SINTEZA 2025
INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND DATA SCIENCE

DATA SCIENCE AND APPLICATIONS SESSION

REVIEW OF METAHEURISTIC APPROACHES FOR BOOSTING MODEL OPTIMIZATION

Miloš Grubješić*,
[0009-0007-3346-8442]

Miodrag Živković,
[0000-0002-4351-068X]

Tamara Živković,
[0000-0003-2969-1709]

Nebojša Bačanin Džakula
[0000-0002-2062-924X]

Singidunum University,
Belgrade, Serbia

Abstract:

Boosting algorithms, recognized for their ability to create strong learners from weak learners, are valuable assets in many areas. Often it is difficult to optimize their hyperparameters, especially when working with high-dimensional datasets. Metaheuristic algorithms, based on natural and evolutionary processes, provide great alternatives to traditional optimization approaches to hyperparameter tuning for boosting models with efficient and robust solutions. This review paper will present an overview of the existing state-of-the-art metaheuristic algorithms utilized in boosting model optimization, highlighting a wide variety of metaheuristic algorithms.

Keywords:

Boosting Algorithms, Metaheuristic Algorithms, Hyperparameter Optimization.

INTRODUCTION

Metaheuristics are a class of stochastic algorithms generated by nature, notably the social behavior of animals. They are successfully deployed to solve NP-hard problems [1]. Metaheuristics provide sufficient outcomes in scenarios where deterministic methods cannot be adopted due to resource or time restrictions, large datasets, nonlinear and/or unstructured problems.

Boosting methods such as XGBoost, LightGBM, CatBoost, and others are capable ranked ML algorithms that are built sequentially upon weak learners [2]. In the end, they create a highly accurate model with defined error minimizations.

Synergizing metaheuristics with boosting methods can vastly improve the model's ability to optimize machine learning models where large complex search space will be searched efficiently, contributing to increased accuracy and robustness of these models.

This paper aims to present a detailed review of metaheuristic approaches merged with boosting methods that showcase their ability to optimize machine learning models regardless of discipline. This review will synthesize takeaways from the critical consideration of selected studies with a metaheuristics-boosting optimization consideration

Correspondence:
Miloš Grubješić

e-mail:
milos.grubjesic.24@singimail.rs

DOI: 10.15308/Sinteza-2025-200-205

and provide a review of the consideration method, problems, disciplines, and results realized to contribute to enhancing understanding and ways of addressing existing issues in research.

2. OVERVIEW OF BOOSTING AND METAHEURISTICS

2.1. BOOSTING ALGORITHMS

Boosting is an ensemble learning strategy that has become an important part of modern machine learning, combining multiple weak learners to create a strong learner with high predictive accuracy. It operates iteratively, with each successive model attempting to correct the mistakes of the models that preceded it, thereby improving the ensemble's overall predictive performance iteratively. This iterative process allows boosting algorithms to achieve higher accuracy than individual models. Boosting methods are particularly valuable in modeling datasets that exhibit large, complex, and nonlinear relationships.

2.2. POPULAR BOOSTING ALGORITHMS

Several boosting algorithms have gained popularity due to their effectiveness and versatility:

- Adaptive Boosting: AdaBoost is one of the most widely used boosting algorithms. It assigns weights to each data point in the training set based on the accuracy of prior models and then trains a new model using the updated weights.
- Gradient Boosting Machines: GBM is a generalization of AdaBoost. It optimiz.
- XGBoost: eXtreme Gradient Boosting is an optimized gradient boosting algorithm known for its efficiency and performance. It includes algorithmic enhancements that contribute to scalability and higher accuracy levels.
- LightGBM: Light Gradient Boosting Machine is a gradient boosting framework designed for high performance and efficiency.
- CatBoost: CatBoost is a gradient boosting algorithm that excels in handling categorical features. It uses a special technique to deal with categorical variables, often leading to better performance, especially when dealing with such data

2.3. HYPERPARAMETER TUNING

There are various types of hyperparameters that can be tuned to achieve optimal model performance. Some include:

- Structural parameters: These control the complexity of individual weak learners, such as the maximum depth of a tree.
- Learning rate parameters: These control how much each weak learner contributes to the final model, such as a learning rate.
- Regularization parameters: These penalize complex models to limit overfitting. Examples include gamma, alpha, and lambda.
- Ensemble parameters: These control the number of weak learners in the ensemble, such as the number of estimators.

