Optimization of Parameter Tuning in the SVM Algorithm Using Metaheuristic Optimization Algorithms for Detecting Stunting Risk in Pregnant Women Based on Urine Test Data

Yudha Wibowo, Agung Mulyo Widodo, Gerry Firmansyah and Budi Tjahjono

February 20, 2025

Optimization of Parameter Tuning in the SVM Algorithm Using Metaheuristic Optimization Algorithms for Detecting Stunting Risk in Pregnant Women Based on Urine Test Data

Yudha Wibowo [1], Agung Mulyo Widodo [2*], Gerry Firmansyah [3], and Budi Tjahjono [4]

[1]University Esa Unggul
[2]University Esa Unggul
[3]University Esa Unggul
[4]University Esa Unggul

yudha.wibowo777@student.esaunggul.ac.id, agung.mulyo@esaunggul.ac.id, gerry@esaunggul.ac.id, budi.tjahyono@esaunggul.ac.id

Abstract

Stunting is a serious health issue affecting children's growth from pregnancy. Early detection of stunting risk in pregnant women is crucial to prevent long-term impacts. This study develops a stunting risk prediction model based on Support Vector Machine (SVM) with parameter optimization using metaheuristic optimization algorithms. The data used is derived from urine test results of pregnant women, encompassing various clinical parameters. The optimization algorithms employed include Grey Wolf Optimizer (GWO), Simulated Annealing (SA), and Firefly Algorithm (FA) to find the optimal C and gamma parameters for SVM. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results show that optimization with GWO increased the model accuracy to 94.15%, compared to the default model, which only achieved 88.46%. SA optimization also improved accuracy to 94.12%, while FA reached 85.71%. These findings indicate that using metaheuristic optimization in SVM parameter tuning can significantly enhance stunting risk prediction performance.

Keywords: Support Vector Machine (SVM), Grey Wolf Optimizer (GWO), Simulated Annealing (SA), Firefly Algorithm (FA), Parameter Tuning, Stunting.

1 Introduction

Stunting is a global public health issue, particularly prevalent in developing nations such as Indonesia. It is characterized by impaired growth due to chronic malnutrition beginning in pregnancy (Black et al., 2013). Early identification of stunting risk factors in pregnant women is crucial to ensure timely intervention. Urine analysis provides valuable clinical insights into maternal and fetal health conditions (Simanjuntak, 2023).

Despite advances in medical diagnostics, traditional methods of stunting detection remain time-consuming and inconsistent across healthcare facilities. Current machine learning models often suffer from suboptimal parameter tuning, reducing their predictive reliability (Syarif et al., 2016). Therefore, our study aims to enhance stunting risk prediction by using metaheuristic optimization to fine-tune SVM hyperparameters, ensuring improved accuracy and efficiency.

The Support Vector Machine (SVM) is widely used in medical classification due to its effectiveness in handling high-dimensional data (Vapnik, 1995). However, SVM performance is highly dependent on parameter tuning, particularly the C and gamma parameters that control the complexity and decision boundaries of the model (Zhang et al., 2023).

Several studies have demonstrated that metaheuristic optimization methods such as the Grey Wolf Optimizer (GWO), Simulated Annealing (SA), and Firefly Algorithm (FA) improve SVM performance (Singh & Agarwal, 2021). This research aims to optimize SVM parameter tuning using these algorithms to enhance stunting risk prediction based on urine test results.

2 Research Methodology

2.1 Dataset and Preprocessing

The dataset used in this study is derived from urine test results of pregnant women, incorporating clinical parameters such as :

1. The patient's age in years.
2. Color of the urine (e.g., Light Yellow, Yellow, Amber, Dark Yellow, etc.).
3. Transparancy is The clarity level of the urine (e.g., Clear, Cloudy, etc.).
4. The glucose level in the urine (Negative, Positive, Trace).
5. The protein level in the urine (Negative, Positive, Trace).
6. pH is The acidity or alkalinity of the urine.
7. Specific Gravity is The urine's specific gravity.
8. WBC is The number of white blood cells in the urine (e.g., 0-2, 2-4, 4-6).
9. RBC is The number of red blood cells in the urine (e.g., 0-2, 2-4, 4-6).
10. Epithelial Cells is The presence of epithelial cells (Rare, Few, Moderate, Many).
11. Mucous Threads is The presence of mucous threads (None Seen, Present).

