

## Article

# Impact of Hyperparameter Optimization to Enhance Machine Learning Performance: A Case Study on Breast Cancer Recurrence Prediction

Lorena González-Castro <sup>1,\*</sup>, Marcela Chávez <sup>2</sup>, Patrick Dufлот <sup>2</sup>, Valérie Bleret <sup>3</sup>, Guilherme Del Fiol <sup>4</sup> and Martín López-Nores <sup>5</sup>

<sup>1</sup> School of Telecommunication Engineering, Universidade de Vigo, 36310 Vigo, Spain

<sup>2</sup> Department of Information System Management, CHU of Liège, 4000 Liège, Belgium

<sup>3</sup> Senology Department, CHU of Liège, 4000 Liège, Belgium

<sup>4</sup> Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT 84132, USA

<sup>5</sup> atlantTtic Research Center, Department of Telematics Engineering, Universidade de Vigo, 36130 Vigo, Spain

\* Correspondence: lorena.gonzalez@uvigo.es

† Most of the work developed by this author has been carried out during her employment with GRADIANT (Galician Research and Development Center in Advanced Telecommunications, Vigo, Spain).

**Abstract:** Accurate and early prediction of breast cancer recurrence is crucial to guide medical decisions and treatment success. Machine learning (ML) has shown promise in this domain. However, its effectiveness critically depends on proper hyperparameter setting, a step that is not always performed systematically in the development of ML models. In this study, we aimed to highlight the impact that this process has on the final performance of ML models through a real-world case study by predicting the five-year recurrence of breast cancer patients. We compared the performance of five ML algorithms (Logistic Regression, Decision Tree, Gradient Boosting, eXtreme Gradient Boost, and Deep Neural Network) before and after optimizing their hyperparameters. Simpler algorithms showed better performance using the default hyperparameters. However, after the optimization process, the more complex algorithms demonstrated superior performance. The AUCs obtained before and after adjustment were 0.7 vs. 0.84 for XGB, 0.64 vs. 0.75 for DNN, 0.7 vs. 0.8 for GB, 0.62 vs. 0.7 for DT, and 0.77 vs. 0.72 for LR. The results underscore the critical importance of hyperparameter selection in the development of ML algorithms for the prediction of cancer recurrence. Neglecting this step can undermine the potential of more powerful algorithms and lead to the choice of suboptimal models.

**Keywords:** hyperparameter optimization; breast cancer; recurrence prediction; machine learning



**Citation:** González-Castro, L.; Chávez, M.; Dufлот, P.; Bleret, V.; Del Fiol, G.; López-Nores, M. Impact of Hyperparameter Optimization to Enhance Machine Learning Performance: A Case Study on Breast Cancer Recurrence Prediction. *Appl. Sci.* **2024**, *14*, 5909. <https://doi.org/10.3390/app14135909>

Academic Editors: Asiya Khan and Gloria Iyawa

Received: 14 May 2024

Revised: 25 June 2024

Accepted: 4 July 2024

Published: 6 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Breast cancer is the most prevalent form of cancer, accounting for 12.5% of total annual cancer cases, and is one of the main causes of mortality among women worldwide [1,2]. Despite the high mortality rates, breast cancer is one of the malignancies with the best prognosis; when diagnosed at an early and localized stage, its five-year survival rate reaches 96% in Europe. However, survival for women diagnosed at an advanced stage is around 38% [3]. It is clear that early detection of breast cancer and its recurrence not only improves treatment opportunities, but it also significantly improves patient survival rates. However, despite notable advances in early detection methods, accurately predicting the prognosis of breast cancer is still a major challenge [4,5].

The use of machine learning (ML) is generating great research interest in the biomedical domain. In particular, ML is increasingly being used in applications such as disease detection and diagnosis [6–8], improvement in patient safety [9,10], and reduction in healthcare costs [11,12]. In cancer prognosis, ML algorithms have been shown to be promising tools capable of assisting in the prediction of cancer recurrence, providing doctors with valuable information for clinical decision-making [13,14].

