Enhancing Ovarian Cancer Detection Using Self-Organizing Maps and Improved Recurrent Neural Networks with Extended Harmony Search Optimization

Priya Vij [1*], Patil Manisha Prashant [1]

Abstract

Background: Ovarian cancer (OC) is a highly fatal malignancy of the female reproductive system, characterized by its high mortality rate and the challenges associated with clinical research due to the disease's complexity and late-stage diagnosis. Advances in technology, such as the Internet of Medical Things (IoMT), offer new opportunities for improving OC detection and diagnosis. Objective: This study aimed to develop and evaluate a novel method for OC detection using IoMT data, leveraging Self-Organizing Maps (SOM) and Improved Recurrent Neural Networks (IRNN) enhanced with the Extended Harmony Search Optimization (EHSO) algorithm to improve feature selection and classification accuracy. Methods: The study utilized OC data from the IoMT and applied SOM for feature selection, which helps in managing and classifying large datasets. SOM was employed to improve data representation and address challenges in labeling and classifying data. The IRNN model, optimized using the EHSO algorithm, was developed to enhance classification performance. The model was tested using a dataset from Kaggle comprising 179 benign and 172 malignant OC images with 50 attributes. Results: The IRNN model with EHSO demonstrated superior performance compared to other methods. In the training dataset, it achieved an accuracy of 95.72% and a Root Mean Square Error (RMSE) of 4.8%. For the testing dataset, the model maintained a high accuracy of 90.4% and an RMSE of 6.8%. The IRNN with EHSO outperformed alternative methods in terms of specificity and sensitivity, while the Genetic Algorithm (GA) showed the lowest performance across all metrics. Conclusion: The proposed method using SOM and IRNN with EHSO significantly improves the detection of ovarian cancer by optimizing feature selection and classification accuracy. This approach offers a promising advancement in utilizing IoMT data for early and accurate OC diagnosis, potentially enhancing patient outcomes through more effective detection and treatment strategies.

Keywords: Ovarian Cancer, Internet of Medical Things, Self-Organizing Maps, Recurrent Neural Networks, Extended Harmony Search Optimization

Significance | This study determined a novel approach combining SOM and IRNN with EHSO to enhance early detection of ovarian cancer, improving diagnostic accuracy and efficiency.

*Correspondence. Priya Vij, Department of CS & IT, Kalinga University, Raipur, India.
E-mail: ku.priyavij@kalingauniversity.ac.in

Editor Surendar Aravindhan, Ph.D., And accepted by the Editorial Board Sep 08, 2024 (received for review Jul 03, 2024)

1. Introduction

Ovarian cancer (OC) is widely recognized as one of the deadliest cancers affecting the female reproductive system. It holds the highest mortality rate among cancers of the reproductive system and is the fifth most common cause of cancer-related death in

Author Affiliation.
[1] Department of CS & IT, Kalinga University, Raipur, India.

Please cite this article.
Priya Vij and Patil Manisha Prashant (2024). Enhancing Ovarian Cancer Detection Using Self-Organizing Maps and Improved Recurrent Neural Networks with Extended Harmony Search Optimization, Journal of Angiotherapy, 8(9), 1-5, 9877

women (Takahashi et al., 2020). The high mortality associated with OC makes it particularly significant among malignancies of the female reproductive tract. Women face a lifetime risk of 1 in 81 of developing OC, with a 1 in 98 chance of dying from it. The majority of those diagnosed are older, with about 50% being 63 years or older. White women are more commonly diagnosed than Black women (Yang et al., 2023).

Advancements in treatment have been driven by a focus on improving patient outcomes through subtype-specific methods (Chatterjee et al., 2024). Accurate subtype identification is essential for personalized therapies, yet this remains challenging due to inconsistencies in diagnosis, reliance on pathologist expertise, and limited access to specialized care, particularly in underserved regions (Ramana & Ravisankar, 2024). With a global incidence rate of 3.35% and a mortality rate of 4.8%, OC poses a significant public health concern, affecting about two million women and causing approximately 150,000 deaths annually (Su et al., 2023). The prognosis is poor, with only 32% of women surviving after diagnosis. Current treatment primarily involves chemotherapy and cytoreductive surgery (Hinchcliff et al., 2022).