2.4. METAHEURISTIC ALGORITHMS

Metaheuristic algorithms can be broadly defined as optimization schemes inspired by deviations in the natural or physical world. "Deviations," in this case, refer to natural inconsistencies or irregularities, deviations that occur in biological, physical, or other natural systems—the deviations from expected or uniform outcomes. Metaheuristic algorithms often use these deviations (such as, for example, random mutations in genetic algorithms or temperature fluctuations in simulated annealing) to diversify the search for solutions and avoid being trapped in sub-optimal areas of the solution space [3]. Their goal is to find a usable solution to a problem that is complex enough that traditional methods would not apply due to constraints of time or resources. Compared to traditional heuristics specific to problems, metaheuristic algorithms provide a higher level of abstraction and are easily applied for wider-reaching optimization problems.

Some notable metaheuristic algorithms include [4]:

- Genetic Algorithms
- Particle Swarm Optimization
- Ant Colony Optimization
- Simulated Annealing
- Tabu Search
- Firefly Algorithm
- Grey Wolf Optimizer, and many more...

These algorithms strive to balance excursion (diversification) and intensification to search for near-optimal solutions within the search space effectively. The general perspective of the excursion is to produce a diverse solution set to examine potential global search space. At the same time, intensification is generally concerned with exploiting the information around the local optimal solution space. Although there is no guarantee that a metaheuristic algorithm will yield an absolute optimal solution each time, a metaheuristic typically produces a near-optimal solution in a reasonable amount of time.

3. REVIEW OF RECENT LITERATURE

A particular emphasis was placed on analysing newer studies from various authors, ensuring the review captures the latest advancements and methodologies in the field.

In [1], the authors integrated XGBoost and AdaBoost with a modified particle swarm optimization algorithm to classify respiratory conditions using a dataset of 920 labelled recordings from 126 patients. This two-tier framework, combining CNN-based feature extraction with boosting and metaheuristics, achieved 98.14% accuracy for binary classification and 81.25% for multi-class classification.

In [5], the researchers introduced an approach that employs XGBoost together with a Genetically Inspired RSA (GIRSA) algorithm for a real-world IoT MQTT dataset comprising six classes of traffic. The method leverages CNNs for feature reduction and obtains an 87.94% multi-class classification accuracy, further enhanced by a modified reptile search algorithm for hyperparameter optimization.

In [6], the authors combined CatBoost and LightGBM with the Chimp optimization algorithm on the CICIoT2023 intrusion detection dataset. By merging CNN-based feature extraction with boosting models fine-tuned via metaheuristics, the study achieved a 99.83% accuracy rate in multi-class IoT intrusion detection.

In [7], the researchers paired XGBoost with a Coyote optimization algorithm to classify 14,878 tweets in a Twitter sexist harassment dataset. The improved Coyote Optimization Algorithm (IBCOA) outperformed other advanced metaheuristics in tuning XGBoost for more accurate detection of sexist content.

In [8], a data-driven train delay prediction framework was developed by combining XGBoost with Genetic Algorithm, Particle Swarm Optimization, Whale Optimization Algorithm, and Grey Wolf Optimization. Using data from the Beijing–Shanghai high-speed railway, the model delivered high accuracy, low prediction errors, and enhanced interpretability by analysing dispatching commands and delay propagation mechanisms.

In [9], the authors used LightGBM optimized by a Whale Optimization (WO) algorithm to diagnose thyroid disease, achieving a 99.75% accuracy. This proves the efficacy of combining boosting models and metaheuristics for medical classification tasks.

