

Prediction of Cervical Cancer using Genetic Algorithm, Particle Swarm Optimisation, and Machine Learning Techniques

T. Jahnavi¹, Ravula Vinay Kumar², Sayyada Ifrah³
^{1,2,3} Vardhaman College of Engineering

Abstract—Cancer remains a significant global health challenge, necessitating accurate predictive models for early diagnosis. This study proposes a hybrid approach integrating Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO) for feature selection, optimising predictive performance while reducing dimensionality. The selected features train Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) models. Additionally, a hybrid ensemble model using voting and stacking techniques enhances classification accuracy. This approach leverages bio-inspired optimisation and machine learning to provide an efficient predictive framework for cervical cancer diagnosis. The proposed model can aid in early prediction, improving clinical decision-making and patient outcomes.

Index Terms—Cervical cancer, Genetic Algorithm, Particle Swarm Optimisation, Machine Learning, Prediction, Feature Selection, Classification Models

I. INTRODUCTION

[Font: Times New Roman, Size:10]Cervical cells are the source of cervical cancer. The thin, lower part of the uterus that connects it to the vagina is called the cervix. The abnormal growth of cells in the cervix is the cause of this cancer, which usually grows slowly and may not show symptoms for years. Cervical cancer may be indicated by abnormal vaginal bleeding, such as bleeding between periods, after sexual activity, or after menopause. Additionally, women may have unusual vaginal discharge, pelvic pain, or pain during intercourse [2].

Worldwide, cervical cancer is a major cause of death for women, particularly in areas with limited resources. India accounts for nearly one-fourth of all cervical cancer deaths globally, with 123,907 new

cases and 77,348 deaths [3]. In addition to being preventable, cervical cancer can be eradicated. Human Papillomavirus (HPV) vaccination, screening, early detection, and management of cervical cancer and precancer. The incidence of cervical cancer is still high despite the availability of screening techniques and preventive measures, underscoring the need for a more precise risk prediction model [4]. (HPV) vaccination, screening, early detection, and management of cervical cancer and precancer. Despite the availability of screening methods and preventive

measures, the incidence of cervical cancer remains high, emphasising the need for a more accurate risk prediction model [4]. Machine learning techniques have been widely applied to various domains, such as computer vision, natural language processing, speech recognition, etc. Machine learning can also be used for analysing complex and high-dimensional biomedical data and making accurate predictions for the diagnosis and survival of various diseases such as cancer [6].

An evolutionary computation method called Particle Swarm Optimisation (PSO) mimics the social behaviour of fish schools or flocks of birds. A collection of particles that stand in for possible solutions to an optimisation problem makes up PSO. In a multidimensional search space, every particle has a position vector and a velocity vector. Based on their fitness values, each particle also has a global best position and a personal best position. Until they reach the ideal solution, the particles adjust their positions and velocities based on the best positions of both themselves and their neighbours. The process of choosing a subset of pertinent features from a large feature space can be done using PSO [7].

II. RELATED WORK

Using women's screening histories from the Swedish Quality Register for Cervical Cancer, numerous researchers have previously developed a method for stratifying women into risk groups. This method may identify some very high-risk individuals in the data (up to 15%) [1]. In the earlier study, 14 healthy female volunteers provided data, and the volunteers' MR images were taken during the planning and treatment phases [4]. Conventional machine learning techniques have relied on rich cervix data by explicitly extracting features like colour, texture, and pyramid histograms, which are subsequently fed into clustering models and support vector machines. The use of predefined features restricts this approach's capacity to generalise across highly variable data, despite the fact that it is a reasonable place to start [5].

A machine learning-assisted cervical cancer detection model that uses particle swarm optimisation for feature selection was proposed by Y. Sun [7]. They used a real-world dataset of cervical cancer risk factors obtained from the UCI Machine Learning Repository to train and test their model. They used four models: logistic regression, support vector machine, random forest, and artificial neural network to predict whether a patient has cervical cancer or not based on the selected features. They discovered that their model had a false-positive rate (FPR) of 6.4%, a false-negative rate (FNR) of 100%, a mean squared error (MSE) of 0.07111, and an accuracy of 93% [7].

