Using MicroRNAs and Deep Learning to Noninvasively Diagnose Gynecologic Conditions

Grant Proposal

Palak Yadav

Massachusetts Academy of Math and Science at WPI

Worcester, MA, USA

Executive Summary

Women's health is a growing public health crisis. 1 out of 10 women will experience some type of chronic gynecological disease in their lifetime, yet confirmative diagnosis can take up to 5-7 years, due to the dismissal of the symptoms, lack of access to adequate resources, and social stigma. MicroRNAs are segments of non-coding RNA that play a vital role in gene expression and have been shown as a noninvasive diagnostic candidate detected in bodily fluids. The goal of this study is to evaluate serum miRNA expression levels to predict gynecologic conditions, specifically ovarian cancer, breast cancer, and endometriosis. A predictive machine-learning model will be trained on public miRNA datasets to generate unique miRNA profiles for each condition. Significant miRNAs in the predictive model willbe extracted through feature selection techniques and inserted in pathway modeling software to determine the pathways most affected in each disease. By identifying the unique and shared pathology of ovarian breast cancer, and endometriosis, miRNA prevalence can be used as a noninvasive diagnostic tool, potentially reduce waiting periods, and guide future therapeutic development.

Keywords: machine learning, microRNAs, gynecology, endometriosis, ovarian cancer, breast cancer, diagnosis.

Using MicroRNAs and Deep Learning to Noninvasively Diagnose Gynecologic Conditions

Women's health has historically been dismissed and stigmatized in society, and even today, painful periods and chronic pelvic pain are deemed as "ladies' problems" (Cook & Dickens, 2014). An analysis of the National Institute of Health funding reported that conditions that negatively affect women receive significantly less funding in proportion to the burden they exert on individuals and society at large (Smith, 2023). For example, ovarian cancer ranks 5th for lethality in a selection of 19 prevalent cancers, yet it ranks 12th in terms of funding. This discrepancy in funding and research has limited the availability of effective and accessible screening tools and therapeutics, resulting in 80% of ovarian cancer cases being diagnosed at an advanced stage (Mogensen et al., 2016). The lack of funding coupled with the challenges of the vagueness of symptoms, lack of awareness and accessibility to resources, and little understanding of the pathology, make it challenging to diagnose detrimental gynecologic conditions at the right time. This study aims to better understand the pathology and etiology of three major gynecological conditions - breast cancer, ovarian cancer, and endometriosis - and work towards developing a noninvasive diagnostic tool.

Diagnosing Gynecological Conditions

Although many past studies have attempted to identify noninvasive diagnostic tools, there is still no screening or non-invasive tool available for several gynecological conditions such as ovarian cancer and endometriosis (Ginsburg et al., 2017). Endometriosis results from the abnormal growth of the uterine lining and may develop into ovarian cancer and breast cancer if not treated effectively. Many studies have speculated a correlation between these three detrimental conditions; however, the results have been varied (Mogensen et al., 2016). A special link between breast and ovarian cancer has been established through the mutations in the BRCA1 and BRCA2 genes (Yoneda et al., 2011). Although endometriosis is a benign condition, it is often managed with oral contraceptives and hormonal therapy,

which can put one at a higher risk for ovarian cancer due to the imbalance of hormone levels. The correlation between breast cancer and endometriosis has been unreliable, as some studies indicate a positive relationship, while others report either the opposite or no significant difference (Ye et al., 2022). Due to the prevalence and detrimental impacts of these conditions, understanding their pathology and etiology is key to identifying potential therapeutics and strategies to diagnose them effectively.

microRNAs and Gene Expression

Harnessing gene expression data as biomarkers for many diseases is an emerging area of research. RNA-seq data examines large datasets of RNA extracted from blood serum, plasma, and tissue biopsies to identify mRNA (protein-coding) or microRNA (non-coding) profiles. MicroRNAs are 22 nucleotides long and play a key role in regulating gene expression and vital biological processes. They function as post-transcription regulators in the 3'untranslated region and cause the degradation of mRNA (Li et al., 2023). There are multiple mechanisms behind the functionality of miRNA, starting from the transcription of the primary miRNA by RNA polymerase II to the migration of the precursor miRNA from the nucleus to the cytoplasm to be transformed by the Dicer to gain more functionality. Ultimately, each miRNA targets specific genes by attaching to the mRNA molecules, leading to the activation or inhibition of the intended protein (O'Brien et al., 2018).

