**REVIEWER COMMENTS ADDRESSING FORM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Corresponding Author Name** | | K. Luxshi | | |
| **EDAS ID** | | 1571137894 | | |
| **Research Title** | | Enhancing Lung Cancer Prediction Using Machine Learning: A Comparative Analysis of Hyperparameter Optimization Techniques | | |
| **Authors/co-authors** | | Luxshi Karunakaran  Senthan Prasanth  A. M Chandrika Malkanthi  Rathnayaka Mudiyanselage Kapila Tharanga Rathnayaka | | |
|  | | | | |
| **Review No.** | **Review Comment** | | **Addressed Solution** | **Section No. Page No. Line No.** |
| 1 | The specific novel contribution of this work needs to be more explicitly highlighted. Many studies exist on ML/DL for lung cancer prediction and hyperparameter tuning. What makes this comparative analysis stand out? | | The study will be unique among the other DL and ML studies discussing the prediction of lung cancer due to its ability to introduce an extensive system of combining Bayesian optimization with a system of hyperparameter tuning of ML and DL models to predict lung cancer using numerical data. In contrast to the existing studies where the interpretation of imaging data is done mainly and validation schemes are non-complex. This paper will be the first to talk systematically about such different types of cross-validations (hold-out, k-fold, stratified k-fold, leave-one-out) and bring strong new light on the accuracy of the model-specific (the XGBoost model resulted in 0.9968, whereas the Swin Transformer resulted in 0.9369, but with a run-time of 400.94 seconds). The preference of numerical data increases the accessibility of health diagnostics in the low-resource regions, which also has a crucial place in regards to healthcare equity. Moreover, the paper also provides a holistic measurement system that is lacking the literature and which exhibits a positive trade-off between performance and computational demands, not just in accuracy but also in precision, recall, F1-score, AUC-ROC, confusion matrix, and training times (0.33s to train the Logistic Regression and 510.02s to train SVM). This has been substantiated by the fact that the prediction has significantly improved where its hyperparameters have been Bayesian tuned; hence, it is now a scalable and efficient model which is ready to be ingrained into the clinical practice, therefore representing a reference point in the quest to predict lung cancer through numerical data. | Section [2.1],  Page [5],  Line [2] |
| Provide a detailed description of the 16 numerical attributes used, including their types, ranges, and any specific characteristics relevant to lung cancer prediction. | | Added the Dataset description as a table. | Section No [3.1]  Page no [7]  Line no [1] |
| Clearly describe the methods used for data preprocessing. | | In addition to mean imputation of missing values, the two important preprocessing methods in lung cancer prediction are encoding categorical variables and feature scaling. Encoding transforms the categories such as Gender (M/F) and Lung Cancer (YES/NO) into numbers (label encoding) so that they can be compatible with algorithms. When numerical features such as Age are standardized or normalized to a similar range, feature scaling balances model training. These procedures enhance the quality of data and the model. | Section No [3.1]  Page No [8]  Line [3] |
| Explain why ensemble models consistently outperformed other models. | | The superiority of ensemble models such as Random Forest, Gradient Boosting, and XGBoost over other models, including deep learning models such as CNN, MobileNet, and Swin Transformer in the lung cancer prediction study can be attributed to their capacity to aggregate multiple weak learners, which in effect helps to reduce overfitting by enhancing generalization. The models produced almost perfect metrics, having an accuracy of 0.9968, an F1-score of 0.9981 and an AUC-ROC of 1.00 with stratified 5-Fold cross-validation. Their advantage is that they make use of different decision trees and iterative error correction which are able to capture the complex patterns in the data more effectively than the simpler models such as Logistic Regression (accuracy of 0.9463) and GNB (accuracy of 0.9150) which were not able to deal with specificity and class imbalance. Although CNN did it equally poorly (0.9838 accuracy, 0.9927 F1-score), it was slightly outperformed by ensembles, and MobileNet (0.9762 ac curacy) and Swin Transformer (0.9421 accuracy) had issues with computational intensity and over-fitting, respectively, thus ensembles are more predictable and efficient on this work | Section No [5]  Page no [20]  Line No [5] |
| Discuss the implications of the varying training times for practical deployment. | | The varying training times of ML models for lung cancer prediction significantly impact their practical employment. Models like XGBoost and CNN, with training times of 15.41 and 13.66 seconds respectively, are more feasible for real time clinical applications due to their efficiency, whereas SVM’s lengthy 510.02 second training time may hinder its use time-sensitive settings. Balancing high accuracy with shorter training durations is crucial for integrating these models into resource-constrained healthcare environments. | Section No [5] Page No [21]  Line No [6] |
| 2 | better to explain data that they used. And correlation among those parameters. If the target is to improve cancer identification, I think higher accuracy may be achieved by using images. | | Added the Dataset description as a table. | Section No [3.1]  Page no [7]  Line no [1] |
| 3 | Properly cite the dataset. Justify the dataset selection. Include detailed dataset description. | | Added the Dataset description as a table. | Section No [3.1]  Page no [7]  Line no [1] |