In [10], researchers explored both XGBoost and LightGBM alongside Genetic Algorithm and Simulated Annealing on the Pima Indian Diabetes dataset. LightGBM with Genetic Algorithm provided the best performance at 86% accuracy, emphasizing the importance of effective feature selection in classification models.

In [11], XGBoost was integrated with a Modified Boxing Match (MBM) algorithm to forecast electricity consumption. The findings show that incorporating MBM significantly improved the model's accuracy for electricity demand prediction.

In [12], the authors combined XGBoost with an Artificial Bee Colony (ABC) algorithm for sales forecasting across three open-source datasets. A new hybrid approach—merging ABC with the Fire Hawk Optimizer (FHO)—yielded superior performance metrics (RMSE and MAPE) compared to using ABC or FHO alone.

In [13], XGBoost was paired with Particle Swarm Optimization and Gray Wolf Optimization to predict failure time for nine shovels at the Gol-Gohar iron ore mine in Iran. The PSO–XGB approach achieved an R^2 of 0.99 and enabled predictive maintenance, saving an estimated $61,189 per month.

In [14], an Enhanced Gradient Boosting Machine (EGBM) was proposed, optimized with a modified Particle Swarm Optimization, for customer churn prediction on seven open-source datasets. This CP-EGBM model outperformed traditional models like GBM and SVM, delivering high accuracy, recall, F1-measure, and AUC values in the telecommunications sector.

In [15], a hybrid SAOA–LightGBM method was presented to improve fault warning accuracy in industrial settings using 1,500 samples from a supervisory information system. With a 90% fault warning accuracy rate, it surpassed comparable state-of-the-art models in both prediction accuracy and generalization ability.

In [16], XGBoost, enhanced by a Boosted Particle Swarm Optimization algorithm, was applied to publicly available AIS data from Kaggle. The approach attained a 99.72% overall accuracy in vessel classification, illustrating the effectiveness of combining boosting and metaheuristics.

In [17], CatBoost was paired with the Grey Wolf Optimizer (GWO) to assess landslide susceptibility in Kerala, India, using data gathered from DEM, data portals, and published maps. The resulting CatBoost–GWO model showed a high AUC value of 0.910, highlighting robust predictive capability.

In [18], researchers optimized XGBoost using a Reptilian search algorithm to analyse a two-year database of hourly pollutant concentrations, including toluene. The study demonstrated how advanced metaheuristics paired with boosting models can significantly improve environmental data analytics.

In [19], a teaching–learning-based (TLB) optimization technique was implemented alongside XGBoost to detect fraud in a synthetic credit card dataset containing transactions from European cardholders. The hybrid model outperformed existing techniques by effectively tuning the boosting classifier and providing superior predictive performance.

In [20], a Hybrid Adaptive Red Fox Optimization algorithm (HARFO) was implemented alongside XGBoost and AdaBoost to detect insider threats through sentiment classification in the Insider Threat Test Dataset provided by the Carnegie Mellon University Software Engineering Institute. This synthetic dataset simulates logs of a large business over a 500-day interval, and the metaheuristic-enhanced approach produced commendable outcomes across various simulated scenarios by effectively tuning the boosting classifiers for superior predictive performance.

In [21], an iterative sine-cosine metaheuristic algorithm was integrated with an XGBoost model to enhance the prediction of external audit opinions. Leveraging a dataset of 12,690 observations from Serbian companies (2016–2019) and incorporating 598 variables, the approach aimed to surpass previously established benchmarks. By comparing six different metaheuristics for hyperparameter tuning and evaluating two distinct test scenarios with varying levels of difficulty, the optimized XGBoost model demonstrated superior predictive performance.

