

Women's Health Recommendation System using Large Language Models

Priyanka Kaushik
Dept. of CSE (AIML), CSE-APEX
Chandigarh University
Chandigarh, India
kaushik.priyanka17@gmail.com
<https://orcid.org/0000-0003-4766-8772>

Priyanka Rawat
Dept. of CSE
Chandigarh University
Chandigarh, India
priyankaktwr@gmail.com

Bommuluri Bhavana Rao
Dept. of Civil Engineering
Galgotias University
Greater Noida, India
bbhavanarao@gmail.com

Mehna Najeem
Dept. of Civil Engineering
Dhanalakshmi Srinivasan College of Eng.
Coimbatore, India

Mamta Bishnoi
Dept. of Pharmacy
Lingayas Vidyapeeth
Haryana, India

S. Kaliappan
Division of R & D
Lovely Professional University
Punjab India
srini.kal_lpu@yahoo.com

Abstract—Globally, numerous systems exist for tracking women's health, recording menstrual cycles, and predicting outcomes. However, most of these applications primarily focus on identifying symptoms and predicting the next menstrual cycle. Few solutions aim to enhance women's overall physical health and immunity. To address this gap, we propose a Women's Health Recommendation System designed to cater to women's diverse health needs. This system leverages crowdsourced data and large language models (LLMs) to provide personalized dietary, exercise, and lifestyle recommendations tailored to the phases of the menstrual cycle. Polycystic Ovary Syndrome (PCOS), a leading cause of anovulation and infertility, serves as a critical focus. While there is no definitive cure for PCOS, symptoms can be managed through medication, lifestyle modifications, and fertility treatments, with prevention supported by maintaining a healthy lifestyle. The proposed system integrates a prediction model and an LLM. The prediction model classifies menstrual cycle phases to inform the generation of customized diet and exercise recommendations, while the LLM addresses common menstruation-related queries. This comprehensive solution empowers users with actionable insights and suggestions to enhance their overall health and well-being.

Index Terms—Menstruation, Large Language Model, Machine Learning, Polycystic Ovary Syndrome, Women's Health

I. INTRODUCTION

Around the world, there are numerous menstrual tracking, prediction, and health tracking systems for women. However, they primarily concentrate on identifying the symptoms and estimating the next menstrual cycle date. There are relatively few applications that concentrate on enhancing women's physical well-being and health [1]. Designing a women's health recommendation system should consider diverse health needs, accurately predict menstrual cycle phases, adapt to hormonal variations, and offer tailored health advice using advanced machine learning and user-centered design. Using a large language model, and machine learning, our proposed system

can predict menstrual phases, empowering users with valuable insights into their health. Additionally, our initiative mainly focuses on lifestyle changes and holistic methods as preventive measures against polycystic ovarian syndrome (PCOS) and polycystic ovarian disease (PCOD). PCOS is a common cause of anovulation and a leading cause of infertility. As it is a chronic illness, there is no known cure. However, medication, lifestyle modifications, and fertility treatments can all help with some symptoms [2]. A healthy diet and active lifestyle play a pivotal role in the

prevention of such hormonal disorders. The proposed solution will equip users with lifestyle suggestions that can help prevent such disorders and promote healthy well-being. The recommendation system will utilize large language models to generate personalized dietary, exercise, and lifestyle recommendations aligned with menstrual cycle phases. Through continuous refinement based on user feedback, the accuracy and relevance of the recommendation system can be enhanced, and user empowerment and engagement can be fostered.

A. Drawbacks of Existing Menstrual Tracking Applications in the market There are several critical issues present in menstrual tracking apps today, as shown in Figure 1, including:

Inadequate Adaptation to Hormonal Fluctuations: Many menstruation tracking applications in the market do not account for the complex hormonal changes that occur during menarche, menopause, and hormonal disorders. [3]. This inaccuracy leads to erroneous dates being forecast, compromising the reliability of tracking systems. In addition, they disregard hormonal imbalances that could impact the menstrual cycle.