The performance of ML models depends critically upon a wise choice of hyperparameter values guided by systematic analyses. One of the main challenges hindering the effective use of ML to predict clinical events is the difficulty in determining the optimal combination of an algorithm and its hyperparameter values. The selection of these hyperparameters is a process that requires specialized knowledge and often involves several time-consuming manual iterations [15]. There is ample evidence that hyperparameter optimization improves overall performance [16,17], and a number of methods have been proposed to automatically choose hyperparameter values in order to make ML more accessible to less experienced users [18]. The goal of these methods is to try to quickly identify a combination of optimal—or, at least, efficient—hyperparameters that maximize a given performance metric for the ML task at hand.

For the past few decades, the grid-search method has been the predominant standard for parameter optimization in ML. This method systematically explores a predefined set of hyperparameter values to find the best combination for a given model. Although other approaches have been proposed, such as Random Search [19], Bayesian Optimization [20], and Gradient Optimization [21], the grid-search method remains the preferred method due to its ease of execution, parallelization capability, and effectiveness in low-dimensional spaces.

However, despite the advances made in the field of ML and the existence of automated hyperparameter optimization techniques, it is still common to find research in the literature that overlooks this crucial step in the process of developing ML algorithms. For example, Ebrahim et al. [22] performed a comparative analysis of seven ML models used to predict breast cancer. They investigated the effect of feature selection but did not optimize the model hyperparameters. Kaushik et al. [23] trained an XGB model to predict cervical cancer. They specified the hyperparameter values used but did not mention how these values were chosen or whether any optimization was performed. A model that combines SVM with an extremely randomized trees classifier for breast cancer diagnosis was proposed by Alfian et al. [24]. They specified the values used for the hyperparameters but did not explain how they were selected, and they also compared their proposed model to other ML algorithms, for which default hyperparameters were used. Lou et al. [25] trained several ML algorithms and performed a statistical analysis to compare their accuracy in predicting breast cancer recurrence in 10 years after surgery. However, no mention was made about whether a hyperparameter optimization process was carried out. In the case of Ganggayah et al. [26], six ML algorithms were compared for breast cancer survival analysis, achieving a similar performance. The authors acknowledged using the default values of the software package used in the study. However, the study could have reached substantially different findings if hyperparameter optimization had been performed. Massafra et al. [27] compared the performance of three ML algorithms in predicting invasive breast cancer recurrence. The authors also omitted a hyperparameter optimization step, but admitted its importance and left it for future work. Failure to adequately address the selection of hyperparameters may weaken the robustness and validity of the findings obtained in these studies.

This article aims to assess the importance of performing the hyperparameter optimization process through a case study that uses real-world data and five modeling approaches to predict five-year recurrence of breast cancer. Through our study, we aim to encourage the adoption of automatic hyperparameter selection methodologies for the analysis of healthcare data, resulting in the development of more efficient and accurate predictive algorithms in future research.

## 2. Materials and Methods

### 2.1. Data Preparation

In this study, we used a cohort of 3839 patients diagnosed with breast cancer between 2010 and 2020 at the CHU de Liège hospital [13]. To normalize the representation of Electronic Health Record (EHR) data, we first mapped patient data to the CASIDE data model [28], adhering to the Fast Healthcare Interoperability Resources (FHIR) stan-

dard. This mapping allowed the data to be represented in a consistent manner, laying the foundations for the subsequent analysis.

Next, a data cleaning process was carried out to improve the quality of the data. We eliminated duplicated records and excluded patients who were missing essential information, including TNM stage, information on the type of treatment administered, and confirmation of survival of at least 5 years after diagnosis or recurrence within that period.

We applied feature transformation to prepare the data for automatic analysis. We transformed dates into numerical representations, such as patient age or age at diagnosis, facilitating their use in quantitative analysis. We then converted nominal features into binary class data. Additionally, some features were aggregated to create more informative representations of them. For example, we mapped diagnosis codes to Elixhauser categories [29], counting the number of different diagnoses per category for each patient. To ensure robust statistical analysis, only categories with 50 or more instances in the dataset were retained.

To address scale variations and ensure uniformity in the impact of different features, we applied scaling to normalize the numerical variables. This process was crucial to mitigate the impact of variable scales on the performance of ML models, promoting more reliable and meaningful results.

The resulting set of features included sex, age at diagnosis, BMI, ECOG, comorbidities, tumor grade and staging, breast cancer biomarkers and treatment. A detailed analysis of the composition and statistical characteristics of this dataset has been described elsewhere [13].