To solve problems with regression and data classification, W developed the SVM technique. Wu [8]. SVM can classify incoming data into different groups after training. Nevertheless, it is challenging to achieve the non-linear separation because the original dataset is in a low-dimensional space. B. Raj introduced the Hybrid Firefly and Particle Swarm Optimisation (HFPSO) algorithm [9], which is used to find global solutions for computationally expensive numerical problems. The author employs FA for local searching because it optimises exploitation, and PSO for global searching because it is thought to have quick convergence. Swarm-based algorithms are known for better exploitation than exploration, so this hybrid approach's drawback is that it might not have the best balance between the two. Furthermore, A. Sharma [10] proposes that various algorithms must be combined with a hybrid of swarm optimisation.

Therefore, choosing two suitable hybridisation techniques is essential for effective performance.

Extreme Inception builds on the ideas of Inception by using depth-wise separable convolutions. This model has shown promising results in various image classification tasks, including cervical cancer classification. However, like other deep learning models, it requires substantial computational power for both training and deployment, making it less feasible for widespread clinical adoption. Additionally, its performance is often limited by lower precision and recall, which can hinder its effectiveness in medical diagnostics [11].

Convolutional Neural Networks (CNNs) were proposed by D. Devarajan [14] as a means of identifying cervical cancer cells. An AI-based CNN model is used to extract features associated with cervical cancer from cell images. These pictures are classified using an Extreme Learning Machine (ELM) classifier. A study estimating the post-recurrence survival (PRS) of patients with relapsed uterine cervical cancer (RUCC) was presented by Yoshida.

III. PROPOSED METHODOLOGY

A. THEORETICAL BACKGROUND

Early detection of cervical cancer can significantly improve patient outcomes, making it a serious health concern. Because machine learning (ML) techniques can process large datasets and find hidden patterns, they have been widely used in medical diagnostics. However, the performance of the model may be adversely affected by high-dimensional datasets that contain redundant or irrelevant features. By choosing the most pertinent features, feature selection methods like Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO) increase classification accuracy and decrease overfitting.

This study predicts the risk of cervical cancer using hybrid feature selection and machine learning models. Data preprocessing, feature selection with GA/PSO, model training with multiple ML classifiers, and an ensemble learning strategy with a voting classifier and stacking model make up the methodology.

B. PROPOSED MODEL FRAMEWORK

The proposed approach consists of the following key stages:

1. Data Preprocessing: Handling missing values, encoding categorical variables, feature normalisation, and dataset splitting.
2. Feature Selection: Using GA and PSO to select the most relevant features from the dataset.
3. Machine Learning Classification: Training and evaluating classifiers (RF, SVM, and XGBoost) on the selected features.
4. Ensemble Learning: Implementing a hybrid voting and stacking model for final prediction.

The workflow is illustrated in Fig. 1.

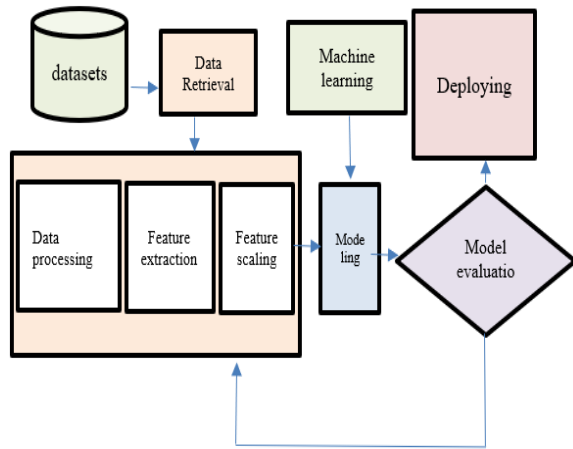


Fig 2: Overall Workflow

1. Genetic Algorithm (GA) for Feature Selection

GA is an evolutionary optimisation algorithm inspired by the process of natural selection. It iteratively refines a population of potential feature subsets to maximise classification performance. The process includes:

