Discussions

This study demonstrates that miRNAs can be used as noninvasive candidates for disease detection and identifying therapeutic targets. In the preliminary stages of binary classification, the deep learning models provided high accuracy (over 90%) and reliability in comparison to the logistic regression model due to the neural network's stronger capabilities for processing high dimensional data. Although each dataset utilized different miRNA extraction techniques impacting the measured miRNAs and resulting in slightly different miRNA profiles for each disease than previous works, this pitfall was mitigated by compiling multiple datasets and using robust normalization methods. The resultant miRNA profiles were cross-validated with works to ensure that there is a consistency of the differentially expressed miRNAs identified and to evaluate the accuracy of the new miRNAs identified in this study. Mitigating potential discrepancies in the miRNA expression profiles is key as the machine learning model and ultimately the pathway analysis rely on the accuracy of the miRNA panels. The aggregated multiclass deep learning model performed with an accuracy greater than 80%. Although the accuracy of the Deep Neural Network was lower than Random Forest, Deep Neural Networks are better apt at balancing complex data, whereas Random Forest is prone to bias prediction due to its inability to accurately balance class sizes. Overall, the models developed performance the same or better than previous models used for miRNA expression levels analysis (Alharbi & Vakanski, 2023). Lastly, the Deep LIFT algorithm was applied to the deep neural network's top features and their contribution to the classification of each disease. The miRNA panel extracted from each of these three models provided a unique profile for each disease as well as insights into their shared pathology. Comparing the panels of miRNAs identified in this study to previous works reveals some similarities and some new miRNAs that should be explored more in future experiments.

Biological significance

The pathway modeling software miRNA-DIANA Tools accepts at most 200 miRNA entries. Approximately 60 miRNAs from the unique disease panel and 20 miRNAs from the shared panel were

used to analyze the miRNA-mediated pathways. miRNA expressed across all three diseases targeted genes found in the pathways of cancer (P=6.04E-18), followed by the Axon Guidance, regulation of the actin cytoskeleton, and the RAP1 signaling pathway, all with significance levels of lower than P=3.15E-08. A PubMed search of the miRNA targeting these pathways revealed the interconnected pathology of these three diseases. miRNA-Let-7d was found across all three diseases and its aberrant expression has been shown to play a key role in female malignancy (Zhang, et. al, 2017). The human lethal-7 (let-7) family of miRNA is found on chromosome 9 and mediates cell proliferation and carcinogenesis. The expression of this family of micro-RNAs is regulated by the transcriptional and post-transcriptional through the OCT4, MYC, and p53 mutations, and the aberrant expression of let-7d ultimately targets mRNAs involved in various tumor hallmarks as well as conditions such as endometriosis and uterine fibroids (Zhang, et. al, 2017). miRNA-let-7d sheds light on the interconnected pathology of these three diseases, indicating that its aberrant expression could predict the presence of one disease and the risks of developing another gynecologic disease.

The underexpression of miR-320a was found to be prevalent in breast cancer samples. The miR-320 family is linked with the EMT process, decreasing the expression of E-cadherin, and increasing the expression of N-cadherin through the targeting of the *FOXM1* gene and other signaling pathways such as P13K/AKT and TGF-β/Smad signaling (Liang et al., 2021). Targeting the underexpression of miR-320a in breast cancer would inhibit the migration and invasiveness of the tumor.

miR-1307-3p upregulation was found with high significance in the ovarian tumor samples (p-value < 10e-8) and slightly lower yet still significance levels in breast cancer and endometriosis. The miR-1307 family of microRNAs is found in the USMG5 gene intron region in chromosome 10, yet there is little understanding of its functionality (Saberianpour & Abkhooie, 2021). Upregulation of this class of miRNAs has been linked to chemoresistance in ovarian tumors, as well as abnormal cell growth, differentiation, and metastasizes (Saberianpour & Abkhooie, 2021). Targeting the genes mediated by

miRNA-1307 could serve as ovarian cancer therapy, and further exploration of its role in other female malignancies could improve survival for a wider range of diseases.

These are just a few examples of pathways and signal mechanisms that demonstrate the interconnected pathology of breast cancer, ovarian cancer, and endometriosis. All these conditions exhibit abnormal behavior in pathways associated with angiogenesis, rapid cell growth, infertility, and insulin levels, indicating that having one of these diseases increases the risk of getting the other two diseases.

Future Research

Future investigation should focus on collecting larger sample tests using similar miRNA extraction methods to improve generalizability and reduce misinterpretation by collecting samples from women of different ethnicities, ages, and stages of the menstruation cycle. To test the effectiveness of serum-based miRNA, healthcare workers should perform miRNA extraction methods on samples of patients and compare the prediction of the machine learning model to current diagnostic and screening practices. Furthermore, the miRNA-mediated pathways and genes identified through this approach should be targeted through *in vitro* and *in vivo* samples to evaluate their effectiveness in suppressing the disease. While this study focused primarily on evaluating deep learning and random forest algorithms as predictive models, ensemble learning models aggregate multiple types of models to improve performance. By collecting larger sets of samples and experimenting with various kinds of machine learning models, miRNAs can be used for prognosis, therapeutics targets, and advancing understanding of the genetic and epigenetic factors of more than just gynecologic diseases. Implementing this diagnostic technique in clinics will expedite diagnostic periods by predicting an array of diseases through simple blood tests, and ultimately provide individuals with the care they need at the right time. This technology will be especially beneficial for marginalized communities who lack access to adequate medical care and provide personalized therapies as the pathology and etiology of each disease may vary from patient to patient.

Conclusion

This study demonstrates the potential use of miRNAs as non-invasive diagnostic markers for breast cancer, ovarian cancer, and endometriosis. Deep learning binary classification and multi-class models were designed to predict the likelihood of a disease based on a patient's serum miRNA expression profile which were selected from publicly available datasets. The models provided strong accuracy, sensitivity, and specificity in predicting each of the diseases, as well as the two stages of ovarian cancer (ovarian tumor and borderline tumor), demonstrating the strengths of machine learning as a predictive models in early detection. Feature extraction techniques applied to the deep learning models provided panels of unique and shared miRNAs found across the three selected diseases. miRNA-Let-7 was found across all three diseases and its dysregulation plays a key role in tumor growth and the onset of several cancer hallmarks. Overall, miRNAs found across all three diseases primarily targeted pathways of cancer, RAP1, RAS, and Hippo signaling, all of which have been spectacled by previous studies in malignancies and found to be a cause of degradation of benign conditions such as endometriosis. These findings pave the way for developing miRNA-based diagnostics and personalized therapies, ultimately improving patient outcomes.