The World Health Organization (WHO) classifies OC into five primary subtypes based on factors such as immune profile, histopathology, and genetic analysis. High-Grade Serous Carcinoma (HGSC) accounts for 75-82% of cases, followed by rarer forms such as endometrioid, clear cell, and mucinous carcinomas (Cree et al., 2020; Chen et al., 2021). OC is a heterogeneous disease with varied origins, metastasis patterns, treatment responses, and survival outcomes (Prat et al., 2018). This study addresses these diagnostic challenges by exploring the application of Internet of Medical Things (IoMT) and deep learning to OC classification.

Though OC initially responds well to chemotherapy, recurrence rates remain high, between 66% and 83%, particularly in advanced cases (Ganguly, 2020). Research has increasingly focused on predicting disease progression and treatment outcomes, utilizing novel data sources to improve detection accuracy (Engqvist et al., 2020). Early detection remains a critical issue, with ongoing efforts to identify biomarkers through serum-level testing (Surendar et al., 2024). Artificial intelligence, specifically computer-aided diagnosis, is also being investigated as a potential tool in enhancing OC diagnostic accuracy (Escorcia-Gutierrez et al., 2022).

Consiglio et al. (2021) conducted a study employing a combination of fuzzy rule modeling and genetic algorithms (GA) to analyze a dataset comprising 23 cases across five categories. This method proved valuable for analyzing gene expression profiles in ovarian cancer (OC) and comparing them to other ovarian diseases (Ramakrishnan et al., 2019). Specifically, their approach utilized GA for feature selection, then constructed a series of if-then rules to identify changes in gene expression that characterize different classes of OC (Consiglio et al., 2021).

Bae et al. (2021) developed a feature selection method that effectively differentiates colorectal cancer patients from healthy individuals using a modified harmony search algorithm (HSA) combined with K-means clustering. This method selected genes based on the Fisher score, after which K-means was used to group them. The modified HSA was then applied to build the file system, representing a streamlined approach for handling large datasets (Bae et al., 2021).

Paik et al. (2019) introduced a gradient boosting (GB)-based predictor that accurately identifies patients with epithelial ovarian cancer (EOC). Their model demonstrated improved accuracy over traditional statistical methods by identifying predictive subgroups among EOC patients (Paik et al., 2019).

In a separate study, Arfiani and Rustam (2019) utilized machine learning techniques, specifically Random Forest (RF), to classify OC as either benign or malignant. Their use of a bagging approach reduced overfitting and optimized the classifier's performance, showing RF as a reliable method for analyzing incomplete data (Arfiani & Rustam, 2019). Elhoseny et al. (2019) further refined OC classification through the integration of a self-organizing map (SOM) and an optimal recurrent neural network (RNN), enhancing detection accuracy via the adaptive harmony search optimization (AHSO) algorithm.

Research in OC continues to uncover its complexity, revealing it as a group of diseases with distinct origins, genetic mutations, and risk factors (Elhoseny et al., 2019). However, challenges remain in addressing the long-term care and psychosocial needs of OC survivors, an area of research still needing more attention.

2. Materials and Methods

Figure 1 illustrates the proposed integration of the Internet of Medical Things (IoMT) into a network system designed to enhance patient monitoring and care. The methodology involves several key components and processes. Initially, biological specimens, critical physiological signals, and other relevant monitoring data are collected from patients using IoMT-enabled devices, which continuously capture and record various health metrics in real-time. This data is then transmitted through routing networks to health advisors or physicians via secure and reliable channels, ensuring the privacy and integrity of patient information. Health advisors or physicians can access this transmitted data remotely through specialized software platforms, allowing them to review patient information, analyze the data, and provide necessary feedback or therapeutic recommendations. To ensure secure access, patients must provide valid identification credentials, which link them to their specific data and enable personalized interactions with the network. The successful expansion and effectiveness of the IoMT network rely on ongoing advancements in digital technology,

