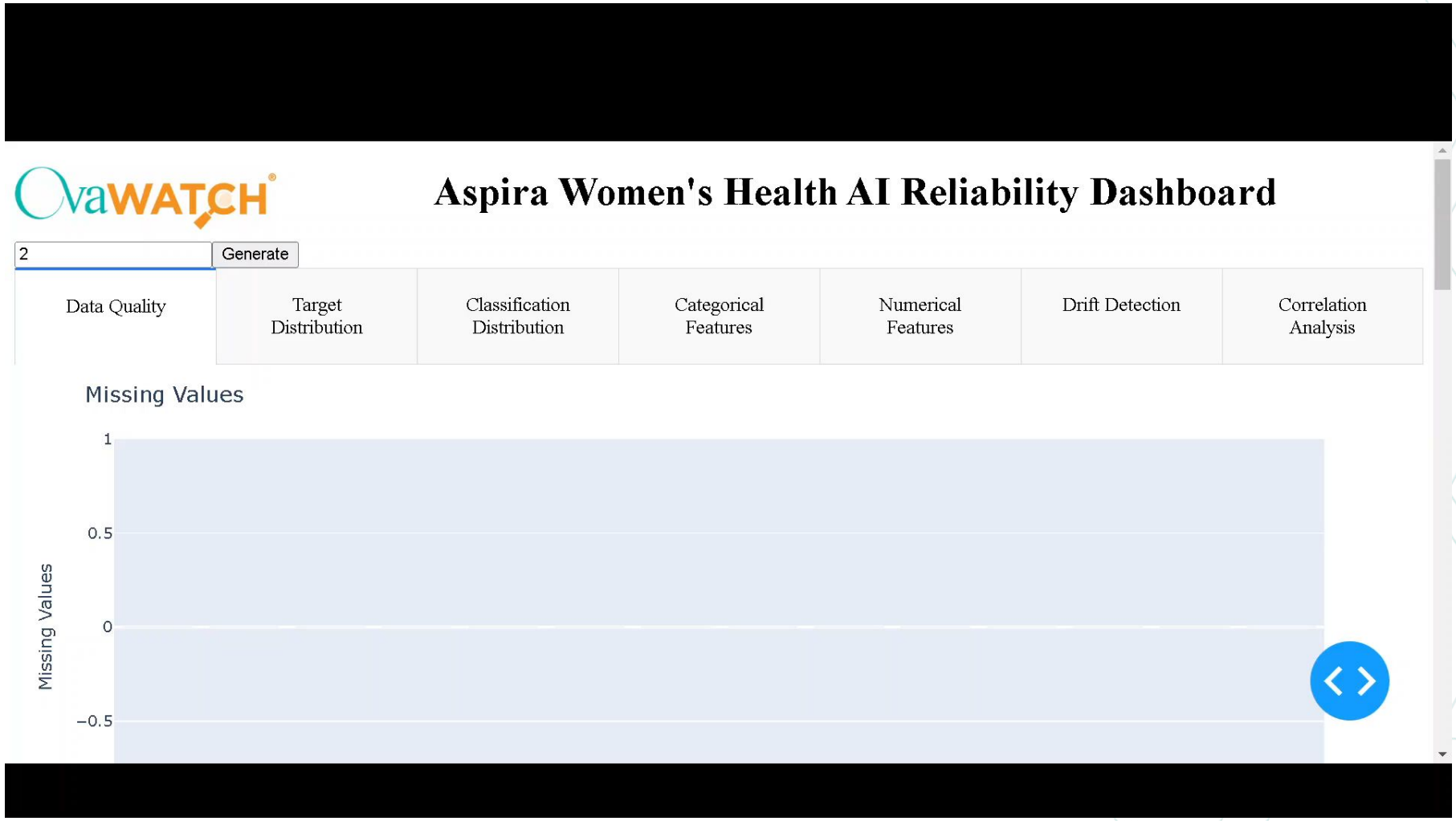


*Individual image components generated by ChatGPT

*Individual image components generated by ChatGPT

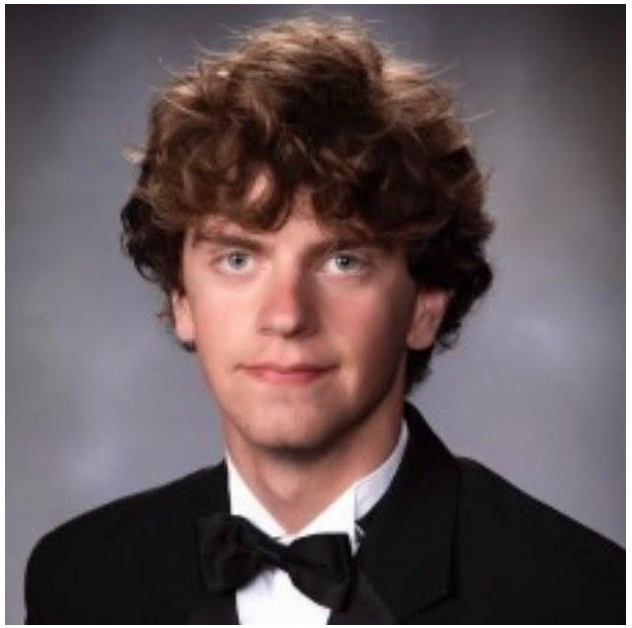
1. Doshi-Velez, F., & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608. 2. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 3. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NIPS).

AI Reliability Dashboard Demo



Acknowledgements

Todd Pappas, Ph.D.
Vice President, R&D
Aspira Women's Health



Cameron Wassmann
Summer Intern
Aspira Women's Health

Christina Kronfel, Ph.D.
Senior Scientist
Aspira Women's Health

Sandy Milligan, MD, JD
President
Aspira Women's Health

Questions?

ASPIRA WOMEN'S HEALTH®