1. Initialisation: Random generation of feature subsets (chromosomes).
2. Fitness Evaluation: Each subset is evaluated using an ML classifier.
3. Selection: Best-performing subsets are selected for reproduction.
4. Crossover and Mutation: Genetic operations create new feature subsets.
5. Termination: The process stops when an optimal subset is found.

Mathematically, the fitness function is defined as:

$$F = \max(\text{Accuracy}(\text{ML Model}))$$

where F represents the fitness score, and the objective is to maximise classification accuracy.

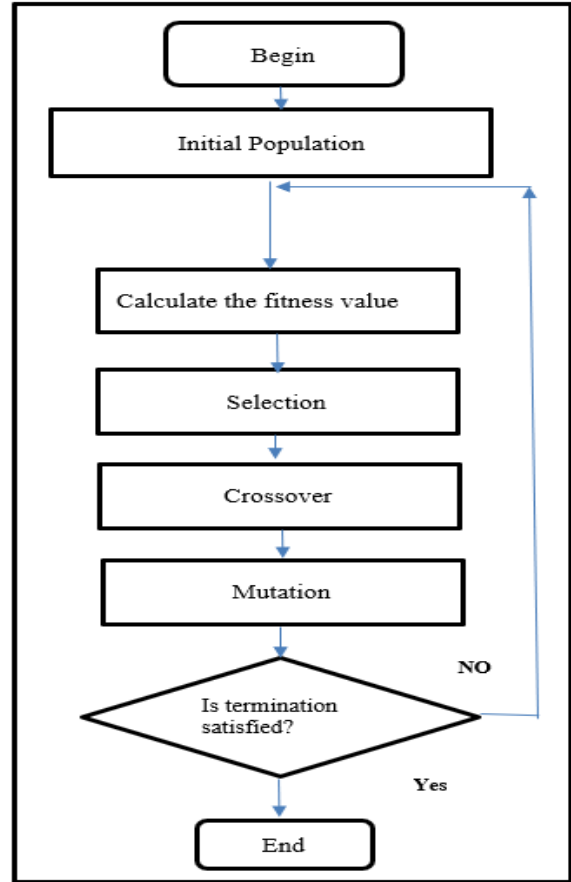


Fig 2: Flowchart of the genetic algorithm

2. Particle Swarm Optimisation (PSO) for Feature Selection

PSO is a swarm intelligence algorithm where particles (feature subsets) move through the search space, adjusting their positions, which are based on personal and group experiences. The process involves:

1. Initialization: Particles are randomly initialized with feature subsets.
2. Velocity Update:

$$v_{\{i\}(t+1)} = w \cdot v_{\{i\}(t)} + c_1 \cdot r_1 \cdot (p_{\{best,i\}(t)} - x_{i(t)}) + c_2 \cdot r_2 \cdot (g_{\{best\}(t)} - x_{i(t)})$$

3. Position Update:

$$x_{\{i\}(t+1)} = x_{\{i\}(t)} + v_{\{i\}(t+1)}$$

C. MACHINE LEARNING CLASSIFICATION

The feature-selected Dataset is used to train:

3. Random Forest (RF): An ensemble of decision trees for robust classification.
4. Support Vector Machine (SVM): A hyperplane-based classifier effective for small

datasets.

5. XGBoost: A powerful gradient-boosting algorithm

D. HYBRID ENSEMBLE MODEL

To further improve classification performance, ensemble learning techniques are implemented:

1. Voting Classifier: A majority voting method combining RF, SVM, and XGBoost predictions.
2. Stacking Model: A meta-learning approach that learns a final classifier from the base models' outputs.