Lack of Health Recommendations: Positive lifestyle interventions are limited by the absence of proactive health suggestions in menstrual tracking applications, which leaves users without direction for cycle-specific good practices.

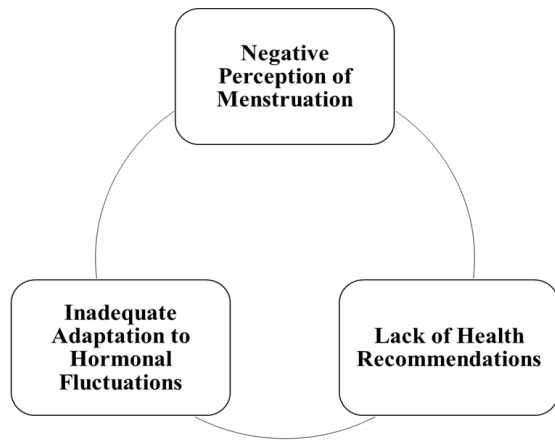


Fig. 1. Drawbacks of existing menstrual tracking applications

By providing individualized dietary, exercise, and nutrition recommendations, the proposed recommendation system can help individuals lead healthier lifestyles.

Negative Perception of Menstruation: Many existing applications inadvertently perpetuate a negative perception of menstruation among users [4]. Due to their inability to promote positive awareness of menstrual health, these applications contribute to the stigma surrounding menstruation, a natural biological occurrence. These apps' designs make sexual and gender minorities feel excluded [5].

Proposed Solution The large language model based Women's Health Recommendation System offers a comprehensive and user-centric approach to tracking menstruation and encouraging well-being to bridge the gaps mentioned in Section I A. Our system combines machine learning algorithms with innovative features to deliver precise phase and cycle forecasts, promote a positive menstruation narrative, and provide personalized advice on healthy behaviors.

rhythms to forecast menstrual cycle phases and adjust to hormonal fluctuations. Proactive nutrition recommendation systems show positive results in users' overall health [5]. Furthermore, this solution will also be able to predict any developing hormonal disorders in an individual in time using machine learning applications. Such existing prediction systems have proven to be reliable for healthy adult females.

- Our recommendation system utilizes a large language model and crowdsourced data to generate personalized recommendations for diet, exercise, and lifestyle practices tailored to individual menstrual cycle phases. This will empower users to make informed choices that optimize their well-being.
- The use of crowdsourced data, which allows us to enhance and optimise recommendation algorithms based on

- shared knowledge and experiences, is essential to our system's effectiveness. Through the utilisation of community knowledge, we can continuously enhance the accuracy and relevance of our health recommendations, fostering a collaborative atmosphere that promotes women's health empowerment.
- A study shows variation in heart rate, QT, QTc intervals in different phases of menstrual cycle where heart rate was significantly increased during menstruation. Variation in phases of menstrual cycle causes physiological alteration of cardiovascular activities and affects Electrocardiogram (ECG) parameters.
- Central to our solution is the promotion of a positive and empowering narrative surrounding menstruation. Through educational resources, supportive content, and community engagement features, we aim to dismantle stigma and foster a culture of informed menstrual health management. Figure 2 explains the various aspects of our proposed solution. Our solution employs user-centric design principles to provide accurate tracking, promote a positive narrative surrounding menstruation, and deliver personalized health recommendations tailored to individual needs. By empowering women with comprehensive insights and actionable strategies, we aim to foster a culture of informed menstrual health management and holistic wellness.

II. LITERATURE REVIEW

Sharma et al. presented a method to segment customers using K-Means and RFM models, enabling businesses to optimize revenue strategies through targeted segmentation. Their approach offers practical insights into data-driven customer behavior analysis for enhancing profitability [6]. Uddin and Hossain examined elders' health through personal networks, highlighting the significance of social interactions in managing physical, mental, and psychological well-being, paving the way for elder-focused health interventions [7]. Hamid and Brohi reviewed large language models in healthcare, categorizing threats and vulnerabilities while proposing a framework to mitigate risks, underscoring their transformative potential in medical applications [8]. Mallya and Kothari proposed an organizational structure for women's health monitoring using OMAS, demonstrating its utility in streamlining health data management for personalized care delivery systems [9]. Ettiyian and Geetha surveyed IoT-based