Compared to other types of non-coding RNAs, microRNAs are easy to analyze due to their abundance and accessibility in bodily fluids such as blood, urine, saliva, and plasma, making them a promising candidate as noninvasive biomarkers for a variety of diseases (Zhao et al., 2014). They also play a significant role in gene expression as mutations in a few miRNAs can have a cascading effect on mRNA transcription and protein production, leading to the dysregulation of numerous biological pathways and signaling (O'Brien et al., 2018). Many previous works have found that certain diseases have different miRNA expression levels in comparison to control samples, however many gynecologic

diseases have shared pathology and therefore overlapping aberrant miRNA expression (Zhao et al., 2014). Therefore, developing unique miRNA profile panels for each disease will allow for the identification of strong biomarkers for diagnostic development.

TABLE 1 | miRNA signaling pathways involved in gynecological cancers.

miRNA	Signaling pathway	Target	Target expression	Action	Pathology	Reference
miR-433	MAPK	RAP1A	Overexpression	Cell migration, proliferation, apoptosis	Breast cancer	(76)
miR-99a	mTOR FGFR3	PI3-AKT	Overexpression	Invasion, proliferation, apoptosis	Cervical cancer Breast cancer	(77)
miR-155	AKT	LKB1	Overexpression	Autophagy	Cervical cancer	(78)
miR-21	TNFR1 PI3K/AKT/mTOR RAS/MEK/ERK	Caspase 3 TNF-alpha PTEN RASA1	Overexpression	Apoptosis	Breast cancer Cervical cancer Ovarian cancer	(79)
miR-200 miR-141 miR-200a miR-200b miR-200c	NOTCH TGF-beta	ZEB1 ZEB2 E cadherin EMT	Overexpression	Invasion, metastasis	Ovarian cancer	(80)
Let-7 *miR let-7d-5p	RAS HGMA1	P53	Overexpression	Apoptosis	Ovarian cancer	(81)
miR-34a	p53	HNRNPA1		Cell proliferation	Breast cancer Endometrial cancer	(82)
miR-424	p53	HNRNPA1	Overexpression	Cell proliferation, apoptosis	Breast cancer	(82)
miR-503	p53	HNRNPA1	Overexpression	Cell proliferation, apoptosis	Breast cancer	(82)
miR-142-3p	Bach-1	EMT	Overexpression	Invasion, migration	Breast cancer	(83)
miR-205	ZEB1, ZEB2	EMT PTEN	Overexpression	Apoptosis, cell differentiation, and proliferation	Endometrial cancer	(84)
miR 4712-5p	PTEn/AKT/GSK3beta/cyclin D1	PTEN	Overexpression	Cell invasion, metastasis	Vulvar cancer	(85)
miR-3147	TGF-β/Smad	TGFβ RII EMT	Overexpression	Invasion, cell proliferation, migration	Vulvar cancer	(86)
miR-146a		BRCA1	Overexpression	Cell proliferation	Breast cancer	(87)

ZEB1 and ZEB2 Zinc finger E-box-binding homeobox 1/2 HNRNPA1 Heterogeneous nuclear ribonucleoprotein A1.

Table 1: miRNA signaling pathways involved in gynecological cancers (Duică et al., 2020). Some miRNAs can be observed in multiple conditions, demonstrating the need to identify unique markers for improved diagnosis.

