Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240538

Multi-Objective Performance Optimization of Machine Learning Models in Healthcare

Maryam MORADPOUR[a,1] , Zully RITTER[a] and Anne-Christin HAUSHILD[a]
[a] *University Medical Center Göttingen, Dep. of Medical Informatics, Göttingen, Germany*
ORCiD: MM 0009-0001-0906-5842, ZR 0000-0002-1330-6178 , ACH 0000-0002-7499-4373

Abstract. Multi-objective optimization holds particular significance for medical applications, wherein enhancing sensitivity is crucial to avoid costly missed diagnoses, and maintaining high specificity is imperative to prevent unnecessary procedures. In particular, when optimizing machine learning architectures for clinical diagnostics, it becomes essential to balance target quality measures such as accuracy, sensitivity, and specificity. Therefore, we developed MOOF, a multi-objective optimization framework that employs NSGA-II and TOPSIS to simultaneously optimize the model parameters of three selected ML algorithms: random forest, support vector machine, and multilayer perceptron. Finally, we evaluated the performance of the optimized MOOF models compared to gold standard methods such as multi-score grid search and single objective optimizations. Our results show that MOOF generally outperforms other approaches by inherently providing optimal solutions, representing the trade-offs between the target objectives. In conclusion, the study supports the importance of multi-objective optimization in medical informatics, with MOOF as a powerful tool for precise ML models, potentially improving patient care and clinical decision support systems.

Keywords. Multi-objective optimization, machine learning, NSGA-II, clinical applications, TOPSIS

1. Introduction

Previous research has shown the great potential of Machine Learning (ML) for providing solutions in the healthcare sector [1–3]. This requires the integration of comprehensive patient data, including medical history, laboratory, vital signs, image, and clinical data. Which, in the best case, are analyzed together to diagnose and treat a patient efficiently. The record of the corresponding final diagnosis of the patient is used mainly to monitor treatment effectiveness but also can be used to develop and train machine learning models. Common goals for ML in healthcare are the prediction of the most probable diagnoses, optimal therapies, and risk estimation [4,5]. Essential for this task is to guarantee the robustness of the model performance when transferred to new data at different locations. In ML, the performance of a trained model is usually estimated by specific metrics, such as accuracy. However, other metrics are of more relevant for clinical applications, such as sensitivity or specificity. For instance, a high-sensitivity

[1] Corresponding Author: Maryam Moradpour, University Medical Center Göttingen, Dep. of Medical Informatics, Göttingen, Germany; E-mail: maryam.moradpour@med.uni-goettingen.de.

ML model can be prioritized over specificity if early detection and treatment of patients with the disease is crucial, boosting recovery chances [6,7]. Therefore, specialized multi-objective optimization (MOOP) techniques have been developed to simultaneously maximize multiple quality measures such as accuracy sensitivity and specificity. However, finding the best solutions can be, complex, and time-consuming. Its nature of providing not only one but a front of optimal solutions has produced growing attention to this method in solving problems in healthcare [6–10], for instance, simultaneous optimization of the allocation of healthcare resources as well as minimizing time and costs, as required during the COVID-19 pandemic [9]. In this work, we aim to evaluate potential performance improvement of the newly developed MOOP framework in comparison against the gold standard approach a multi-scoring grid search. Therefore, we compare the performance of both approaches based on patient data from the Cleveland heart failure data set [11] focusing on the three main metrics, accuracy, sensitivity and specificity.

2. Material and Methods

To benchmark our hypothesis that the employment of the developed multi-objective optimization strategy allows improved performance we evaluated it on the Cleveland heart failure data set [11] which was also employed in the benchmark study evaluating MO techniques [12] and used for comparisons with the approach proposed in this work. The Cleveland data set consists of 303 samples and 14 attributes of patient clinical data, and their corresponding outcome of having or not a heart attack. We developed the Multi-Objective Optimization Framework (MOOF) that consists of three phases, (1) modeling, (2) optimization, and (3) the selection phase (Figure 1).