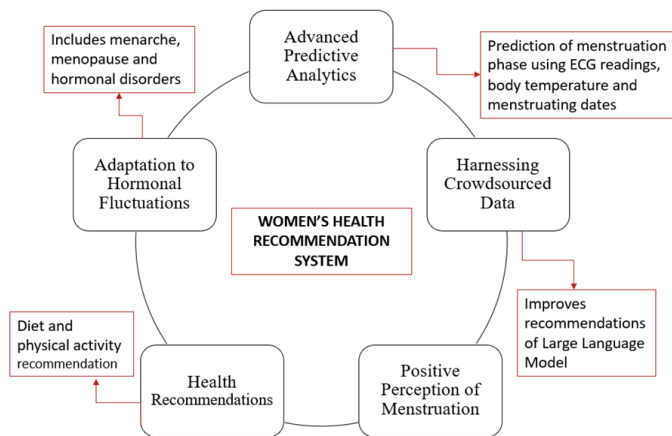


Fig. 2. Five Pillars of Women's Health Recommendation System

Our proposed system employs machine learning algo-

maternity health monitoring, emphasizing its potential to enhance maternal care accessibility and real-time monitoring capabilities, ensuring timely medical intervention [10]. Rathore et al. developed a predictive model for Black Friday sales using smart analytics, enabling businesses to forecast sales trends and improve inventory and marketing strategies effectively [11]. Cetera et al. explored AI applications in managing endometriosis and adenomyosis, asserting that machine-based approaches could surpass human expertise in diagnostic accuracy and treatment outcomes [12]. Mallya et al. designed an adaptive information system for women's health monitoring, emphasizing dynamic requirement modeling to enhance system efficiency and adaptability to user-specific needs [13]. Alamsyah et al. proposed a maternal health risk classification system using Random Forest and Evolutionary Algorithms, showcasing its efficacy in predicting health risks for improved maternal care [14]. Kaushik et al. investigated advanced data analytics for fraud detection in insurance claims, presenting techniques that enhance fraud identification accuracy and bolster financial sustainability in the insurance sector [15]. Nazi and Peng reviewed large language models in medical domains, outlining their applications, limitations, and future potential, emphasizing their capability to revolutionize healthcare delivery and research [16].

S. K. Rajamani and R. S. Iyer (2023) explored the potential of Large Language Models (LLMs) in enhancing women's safety through computational intelligence. The study presents innovative applications of LLMs to address safety concerns, underscoring the transformative role of AI in societal welfare [17]. Kaushik (2023) introduced a multi-agent deep learning approach for detecting cyber-attacks in IoT systems. This work highlights the robustness of AI frameworks in safeguarding IoT networks, emphasizing the integration of security protocols in smart environments [18]. K. Nassiri and M. A. Akhlofi (2024) reviewed advancements in LLMs within healthcare. The authors identified breakthroughs in medical diagnostics and patient engagement, showcasing how LLMs are revolutionizing healthcare delivery through intelligent data processing [19]. T. Yoldemir (2020) examined AI's role in women's health, presenting insights into its capabilities for disease prediction and personalized treatment. The study advocates for incorporating AI to advance healthcare accessibility for women [20]. P. L. Lau et al. (2023) conducted a systematic review on AI's contributions toward achieving UN Sustainable Development Goals, particularly in women's healthcare. Their findings demonstrate the integration of AI-driven solutions to address healthcare disparities and improve global health outcomes [21]. Kaushik et al. (2023) developed a framework using multiscale adaptive object detection and contrastive feature learning for analyzing customer behavior in retail. This study bridges AI technology with business insights, driving efficiency in customer interactions [22]. M. Geantă et al. (2024) highlighted LLMs' emerging role in enhancing prostate cancer literacy. Their study emphasizes AI's potential to simplify complex medical information, empowering patients with better health

literacy [23]. K. Sahoo et al. (2019) proposed DeepReco, a health recommender system leveraging deep learning and collaborative filtering. The system demonstrated effectiveness in providing personalized health recommendations, advancing AI's role in proactive healthcare [24]. H.-K. Kim (2024) conducted a systematic review and meta-analysis on AI chatbots' effects on women's health. The research validated the effectiveness of chatbots in providing accurate health information and improving women's healthcare engagement [25].