## 2.2. Model Training, Hyperparameter Optimization, and Evaluation

To assess the effect of hyperparameter optimization on model performance, we trained, optimized, and evaluated five distinct ML algorithms for predicting breast cancer recurrence within 5 years, namely Logistic Regression (LR), Decision Tree (DT), Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), and Deep Neural Network (DNN). The implementation of the ML algorithms and hyperparameter tuning has been carried out in Python, using the Scikit-Learn [30], XGBoost [31], and Tensorflow [32] libraries.

To analyze the effectiveness of hyperparameter optimization on model prediction, we compared the performance of each model before (i.e., using default hyperparameters of the respective ML libraries) versus after optimization. To build and evaluate the models, we adopted a combination of hold-out and cross-validation strategies, thereby ensuring a robust evaluation by assessing model performance on unseen data [33].

The dataset was randomly divided into two mutually exclusive sets with 90% of the samples allocated for training and fine-tuning and 10% for testing. First, we trained all the algorithms using the default value for hyperparameters specified in the corresponding software package based on experience and general recommendations. In the case of the DNN, some extra configuration was required. A network with a single hidden layer with as many neurons as input features was chosen as the basis. In addition, as recommended for classification problems [34], binary cross-entropy was set as the loss function.

Subsequently, hyperparameter optimization was performed using the grid-search method, implemented through three rounds of stratified 6-fold cross-validation on the training set. Each pass of cross-validation utilized 75% of the whole dataset for training and 15% for validation. This strategy allowed us to systematically explore multiple combinations of hyperparameters and select the most effective set for each algorithm. Table 1 describes the parameters that were optimized in each of the algorithms, along with the search space used.

Table 1. Optimized parameters and search space.

Algorithm	Package Version	Parameter	Search Space
LR	scikit-learn 1.0.2	solver	'newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga'
		penalty	'none', 'l1', 'l2', 'elasticnet'
		C	$1 \times 10^{-5}$ , $1 \times 10^{-4}$ , $1 \times 10^{-3}$ , $1 \times 10^{-2}$ , $1 \times 10^{-1}$ , 1, 10, 100
		l1_ratio	0.1, 0.3, 0.5, 0.7, 0.9
		class_weight	None, 'balanced'
DT	scikit-learn 1.0.2	criterion	'gini', 'entropy'
		splitter	'best', 'random'
		max_depth	2, 3, 5, 8, 12, 20, None
		min_samples_split	2, 3, 4, 5, 6, 7, 8
		min_samples_leaf	1, 5, 10, 20, 30, 40, 50
		max_leaf_nodes	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50
		max_features	'sqrt', 'log2', None
GB	scikit-learn 1.0.2	min_samples_split	0.1, 0.3, 0.5, 0.7, 0.9, 1
		min_samples_leaf	0.1, 0.2, 0.3, 0.4, 0.5, 1
		max_features	'auto', 'sqrt', 'log2', None
		max_leaf_nodes	8, 16, 64, 100, None
		learning_rate	0.01, 0.05, 0.1, 0.25
		n_estimators	8, 16, 32, 64, 100, 200
		max_depth	2, 3, 5, 8, 12, 20
XGB	xgboost 1.5.2	n_estimators	35, 50, 65, 80, 100, 115, 130, 150, 300
		learning_rate	0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3
		max_depth	4, 6, 8, 10
		min_child_weight	1, 4, 6, 8
		subsample	0.5, 0.8, 1.0
		colsample_bytree	0.5, 0.8, 1.0
		gamma	0, 0.01, 0.25, 0.5, 1
		scale_pos_weight	1, 4, 7
		reg_alpha	0, 0.001, 0.01, 0.1, 0.5, 1, 2, 5, 10
		reg_lambda	0, 0.001, 0.01, 0.1, 0.5, 1, 2, 5, 10
DNN	tensorflow 2.7.0	number hidden layer	1, 2, 3
		epochs	10, 20, 30, 40, 50, 70, 90, 110, 150
		batch_size	1, 16, 32, 64
		dropout	0.0, 0.25, 0.5
		units	40, 80, 100
		kernel_initializer	'uniform', 'lecun_uniform', 'normal', 'zero', 'glorot_normal', 'glorot_uniform', 'he_normal', 'he_uniform'
		activation	'softmax', 'softplus', 'softsign', 'relu', 'tanh', 'sigmoid', 'hard_sigmoid', 'linear'
		kernel_constraint	1, 2, 3, 4, 5
		learning_rate	0.001, 0.01, 0.1, 0.2, 0.3
		optimizer	SGD, RMSprop, Adagrad, Adadelata, Adam, Adamax, Nadam

