A Novel Deep Learning Pipeline to Noninvasively Detect Gynecologic Diseases Using MiRNA Expression and In Silico Modeling

Women's health is a growing public health issue due to a lack of funding, accessibility, and available diagnostic tools. Unfortunately, 1 out of 10 women will experience a chronic gynecological condition in their lifetime, yet the average diagnostic period is about 5-7 years due to the dismissal of symptoms, lack of access to adequate resources, social stigma, and other factors (Ahn et al., 2017). There is an urgent need to address this inequity in healthcare and provide female-identifying patients with the right services at the right time.

Current State of Gynecological Diseases

An analysis of the National Institute of Health (NIH) funding reported that health conditions negatively affect women and receive significantly less funding than the burden they exert on individuals and society at large (Smith, 2023). For example, ovarian cancer ranks 5th in lethality amongst a panel 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 (Smith, 2023; Mogensen et al., 2016). The vagueness of symptoms of gynecological conditions also contributes to the challenges in accurately diagnosing female patients (Cook & Dickens, 2014). Common symptoms associated with gynecological conditions, including but not limited to pelvic pain, painful periods, and fatigue, are difficult to classify due to their subjectivity, thereby preventing a clear diagnosis (Ahn et al., 2017). The lack of funding coupled with the challenges of the vagueness of symptoms, lack of awareness, understanding of the pathology, and accessibility to resources, contribute to the delays in accurate diagnoses in female patients. 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 that provides patients and healthcare providers with a stronger indication of what condition they may have through only a blood test. This method would increase accessibility and precision of identifying gynecological conditions since current screening practices generally involve laparoscopy and biopsies, which are often expensive and inaccessible for women from marginalized backgrounds (Ginsburg et al., 2017)

Breast Cancer, Ovarian Cancer, and Endometriosis

Endometriosis results from the abnormal growth of the uterine lining outside of the uterus and is a risk factor for ovarian cancer and breast cancer. 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). There is also a genetic risk between endometriosis and ovarian cancer due to mutations in the PI3K/AKT pathway (Brilhante et al., 2017). Additionally, endometriosis is often managed with oral contraceptives and hormonal therapy, which further increases the risk of ovarian cancer due to the estrogen imbalance (Mogensen et al., 2016). However, 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). There is a need to better understand the correlation between these three conditions beyond population studies, as a molecular understanding is key to identifying potential therapeutic and diagnostic strategies.

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.

MicroRNA

Given the current state of gynecologic conditions, there is a need to find a noninvasive way to diagnose such diseases quickly and accurately. Tracing gene expression data as indicators of disease is an emerging area of research. RNA-seq data examines large datasets of RNA extracted from patient samples of blood, saliva, urine, and other bodily fluids. MicroRNAs are 22 nucleotides-long non-coding segments of the RNA. They derive from the transcription of DNA, and the interaction between 3' untranslated regions (3' UTR) of target mRNAs is regulated by the interaction between various miRNA transcriptional

factors to either suppress or express specific genes (Gilabert-Estelles et al., 2012). They 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 dysregulation in numerous biological pathways and signaling (O'Brien et al., 2018). MicroRNAs are easy to analyze due to their abundance and accessibility in bodily fluids and hold promise as noninvasive diagnostic candidates. Previous research has found that certain diseases have unique miRNA expression profiles in comparison to other samples; however, many miRNAs are found across several diseases. Discovering miRNA-specific profiles between gynecologic diseases can provide insight into the biological interconnectedness and provide potential therapeutics (Zhao et al., 2014).

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 based on their miRNA expression profiles (Alharbi & Vakanski, 2023). Unlike previous works, this study aims to target gynecologic conditions specifically and evaluate the biological significance between three prevalent diseases. A greater understanding of the miRNA profiles will shed light on their unique and shared pathology. All of this will contribute to identifying promising diagnostic and therapeutic targets.

Problem Statement: Diagnosing gynecological conditions quickly and effectively is a challenge due to the lack of funding, knowledge gaps, and accessibility to screening tools.

Research Question: How can miRNA expression data and machine learning be used to diagnose and provide a greater understanding of the pathology of three major gynecological conditions: ovarian cancer, breast cancer, and endometriosis?

Objective:

Obj. 1: Collect miRNA expression samples associated with the target diseases and healthy control to train the machine learning models. For the scope of this project, the diseases will include breast cancer, ovarian cancer, and endometriosis.

Obj. 2: Design a deep learning binary classification model to differentiate miRNA expression levels of specific disease samples and control. The goal is to achieve a greater 80% accuracy to prove that each disease has a unique miRNA expression profile and to determine the significance levels (p-value) of the top feature miRNA to identify miRNA unique to each disease.

Obj. 3: Implement multiclass predictive models to classify all the target diseases with an accuracy of at least 75% and extract a panel of miRNA that is significantly found across all three diseases.

Obj. 4: Use the miRNA profiles developed to model miRNA-mediated pathways and draw biological significance about the pathology and etiology of the three diseases.

Hypothesis

Given the effectiveness of miRNA expression data and deep learning models in classifying various types of cancer in previous works, it can be hypothesized that deep learning models will produce an accurate classification of gynecologic diseases for early disease prognosis and identify potential therapeutic targets.