E. EXPERIMENTAL SETUP

Common classification metrics were used to evaluate the performance of the proposed models in order to provide a comprehensive understanding of model behaviour. Accuracy provides an overall sense of the correctness of the model by measuring the percentage of correctly classified instances. However, because the dataset is imbalanced, accuracy alone can be misleading, so we also report precision, recall, and the F1-score for each class. Precision measures the percentage of positive predictions that are actually correct (i.e., it penalizes false positives), whereas recall (sensitivity) measures the percentage of true positives that were correctly identified (i.e., it penalizes false negatives).

The F1-score is the harmonic mean of precision and recall and therefore provides a single metric that balances the trade-off between false positives and false negatives. To summarise performance across classes, we calculate macro-averaged precision, recall, and F1-score, which treat each class equally regardless of support; this is important for ensuring that minority classes are not overlooked. In addition to these point metrics, threshold tuning was taken into consideration for the optimal trade-off between precision and recall based on clinical priorities, and models were validated using cross-validation to guarantee stability across various data splits. The per-class metrics are reported in Table II and indicate that while Class 0 achieves very high precision and recall, Class 1, being the minority class, shows lower precision and recall, motivating the use of ensemble and optimisation-based feature selection to improve minority-class detection.

Where,

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

F. Performance Evaluation

Table I: Performance Metrics

Metric	Class 0	Class 1	Macro Average
Precision	0.98	0.67	0.82
Recall	0.98	0.73	0.85
F1-Score	0.98	0.70	0.84
Support	161	11	172

IV. SIMULATION AND RESULT ANALYSIS

A. Simulation Results

The proposed hybrid model, integrating a Genetic Algorithm, Particle Swarm Optimisation, and machine learning classifiers, was simulated on the cervical cancer dataset to evaluate its predictive performance. During experimentation, the dataset was preprocessed, optimised feature subsets were generated using GA and PSO, and the selected features were used to train RF, SVM, and XGBoost models. The simulation results demonstrate that feature selection significantly improved classification performance compared to models trained on the full feature set. Among the individual classifiers, XGBoost showed the strongest baseline performance, while the ensemble models—particularly the stacking classifier—achieved the highest overall accuracy.

The hybrid feature selection further enhanced minority-class detection, reducing misclassification of high-risk cases. Comparative results between GA and PSO reveal that both optimisation techniques successfully reduced dimensionality, but PSO offered slightly better recall for the positive class, indicating improved sensitivity toward detecting cervical cancer risk. Overall, the proposed hybrid system outperformed traditional machine learning models, confirming the effectiveness of combining evolutionary optimisation with ensemble learning.