A critical problem in training ML algorithms on healthcare data is the inherent imbalance of the datasets. The class imbalance distribution biases classifiers towards the majority class, which results in an unsatisfactory prediction performance for the minority class [35]. In our dataset, we found that the number of patients who eventually had cancer recurrence in five years was 13% of the total sample. To mitigate the impact of imbalanced data, we applied the Synthetic Minority Over-sampling Technique (SMOTE) [36] in each cross-validation pass. This technique oversampled the minority class, ensuring equal representation and reducing bias in the model towards the majority class. A k-value of 5 was employed to generate synthetic samples, contributing to a more representative training set.

Once the best hyperparameters for each algorithm were identified, we refitted the models using the entire training partition. Subsequently, we evaluated the performance of

the optimized models on the test hold-out, providing an unbiased assessment using the data that had not been previously seen in both the training and optimization phases.

Performance evaluation was conducted using the area under the ROC curve (AUC) as the primary outcome, and precision, recall, and F1 measure as the secondary outcomes.

### 3. Results

The five algorithms analyzed in this study have been adjusted to determine their optimal combination of hyperparameters, as described in Table 2.

**Table 2.** Optimal combination of hyperparameters for each algorithm.

Algorithm	Optimized Hyperparameters
R	solver = 'saga', penalty = 'l1', C = 1, l1_ratio = 0.1, class_weight = None
DT	criterion = 'gini', splitter = 'best', max_depth = 3, min_samples_split = 2, min_samples_leaf = 1, max_leaf_nodes = 6, max_features = None
GB	min_samples_split = 0.1, min_samples_leaf = 1, max_features = 'log2', max_leaf_nodes = 8, learning_rate = 0.05, n_estimators = 16, max_depth = 5
XGB	n_estimators = 50, learning_rate = 0.1, max_depth = 4, min_child_weight = 1, subsample = 0.5, colsample_bytree = 1, gamma = 0, scale_pos_weight = 1, reg_alpha = 0, reg_lambda = 1
DNN	number_hidden_layer = 1, epochs = 30, batch_size = 64, dropout = 0.5, units = 100, kernel_initializer = 'he_normal', activation = 'relu', kernel_constraint = 1, learning_rate = 0.001, optimizer = 'Adam'

Table 3 shows a comparison of the predictive performance of the five ML algorithms before optimization (using the default values of the corresponding software package) and after performing hyperparameter tuning. The results before hyperparameter optimization show that simpler models, such as LR and DT, outperformed more complex algorithms like XGB and DNN in various metrics. LR achieved the best precision and AUC (0.87 and 0.77, respectively), while GB exhibited the highest performance for recall and F1 score (0.88 and 0.87, respectively).

**Table 3.** Performance of ML models before and after Hyperparameter (HP) Optimization.

	Before HP Optimization				After HP Optimization			
	Precision	Recall	F1	AUC	Precision	Recall	F1	AUC
LR	<b>0.87</b>	0.83	0.85	<b>0.77</b>	0.86	0.8	0.82	0.72
DT	0.83	0.78	0.8	0.62	0.87	0.86	0.86	0.7
GB	0.86	<b>0.88</b>	<b>0.87</b>	0.7	0.91	0.9	0.91	0.8
XGB	0.81	0.86	0.83	0.7	<b>0.92</b>	<b>0.93</b>	<b>0.92</b>	<b>0.84</b>
DNN	0.82	0.72	0.76	0.64	0.91	0.92	0.91	0.75

After hyperparameter optimization, all algorithms demonstrated improved performance across every metric, except for LR, which exhibited a decrease in overall performance. The AUCs obtained before and after adjustment were 0.7 vs. 0.84 for XGB, 0.64 vs. 0.75 for DNN, 0.7 vs. 0.8 for GB, 0.62 vs. 0.7 for DT, and 0.77 vs. 0.72 for LR. This decrease in the results of LR may have happened because the optimal hyperparameters in each fold did not average to values that improve performance on the unseen test partition. This unexpected outcome could be attributed to collinearity issues, a phenomenon well known to regression models but less prevalent in other algorithms.