Role of Machine Learning

Machine learning models can be trained on large sets of patient miRNA data to build predictive models. Several studies have attempted to use this approach for classifying various types of cancers and have proved successful in classifying diseases based on their miRNA expression (Alharbi & Vakanski, 2023). A systematic review of machine learning and miRNAs in cancer classification stated that the following algorithms are most used in machine learning models: Decision Trees, Support Vector Machines, Random Forest Trees, KNN algorithms, and Artificial Neural Networks (Sivajohan et al., 2022). Such algorithms allow for binary classification of having a condition or not having a condition and subtype differentiation. Deep Neural Networks allow for more sophisticated computation and can

differentiate between multiple input classes and determine the weight of each input in the final prediction (Alharbi & Vakanski, 2023).

Section II: Specific Aims

This proposal's objective is to develop a machine-learning model to classify serum-based microRNAs from patients with endometriosis, ovarian cancer, and breast cancer. The long-term goal is to predict more diseases from multiple types of miRNA samples (blood, urine, and saliva-based), and implement this in clinical settings to expedite the diagnostic period of prevalent diseases and make it accessible and safe for patients. The central hypothesis is that serum from patients with different gynecological conditions will contain unique miRNA profiles that will allow for differentiation due to the tissue-specificity and mechanism of each condition. The work we propose here will decrease the diagnosis wait period for females and provide miRNA panels for future diagnosis and treatments.

- **Specific Aim 1:** Identify differentially expressed miRNAs for breast cancer, ovarian cancer, and endometriosis.
- **Specific Aim 2:** Improve the accuracy of machine learning models for binary classification and multiclass identification.
- **Specific Aim 3:** Use a pathway modeling platform to evaluate the significant miRNAs' biological significance and provide potential therapeutic and diagnostic targets.

The expected outcome of this work is a computational model that provides a binary classification of disease, followed by a Deep Neural Network model for multiclass identification of the gynecological condition with at least 80% accuracy.

Section III: Project Goals and Methodology

Relevance/Significance

Around 80% of ovarian cancers are diagnosed in the later stages, and endometriosis diagnosis can take up to 5-7 years while also putting one at higher risk for future malignancy (Mogensen et al., 2016). Furthermore, breast cancer is the most common type of cancer in women and the second leading cancer-related death source (Yoneda et al., 2011). A definitive diagnosis for gynecological conditions requires invasive and often expensive procedures such as laparoscopy or biopsy and can be especially inaccessible for patients from underprivileged backgrounds (Duică et al., 2020). Therefore, there is a need to develop a noninvasive method to predict what condition a patient has based on accessible samples, such as blood and urine. This study utilizes serum-based microRNAs and artificial intelligence to improve diagnostic procedures.

Current Knowledge Gaps

Diseases predominantly affecting the female population are challenging to diagnose due to overlapping symptoms, lack of awareness, improper screening, and underfunding in research. Recent studies on the role of microRNA in gynecological conditions hold promise as non-invasive tests (Gilabert-Estelles et al., 2012). Although there have been machine learning models applied to other forms of cancers, there is a need to develop a robust and accurate model for common gynecological conditions. Although miRNA markers for each of the selected conditions have been identified, many miRNAs are common across diseases, and more computation must be performed to identify unique markers that can serve as diagnostic candidates and advance the current understanding of the pathology and etiology of these diseases (Sivajohan et al., 2022).

Innovation

Current models that use miRNA as a diagnostic biomarker are aimed at identifying one specific condition and have not yet compared miRNA panels for multiple female conditions. This study aims to

utilize multiple datasets for each condition and identify differentially expressed genes and unique microRNAs to be used in the machine-learning model. A systematic review of machine learning models in cancer classification described that a standard Deep Neural Network resulted in around mid-70% accuracy for breast cancer, whereas a novel multilayer perceptron network resulted in a 98.74% accuracy (Alharbi & Vakanski, 2023). This study aims to modify past methods to generate a more robust neural network model for cancerous as well as non-cancerous gynecological conditions in hopes of achieving 90% accuracy for cancerous conditions and 80% for benign conditions.