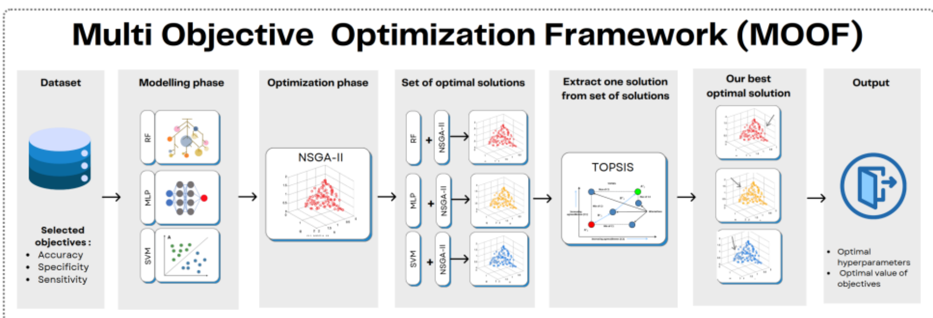

Figure 1. Overview of the proposed framework approach

In the modeling phase, we are using three ML models, including random forest, support vector machine, and multilayer perceptron. In the second phase, we optimized the ML models using NSGA-II [13], which is well-known to handle complex optimization problems. Our setting allows us to identify a set of optimal solutions amongst many models while simultaneously optimizing the model parameters for multiple objectives. This property makes this method a powerful tool for tailoring the design of solutions in healthcare [14]. We focus on improving three objectives: accuracy, sensitivity, and specificity. These criteria are important for evaluating the performance of predictive models in medical diagnosis, where false negatives and positives can have significant consequences. The optimization process identifies a set of optimal solutions that provide a trade-off between the three predefined objectives. Each solution corresponds to a specific set of model parameters and resulting quality metrics; these represent the so-called Pareto front. These optimal solutions provide a comprehensive understanding of the behavior of model and can be analyzed based on specific research goals. There are multiple ways to extract the best solution from a set of optimal solutions. In phase three, we employed the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method [15], which evaluates the relative closeness of each solution to an ideal solution based on a predefined criteria. We used it to rank solutions and select the one that is closest to the ideal solution. For comparability, we selected the same metric weights based on the study by Lin and Yeh [15]. Finally, we compare the quality of the optimal solution identified by MOOF with multi-scoring grid search (GS) as the gold standard and the results of a benchmark study by Nalluri et al. employing Particle Swarm Optimization (PSO), Gravitational Search Algorithm (GSA), and Firefly Algorithm (FA) that employed a weighted metrics approach to address several objectives simultaneously [12].

3. Results

In this study, we utilized the NSGA-II algorithm to enhance the performance of three distinct ML models, namely RF, SVM, and MLP, focusing on concurrently improving accuracy, sensitivity, and specificity. For each ML method, NSGA-II identified a set of optimal solutions. Demonstrating the effectiveness of this optimization process in comparison to other approaches, we provide a detailed examination of the results based on the MLP model (Figures 2 and 3), noting that similar patterns and compromises were evident in the RF and SVM models as well.