12. Amorphous Urates is The presence of amorphous urates (None Seen, Few, Present).
13. Bacteria is The presence of bacteria in the urine (None Seen, Rare, Present).
14. Diagnosis is The diagnostic result (Negative = no health issues, Positive = presence of a health issue, such as an infection or kidney disorder).

Data preprocessing involves:
1. Handling missing values through imputation techniques.
2. Normalizing the dataset using Min-Max Scaling.
3. Encoding categorical variables such as urine color and transparency.
4. Splitting the dataset into 70% training data and 30% testing data.

2.2 Machine Learning Model: Support Vector Machine (SVM)

SVM, utilizing the Radial Basis Function (RBF) kernel, is implemented as the primary predictive model. The parameter **C** controls misclassification tolerance, while **gamma** influences the importance of data points in the feature space (Vapnik, 1995).

2.3 Metaheuristic Optimization Algorithms

While traditional grid search and Bayesian optimization methods are commonly used for SVM hyperparameter tuning, they often fail to navigate large search spaces effectively.
Three metaheuristic optimization algorithms are applied to optimize SVM parameters:
1. Grey Wolf Optimizer (GWO): Mimics the hierarchical leadership and hunting strategies of grey wolves to find optimal solutions (Zhang & Chen, 2021).
2. Simulated Annealing (SA): Utilizes probabilistic mechanisms for escaping local optima during optimization (Mahareek et al., 2021).
3. Firefly Algorithm (FA): Uses the attraction mechanism of fireflies, where brightness determines the suitability of solutions (Sharma et al., 2013).

Metaheuristic optimization methods, including GWO, SA, and FA, are selected for their ability to efficiently explore diverse parameter settings while avoiding local optima (Russell & Norvig, 2020).

2.4 Model Evaluation Metrics

The optimized models are assessed based on the following performance metrics :
1. Accuracy : The proportion of correctly classified instances.
2. Precision : The ability to avoid false positives
3. Recall : The capability to all positive cases
4. F1-score : Ther harmonic mean of precision and recall

3 Results and Discussion

This section presents the results of the experiments conducted to evaluate the effectiveness of Support Vector Machine (SVM) optimized with metaheuristic algorithms in detecting stunting risk in pregnant women. The evaluation metrics used include accuracy, precision, recall, and F1-score. The results obtained from different optimization techniques are analyzed to determine their impact on model performance.

3.1 Performance Comparison of Models

Table 3.1 summarizes the performance comparison of different optimization algorithms applied to SVM. The comparison is based on the key metrics: accuracy, precision, recall, and F1-score. The results indicate that GWO outperforms other optimization methods, followed by SA and FA.

Method	Accuracy (%)	Precision (%)	Recall (%)
Default SVM	88.46	85.71	82.14
GWO-SVM	94.15	93.22	91.79
SA-SVM	94.12	100.00	75.00
FA-SVM	85.71	88.12	83.00

Table 1: Performance Metrics of Optimization Methods

3.2 Analysis of Results

1. GWO achieved the highest recall (91.79%), making it the most effective for stunting risk detection. The Grey Wolf Optimizer (GWO) demonstrated superior performance in terms of recall. Recall is a critical metric in medical diagnostics, as it indicates the model's ability to correctly identify positive cases. A high recall means fewer false negatives, ensuring that high-risk individuals are detected effectively. The adaptive search mechanism of GWO enables it to explore a diverse range of hyperparameter values, improving the decision boundary of the SVM model.

2. SA achieved the highest precision (100%), reducing false positives but with a lower recall. Simulated Annealing (SA) exhibited the highest precision, which implies that it minimized false positives. This is particularly important in healthcare applications where false alarms can lead to unnecessary medical interventions. However, its recall was lower compared to GWO, meaning that while it effectively identified true positives, it also missed some high-risk cases. The lower recall suggests that SA might overfit to specific patterns in the training data, limiting its generalizability.

3. FA maintained a balanced precision and recall, though its overall accuracy was lower than GWO and SA. The Firefly Algorithm (FA) achieved a moderate balance between precision and recall. Although its accuracy was lower than GWO and SA, it still provided stable performance across different evaluation metrics. FA relies on the attractiveness and light intensity mechanism, which enables it to avoid premature convergence. However, its slightly lower accuracy indicates that it might not explore the optimal hyperparameters as efficiently as GWO and SA.