Methodology

The following steps will be taken to meet the specified aims of the project:

1. Accurate datasets and miRNA profiles from at least 50 samples of each condition must be obtained. The National Institute of Health (NIH) Gene Expression Omnibus and the Cancer Atlas TCGA contain publicly available miRNA profiles for several diseases. The GEOR2 platform can be used to apply statistical tests to determine the differential expression of miRNAs.

2. All the miRNA sample profiles will be compiled into one large dataset to allow for efficient computation. Firstly, all data must be normalized using the mix-max normalization method and clean irrelevant features for efficient and accurate evaluation. Previous studies have developed a thorough feature selection procedure to prevent overfitting and preserve necessary values to reduce complexity and time (Hamidi et al., 2023).

3. Develop a model that performs binary classification on each condition dataset. Utilize Logistic Regression, Random Forest, and K-Nearest Neighbor Algorithm to differentiate the miRNA markers of samples with the conditions and healthy samples.

4. miRNA markers identified for each gynecological condition from the last step will be utilized to design a Deep-Learning Neural Network that can classify three different gynecological conditions. Python packages such as TensorFlow, NumPy, Pandas, Matplotlib, and Scikit Learn are useful in developing neural networks.
5. Identify which miRNA markers are most influential in the onset of each gynecological condition using feature extraction algorithms (Srinivasulu et al., 2023).
6. Evaluate the biological significance of the differential expressed microRNAs by modeling pathways in the KEGG Pathway Database to draw biological significance, provide insights for potential therapeutic and diagnostic targets, and advance the current understanding of the pathology of gynecological diseases (Alharbi & Vakanski, 2023).
7. The model will be validated through comparison to previous studies to verify which miRNA and pathways were found to be common and which were newly identified.

Specific Aim #1: Identify differentially expressed miRNA for each disease.

Justification and Feasibility. The first objective is to develop a panel of unique miRNAs for each condition from publicly available datasets from the Gene Expression Omnibus and Cancer Atlas TCGA. The GEOR2 platform can be used to perform statistical tests from the Bioconductor package and generate the top differentially expressed miRNAs in each condition. By determining the top differentially expressed miRNAs, a binary classification machine learning model can be developed (Alharbi & Vakanski, 2023).

Summary of Preliminary Data. Datasets selected from the Gene Expression Omnibus were processed and normalized for accurate analysis. The *limma* package and libraries from the Bioconductor Project in R-Studio were used to perform statistical tests such as t-tests and log2 fold change to identify which miRNAs were differentially expressed in each disease in comparison to the control sample. Two-sample t-tests were performed to determine which miRNAs are differentially expressed between multiple diseases and found common across all three. The Venn Diagram visualizes the distribution of the different groups and proves the claim that each disease has a unique miRNA profile that can be used for future classification models.

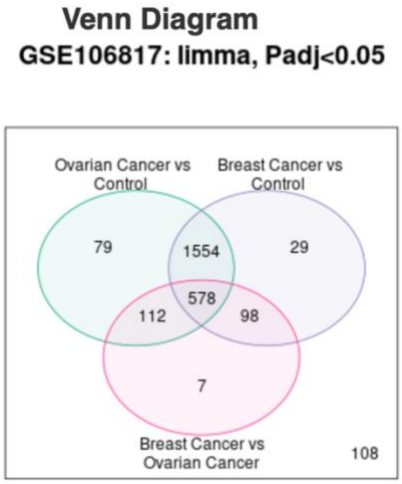

Figure 1: The distribution of miRNAs that are unique and shared amongst breast cancer, ovarian cancer, and control in one dataset. The center highlights 578 miRNAs found across all three cases, 7 miRNAs to differentiate breast cancer from ovarian cancer, 79 to classify ovarian cancer from control, and 29 miRNAs to classify breast

Expected Outcomes. Identifying the differentially expressed miRNAs through statistical tests will create a unique miRNA profile for each disease and help with the feature selection for the machine learning models. The miRNAs that are identified as most significant for each disease can be assigned greater weight to improve the accuracy of the predictive models.