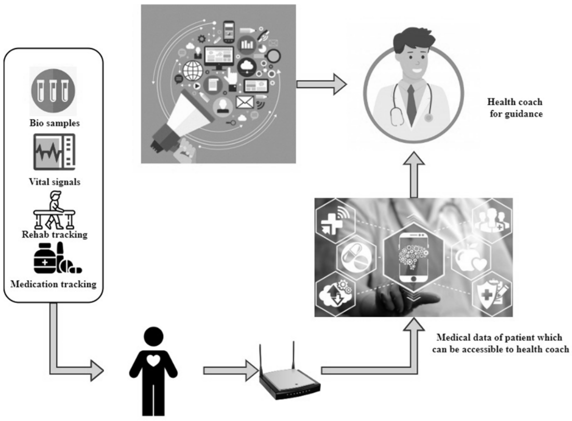

Figure 1 IoMT framework

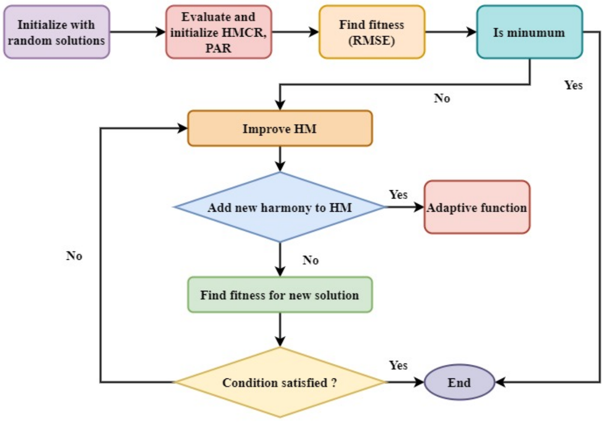

Figure 2 Extended Harmony Search Optimization (EHSO) Algorithm in IRNN for effective FS and classifying OC

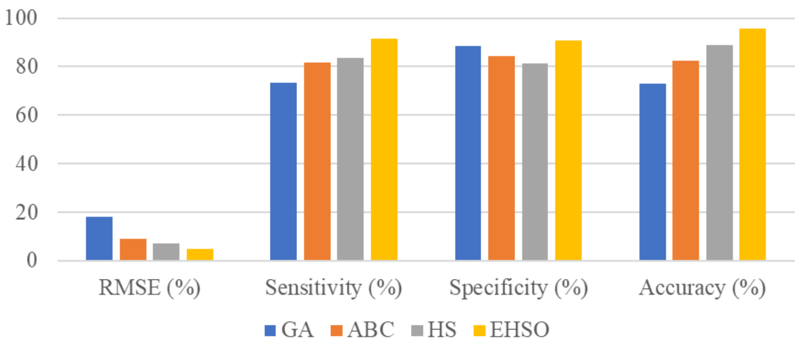

Figure 3 Performance of the various optimization methods for the classification of OC for training dataset

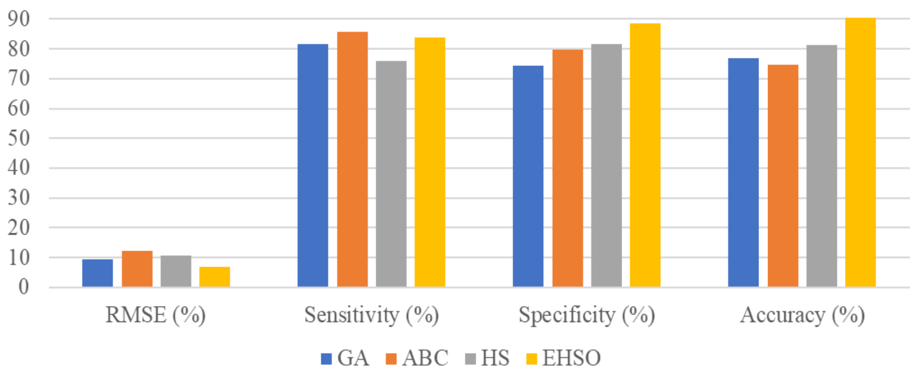

Figure 4 Performance of the various optimization methods for the classification of OC for testing dataset

including improvements in data collection devices, network infrastructure, and cybersecurity measures. By leveraging these components, the IoMT network aims to create a more efficient and accessible healthcare system, facilitating timely medical interventions and improving patient outcomes.

2.1 Self-Organizing Maps

The study used SOM to classify and group datasets with large dimensions. The concept of perception, as managed by SOM procedures, is used to convert data estimates into relationships within a flexible and unique representative sample. The SOM was assessed in FS using several adjustable variable choices based on a schedule derived from well-known examples and sensitivity testing. Specific attributes were chosen. SOM's method involves the regulation of the training process. The approach utilizes class association in testing instead of information in the neuron's category to modify the neighborhood variable. The primary concern in FS is to include the characteristics that provide the most optimal classification of OC. Therefore, the likelihood metric involves selecting the optimal number of markers that exhibit the most significant mapping from the SOM method. The trial opinions were primarily chosen based on the FS.

