A Conceptual Framework for Predicting Premenstrual Syndrome (PMS) Symptoms in Professional Women Using SVM and ARIMA

Kay Thwe Min Han[1], Paing Moe Kyaw[2]

[1,2] Siam University, Thailand

[1]*kay.tmhan@siam.edu@siam.edu,* [2]*6417193010@siam.edu*

ABSTRACT

Premenstrual Syndrome (PMS) is a common condition affecting women, particularly in professional settings where stress and workload can exacerbate symptoms. Effective prediction of PMS symptom onset and severity is critical for personalized management, yet existing approaches are largely limited to retrospective tracking. This paper proposes a conceptual framework that integrates Support Vector Machines (SVM) for symptom classification and AutoRegressive Integrated Moving Average (ARIMA) for time-series forecasting. By leveraging data from mobile health applications, the model aims to predict the timing and intensity of PMS symptoms in professional women, factoring in lifestyle variables such as stress, sleep, and exercise. The dataset used in this study is derived from publicly available menstrual cycle data from Marquette University, which provides comprehensive tracking of symptoms and lifestyle factors. This predictive framework offers a novel approach to PMS management, empowering women to take proactive measures to mitigate symptom impact and improve both personal well-being and workplace productivity. The proposed model seeks to address a significant gap in PMS research by moving beyond symptom logging to provide a forward-looking, personalized solution for professional women.

Keywords: Predictive Modeling, Premenstrual Syndrome (PMS), Support Vector Machines (SVM), Auto Regressive Integrated Moving Average (ARIMA), Classification, Time

INTRODUCTION

Premenstrual Syndrome (PMS) is a prevalent health condition characterized by a wide array of physical, emotional, and behavioral symptoms that affect many women during their reproductive years. These symptoms typically occur during the luteal phase of the menstrual cycle and resolve shortly after menstruation begins (Tsegaye & Getachew, 2019). Additionally, research by Yonkers et al. (2008) emphasizes that mood and behavioral symptoms, influenced by hormonal levels, can be significantly affected by external stressors and lifestyle choices, further complicating the symptomatology of PMS(Yonkers et al., 2008). For professional women, the demands of work-related stress and time constraints can exacerbate PMS symptoms, leading to reduced productivity and overall quality of life (Kahyaoglu Sut &

ISBN

Mestogullari, 2016).

Extensive research has explored PMS, with studies investigating its prevalence, symptom variations, and influencing factors. (Gao et al., 2021) conducted a bibliometric analysis highlighting the cyclical and recurrent nature of PMS, detailing its psychological and somatic symptoms during the luteal phase of the menstrual cycle, and underscoring its significant prevalence among women. (Sistiarani et al., 2023)examined the relationship between stress levels and PMS types in women of childbearing age, indicating that lifestyle factors significantly impact symptomatology. However, a gap remains in developing personalized and predictive solutions for effectively managing symptoms. Existing approaches often rely on self-reported data or subjective assessments, leading to limitations in accuracy and individual-specific predictions.

Machine learning has emerged as a promising tool for analyzing complex health data and generating predictive models. Support Vector Machines (SVM), known for its efficiency in high-dimensional data classification, holds potential for identifying PMS presence and symptom severity. ARIMA, a powerful time series forecasting model, can be utilized to predict the onset or severity of identified symptoms. This research proposes a predictive model utilizing Support Vector Machines (SVM) and AutoRegressive Integrated Moving Average (ARIMA) to forecast PMS symptoms in professional women. By leveraging data from publicly available menstrual cycle datasets like the one provided by Marquette University (Fehring, 2013), this model aims to empower women with proactive approaches to PMS management.

RESEARCH METHODOLOGY

This research employs a data-driven approach to develop a predictive model for Premenstrual Syndrome (PMS) symptoms in professional women. The methodology involves several key steps: data collection, preprocessing, feature extraction, model development, and evaluation. By leveraging machine learning techniques and a publicly available dataset, this study aims to provide personalized insights and support for women experiencing PMS.