Notably, the improvement in performance was substantial for XGB and DNN in all metrics. DNN showed substantial improvements over the results obtained using the predefined parameters, achieving up to a 20% increase in recall. For its part, XGB was the algorithm that achieved the highest performance among all, reaching precision = 0.92,

recall = 0.93, F1 = 0.92, and AUC = 0.84. These results suggest that algorithms with a higher number of hyperparameters, such as XGB and DNN, may derive greater benefits from hyperparameter optimization.

#### 4. Discussion

The development of ML models for solving practical problems with real-world data requires careful attention to the appropriate selection of hyperparameters to handle specific datasets. In particular, our study highlights the critical role of hyperparameter optimization in ML algorithms trained for predicting breast cancer recurrence.

Looking at the optimal combinations of the resulting hyperparameters in Table 2, we can see that for LR, the choice of the 'saga' solver and the 'l1' penalty suggests that L1 regularization is effective in handling high dimensionality and multicollinearity in the data. As for DT, a reduced maximum depth of estimators and limited number of maximum leaf nodes reflect a preference for simpler, shallower models, which are less prone to overfitting and can generalize better on our dataset. In the cases of GB and XGB, a low learning rate and a moderate maximum depth of estimators show that models with slower learning and less depth tend to improve performance, possibly by avoiding overfitting. Finally, for DNN, the choice of a single hidden layer with 100 neurons, together with the Adam optimizer and a low learning rate, turned out to be the most effective, suggesting that even in shallow architectures, it is possible to obtain good performance when they are properly optimized.

Not all the models in this study benefited equally from hyperparameter optimization. Simpler models, such as LR, experienced virtually no benefit, suggesting that using their default settings might be a reasonable approach for the task at hand. However, the improvements obtained with hyperparameter optimization become more significant as the models grow in complexity—that is, a greater number of hyperparameters. In our study, the improvement in performance was greater in the more complex models, i.e., GB, XGB, and DNN. For example, the DNN model showed a substantial increase, from 0.72 in recall and 0.76 in F1, using the default values, to 0.92 and 0.91 with the hyperparameter values optimized, respectively.

The results also highlight the crucial role that hyperparameter selection has in determining the final performance of the models, as well as hint at the possible consequences of not investing sufficient effort in the optimization phase. Before optimization, one would have chosen to use LR or GB, depending on the metric to maximize. However, after hyperparameter tuning, XGB is clearly positioned as the optimal choice. Therefore, the lack of systematic hyperparameter optimization risks discarding a more efficient and robust algorithm in favor of a simpler one that requires minimal adjustment, but that is incapable of modeling complex non-linear data, as is often the case in the healthcare domain [37].

In the context of predicting breast cancer recurrence, the absence of hyperparameter optimization in prior studies, like Lou et al. [25], Ganggayah et al. [26], and Massafra et al. [27], raises concerns about the validity of their conclusions. Hyperparameter optimization could have led to a significantly higher performance for the models used or even revealed a different superior model, one better suited to capturing the intricacies of health data.

Our observations add to a growing body of research supporting the importance of hyperparameter optimization in the development of ML models for healthcare. Several previous studies have shown substantial performance gains after implementing this technique [38–40]. This reinforces our findings that prioritizing hyperparameter optimization can significantly enhance model accuracy and robustness. Therefore, rigorous hyperparameter optimization should be a standard practice in future research to ensure a fair comparison of ML models and fully leverage their capabilities in healthcare.

The improvement in the performance of the algorithm selected after optimization has significant implications as it can have profound consequences in decision-making regarding prevention, early detection, and treatment strategies for breast cancer recurrence [41]. A higher precision algorithm ensures that recurrence predictions are more accurate, reducing

the likelihood of false positives. An algorithm that allows for the identification of high-risk patients accurately becomes a powerful tool to help healthcare providers in decision-making. Clinicians could confidently tailor interventions based on the specific risk profile and prioritize follow-up and surveillance for those patients at highest risk of recurrence. For low-risk patients, the application of unnecessary and burdensome treatments or techniques could be avoided. Conversely, patients at higher risk could receive more aggressive interventions, potentially improving survival rates and avoiding potential delays in critical therapeutic measures.