Potential Pitfalls and Alternative Strategies. Each dataset utilized different miRNA extraction techniques impacting the measured miRNAs and resulting in slightly different miRNA profiles for each disease than past works. To mitigate this pitfall, multiple datasets can be compiled and normalized when performing the statistical tests. The resultant miRNA profiles can be cross validated by previous works to ensure that there is a consistency of the differentially expressed miRNAs identified and evaluate the accuracy of the new miRNAs in this study.

Specific Aim #2: Improve the accuracy of predictive machine learning models to 90% for binary classification and 80% for multiclass models.

Justification and Feasibility.

Recently, Zhang et al. (2022) developed a micro-RNA-based Diagnostic Model for Multi-Cancer Detection, providing evidence that multiple diseases can be classified in one model using a weighted diagnostic ranking for the top four differentially expressed miRNAs. This approach must be improved to perform on gynecological conditions as it is harder to classify them due to the shared pathology and biological nature. New strategies must be determined to pick the top differentially expressed miRNAs and harness feature selection techniques to improve the accuracy of gynecologic predictive models. Deep Neural Network and Random Forest predictive models have demonstrated the best performance for previous miRNA classification models, and such models can be tailored through accurate feature selection techniques to perform on gynecological disease (Alharbi & Vakanski, 2023). It is feasible to do feature selection for complex models as the GEO2R platform generates the significance levels of each miRNA in a disease.

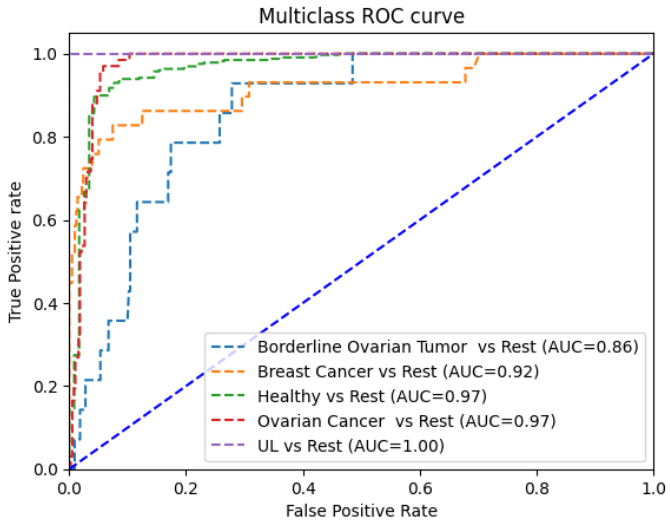

Figure 2: A Receiving Operator Curve (ROC) illustrates the performance of binary classification models. In this case, each line demonstrates the accuracy levels of classifying the condition out of all the possible outcomes in that dataset.

Summary of Preliminary Data. A simple multiclass Random Forest algorithm was applied to classify 4 different types of gynecological diseases and control samples. The overall model achieved an accuracy of 92%, with the sub-class accuracy as described in Figure 2. The varying accuracy levels for each class validate that each disease contains a unique miRNA profile that can be harnessed to classify patient samples. It can also be

hypothesized that Borderline Ovarian Tumor had the lowest accuracy due to its biological nature and miRNA expression levels being like control samples and ovarian cancer samples.

Expected Outcomes. The overall outcome of this aim is to develop machine learning models that can successfully classify multiple gynecologic diseases. Next, a Deep Neural Network will be implemented, and each miRNA will be assigned a weight through feature extraction of the Random Forest model to improve the accuracy.