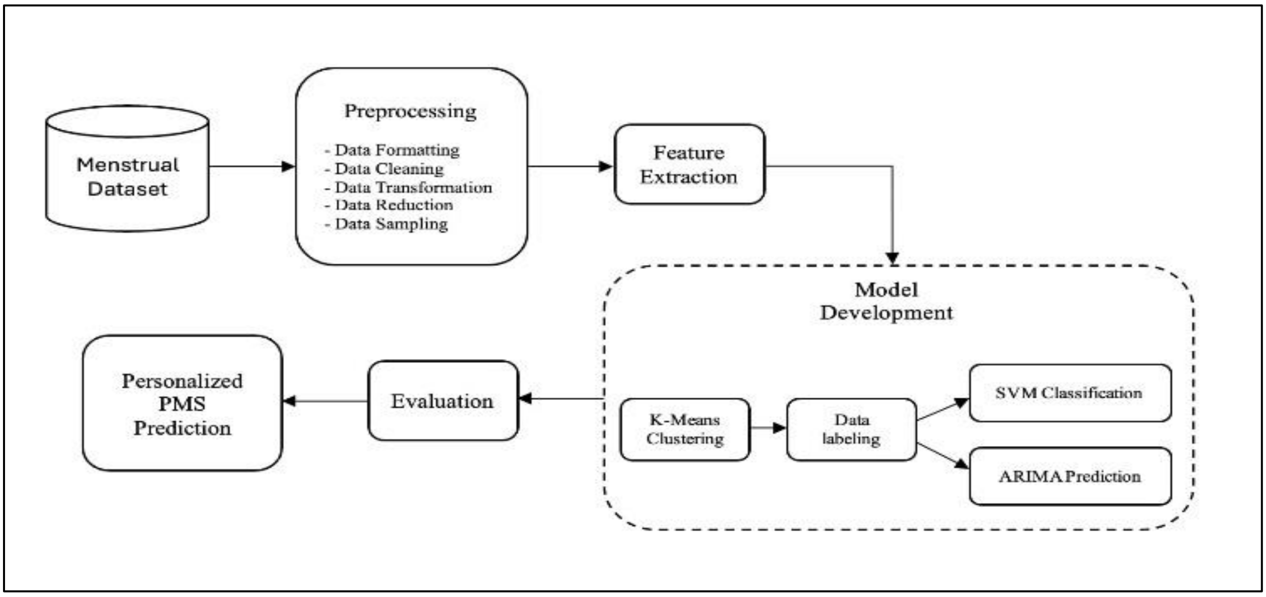

Figure 1: Block Diagram of the PMS Prediction Framework

DATA COLLECTION

ISBN

The dataset from Marquette University, as described by (Richard J. Fehring, 2012)encompasses a wide array of menstrual cycle data, including symptom tracking and lifestyle factors, collected from a diverse sample of women. This publicly available dataset is instrumental for researchers aiming to analyze menstrual health and its implications on women's overall well-being. The data includes detailed records of menstrual cycle lengths, symptom severity, and various lifestyle factors such as stress levels, physical activity, and dietary habits. Such comprehensive data collection is crucial for understanding the multifactorial nature of PMS and its impact on women's health.

DATA PREPROCESSING

Since data has been collected, the next step involves cleaning the dataset to handle missing values, outliers, and inconsistencies. Techniques such as imputation for missing data and normalization of numerical features will be employed to ensure that the dataset is suitable for analysis. Additionally, categorical variables will be encoded appropriately to facilitate their use in machine learning algorithms. This step is crucial for enhancing the quality of the data and ensuring that the predictive model is trained on a robust dataset.

FEATURE EXTRACTION

When the data is preprocessed, the next phase involves feature extraction. This step focuses on identifying and selecting the most relevant features that contribute to the prediction of PMS symptoms. Key features from the dataset were selected based on their relevance to predicting Premenstrual Syndrome (PMS) symptoms in professional women. These include menstrual cycle characteristics such as cycle number, length of cycle, and length of luteal phase, which provide insights into cycle regularity and phase durations, critical for understanding pms symptom onset. additionally, symptom severity data, including menses score day one to menses score day fifth and total menses score, track daily and overall PMS symptom severity across cycles. fertility-related variables, such as cycle with peak or not and total number of peak days, are also incorporated to capture hormonal fluctuations that may influence symptom patterns.