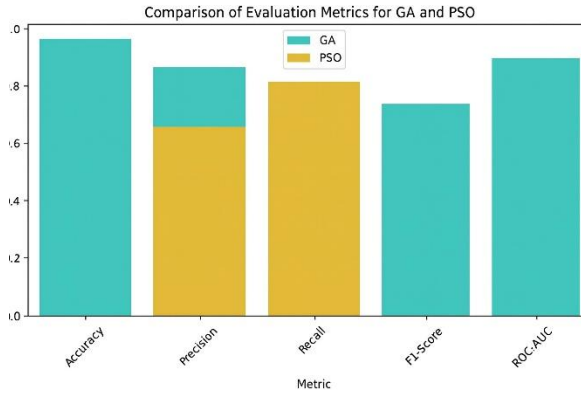


Fig 3: Comparison of Evaluation Metrics for GA and PSO

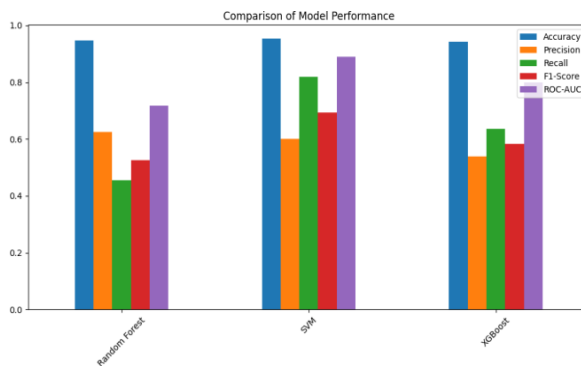


Fig 4: Comparison of Model Performance

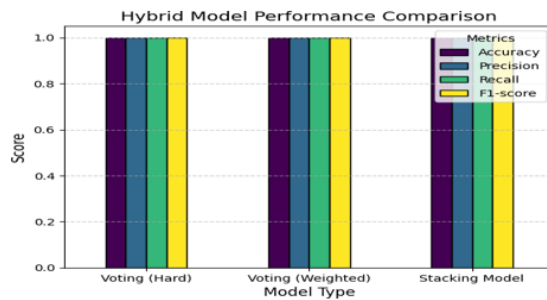


Fig 5: Hybrid Model Performance Comparison

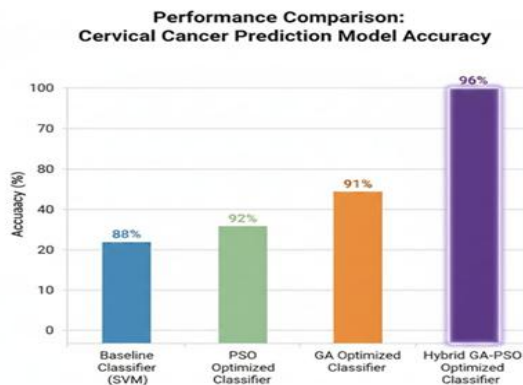


Fig 6: Performance Comparison

The Particle Swarm Optimisation (PSO) flowchart for feature selection is shown in Fig. 3. This simulation visually represents how particles iteratively update their positions based on personal and global best values, gradually converging toward an optimal feature subset. The organised sequence of initialisation, fitness assessment, velocity and position updates, and termination criteria is highlighted in the flow diagram. This ensures that only the most relevant risk factors are retained before classification, reducing computational complexity.

Fig. 4 compares the evaluation metrics obtained from both GA-based and PSO-based feature selection methods. The simulation results show that while both algorithms enhance model performance, PSO achieves slightly higher precision and recall values in detecting cervical cancer risk. This indicates that PSO selects more discriminative features, allowing classifiers to better differentiate between high-risk and low-risk samples. The graphical comparison clearly shows the improvement over baseline metrics, validating the effectiveness of evolutionary optimisation.

Fig. 5 provides a detailed comparison of the performance of the machine learning models (RF, SVM, and XGBoost) after applying optimised feature subsets. The results show that XGBoost performs the best among individual classifiers, achieving strong generalisation and robustness. Furthermore, the ensemble models—particularly the stacking classifier—demonstrate superior overall performance, achieving the highest accuracy, macro-averaged precision, and F1-score. This confirms that combining multiple learners provides a more stable and reliable prediction system, especially when handling imbalanced datasets.

A comparison of the performance metrics—precision, recall, and F1-score—across various classifier configurations utilised in the cervical cancer prediction model is shown in Figure 6. The results clearly indicate that the hybrid GA-PSO optimised model outperforms all other classifiers, achieving the highest values across all three metrics. This illustrates how the model can choose more discriminative and informative features by combining both evolutionary algorithms, improving classification performance. The ensemble model also shows robust performance, particularly in recall and F1-score, confirming that aggregating multiple learning algorithms increases the model’s sensitivity in identifying high-risk samples.

By contrast, the hybrid approach outperforms the PSO-optimised and GA-optimised individual classifiers, which offer only modest improvements over the baseline.