Potential Pitfalls and Alternative Strategies. Identifying publicly available datasets that contain accurate and ample samples is key and is often a challenge. Random Forest algorithms can become ineffective in predicting large sets of data (Srinivasulu et al., 2023). However, they can be helpful in identifying the feature significance and developing robust Neural Networks for predicting larger sets of data. Additionally, as the dataset increases, machine learning models are likely to overtrain; however, this problem can be mitigated by adjusting the hyperparameters. The datasets are also aggregated from multiple studies so different methods of collection were utilized, resulting in varying levels of measured microRNAs. This potential pitfall can be addressed by using effective normalizing and cleaning filters (Alharbi & Vakanski, 2023).

Specific Aim #3: Use a pathway modeling platform to evaluate significant miRNAs' biological significance and provide potential therapeutic and diagnostic targets.

Justification and Feasibility. miRNAs control gene expression, therefore aberrant expression of these regulatory molecules can result in the disruption of biological pathways and signaling (O'Brien et al., 2018). The miRNA panels generated from Specific Aim 1 and 2 will provide a unique biological serum-based profile that can be used to model dysregulated biological processes, such as apoptosis, proliferation, angiogenesis, etc. (Zhao et al., 2014). Previous studies have applied feature extraction techniques such as the Deep LIFT algorithm to determine which miRNAs played a significant role in the classification of cancer and inserted that into accessible databases such as the Kyoto Encyclopedia of Genes and Genomes (Hamidi et al., 2023). Understanding the biological pathways involved in each disease will advance the current understanding of the interconnected pathology of gynecological diseases.

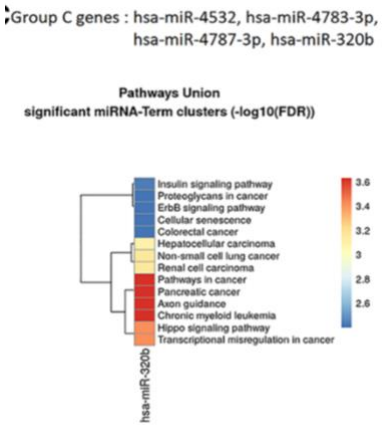

Figure 3: A heatmap highlighting the major pathways influenced by a differentially expressed miRNA in ovarian cancer (Hamidi et al., 2023).

Summary of Preliminary Data. The Deep LIFT Algorithm was applied to the Random Forest Multiclassification model. The miRNAs featured in Figure 4 highlight the most significant input features and each of their influence in predicting a certain disease. These miRNAs can be cross verified by the unique miRNA panels developed in Aim 1 and be used to model the biological pathways in the three diseases being studied.

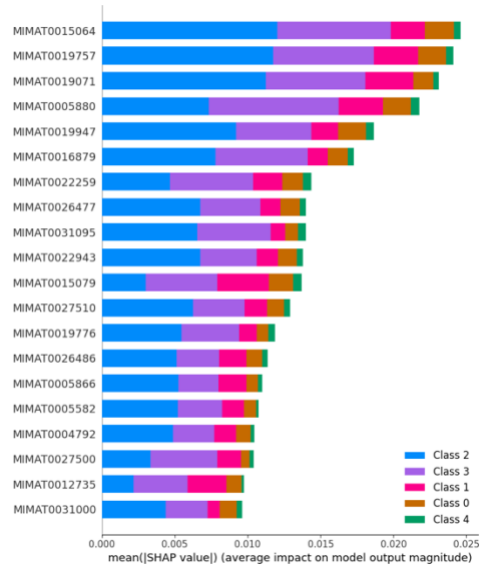

Figure 4: Results of the Deep LIFT algorithm, highlighting the top miRNAs that play a key role in prediction of each class.

Expected Outcomes. The expected outcome for this stage is to generate a biological pathway-profile for each gynecological condition based on the significant feature miRNA. Understanding the pathology of each disease will allow for identifying potential diagnostic and therapeutic targets.

Potential Pitfalls and Alternative Strategies. The significance of this stage relies on the accuracy of the model from the previous two stages. If an inaccurate miRNA panel is generated for each condition, an inaccurate biological significance analysis will be performed. Such errors can be avoided by cross-referencing each stage's results from previous studies and adjusting the model as necessary.