In [22], a framework based on a two-tier structure was presented, which uses a convolutional neural network (CNN) throughout the first layer and AdaBoost/XGBoost classifiers in the second layer, with PSO (particle swarm optimisation) implemented for hyperparameter optimisation. 5 data sets from NASA's Promise repository (KC1, JM1, CM1, KC2, and PC1) and approximately 1000 crowd-sourced Python programming problems were used. The two-tier structure produced accuracies of the classifiers of 0.768799 and 0.772166 while NLP solutions demonstrated even higher accuracies of 0.979781 and 0.983893 from the AdaBoost and XGBoost, respectively, thereby surpassing existing solutions.

4. CONCLUSION

These studies highlight the powerful combination of boosting algorithms and metaheuristic optimization for various tasks. Whether applied to IoT security, medical diagnosis, infrastructure maintenance, or sales and demand forecasting, this combined approach consistently demonstrates notable improvements in predictive performance.

Metaheuristic algorithms (e.g., modified reptile search, improved Coyote Optimization Algorithm, hybrid ABC-FHO, MBM, PSO, GWO) have been shown to be instrumental in fine-tuning the hyperparameters of boosting models, offering clear gains over traditional or standalone optimization methods.

Although the studies provided evidence of strengthening boosting by metaheuristic optimization, there are still areas to further investigate. For instance, the effectiveness of approaches may depend on the context in which the approach is utilized or the nature of the datasets that were used. By conducting tests on the same dataset and establishing benchmarks, one could determine whether the improvements were consistent across the varying contexts, allowing for a clearer identification of the best-performing approaches.

In general, the evidence clearly indicates that integrating metaheuristic-based hyperparameter tuning with advanced boosting algorithms generates robust, high-performing models that can adapt to a range of complex real-world problems.

REFERENCES

[1] S. Purkovic et al., “Audio analysis with convolutional neural networks and boosting algorithms tuned by metaheuristics for respiratory condition classification,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 10, p. 102261, Dec. 2024, doi: 10.1016/j.jksuci.2024.102261.

[2] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in KDD ’16*. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[3] I. Boussaïd, J. Lepagnot, and P. Siarry, “A survey on optimization metaheuristics,” *Elsevier BV*, vol. 237, pp. 82–117, Mar. 2013, doi: 10.1016/j.ins.2013.02.041.

[4] S. Almufti, R. Ali, and J. Fuente, “Overview of Metaheuristic Algorithms,” *Polaris Global Journal of Scholarly Research and Trends*, vol. 2, pp. 10–32, Apr. 2023, doi: 10.58429/pgjsrt.v2n2a144.

[5] M. Salb et al., “Enhancing Internet of Things Network Security Using Hybrid CNN and XGBoost Model Tuned via Modified Reptile Search Algorithm,” *Applied Sciences*, vol. 13, no. 23, p. 12687, Jan. 2023, doi: 10.3390/app132312687.

[6] M. Antonijevic et al., “Intrusion detection in metaverse environment internet of things systems by metaheuristics tuned two level framework,” *Sci Rep*, vol. 15, no. 1, p. 3555, Jan. 2025, doi: 10.1038/s41598-025-88135-9.

[7] M. Dobrojevic et al., “Cyberbullying Sexism Harassment Identification by Metaheurustics-Tuned eXtreme Gradient Boosting. *EBSCOhost*,” Sep. 2024. doi: 10.32604/cmc.2024.054459.

[8] T. Gao, J. Chen, and H. Xu, “Data-driven train delay prediction incorporating dispatching commands: An XGBoost-metaheuristic framework,” *IET Intelligent Transport Systems,* vol. 18, no. 10, pp. 1777–1796, 2024, doi: 10.1049/itr2.12461.

[9] B. B. Sinha, M. Ahsan, and R. Dhanalakshmi, “LightGBM empowered by whale optimization for thyroid disease detection,” *International Journal of Information Technology*, vol. 15, no. 4, pp. 2053–2062, Apr. 2023, doi: 10.1007/s41870-023-01261-3.