Some limitations need to be acknowledged in our study. First, the findings are specific to the prediction of breast cancer recurrence and may not be directly applicable to other predictive modelling tasks. Additionally, the scope of the study is limited only to a cohort of patients from the CHU hospital in Liège, which is not representative of a wider population. Therefore, generalization to other populations, patient demographics, and healthcare settings is uncertain. It would be of great interest to extend this study to a variety of centers to verify if the results obtained are generalizable to other cohorts of patients. Furthermore, only five ML algorithms were analyzed in this study, and the applicability of the results to other algorithms remains unexplored. Moreover, in the field of survival analysis, there are methods such as models based on Cox proportional hazards that offer greater granularity in predictions regarding the time of recurrence. Although binary classifiers such as those used in this study have the potential to achieve greater predictive accuracy and can be useful in situations where a quick and clear decision is critical, they lack the interpretability and flexibility provided by modeling the event probabilities as a function of time [42]. Future research should explore the implementation and comparison of time-sensitive methods for the prediction of breast cancer recurrence. Finally, this study was based on a single hyperparameter selection method, and exploring a broader range of methods could provide additional insights. Future work should compare the search efficiency and overall ML model performance using different hyperparameter selection methods, such as random search, Bayesian optimization, or hyperparameter search using genetic algorithms. Furthermore, it would also be interesting to investigate the interactions among different hyperparameters, and how these affect the selection process and the ability to find optimal values. Understanding how hyperparameters influence each other could provide valuable information to improve the efficiency and efficacy of automatic selection methods.

## 5. Conclusions

The results obtained in this study show that the adequate selection of hyperparameter values significantly improved the performance of the evaluated models. This is especially relevant for more complex algorithms like XGB and DNN, which translates into greater performance in the prediction of breast cancer recurrence. The study demonstrated that skipping hyperparameter optimization can lead to the selection of less accurate models with suboptimal predictive capabilities for a given task. Therefore, our findings confirm the significance of performing hyperparameter optimization as an essential step in the development of ML models, especially in healthcare, where the data complexity requires more careful modeling.

Optimized ML models have the potential to achieve higher accuracy levels, becoming valuable tools that can help clinicians make more informed decisions. The findings presented in this study contribute to the ongoing efforts to leverage ML in healthcare, moving the medical community one step closer to more precise, effective, and patient-centered approaches for cancer treatment and management, ultimately contributing to improved patient outcomes.

**Author Contributions:** Conceptualization and methodology, L.G.-C.; data curation, L.G.-C. and P.D.; software, L.G.-C.; formal analysis, L.G.-C.; investigation, L.G.-C.; resources, M.C. and V.B.; supervision, G.D.F. and M.L.-N.; project administration, L.G.-C.; writing—original draft, L.G.-C.; writing—review and editing, L.G.-C., M.L.-N. and G.D.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** Part of this work was supported by the European Union’s Horizon 2020 research and innovation program under Grant Agreement No. 875406. The authors from the University of Vigo received support from the European Regional Development Fund (ERDF) and the Galician Regional Government under an agreement to fund the atlantTic Research Center for Telecommunication Technologies.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Centre Hospitalier Universitaire de Liège (protocol code 2020/248, approved on 25 August 2020) for studies involving humans.

**Informed Consent Statement:** Patient consent was waived due to the retrospective nature of this study.

**Data Availability Statement:** The datasets analyzed during the current project are not publicly available due to legal agreements made with the providing institution. Aggregated data in the form of tables are available from the corresponding author on reasonable request and subject to institutional approval.