4 Conclusion

4.1 Key Findings

This study demonstrates that metaheuristic optimization significantly improves the performance of SVM in detecting stunting risk. The following key findings were observed :

1. Metaheuristic optimization significantly enhances SVM performance in detecting stunting risk. Compared to the default SVM model, optimized models exhibit improved classification performance.

2. GWO achieved the best accuracy (94.15%) and recall (91.79%), making it the most effective overall. The ability of GWO to explore a wide search space and avoid local optima contributed to its superior classification performance.
3. SA was effective in maximizing precision, although its recall was lower. The high precision achieved by SA indicates its strength in minimizing false positives, making it suitable for applications where false alarms need to be avoided.
4. FA provided a balance between precision and recall, but with lower accuracy. FA's ability to maintain balanced performance across metrics suggests it is a viable alternative but not as effective as GWO.

4.2 Practical Implications

The findings of this study have practical applications in the healthcare sector. The integration of optimized SVM models into decision-support systems can assist medical professionals in early detection and intervention.

1. The proposed SVM metaheuristic optimization can be implemented in healthcare decision-support systems. By embedding these models into hospital management software, real-time patient screening for stunting risk can be automated, improving efficiency and decision-making.
2. Real-time stunting risk prediction models can be developed for hospitals and clinics. A real-time risk assessment tool can be developed, allowing doctors and medical practitioners to input patient data and receive instant risk analysis.

4.3 Recommendations for Future Research

While this study has demonstrated the effectiveness of metaheuristic-optimized SVM models for stunting risk detection, further research is needed to address some limitations and improve the model's applicability.

1. Expanding the dataset to improve generalization across diverse populations (Sulastri et al., 2021). Future research should incorporate larger datasets with diverse demographic distributions to enhance model robustness and reduce potential biases.
2. Integrating additional clinical factors such as dietary intake and maternal health history (Simanjuntak, 2023). Including more physiological and lifestyle factors can provide a more comprehensive risk assessment model.
3. Developing a web or mobile-based application for real-time stunting risk assessment in healthcare settings (Santoso & Laila, 2019). A mobile or web-based system could be developed to facilitate easier access for both medical practitioners and patients, increasing the accessibility of early screening.

References

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., ... & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet, 382*(9890), 427-451.

Mahareek, E. A., Desuky, A. S., & El-Zhni, H. A. (2021). Simulated annealing for SVM parameters optimization in student's performance prediction. *Bulletin of Electrical Engineering and Informatics, 10*(3), 1211–1219.

Polikar, R. (2012). Ensemble learning. *IEEE Signal Processing Magazine, 29*(5), 45–62.

Santoso, A., & Laila, L. (2019). Skrining kesehatan melalui pemeriksaan protein urine di kompleks Aditarina Kota Makassar. *Jurnal Medika, 3*(2), 227–233.

Scikit-learn. (n.d.). RBF SVM parameters. Retrieved from https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html.

Sharma, A., Zaidi, A., Singh, R., Jain, S., & Sahoo, A. (2013). Optimization of SVM classifier using firefly algorithm. *2013 IEEE 2nd International Conference on Image Information Processing (ICIIP-2013),* 198–202.

Simanjuntak, I. K. (2023). Analisis kadar protein dan glukosa urine pada ibu hamil di Puskesmas Teladan Medan tahun 2023. *Universitas Medan Area.*

Singh, M., & Agarwal, R. (2021). Particle swarm optimization for SVM in classifying medical data. *Elsevier.*

Sulastri, D., Sari, D. P., & Lestari, D. (2021). Mencegah stunting pada 1000 hari pertama kehidupan pada masa pandemi Covid-19. *Jurnal Ilmiah Hesti Wira Sakti, 9*(1), 78–83.

Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *TELKOMNIKA (Telecommunication Computing Electronics and Control), 14*(4), 1502–1509.

Vapnik, V. (1995). *The nature of statistical learning theory.* Springer.

Zhang, X., & Chen, Y. (2021). Grey wolf optimization for support vector machine in disease classification. *Elsevier.*

Zhang, X., Liu, Y., & Wang, T. (2023). Optimization of SVM parameters using genetic algorithm for medical diagnosis. *Springer.*

Zhao, F., Wang, J., & Li, X. (2022). Firefly algorithm for support vector machine in healthcare data. *IEEE.*

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint. https://arxiv.org/abs/1802.07228

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group.

Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.Voronkov, A. (2004). *EasyChair conference system*. Retrieved from easychair.org