III. METHODOLOGY

The Women's Health Recommendation System's main goal is to create a novel solution that promotes users' overall wellness and addresses the drawbacks of existing menstrual tracking apps. It seeks to accomplish the following objectives: **Accurate Menstrual Cycle Tracking:** Create predictive analytics for reliable menstrual cycle monitoring that adjusts to hormonal swings throughout life.

Personalised Health advice using Large Language Model: Utilise sophisticated language models to provide personalised food, exercise, and lifestyle advice that coincide with menstrual cycle stages.

Positive Menstrual Health Promotion:

Promote positive menstrual narratives by providing educational materials and supporting content, to reduce stigma and foster educated period health management. Our solution performs fine-tuning on large language models [26]. LLM fine-tuning is a technique that involves adapting the parameters of a pre-trained language model (like BERT, Flan T5, GPT) to train it on a specific domain or task. This means training the model on a smaller, task-specific dataset to tailor its learned representations to the specifics of the target task. Fine-tuning takes advantage of the information gained during pre-training, resulting in increased performance for the specific task or domain while decreasing the requirement for considerable computing resources. Figure 3 shows the process of fine-tuning a large language model, Flan T5 base.

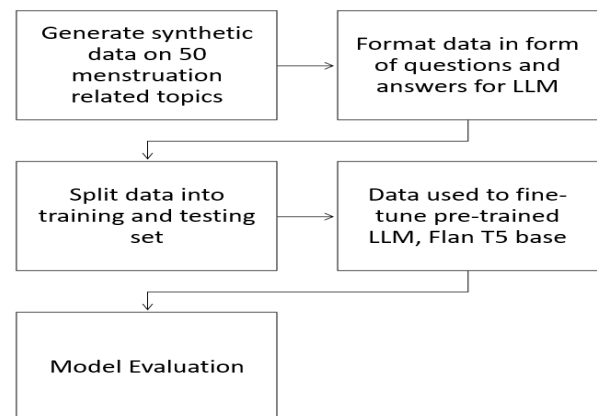


Fig. 3. Fine-tuning the Large Language Model

Datasets Used

The project's dataset, gathered in 2012 by Richard J. Fehring at Marquette University, includes a wide range of variables related to the menstrual cycle as well as several demographic, health, and lifestyle factors. The dataset includes information such as cycle number, study group, peak fertility indicators, cycle length, menstrual intensity, reproductive history, age, marital status, education, ethnicity, medical history, contraceptive use, fertility intentions amongst 80 attributes. It records 1665 cycles of 159 individuals over a period of time [27]. With its extensive coverage of menstrual cycle parameters, this dataset serves as a valuable resource for predicting menstruation phases and understanding the relationships between menstrual health and individual characteristics. The data has been used to predict the phase of the menstrual cycle that a person is undergoing at a given time.

For fine-tuning, Google Flan T5 Base large language model is used. The model has been pre-trained on 250 million parameters as of March 2024. We have curated a small yet impactful dataset consisting of 530 questions and corresponding answers related to over 50 distinct topics on menstruation and menstrual health. The data covers many domains, including but not limited to, menstrual hygiene management, menopause and menarche, cultural and technological aspects of menstruation, menstrual health education, environmental impact, treatments, taboos and stigma, menstrual products, health and fitness, nutrition, sanitation, phases and hormonal changes throughout menstruation and related disorders. This LLM seeks to raise people's knowledge of menstrual health and promote awareness about it among individuals [28]. By using this dataset, we hope to improve the model's capacity to produce insightful and beneficial responses to a variety of menstrual health-related inquiries, which will further the more general objective of fostering menstrual health literacy and wellbeing.