**Acknowledgments:** The authors want to thank 3M for the free usage of their 360 Encompass™ anonymization tool. This study is part of the H2020 PERSIST project coordinated by GRADIANT. The content of this article is the sole responsibility of its authors, and it does not represent the opinion of the EC.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Breast Cancer Facts and Statistics. Available online: <https://www.breastcancer.org/facts-statistics> (accessed on 8 April 2024).
3. Breast Cancer Outcomes. Available online: <https://www.oecd-ilibrary.org/sites/c63a671a-en/index.html?itemId=/content/component/c63a671a-en#> (accessed on 8 April 2024).
4. Tufail, A.B.; Ma, Y.K.; Kaabar, M.K.; Martínez, F.; Junejo, A.R.; Ullah, I.; Khan, R. Deep learning in cancer diagnosis and prognosis prediction: A minireview on challenges, recent trends, and future directions. *Comput. Math. Methods Med.* **2021**, *2021*, 9025470. [[CrossRef](#)] [[PubMed](#)]
5. Madani, M.; Behzadi, M.M.; Nabavi, S. The role of deep learning in advancing breast cancer detection using different imaging modalities: A systematic review. *Cancers* **2022**, *14*, 5334. [[CrossRef](#)] [[PubMed](#)]
6. Zheng, T.; Xie, W.; Xu, L.; He, X.; Zhang, Y.; You, M.; Yang, G.; Chen, Y. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int. J. Med. Inform.* **2017**, *97*, 120–127. [[CrossRef](#)] [[PubMed](#)]
7. Mahmood, T.; Arsalan, M.; Owais, M.; Lee, M.B.; Park, K.R. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *J. Clin. Med.* **2020**, *9*, 749. [[CrossRef](#)] [[PubMed](#)]
8. Gupta, R.; Kumari, S.; Senapati, A.; Ambasta, R.K.; Kumar, P. New era of artificial intelligence and machine learning-based detection, diagnosis, and therapeutics in Parkinson’s disease. *Ageing Res. Rev.* **2023**, *90*, 102013. [[CrossRef](#)]
9. Lundberg, S.M.; Nair, B.; Vavilala, M.S.; Horibe, M.; Eisses, M.J.; Adams, T.; Liston, D.E.; Low, D.K.-W.; Newman, S.-F.; Kim, J.; et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2018**, *2*, 749–760. [[CrossRef](#)]
10. Divya, P.; Susmita, A.; Sushmitha, P.; Ch, R.; Chandini, K. Automation in pharmacovigilance: Artificial intelligence and machine learning for patient safety. *J. Innov. Appl. Pharm. Sci.* **2022**, *7*, 118–122. [[CrossRef](#)]
11. Lo-Ciganic, W.H.; Donohue, J.M.; Thorpe, J.M.; Perera, S.; Thorpe, C.T.; Marcum, Z.A.; Gellad, W.F. Using machine learning to examine medication adherence thresholds and risk of hospitalization. *Med. Care* **2015**, *53*, 720. [[CrossRef](#)]
12. Huang, Y.; Talwar, A.; Chatterjee, S.; Aparasu, R.R. Application of machine learning in predicting hospital readmissions: A scoping review of the literature. *BMC Med. Res. Methodol.* **2021**, *21*, 96. [[CrossRef](#)]
13. González-Castro, L.; Chávez, M.; Duflo, P.; Bleret, V.; Martin, A.G.; Zobel, M.; Nateqi, J.; Lin, S.; Pazos-Arias, J.J.; Del Fiol, G.; et al. Machine Learning Algorithms to Predict Breast Cancer Recurrence Using Structured and Unstructured Sources from Electronic Health Records. *Cancers* **2023**, *15*, 2741. [[CrossRef](#)] [[PubMed](#)]
14. Alzu’bi, A.; Najadat, H.; Doulat, W.; Al-Shari, O.; Zhou, L. Predicting the recurrence of breast cancer using machine learning algorithms. *Multimed. Tools Appl.* **2021**, *80*, 13787–13800. [[CrossRef](#)]
15. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [[CrossRef](#)]
16. Probst, P.; Boulesteix, A.L.; Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1934–1965.
17. Van Rijn, J.N.; Hutter, F. Hyperparameter importance across datasets. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2367–2376. [[CrossRef](#)]