Incorporating lifestyle and behavioral factors, such as number of days of intercourse, unusual bleeding, and intercourse in fertile window, allows for a deeper exploration of how sexual activity, stress, and unusual cycle events correlate with PMS symptoms. Demographic information, including age, BMI, and marital status, alongside reproductive histories, such as number of pregnancies, miscarriages, and abortions, further enriches the model by accounting for personal health factors. However, these features provide a comprehensive view of the variables influencing PMS symptoms, serving as critical inputs for the Support Vector Machine (SVM) classification model and the Autoregressive Integrated Moving Average (ARIMA) forecasting system.

MODEL DEVELOPMENT

This study revolves around the development of predictive models to classify and forecast Premenstrual Syndrome (PMS) symptoms. Two primary machine learning techniques—K-means clustering and Support Vector Machines (SVM) will be employed for classification tasks, alongside AutoRegressive Integrated Moving Average (ARIMA) for time series forecasting.

K-MEANS CLUSTERING

ISBN

K-Means clustering, an unsupervised machine learning technique(Mehar et al., 2013), is applied in this study to identify distinct groups of women based on similar patterns of Premenstrual Syndrome (PMS) symptoms. The decision to use unsupervised learning stems from the absence of explicit PMS labels in the dataset. Since there are no predefined indicators of whether a woman experiences PMS or not, K-Means explore and uncover inherent groupings within the data based solely on feature similarities(Oti et al., 2021), such as menstrual cycle characteristics, symptom severity scores, lifestyle behaviors, and reproductive history.

By applying the K-Means algorithm, two distinct clusters emerged: one group predominantly exhibiting features associated with PMS, and another characterized by non-PMS symptom patterns. This unsupervised approach is particularly valuable as it provides a means to segment the population without needing labeled data.

DATA LABELING

Following the results of the K-Means clustering, which identified two distinct clusters, a labeling process will be employed to categorize the data. Each individual in the dataset will be assigned a label based on their cluster membership: PMS=1 for women in the cluster exhibiting PMS-related patterns and Non-PMS=0 for those in the cluster without significant PMS symptoms. This labeled data will then serve as the foundation for further supervised learning models, enabling more accurate prediction of PMS presence and severity in future analyses. By transforming the previously unsupervised dataset into a labeled one, this approach allows for more targeted predictive modeling using machine learning algorithms with Support Vector Machines (SVM).

SVM CLASSIFICATION

Support Vector Machines (SVM) (Cortes & Vapnik, 1995)are employed for classification tasks to categorize PMS symptoms based on the labeled data derived from the K-Means clustering process. SVM is particularly effective for high-dimensional datasets, allowing for the direct prediction of class labels (e.g., PMS or non-PMS) for new data points. The algorithm works by finding the optimal hyperplane that separates the classes in the feature space, ensuring robust classification performance even in the presence of noise and complex data(Sun et al., 2010). Hyperparameter tuning will be performed to optimize the model's performance, ensuring accurate predictions tailored to individual profiles (Yenaeng et al., 2014).

ARIMA PREDICTION

In addition to SVM, AutoRegressive Integrated Moving Average (ARIMA) is utilized for time series forecasting of PMS symptoms. ARIMA leverages historical data to predict future occurrences and severity of symptoms, providing valuable insights into potential symptom trajectories over time(Beaumont et al., 1984). This model is particularly suited for time-dependent data(Lund, 2007), allowing for the identification of trends and seasonal patterns in PMS symptom presentation. By combining SVM for classification with ARIMA for forecasting, this methodology aims to deliver a holistic understanding of PMS, facilitating improved management strategies for women experiencing these symptoms.