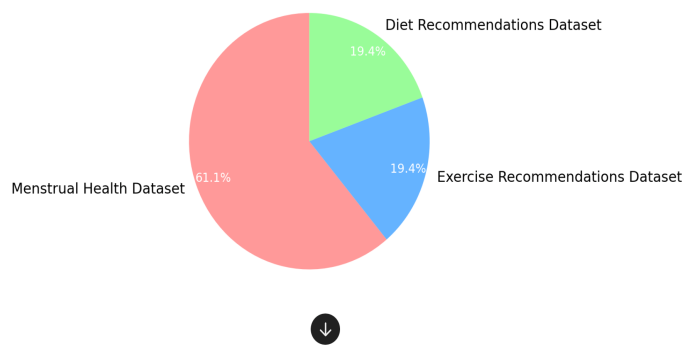


Fig. 4. Datasets used in the women's health recommendations system

In our project development process, we crafted two mini

datasets, which included exercise recommendations and diet recommendations correlating to different menstrual cycle phases. These datasets serve as valuable resources for training and fine-tuning our models to provide personalized guidance on exercise regimens and dietary choices tailored to each phase of the menstrual cycle. By leveraging these datasets, our goal is to empower individuals with actionable insights that optimize their physical well-being throughout their menstrual cycle.

IV. RESULTS

We examined the finer details of menstrual cycle patterns and found that there is a significant standard deviation of 3.92 days in when ovulation occurs in individuals. In order to successfully address this variability, we leveraged the predictive power of a random forest regressor model. The random forest regressor is a machine learning technique that creates different decision trees during training and provides their average prediction for regression problems. By training each tree on a random subset of the training data and a random selection of features, the model reduces overfitting and increases resilience. Impressively, our model exhibited a remarkable accuracy rate of 97% in forecasting the length of menstruation, showcasing its proficiency in this aspect. However, when it came to estimating the day of ovulation, our model demonstrated a comparatively lower accuracy rate of 56%. While this discrepancy suggests inherent challenges in precisely predicting ovulation, it also highlights the complexity of menstrual cycle dynamics and the need for further research and development of predictive models in order to increase their effectiveness in this domain.

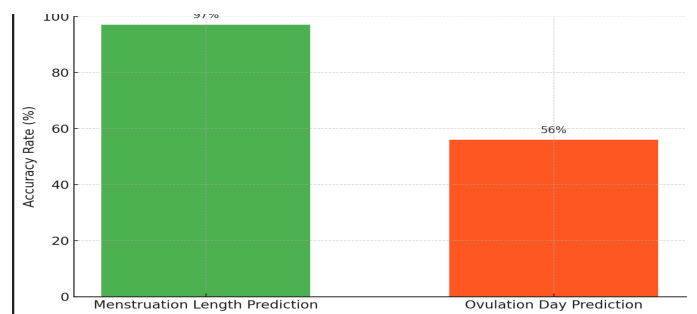


Fig. 5. Prediction accuracy rates of random forest model

V. DISCUSSION

Our analysis found that, while our dataset provided useful insights, its capacity to properly forecast ovulation dates was restricted due to the inherent variability in menstrual cycles between people. However, we saw chances for development and expansion in numerous critical areas. To begin with, there is a scarcity of open access to standardised menstrual datasets. While this might be due to a variety of causes, including privacy concerns, cultural taboos, and the sensitive nature of the subject, a lack of relevant data impedes research and development activities targeted at improving menstrual health outcomes and understanding associated

phenomena. To overcome data scarcity, we may increase phase prediction model performance by using datasets with more data points, more diversified menstrual cycle patterns, or curating synthetic datasets that emulate the properties of real menstrual data. Furthermore, increasing the recommendation system dataset to include a broader variety of exercise and food advice targeted to various stages of the menstrual cycle can improve our model's suggestive capabilities. Furthermore, expanding our large language model with more comprehensive training data can allow it to provide more nuanced and useful responses. While predicting ovulation dates may remain challenging, these improvements will surely help to refine our understanding of menstrual health and improve the effectiveness of our predictive algorithms. While our dataset provided useful insights, reliably predicting ovulation dates proved challenging due to menstrual cycle variability. However, there is room for improvement. Expanding exercise and nutrition recommendation data, scaling up the language model, and integrating Electrocardiogram (ECG) and Basal Body Temperature (BBT) data can all improve the predictive capabilities of the model. Furthermore, by increasing and enhancing exercise and diet recommendations to incorporate regional recipes, possibly through crowd sourcing, we can further enhance the recommendation system.