18. Luo, G. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Netw. Model. Anal. Health Inform. Bioinform.* **2016**, *5*, 18. [[CrossRef](#)]
19. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
20. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40. [[CrossRef](#)]
21. Bakhteev, O.Y.; Strijov, V.V. Comprehensive analysis of gradient-based hyperparameter optimization algorithms. *Ann. Oper. Res.* **2020**, *289*, 51–65. [[CrossRef](#)]
22. Ebrahim, M.; Sedky, A.A.H.; Mesbah, S. Accuracy Assessment of Machine Learning Algorithms Used to Predict Breast Cancer. *Data* **2023**, *8*, 35. [[CrossRef](#)]
23. Kaushik, K.; Bhardwaj, A.; Bharany, S.; Alsharabi, N.; Rehman, A.U.; Eldin, E.T.; Ghamry, N.A. A machine learning-based framework for the prediction of cervical cancer risk in women. *Sustainability* **2022**, *14*, 11947. [[CrossRef](#)]
24. Alfian, G.; Syafrudin, M.; Fahrurrozi, I.; Fitriyani, N.L.; Atmaji, F.T.D.; Widodo, T.; Bahiyah, N.; Benes, F.; Rhee, J. Predicting breast cancer from risk factors using SVM and extra-trees-based feature selection method. *Computers* **2022**, *11*, 136. [[CrossRef](#)]
25. Lou, S.J.; Hou, M.F.; Chang, H.T.; Chiu, C.C.; Lee, H.H.; Yeh, S.C.J.; Shi, H.Y. Machine learning algorithms to predict recurrence within 10 years after breast cancer surgery: A prospective cohort study. *Cancers* **2020**, *12*, 3817. [[CrossRef](#)] [[PubMed](#)]
26. Ganggayah, M.D.; Taib, N.A.; Har, Y.C.; Lio, P.; Dhillon, S.K. Predicting factors for survival of breast cancer patients using machine learning techniques. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 48. [[CrossRef](#)]
27. Massafra, R.; Latorre, A.; Fanizzi, A.; Bellotti, R.; Didonna, V.; Giotta, F.; La Forgia, D.; Nardone, A.; Pastena, M.; Ressa, C.M.; et al. A clinical decision support system for predicting invasive breast cancer recurrence: Preliminary results. *Front. Oncol.* **2021**, *11*, 576007. [[CrossRef](#)] [[PubMed](#)]
28. González-Castro, L.; Cal-González, V.M.; Del Fiol, G.; López-Nores, M. CASIDE: A data model for interoperable cancer survivorship information based on FHIR. *J. Biomed. Inform.* **2021**, *124*, 103953. [[CrossRef](#)] [[PubMed](#)]
29. Quan, H.; Sundararajan, V.; Halfon, P.; Fong, A.; Burnand, B.; Luthi, J.-C.; Saunders, L.D.; Beck, C.A.; Feasby, T.E.; Ghali, W.A. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care* **2005**, *43*, 1130–1139. [[CrossRef](#)]
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. XGBoost. Available online: <https://xgboost.readthedocs.io/> (accessed on 8 April 2024).
32. Tensorflow. Available online: <https://www.tensorflow.org/> (accessed on 8 April 2024).
33. Varma, S.; Simon, R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinform.* **2006**, *7*, 91. [[CrossRef](#)]
34. Le, P.B.; Nguyen, Z.T. ROC curves, loss functions, and distorted probabilities in binary classification. *Mathematics* **2022**, *10*, 1410. [[CrossRef](#)]
35. Fotouhi, S.; Asadi, S.; Kattan, M.W. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *J. Biomed. Inform.* **2019**, *90*, 103089. [[CrossRef](#)]
36. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
37. Rahman, S.A.; Walker, R.C.; A Lloyd, M.; Grace, B.L.; I van Boxel, G.; Kingma, B.F.; Ruurda, J.P.; van Hilleegersberg, R.; Harris, S.; Mercer, S.; et al. Machine learning to predict early recurrence after oesophageal cancer surgery. *J. Br. Surg.* **2020**, *107*, 1042–1052. [[CrossRef](#)] [[PubMed](#)]
38. Belete, D.M.; Huchaiah, M.D. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *Int. J. Comput. Appl.* **2022**, *44*, 875–886. [[CrossRef](#)]
39. Ratul, I.J.; Al-Monsur, A.; Tabassum, B.; Ar-Rafi, A.M.; Nishat, M.M.; Faisal, F. Early risk prediction of cervical cancer: A machine learning approach. In Proceedings of the 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Prachuap Khiri Khan, Thailand, 24–27 May 2022; pp. 1–4. [[CrossRef](#)]
40. Talukder, M.S.H.; Akter, S. An improved ensemble model of hyper parameter tuned ML algorithms for fetal health prediction. *Int. J. Inf. Technol.* **2024**, *16*, 1831–1840. [[CrossRef](#)]
41. Siddiq, M. Integration of Machine Learning in Clinical Decision Support Systems. *Eduvest-J. Univers. Stud.* **2021**, *1*, 1579–1591. [[CrossRef](#)]
42. Kvamme, H.; Borgan, Ø.; Scheel, I. Time-to-event prediction with neural networks and Cox regression. *J. Mach. Learn. Res.* **2019**, *20*, 1–30.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.