EVALUATION

ISBN

The performance of the developed models will be assessed using several appropriate metrics, including accuracy, precision, recall, F1-score, and correlation, providing a comprehensive evaluation of their predictive capabilities for both classification and forecasting. Accuracy will measure the overall correctness of the model's predictions, while precision and recall will offer insights into the model's ability to identify true positives (PMS cases) and minimize false positives. The F1-score will be particularly valuable as it balances precision and recall, serving as a single metric to evaluate performance in scenarios with imbalanced class distributions. Additionally, correlation analysis will assess the relationship between predicted symptom severity and actual symptom data, ensuring that the models align with real-world experiences.

Comparisons will be made between the model predictions and self-reported symptom data obtained from mobile health applications used by professional women. This alignment of model outputs with real-world symptom tracking will enhance the validity of the findings and facilitate the identification of effective intervention strategies for managing PMS. By integrating self-reported data into the evaluation process, the study aims to ensure that the predictive models not only achieve statistical significance but also relevant to the real experiences of women dealing with PMS.

PERSONALIZED PMS PREDICTION

Based on the predictions generated by the models, personalized recommendations for symptom management will be developed to empower women to take proactive measures in alleviating their Premenstrual Syndrome (PMS) symptoms. By tailoring these recommendations to individual profiles and predicted symptom severity, the approach aims to enhance women's capacity to manage their health effectively. This personalized strategy not only promotes greater engagement with symptom management but also provides women with actionable insights, enabling them to make informed decisions regarding lifestyle modifications, self-care practices, and the appropriate timing for seeking medical advice. Finally, this study seeks to improve the overall quality of life for women experiencing PMS by equipping them with the necessary tools and knowledge to navigate their symptoms more effectively.

EXPECTED RESULT

The expected outcome of this research is the establishment of a conceptual framework for developing a predictive model for Premenstrual Syndrome (PMS) that aims to effectively classify and forecast symptoms in professional women.

Firstly, the study intends to develop the framework for a predictive model that can accurately classify PMS symptoms using machine learning techniques. The proposed methodologies, Support Vector Machines (SVM), seeks to enhance the model's reliability and performance in predicting the presence and severity of PMS symptoms. While definitive classification results will not be available at this stage, the framework is expected to provide healthcare practitioners with insights on identifying at-risk individuals, thus informing future targeted intervention strategies.

Next, the research intends to outline a process for forecasting PMS symptoms based on historical data using AutoRegressive Integrated Moving Average (ARIMA). This time series forecasting approach will help establish methodologies for predicting the onset and severity of symptoms over time, empowering women to anticipate and manage their symptoms proactively. The conceptual framework aims to offer guidance on how to provide insights into symptom

ISBN

trajectories, equipping women with knowledge for implementing timely and effective management strategies.

Lastly, the study seeks to propose a pathway for offering personalized recommendations for symptom management derived from model predictions. By integrating individual symptom profiles with tailored management strategies, the research aspires to empower women to take proactive steps in alleviating their PMS symptoms. This personalized approach is envisioned to enhance women's engagement in their health management, ultimately contributing to improved quality of life and overall well-being. Through these conceptual outcomes, the research aims to make meaningful contributions to the field of women's health and lay the foundation for future clinical practices in PMS management.

DISCUSSION AND FUTURE WORK

The findings of this study underscore the potential of machine learning techniques, particularly Support Vector Machines (SVM), K-Means clustering, and AutoRegressive Integrated Moving Average (ARIMA), in enhancing the understanding and management of Premenstrual Syndrome (PMS). By identifying distinct clusters of women based on their symptom profiles, the research provides valuable insights into the variability of PMS experiences among individuals. This clustering approach not only reveals patterns in symptomatology but also highlights the importance of recognizing diverse manifestations of PMS across different populations. The proposed conceptual framework aims to facilitate the development of personalized recommendations for symptom management, empowering women to take proactive steps in alleviating their symptoms and fostering greater engagement with their health management strategies.

Despite these advancements, PMS remains a condition that is often overlooked in professional environments, where the interplay of work-related stress and hormonal fluctuations can exacerbate symptoms. Many women may not fully recognize the impact of PMS on their productivity and overall well-being, leading to a lack of appropriate support systems in the workplace. Addressing this gap is crucial for improving health outcomes for women in professional settings.