VI. CONCLUSION

The Women's Health Recommendation System provides a ground-breaking solution to significant flaws in existing period monitoring applications while also boosting overall well-being among users. Our effort transforms the menstruation tracking experience by leveraging machine learning algorithms, personalized health suggestions based on user's current menstrual phase, and creating a positive menstrual health narrative, encouraging women to prioritize their health and well-being. Personalized health recommendations based on large language models allow users to make more educated decisions, promoting proactive health management as well as positive lifestyle changes. While our dataset provided useful insights, its ability to properly predict ovulation dates was restricted due to menstrual cycle variability. However, there are areas for improvement. Expanding data on menstrual cycles can improve phase prediction accuracy. The large language model training dataset can be scaled up to get better results. While precise ovulation prediction remains difficult, these advancements will improve menstrual health understanding and predictive algorithms. To summarise, the Women's Health Recommendation System provides an in-depth solution for complicated women's health issues by leveraging the power of large language models and advanced machine learning algorithms. Using technology, data insights, and user-centric design principles, our system aims to enable women worldwide to live healthier, happier lives and encourage positive attitudes towards menstruation health and well-being.

REFERENCES

- [1] S. F. Aktar, P. B. Upama and S. I. Ahamed, "Leveraging Technology to Address Women's Health Challenges: A Comprehensive Survey," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 941-949.
- [2] A. Gupta, S. Kumar and P. M. Pattanaik, "A Novel HISMO Solution to Coverage Hole Mitigation in 6G-IoT," 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2022, pp. 1-5.
- [3] N. N. Jones and M. F. Williams, "Archives, Rhetorical Absence, and Critical Imagination: Examining Black Women's Mental Health Narratives at Virginia's Central State Hospital," in IEEE Transactions on Professional Communication, vol. 65, no. 1, pp. 179-196, March 2022.
- [4] K. Denecke, R. May, and O. Rivera Romero, "Potential of Large Language Models in Health Care: Delphi Study," Journal of Medical Internet Research, vol. 26. JMIR Publications Inc., p. e52399, May 13, 2024.
- [5] N. R. P. N, P. S and J. K B, "FemmeVibe: Device to amplify Women's Health and Harmony," 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India, 2024, pp. 1-6.
- [6] V. Sharma, P. Agarwal, H. Y. Shaikh, R. M. Lenka, S. K. Manjhi and R. Rathore, "Smart Next-Generation Revenue Growth: A Methodology for Partitioning Customers Utilizing the K-Means Algorithm and RFM Model," 2023 International Conference on Smart Devices (ICSD), Dehradun, India, 2024, pp. 1-6.
- [7] M. S. Uddin and L. Hossain, "Exploring physical, mental and psychological health for elders through their personal networks," 2009 11th International Conference on e-Health Networking, Applications and Services (Healthcom), Sydney, NSW, Australia, 2009, pp. 29-35.
- [8] R. Hamid and S. Brohi, "A Review of Large Language Models in Healthcare: Taxonomy, Threats, Vulnerabilities, and Framework," Big Data and Cognitive Computing, vol. 8, no. 11. MDPI AG, p. 161, Nov. 18, 2024.
- [9] R. Mallya and S. Kothari, "Organizational structure for woman's health monitoring system using OMAS," 2017 Conference on Information and Communication Technology (CICT), Gwalior, India, 2017, pp. 1-6.
- [10] R. Ettiyan and V. Geetha, "A Survey of Health Care Monitoring System for Maternity Women Using Internet-of-Things," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1290-1296.
- [11] Rathore, S. P. S., Bihade, V. M., Sane, S., Limbore, N. V., Lenka, R., & Priyanka. (2024). Smart Model For Black Friday Sales Prediction. In 2023 International Conference on Smart Devices (ICSD) (pp. 1-5). 2024 International Conference on Smart Devices (ICSD).
- [12] G. E. Cetera, A. E. Tozzi, V. Chiappa, I. Castiglioni, C. E. M. Merli, and P. Vercellini, "Artificial Intelligence in the Management of Women with Endometriosis and Adenomyosis: Can Machines Ever Be Worse Than Humans?," Journal of Clinical Medicine, vol. 13, no. 10. MDPI AG, p. 2950, May 16, 2024.
- [13] R. Mallya, S. Kothari and N. Mehta, "Requirements model for adaptive information system for woman health monitoring system," 2018 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2018, pp. 1-6.
- [14] D. P. Alamsyah, Y. Ramdhani and L. Susanti, "Maternal Health Risk Classification: Random Forest and Evolutionary Algorithms," 2023 10th International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2023, pp. 1-6.
- [15] P. Kaushik, S. P. S. Rathore, A. S. Bisen and R. Rathore, "Enhancing Insurance Claim Fraud Detection Through Advanced Data Analytics Techniques," 2024 IEEE Region 10 Symposium (TENSYP), New Delhi, India, 2024, pp. 1-5.
- [16] Z. A. Nazi and W. Peng, "Large Language Models in Healthcare and Medical Domain: A Review," Informatics, vol. 11, no. 3. MDPI AG, p. 57, Aug. 07, 2024.
- [17] S. K. Rajamani and R. S. Iyer, "Application of Large language Models (LLMs) in Women's Safety," Advances in Computational Intelligence and Robotics. IGI Global, pp. 80-103, Dec. 01, 2023.
- [18] Kaushik, P. (2023). Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT. International Journal for Global Academic & Scientific Research, 2(2), 15-29