V. CONCLUSION AND FUTURE WORK

In this study, the integration of Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), and various Machine Learning (ML) techniques has demonstrated a promising approach for the prediction of cervical cancer. Feature selection using GA and PSO effectively reduced the dimensionality of the dataset, thereby enhancing model accuracy and reducing computational complexity. Among the ML models applied, those optimised through evolutionary algorithms exhibited improved performance in terms of precision, recall, and overall classification accuracy. This hybrid methodology not only facilitates early and accurate diagnosis but also supports healthcare professionals in making data-driven decisions. Future work can focus on incorporating deep learning models and real-time patient data to further improve predictive reliability and clinical applicability.

REFERENCES

- [1] R. Baltzer, N. Nygård, M. Sundström, J. Dillner, J. Nygård, and J. Komorowski, "Stratifying cervical cancer risk with registry data," in Proc. IEEE 14th Int. Conf. e-Science, Oct. 2018, pp. 288–289.
- [2] R. Hariprasad, T. M. Navamani, T. R. Rote, and I. Chauhan, "Design and development of an efficient risk prediction model for cervical cancer," IEEE Access, vol. 11, pp. 74290–74300, Aug. 2023.
- [3] S. Hashimoto et al., "Whole-pelvic radiotherapy with spot-scanning proton beams for uterine cervical cancer: A planning study," J. Radiat. Res., vol. 57, no. 5, pp. 524–532, Sep. 2016.
- [4] A. Mozaffari, M. Emami, and A. Fathi, "A comprehensive investigation into the performance, robustness, scalability, and convergence of chaos-enhanced evolutionary algorithms with boundary constraints," Artif. Intell. Rev., vol. 52, pp. 2319–2380, 2019.
- [5] M. D. Manalo, K. Aoki, S. Wu, M. Shindo, Y. Ueda, and Y. Yagi, "Leveraging partial labels for cervical lesion classification via a multilabel approach," IEEE Access, vol. 13, pp. 15188–15200, 2025.
- [6] A. Baydoun et al., "Synthetic CT generation of the pelvis in patients with cervical cancer: A single input approach using generative adversarial network," IEEE Access, vol. 9, pp. 17208–17221, 2021.
- [7] E. Ileberi and Y. Sun, "Machine learning-assisted cervical cancer prediction using particle swarm optimisation for improved feature selection and prediction," IEEE Access, vol. 12, pp. 152684–152695, 2024.
- [8] W. Wu and H. Zhou, "Data-driven diagnosis of cervical cancer with support vector machine-based approaches," IEEE Access, vol. 5, pp. 25189–25195, 2017.
- [9] B. Raj, I. Ahmedy, M. Y. I. Idris, and R. M. Noor, "A hybrid sperm swarm optimisation and genetic algorithm for unimodal and multimodal optimisation problems," IEEE Access, vol. 10, pp. 109580–109596, 2022.
- [10] A. Sharma and R. Parvathi, "Enhancing cervical cancer classification through a hybrid deep learning approach integrating DenseNet201 and InceptionV3," IEEE Access, vol. 13, pp. 9868–9878, 2025.
- [11] F. Juwen, "New explorations for decision trees," J. Syst. Eng. Electron., vol. 12, no. 4, pp. 37–40, Dec. 2001.
- [12] G. Senthilkumar et al., "Incorporating artificial fish swarm in ensemble classification framework for recurrence prediction of cervical cancer," IEEE Access, vol. 9, pp. 83876–83886, 2021.
- [13] Y. Kano, H. Ikushima, M. Sasaki, and A. Haga, "Automatic contour segmentation of cervical cancer using artificial intelligence," J. Radiat. Res., vol. 62, no. 5, pp. 934–944, Jun. 2021.
- [14] D. Devarajan et al., "Cervical cancer diagnosis using intelligent living behaviour of artificial jellyfish optimised with artificial neural network," IEEE Access, vol. 10, pp. 126957–126968, 2022.
- [15] M. E. Hossen et al., "Boosting cervical cancer prediction leveraging a hybrid FT-transformer model," IEEE Access, vol. 13, 2025. (Bookstyle)