Section IV: Resources/Equipment

Datasets

Publicly available datasets will be obtained from the Gene Expression Omnibus and the Cancer Atlas TCGA for endometriosis, ovarian, and breast cancer. These datasets contain miRNA and other types of RNA-seq data on samples collected from patients. Comprehensive analytic tools are also available to evaluate meaning from large datasets.

Computational Model

Various computational software will be utilized to construct the machine learning model, including but not limited to, Google Collaboratory, RStudio, and the KEGG Database (Alharbi & Vakanski, 2023).

Section V: Ethical Considerations

This model analyzes publicly available data to address the research problem. This model cannot be applied in a clinical setting due to the nature of artificial intelligence and lack of reliability, and physicians may still need to use biopsy and other procedures to confirm accurate diagnosis.

Section VI: Timeline

Link to Ghent Chart with outlined steps: https://bit.ly/ghanttChartPalak

Major Milestones:

- Understanding the current state of gynecologic conditions and previous works done with miRNA and machine learning: October 30th, 2023.
- Collecting the main datasets that will be used for this project: November 20th.
- Identify differentially expressed genes for each disease: November 30th.
- Develop binary classification models using Logistic Regression: December 30th.
- Develop multi-classification models using Random Forest and Deep Neural Network: January 15th.
- Perform feature extraction to determine the most significant miRNAs and the biological pathways most affected: January 30th.

Section VII: Appendix

Appendix 1a.

A list of programming languages, applications, and libraries required to build, train, and test the intended machine learning models.

Applications	**Description**
Google Collaboratory Python 3	A product from Google Research that allows for accessible and efficient Python code with support for important libraries and functions for machine learning models.
TensorFlow/Keras	A Python-based framework for building, training, and testing machine learning models
Matplotlib	Python library for visualization and graphics.
Sci-Kit Learn	Python-based machine learning library that consists of commands and functions to create a wide range of models.
R/R-Studio	An integrated platform to use R and R-based packages and libraries for statistical analysis.
GEO2R	A web-based NCBI Gene Expression Omnibus (GEO) analytical tool with an in-built limma package, DESEQ-2 commands, and other Bioconductor projects to identify differentially expressed genes and miRNAs.

Appendix 1b.

A list of datasets selected from Gene Expression Omnibus and the Cancer Atlas TCGA used for the training and testing phase.

GEO Access Link	**Description**
GSE106817	333 ovarian cancers, 66 benign tumors, 29 of ben ovarian, 143 breast cancer, and 275f non-car controls.
GSE235525	34 high-grade serous ovarian cancer, 36 samples
GSE201712	64 ovarian cancer samples
GSE226445	350 samples of women with known BRCA mutation, which is known to cause breast and ovarian cancer and 30303 wild types.
GSE230956	4 endometriosis samples and 4 benign samples.
GSE113486	100 ovarian cancers and 100 breast cancer samples.

Appendix 2.

A decision matrix to evaluate the model in comparison to previous studies' models (Alharbi & Vakanski, 2023).

Criteria	**Rank**	**Logistic Regression**	**Expected %**	**Random Forest**	**Expected %**	**Neural Network**	**Expected %**
AUC	8	9	95%	9	95%	9	95%
Accuracy	9	9	80%	8	90%	9	87%
Sensitivity	8	7	82%	7	88%	8	90%

Specificity	7	7	85%	8	83%	8	90%
Total		**258**		**256**		**273**	

Section VIII: References

Ahn, S. H., Singh, V., & Tayade, C. (2017). Biomarkers in endometriosis: challenges and opportunities. *Fertility and Sterility*, *107*(3), 523–532. https://doi.org/10.1016/j.fertnstert.2017.01.009

Alharbi, F., & Vakanski, A. (2023). Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. *Bioengineering*, *10*(2), 173. https://doi.org/10.3390/bioengineering10020173