The concept of topological order protection entails the need to designate a related block or neighboring entities as examination points for the attribute space of the neighbors. This may be achieved by displaying the crucial region below. The k-clustering feature is represented as $F_{(P,S)}$ and is defined by Equation (1).

$$F_{(P,S)} = \sum_{i=0}^{I} \sum_{j=0}^{J} S_{ij} * \{h_j * d^2(x_i, s_i)\} \qquad (1)$$

If d and s are the path and the path section groups, h_j is the portion of the area among j. The input information gathered from the patient is represented as x_i. The chosen vector s is stated as follows:

$$s = \arg min_i\{|h_i - x_i|\} \qquad (2)$$

h_j is the area among j. The input data from the patient is given as x_i. At this juncture, the source variables of the learning phase examined the differentiation between each matrix using the distance calculated by Euclidean geometry.

3.2 IRNN using Extended Harmony Search Optimization (EHSO) Algorithm

Each artist or musician can be described as a decision factor. The preferences of the audience are indistinguishable from the objective function. The range of pitches in a musical instrument corresponds to the range of values in decision factors. Artists' spontaneous actions are analogous to regional and global search strategies. Musical harmony can sometimes be contrasted to an address vector with specific emphasis. An adaptive function was used to reduce the error rate of the RNN. The research specifically focused on using the crossover rate to enhance the performance of the EHSO technique.

The input function may be seen as the collection of inputs, the layer order, and the matrix value of the activation function. Next, the optimization variables need to be set. Harmonic Storage Size (HSS), Harmony Memory (HM), HM Contemplation Rate (HMCR), and Pitch Adaptation Rate (PAR). Next, the improved attributes need to be prepared.

Figure 2 shows the Extended Harmony Search Optimization (EHSO) Algorithm in IRNN for effective FS and classifying OC. Although initialization certainly impacts the success of the technique, it mostly serves to delay the acceleration of gradients during optimization. Every search localization technique is inherently associated with a concept of geometry inside the search space. In the last decade, RNN was developed to address linear and nonlinear restricted optimization problems. Compared to traditional numerical optimization techniques, the IRNN approach offers many advantages over time compared to EHSO. The flowchart of EHSO is seen in Figure 2.

3. Results and discussion

The validity of the proposed model's experimental results is supported by utilizing the ovarian cancer (OC) dataset from the Kaggle repository (Shahane, 2024). This dataset comprises 179 images classified as benign OC and 172 images classified as malignant OC, along with a total of 50 attributes.

Figure 3 illustrates the performance of various optimization methods for classifying OC within the training dataset. The proposed Improved Recurrent Neural Network (IRNN) with Enhanced Harmony Search Optimization (EHSO) achieved the highest performance, with an accuracy of 95.72% and the lowest Root Mean Square Error (RMSE) of 4.8%. In terms of specificity and sensitivity, the IRNN with EHSO outperformed other methods, including Harmony Search (HS) and Artificial Bee Colony (ABC), while Genetic Algorithm (GA) showed the lowest performance across all metrics.

Similarly, Figure 4 shows the performance of these optimization methods on the testing dataset. The IRNN with EHSO again delivered the highest accuracy of 90.4% and the lowest RMSE of 6.8%. It also exhibited superior specificity and sensitivity compared to HS and ABC. The GA method, once again, demonstrated the lowest performance in terms of accuracy, sensitivity, and specificity for the testing dataset.

4. Conclusion

The research demonstrated that integrating Self-Organizing Maps (SOM) and Improved Recurrent Neural Networks (IRNN) with the Extended Harmony Search Optimization (EHSO) algorithm significantly enhances ovarian cancer detection. By utilizing data from the Internet of Medical Things (IoMT), the proposed model achieved an accuracy of 95.72% and a low Root Mean Square Error (RMSE) of 4.8% on training datasets. Testing results confirmed its robustness, with an accuracy of 90.4% and an RMSE of 6.8%. This method overcomes challenges related to feature selection and classification label scarcity, providing a more effective tool for early diagnosis and improving patient outcomes in ovarian cancer care.

Author contributions

PV and PMP contributed to conceptualization, fieldwork, data analysis, drafting the original manuscript, editing, funding acquisition, and manuscript review. Both PV and PMP were also involved in research design, methodology validation, data analysis, visualization, and manuscript review and editing. Additionally, PV took lead in methodology validation, investigation, funding acquisition, supervision, and final revisions. All authors have reviewed and approved the final version of the manuscript.

Acknowledgment

The authors were grateful to their department.

Competing financial interests

The authors have no conflict of interest.