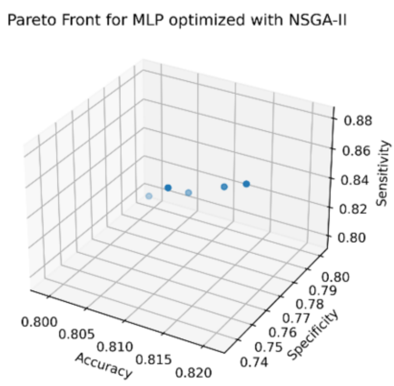

Figure 2. Pareto front plot for MLP

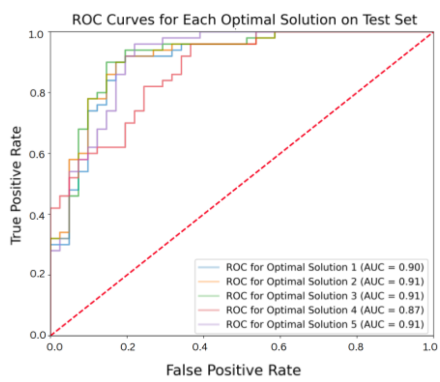

Figure 3. ROC curve plot for MLP

The Pareto front in Figure 2, effectively illustrates the multi-faceted trade-offs involved, offering a clear visual representation of how various optimized solutions manage to balance the three key objectives. Moreover, the robustness of these solutions is quantitatively validated by the Area Under the Receiver Operating Characteristics (ROC) curves (AUC), indicating a strong performance in the range of 0.87 to 0.91. This demonstrates NSGA-IIs ability to accurately differentiate patient outcomes in heart failure scenarios. For MLP, MOOF showed the best accuracy of 86.81% compared to the baseline model (79.10%) and GS (83.06%) see Table 1. Sensitivity reached 90.00%, surpassing outcomes from PSO, GSA, and FA algorithms. This underlines MOOF's efficiency in finely tuning the balance between sensitivity and specificity, critical for clinical decision-making processes. Similar to the MLP analysis MOOF showed superior accuracy of 83.52% for the optimized SVM model, see Table 2. Remarkably, when compared to PSO, GSA and FA, MOOF showed a higher sensitivity of 92%, demonstrating a greater increase than the single-objective approaches. The performance of the RF-MOOF model of (ACC 89.01%) exceeded the baseline (ACC 80.84%) and GS (ACC 81.95%) optimized models. However, the GS demonstrated the higher sensitivity. The specificity remained strong at 85.37%, despite challenges in maintaining a high true negative rate alongside improvements in true positive detection.

Table 1. Performance Metrics for MLP with MOOF with GS and reference paper

Objective	Baseline	GS	PSO	GSA	FA	MOOF
ACC	79.10	83.06	85.14	84.15	85.80	86.81
SEN	79.42	93.33	85.60	84.21	87.5	90.00
SPE	78.61	79.35	84.79	84.11	84.57	82.93

Table 2. Performance Metrics for SVM with MOOF with GS and reference paper

Objective	Baseline	GS	PSO	GSA	FA	MOOF
ACC	78.50	82.70	83.49	82.83	83.49	83.52
SEN	82.61	99.73	82.38	82.18	82.38	92.00
SPE	73.58	79.20	85.03	83.72	85.03	73.17

Table 3. Performance Comparison for MOOF

Objective	Baseline	GS	MOOF
ACC	80.84	81.95	89.01
SEN	85.22	76.39	92.00
SPE	75.67	87.83	85.37

4. Discussion and Conclusions

In this study, we designed MOOF, a framework for multi-objective optimization to refine ML models for clinical predictions by simultaneously tuning accuracy, sensitivity, and specificity and allows selecting an optimal solution from the Pareto front. This is particularly relevant for medical diagnostics, where e.g., sensitivity must be improved in situations where missed diagnoses are costly, and high specificity is essential to avoid unnecessary procedures. Our results demonstrate that NSGA-II generally outperforms other approaches such as multi-scoring GS, PSO, GSA and FA by inherently providing optimal solutions, representing the trade-offs between the target objectives. For instance, it achieved high sensitivity without degrading model performance, proving its superiority over single-objective optimizers as presented by Nalluri et al [12]. In contrast, multi-scoring GS outperformed the MOOF selected MLP and SVM models concerning sensitivity, however, generally showed worse specificity, while the MOOF selected RF model showed superior accuracy and sensitivity. Furthermore, the principal benefit of the framework lies in its ability to balance between various performances metrics and provide a range of solutions to meet different medical conditions. NSGA-II's flexibility across different models highlighted its potential in handling complex clinical diagnosis and treatment planning issues. We expect to improve this property by using NSGA-III, as shown in other studies [16]. Moreover, the distinct sets of optimal solutions found for the SVM, RF, and MLP models highlight flexibility of NSGA-II across different types of model architectures and their unique ways of processing clinical data. On the one hand, this flexibility can be a strength for overcoming the complex clinical diagnosis and treatment planning issues. On the other hand, MOOF's performance depends on fine-tuning a multitude of parameters like population size and crossover probability in NSGA-II as well as model specific parameters, introducing additional complexity and potentially delaying rapid deployment. Therefore, in future work we aim to test MOOF on various datasets and incorporate NSGA-III to improve efficiency and solution. In conclusion, the study shows the importance of multi-objective optimization in medical informatics, potentially improving patient care and clinical decision support systems. In conclusion, the study shows the importance of multi-objective optimization in medical informatics, potentially improving patient care and clinical decision support systems.