| Enhance literature review | | The related work discusses recent developments in ML and DL for predicting lung cancer and the application of such approaches to numerical data. It also examines how the hyperparameter optimization strategy and cross-validation methods are utilized to ensure the developed models are robust and reliable. ML and DL methods have gained rapid deployment in lung cancer prediction over the last decade [3]. Ensemble methods such as GNB, SVM, Logistic Regression, Decision tree, Random forest, Gradient boosting and XGBoost use features such as age, smoking history and symptoms to identify cancerous and non-cancerous patients and the results have been accurate and stable with ensemble techniques in traditional models [4, 5]. DL models such as CNNs, MobileNet and Swin Transformer perform well in extracting complex patterns, whereby CNNs only capture hierarchical features, MobileNet is capable of capturing low-resource efficient tasks, and Swin Transformer models long-range dependencies using an attention mechanism [1]. The transfer learning also enhances the performance, particularly where there is insufficient training data available [6]. Hyperparam eter tuning techniques, including Bayesian optimization have been used to optimize these models by sensitivity adjustment of parameters that improve on model generalization and avoid overfitting (e.g., learning rates, dropout rates). Besides, cross-validation procedures such as 5-fold, stratified 5-fold and Leave One-Out (LOO) are needed to gain information on the robustness of a model to ensure that reliable and precise clinical decision support systems. [7].  However, there are still limitations and challenges in the previous research to predict lung cancer. Homogeneous or small datasets are hardly representative of the diverse populations of patients that make the models less generalizable [2]. The excessive focus on accuracy as an essential indicator might ignore other vital clinical indicators such as sensitivity and specificity in the predisposition to unreliable forecasts. Transformers exhibit computational complexity, requiring large resources, an aspect that increases their limitation to low-resource envi ronments [6, 8]. Features such as poor cross-validation strategies and inability to interpret the model weaken the trust placed in a given model by physicians, particularly due to the risk of overfitting involved [9]. Real-world applications are further complicated by poor integration into clinical workflows and adjustable data preprocessing, including processing of missing values or normalization [10].  The paper applied deep learning to detect lung cancer using a U-Net model to process the histopathological images and ResNet-50, VGG-16, EfficientNet-B5 and an Ensemble model as the models to classify the normal tissues, encompass ing adenocarcinoma of the lung and squamous-cell Carcinoma of the lung. Of all the models tested, the Ensemble model had the best accuracy of 99 and the Effi cientNet-B5 coming in as the second-best with an accuracy of 97. The strategy enhances early diagnosis, where feature extraction is done by automating 15,000 images and making a precision diagnosis [5].  This paper analyzes the process of lung cancer detection involving hybrid feature extraction of not only Gray-level Co-occurrence Matrix (GLCM), Haral ick and autoencoder features, but also using optimized machine learning models. Being very accurate with 99.89, the SVM RBF and SVM Gaussian models with hybrid features and SVM polynomial with single Haralick features obtain high accuracy. The study outlines the prospect of incorporating various features and refined hyperparameters to achieve diagnostic precision in non-small cell lung cancer (NSCLC) precisely and small cell lung cancer (SCLC) [7]. | Section No [2]  Page No [2]  Line No [1,2]  Section [2],  Page No [4],  Line No [6]  Section [2],  Page No [4],  Line No [8] |
| Describe in detail how cross-validation was used in both hyperparameter tuning and final model evaluation | | Both hyperparameter optimization and evaluation of the final model were done using stratified 5-Fold cross-validation with k=5. To hyperparameter tune, the data was stratified into five folds to maintain the balance between classes (YES/NO Lung Cancer) to mitigate the imbalance issue. Each fold was used as a validation set once, and the other four as training data to tune the hyperparameters using Bayesian optimization to maximize the accuracy, and obtained 0.9968 when using XGBoost. To evaluate the final models, stratified 5-Fold cross validation with optimized hyperparameters was used, and such performance measures as accuracy, F1-score (0.9981 in case of XGBoost), or AUC-ROC were averaged across folds to provide the stable generalization on unseen data. | Section No [5]  Page No [20]  Line No [2] |
| Justify the choice of k to 5 | | The choice of k=5 in stratified 5-Fold cross-validation is its compromise between reliability and efficiency. It allows dividing the data into 80 training and 20 validation per fold, which is enough to train on, and it gives stable performance estimates with low variance compared to k=3 and Holdout. k=5 is computation ally efficient in comparison to k=10 and LOOCV. In the case of models such as SVM (510.02s training time) and appropriate to the size of the dataset, as demonstrated by XGBoost rapid 15.41s training time and stable accuracy of 0.9968. Stratified k=5 also preserves the proportion of classes, which is best in terms of reliability with imbalanced data about lung cancer, and thus it is the best option in this study. | Section No [5]  Page 20,  Line No [3] |