References

Arfiani, A., & Rustam, Z. (2019, November). Ovarian cancer data classification using bagging and random forest. In AIP Conference Proceedings (Vol. 2168, No. 1). AIP Publishing.

Bae, J. H., Kim, M., Lim, J. S., & Geem, Z. W. (2021). Feature selection for colon cancer detection using k-means clustering and modified harmony search algorithm. Mathematics, 9(5), 570.

Chatterjee, P., Siddiqui, S., Granata, G., Dey, P., & Abdul Kareem, R. S. (2024). Performance analysis of five U-Nets on cervical cancer datasets. Indian Journal of Information Sources and Services, 14(1), 17–28.

Chen, S., Chen, Y., Yu, L., & Hu, X. (2021). Overexpression of SOCS4 inhibits proliferation and migration of cervical cancer cells by regulating JAK1/STAT3 signaling pathway. European Journal of Gynaecological Oncology, 42(3), 554-560.

Consiglio, A., Casalino, G., Castellano, G., Grillo, G., Perlino, E., Vessio, G., & Licciulli, F. (2021). Explaining ovarian cancer gene expression profiles with fuzzy rules and genetic algorithms. Electronics, 10(4), 375.

Cree, I. A., White, V. A., Indave, B. I., & Lokuhetty, D. (2020). Revising the WHO classification: female genital tract tumours. Histopathology, 76(1), 151-156.

Elhoseny, M., Bian, G. B., Lakshmanaprabu, S. K., Shankar, K., Singh, A. K., & Wu, W. (2019). Effective features to classify ovarian cancer data in internet of medical things. Computer Networks, 159, 147-156.

Engqvist, H., Parris, T. Z., Biermann, J., Rönnerman, E. W., Larsson, P., Sundfeldt, K., & Helou, K. (2020). Integrative genomics approach identifies molecular features associated with early-stage ovarian carcinoma histotypes. Scientific Reports, 10(1), 7946.

Escorcia-Gutierrez, J., Torrents-Barrena, J., Gamarra, M., Madera, N., Romero-Aroca, P., Valls, A., & Puig, D. (2022). A feature selection strategy to optimize retinal vasculature segmentation. Computers, Materials and Continua, 70(2), 1-15.

Ganguly, S. (2020). Multi-objective distributed generation penetration planning with load model using particle swarm optimization. Decision Making: Applications in Management and Engineering, 3(1), 30-42.

Hinchcliff, E., Westin, S. N., & Herzog, T. J. (2022). State of the science: Contemporary front-line treatment of advanced ovarian cancer. Gynecologic Oncology, 166(1), 18-24.

Paik, E. S., Lee, J. W., Park, J. Y., Kim, J. H., Kim, M., Kim, T. J., ... & Seo, S. W. (2019). Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods. Journal of Gynecologic Oncology, 30(4).

Prat, J., D'Angelo, E., & Espinosa, I. (2018). Ovarian carcinomas: at least five different diseases with distinct histological features and molecular genetics. Human Pathology, 80, 11-27.

Ramakrishnan, J., Ravi Sankar, G., & Thavamani, K. (2019). Publication growth and research in India on lung cancer literature: A bibliometric study. Indian Journal of Information Sources and Services, 9(S1), 44-47.

Ramana, R. H. V., & Ravisankar, V. (2024). Precision in prostate cancer diagnosis: A comprehensive study on neural networks. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15(2), 109-122. https://doi.org/10.58346/JOWUA.2024.I2.008

Shahane, S. (2024). Predict ovarian cancer [Dataset]. Kaggle. https://www.kaggle.com/saurabhshahane/predict-ovarian-cancer

Su, Y. N., Wang, M. J., Yang, J. P., Wu, X. L., Xia, M., Bao, M. H., & Fu, L. J. (2023). Effects of Yulin Tong Bu formula on modulating gut microbiota and fecal metabolite interactions in mice with polycystic ovary syndrome. Frontiers in Endocrinology, 14, 1122709.

Takahashi, A., Hong, L., & Chefetz, I. (2020). How to win the ovarian cancer stem cell battle: Destroying the roots. Cancer Drug Resistance, 3(4), 1021.

Yang, J. P., Ullah, A., Su, Y. N., Otoo, A., Adu-Gyamfi, E. A., Feng, Q., & Ding, Y. B. (2023). Glycyrrhizin ameliorates impaired glucose metabolism and ovarian dysfunction in a polycystic ovary syndrome mouse model. Biology of Reproduction, 109(1), 83-96.