Future work should focus on several key areas to enhance the conceptual framework and its applications. First, expanding the dataset to include a more diverse population will enhance the generalizability of the predictive models. By capturing a wider range of demographic variables, such as age, ethnicity, and lifestyle factors, the framework can better reflect the experiences of all women affected by PMS.

Second, tracking changes in symptom patterns over time will be implemented in the future. This approach would allow for the refinement of the predictive framework, enabling it to adapt to individual variations and shifts in symptom severity. Incorporating qualitative data, such as personal experiences and coping strategies, could further enrich the understanding of PMS and inform the development of tailored interventions that resonate with women’s lived experiences.

Additionally, exploring the integration of wearable technology and mobile health applications could facilitate real-time symptom tracking and provide immediate feedback to users. This capability would empower women to monitor their symptoms proactively and receive timely notifications when PMS is anticipated. Such a system could enhance adherence to personalized

ISBN

management strategies, ultimately improving the effectiveness of interventions.

By addressing these areas, future research can contribute to a more comprehensive understanding of PMS and its management, ultimately improving the quality of life for women affected by this condition. These developments will not only aid in symptom management but also help raise awareness about PMS in professional settings, fostering a supportive environment for women facing these challenges.

REFERENCES

Beaumont, C., Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1984). Forecasting: Methods and Applications. *The Journal of the Operational Research Society*, *35*(1). https://doi.org/10.2307/2581936

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3). https://doi.org/10.1023/A:1022627411411

Gao, M., Gao, D., Sun, H., Cheng, X., An, L., & Qiao, M. (2021). Trends in Research Related to Premenstrual Syndrome and Premenstrual Dysphoric Disorder From 1945 to 2018: A Bibliometric Analysis. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.596128

Kahyaoglu Sut, H., & Mestogullari, E. (2016). Effect of Premenstrual Syndrome on Work-Related Quality of Life in Turkish Nurses. *Safety and Health at Work*, *7*(1). https://doi.org/10.1016/j.shaw.2015.09.001

Lund, R. (2007). Time Series Analysis and Its Applications: With R Examples. *Journal of the American Statistical Association*, *102*(479). https://doi.org/10.1198/jasa.2007.s209

Mehar, A. M., Matawie, K., & Maeder, A. (2013). Determining an optimal value of K in K-means clustering. *Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013*. https://doi.org/10.1109/BIBM.2013.6732734

Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, *07*(08). https://doi.org/10.31695/ijasre.2021.34050

Richard J. Fehring. (2012). *Menstrual Cycle Data" (2012). Randomized Comparison of Two Internet-Supported Methods of Natural Family Planning. 7*.

Sistiarani, C., Hariyadi, B., Wahyuningsih, E., & Nurhayati, S. (2023). Stress Level and Length of Use Contraception to Premenstrual Syndrome (PMS) Type in Childbearing Age Women. *Global Journal of Reproductive Medicine*, *10*(1). https://doi.org/10.19080/gjorm.2023.10.555779

ISBN

Sun, J., Zheng, C., Li, X., & Zhou, Y. (2010). Analysis of the distance between two classes for tuning SVM hyperparameters. *IEEE Transactions on Neural Networks*, *21*(2). https://doi.org/10.1109/TNN.2009.2036999

Tsegaye, D., & Getachew, Y. (2019). Premenstrual dysphoric disorder and associated factors among female health science students in Wollo University, Ethiopia, 2017/18. *Maternal Health, Neonatology and Perinatology*, *5*(1). https://doi.org/10.1186/s40748-019-0102-z

Yenaeng, S., Saelee, S., & Samai, W. (2014). Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms. *International Journal of Information and Education Technology*, *4*(2). https://doi.org/10.7763/ijiet.2014.v4.384

Yonkers, K. A., O'Brien, P. M. S., & Eriksson, E. (2008). Premenstrual syndrome Seminer Lancet 05_04p1200_1210G.indd. *Lancet*, *371*.

ISBN