Cook, R. J., & Dickens, B. M. (2014). Reducing stigma in reproductive health. *International Journal of Gynecology & Obstetrics*, *125*(1), 89–92. https://doi.org/10.1016/j.ijgo.2014.01.002

Duică, F., Carmen Elena Condrat, Cezara Alina Dănilă, Andreea Elena Boboc, Radu, M., Xiao, J., Li, X., Sanda Maria Crețoiu, Suciu, N., Dragoş Cretoiu, & Dragoş Predescu. (2020). MiRNAs: A Powerful Tool in Deciphering Gynecological Malignancies. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.591181

Ginsburg, O., Bray, F., Coleman, M. P., Vanderpuye, V., Eniu, A., Kotha, S. R., Sarker, M., Huong, T. T., Allemani, C., Dvaladze, A., Gralow, J., Yeates, K., Taylor, C., Oomman, N., Krishnan, S., Sullivan, R., Kombe, D., Blas, M. M., Parham, G., & Kassami, N. (2017). The global burden of women's cancers: a grand challenge in global health. *The Lancet*, *389*(10071), 847–860. https://doi.org/10.1016/s0140-6736(16)31392-7

Hamidi, F., Gilani, N., Reza, B., Hanif, Y. Esmaeil, B., Parvin, S., & Jamileh, M. (2023). Identifying potential circulating miRNA biomarkers for the diagnosis and prediction of ovarian cancer using the machine-learning approach: application of Boruta. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1187578

Horne, A. W., & Missmer, S. A. (2022). Pathophysiology, diagnosis, and management of endometriosis. *BMJ*, *379*, e070750. https://doi.org/10.1136/bmj-2022-070750

Moustafa, S., Burn, M., Mamillapalli, R., Nematian, S., Flores, V., & Taylor, H. S. (2020). Accurate diagnosis of endometriosis using serum microRNAs. *American Journal of Obstetrics and Gynecology*, *223*(4), 557.e1–557.e11. https://doi.org/10.1016/j.ajog.2020.02.050

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, *9*(402). https://doi.org/10.3389/fendo.2018.00402

Sivajohan, B., Elgendi, M., Menon, C., Allaire, C., Yong, P., & Bedaiwy, M. A. (2022). Clinical use of artificial intelligence in endometriosis: a scoping review. *Npj Digital Medicine*, *5*(1). https://doi.org/10.1038/s41746-022-00638-1

Sathipati, S. Y., Tsai, M.-J., Shukla, S. K., & Ho, S.-Y. (2023). Artificial intelligence-driven pan-cancer analysis reveals miRNA signatures for cancer stage prediction. *Human Genetics and Genomics Advances, 4*(3), 100190. https://doi.org/10.1016/j.xhgg.2023.100190

Smith, K. (2023, May 3). *Women's health research lacks funding – these charts show how*. www.nature.com. https://www.nature.com/immersive/d41586-023-01475-2/index.html

Ye, J., Peng, H., Huang, X., & Qi, X. (2022). The association between endometriosis and risk of endometrial cancer and breast cancer: a meta-analysis. *BMC Women's Health*, *22*(1). https://doi.org/10.1186/s12905-022-02028-x

Yoneda, A., Lendorf, M. E., Couchman, J. R., & Multhaupt, H. A. B. (2011). Breast and Ovarian Cancers. *Journal of Histochemistry & Cytochemistry*, *60*(1), 9–21. https://doi.org/10.1369/0022155411428469

Zhang, A., & Hu, H. (2022). A Novel Blood-Based microRNA Diagnostic Model with High Accuracy for Multi-Cancer Early Detection. *Cancers*, *14*(6), 1450. https://doi.org/10.3390/cancers14061450

Zhao, Y.-N., Chen, G.-S., & Hong, S.-J. (2014). Circulating MicroRNAs in gynecological malignancies: from detection to prediction. *Experimental Hematology & Oncology*, *3*(1), 14. https://doi.org/10.1186/2162-3619-3-14