- [19] K. Nassiri and M. A. Akhloufi, "Recent Advances in Large Language Models for Healthcare," *BioMedInformatics*, vol. 4, no. 2. MDPI AG, pp. 1097–1143, Apr. 16, 2024.
- [20] T. Yoldemir, "Artificial intelligence and women's health," *Climacteric*, vol. 23, no. 1. Informa UK Limited, pp. 1–2, Jan. 02, 2020.
- [21] P. L. Lau, M. Nandy, and S. Chakraborty, "Accelerating UN Sustainable Development Goals with AI-Driven Technologies: A Systematic Literature Review of Women's Healthcare," *Healthcare*, vol. 11, no. 3. MDPI AG, p. 401, Jan. 31, 2023.
- [22] Kaushik, P., Singh Rathore, S. P., Kaur, P., Kumar, H., & Tyagi, N. (2023). Leveraging Multiscale Adaptive Object Detection and Contrastive Feature Learning for Customer Behavior Analysis in Retail Settings. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(6s), 326–343.
- [23] M. Geantă et al., "The Emerging Role of Large Language Models in Improving Prostate Cancer Literacy," *Bioengineering*, vol. 11, no. 7. MDPI AG, p. 654, Jun. 27, 2024.
- [24] K. Sahoo, C. Pradhan, R. K. Barik, and H. Dubey, "DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering," *Computation*, vol. 7, no. 2. MDPI AG, p. 25, May 22, 2019.
- [25] H.-K. Kim, "The Effects of Artificial Intelligence Chatbots on Women's Health: A Systematic Review and Meta-Analysis," *Healthcare*, vol. 12, no. 5. MDPI AG, p. 534, Feb. 23, 2024.
- [26] M. Verma, A. Kumar and S. Kumar, "Medical Waste Classification using Deep Learning and Convolutional Neural Networks," 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2022, pp. 1-5.
- [27] W. Fan et al., "A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models," *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, pp. 6491–6501, Aug. 24, 2024.
- [28] S. Pathania et al., "An Efficient Electrical-Thermal Co-Design Methodology for Analysis of High-Speed PCB Interconnects," 2023 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), Winnipeg, MB, Canada, 2023, pp. 154-157.