[10] G. Alp and F. Soygazi, “Meta-Heuristic Supported Feature Selection in Classification Algorithms for Diabetes Diagnosis,” in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, Mersin, Turkiye: IEEE, May 2024, pp. 1–4. doi: 10.1109/SIU61531.2024.10601062.

[11] N. Liao, Z. Hu, and D. Magami, “A metaheuristic approach to model the effect of temperature on urban electricity need utilizing XGBoost and modified boxing match algorithm,” *AIP Adv*, vol. 14, no. 11, p. 115318, Nov. 2024, doi: 10.1063/5.0219713.

[12] B. Gülsün, “Optimizing a Machine Learning Algorithm by Novel Metaheuristic Approach: Case Study in Forecasting,” Nov. 13, 2024, Preprints: 2024110907. doi: 10.20944/preprints202411.0907.v1.

[13] M. Mirzehi Kalateh Kazemi, Z. Nabavi, M. Rezakhah, and A. Masoudi, “Application of XGB-based metaheuristic techniques for prediction time-to-failure of mining machinery,” *Systems and Soft Computing,* vol. 5, p. 200061, Dec. 2023, doi: 10.1016/j.sasc.2023.200061.

[14] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnaim, “An efficient churn prediction model using gradient boosting machine and metaheuristic optimization,” *Sci Rep*, vol. 13, no. 1, p. 14441, Sep. 2023, doi: 10.1038/s41598-023-41093-6.

[15] S. Li, N. Jin, A. Dogani, Y. Yang, M. Zhang, and X. Gu, “Enhancing LightGBM for Industrial Fault Warning: An Innovative Hybrid Algorithm,” *Processes*, vol. 12, no. 1, p. 221, Jan. 2024, doi: 10.3390/pr12010221.

[16] A. Petrovic et al., “Marine Vessel Classification and Multivariate Trajectories Forecasting Using Metaheuristics-Optimized eXtreme Gradient Boosting and Recurrent Neural Networks,” *Applied Sciences,* vol. 13, no. 16, p. 9181, Jan. 2023, doi: 10.3390/app13169181.

[17] R. S. Ajin, S. Segoni, and R. Fanti, “Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility,” *Sci Rep*, vol. 14, no. 1, p. 24851, Oct. 2024, doi: 10.1038/s41598-024-72663-x.

[18] A. Stojić, “Revealing Toluene Behaviour in the Atmosphere Based on Coupling of Metaheuristics, Xgboost, and Shap,” *Sinteza 2023 - International Scientific Conference on Information Technology, Computer Science, and Data Science*, pp. 17–22, 2023, doi: 10.15308/Sinteza-2023-17-22

[19] A. Petrovic, M. Antonijevic, I. Strumberger, L. Jovanovic, N. Savanovic, and S. Janicijevic, “The XGBoost Approach Tuned by TLB Metaheuristics for Fraud Detection,” in *Proceedings of the 1st International Conference on Innovation in Information Technology and Business (ICIITB 2022)*, N. Bacanin and H. Shaker, Eds., Dordrecht: Atlantis Press International BV, 2023, pp. 219–234. doi: 10.2991/978-94-6463-110-4_16.

[20] D. Mladenovic et al., “Sentiment classification for insider threat identification using metaheuristic optimized machine learning classifiers,” *Sci Rep*, vol. 14, no. 1, p. 25731, Oct. 2024, doi: 10.1038/s41598-024-77240-w.

[21] M. Todorovic, N. Stanisic, M. Zivkovic, N. Bacanin, V. Simic, and E. B. Tirkolaee, “Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis,” *Applied Soft Computing*, vol. 149, p. 110955, Dec. 2023, doi: 10.1016/j.asoc.2023.110955.

[22] A. Petrovic et al., “Exploring Metaheuristic Optimized Machine Learning for Software Defect Detection on Natural Language and Classical Datasets,” *Mathematics*, vol. 12, no. 18, p. 2918, Sep. 2024, doi: 10.3390/math12182918.