Acknowledgments and Funding This work is supported by the German Ministry of Education and Research (BMBF) grant No. 01KD2208A (FAIrPaCT), and KISSKI No. 01 IS 22 093 A-E, and by the Innovation Committee at the Federal Joint Committee (No. 01VSF20014, KI-Thrust).

References

[1] Okada Y, Mertens M, Liu N, Lam SSW, Ong MEH. AI and machine learning in resuscitation: Ongoing research, new concepts, and key challenges. Resusc Plus. 2023 Jul 28;15:100435. doi: 10.1371/journal.pone.0245157

[2] Elendu C, et al. Ethical implications of AI and robotics in healthcare: A review. Medicine (Baltimore). 2023 Dec 15;102(50):e36671. doi: 10.1097/MD.0000000000036671

[3] Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2019 Nov 4;27(3):491–7. doi: 10.1093/jamia/ocz192.

[4] van Doorn WPTM, Stassen PM, Borggreve HF, Schalkwijk MJ, Stoffers J, Bekers O, et al. A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. PLoS One. 2021 Jan 19;16(1):e0245157. doi: 10.1371/journal.pone.0245157.

[5] Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, et al. A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. Front Public Health. 2021 Oct 15;9:754348. doi: 10.3389/fpubh.2021.754348.

[6] Steiner JM, Morse C, Lee RY, Curtis JR, Engelberg RA. Sensitivity and Specificity of a Machine Learning Algorithm to Identify Goals-of-care Documentation for Adults With Congenital Heart Disease at the End of Life. J Pain Symptom Manage. 2020 Sep;60(3):e33–6. doi: 10.1016/j.jpainsymman.2020.06.018.

[7] Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health. 2017 Nov 20;5:307. doi: 10.3389/fpubh.2017.00307.

[8] Angelo JS, Guedes IA, Barbosa HJ, Dardenne LE. Multi-and many-objective optimization: present and future in de novo drug design. Frontiers in Chemistry. 2023;11. doi: 10.3389/fchem.2023.1288626.

[9] AbdelAziz AM, et al. A Multi-Objective Optimization Method for Hospital Admission Problem—A Case Study on Covid-19 Patients. Algorithms. 2021 Feb;14(2):38. doi: 10.3390/a14020038.

[10] Eriskin L, Karatas M, Zheng YJ. A robust multi-objective model for healthcare resource management and location planning during pandemics. Ann Oper Res. 2022 May 21;1–48. doi: 10.1007/s10479-022-04760-x

[11] Janosi A, Steinbrunn W, Pfisterer M, Detrano R. UCI machine learning repository-heart disease data set. School Inf. Comput. Sci., Univ. California, Irvine, CA, USA. 1988. doi: 10.24432/C52P4X.

[12] Nalluri MR. Hybrid Disease Diagnosis Using Multiobjective Optimization with Evolutionary Parameter Optimization. Journal of Healthcare Engineering. doi: 10.1155/2017/5907264.

[13] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 2002 Apr;6(2):182–97. doi: 10.1109/4235.996017.

[14] Wan F, et al. Two-stage multi-objective optimization for ICU bed allocation under multiple sources of uncertainty. Sci Rep. 2023 Nov 2;13:18925. doi: 10.1038/s41598-023-45777-x

[15] Lin YK, Yeh CT. Multi-objective optimization for stochastic computer networks using NSGA-II and TOPSIS. European Journal of Operational Research. 2012 May 1;218(3):735–46. doi: 10.1016/j.ejor.2011.11.028.

[16] Manupati VK, Schoenherr T, Wagner SM, Soni B, Panigrahi S, Ramkumar M. Convalescent plasma bank facility location-allocation problem for COVID-19. Transportation Research Part E: Logistics and Transportation Review. 2021 Dec;156:102517. doi: 10.1016/j.tre.2021.102517.