

Dear Reviewer, First of all, we would like to explain that the editor requested us to make changes and send the result as a new submission, and we wrongly understood that we don't have to make the document with a detailed answer for every comment. Sorry for having made you to apply unnecessary efforts. And, of course, thank you for your comments. We have carefully reviewed all the points raised, and have made the adjustments and incorporated clarifications in this regard. We hope that these changes will bring the manuscript to the quality required for publication. Thank you very much for your help to improve the quality of the work. Kind regards

Thanks to the authors for the efforts spent in addressing the raised issues. I only retain some concerns about comments 1.2 and 1.3.

The provided reviewed version of the manuscript is significantly improved and can be considered as the final draft. Being so, it is now vital to rework the abstract.

In particular, it is vital to clearly state what is the main output of the study, which (to the best of my understanding, based on authors responses) is to prove that the automatized machine learning approach is surpassing the current human-based performance. This must be stated clearly in the abstract. In order to give the right robustness to this statement, it is equally important to also provide a quantitative assessment, which implies reporting a reference performance indicator value for the human-based estimation and the corresponding performance indicator value for the ML-based model.

Thank you very much for the review, the comments provided and the opportunity to improve the work. Please find attached the description of the changes made, as well as the corresponding clarifications:

We have reworked the abstract, elaborating on how the machine learning-based model surpasses the results obtained by scoring models based on clinical experience. We indicate specific quantifications of an illustrative metric (f-1 score):

This investigation substantiates that automated machine learning models significantly surpass traditional prediction techniques for patient outcomes within critical care settings. Specifically, the machine learning-based model attained an F1 score of 0.351 for mortality forecast and 0.615 for length of stay, in contrast to the traditional scoring model's F1 scores of 0.112 for mortality and 0.412 for length of stay. These results strongly support the advantages of integrating advanced computational techniques in critical healthcare environments.

We perform the detailed deepening and justification in a comparative way by analyzing the results of the intelligent model with respect to the scoring model presented in tables 7 and 8 in section 4.3. and in figure 6 we visually compare both performances.

Regarding point 1.2, we have incorporated the breakdown by age group and sex, as well as indicated the geographic areas of data collection and clinical center location. We have also corrected the poor explanation of data quality.

Regarding point 1.3 we have incorporated the application of the actual case study within the methodology, as well as the comparative analysis with the scoring system.

The typo in Figure 4 has been corrected.

We have included the clarifications in this document, and highlighted the changes in the manuscript.

We hope that the changes will improve the quality of the manuscript, and enable its publication.

Thank you very much for your help to improve the work done. Best regards

ORIGINAL COMMENT

GENERAL I appreciate the efforts made by authors in addressing the issues raised. However, there is a major problem in the provided review: responses are very generic and the direct reference to the new version of the manuscript (punctual indication of parts of the text removed, added or reworked) is not provided. As a consequence, it is very hard to assess if some of the comments were actually solved in the new version of the manuscript. Additionally, I still have concerns about many of the raised issues, which, in my opinion, were not treated with the necessary attention. Specifically:

RESPONSE - R3

Thank you for the comments, and the opportunity to improve the quality of the manuscript. We have made considerable changes to the manuscript, including increasing the size of the dataset, using a single dataset for both models, simplifying the structure of methods and materials, reorganizing the information between results, discussion and conclusions, and, above all, making the section on artificial intelligence more comprehensive, rigorous and clear.

We have tried to respond to comments by indicating, whenever possible, where the manuscript has been updated.

NEW COMMENT - R3

R.3 A punctual indication of parts of the text removed, added or reworked is again not provided in this review, so the main issue still remains. This made the review activity a lot harder to perform.

I appreciate authors effort in addressing some of the raised points, but many of them are still unanswered. Also, there are some additional concerns relevant to the new version.

RESPONSE - R4

We are very sorry for the error made with the lack of indication of the changes made. On this occasion, as the update has been made as a new shipment, the changes with respect to the previous version are indicated. Additionally, we have included the modified text in the attached answers. We have made adjustments according to the indications transmitted, both to the previous comments and to the new comments.

NEW COMMENT - R4

R4. Thank you for your efforts.

ORIGINAL COMMENT

Comment 1.1):

R1. I believe that, regardless the evaluated performance of the proposed model, its use as a primary decision maker is not to be recommended. Until extensive and rigorous on-field assessments are performed, the role for the model indicated by authors should be as a support for human decisionmakers. This is an ethical issue of primary importance. I therefore suggest to re-consider a modify all statements in the text relevant to this matter, (such as, for example, <<The result of a model with these characteristics would provide security in decisions, reducing their subjectivity>>) and to clarify this aspect."

R2. Please indicate punctually the changes (parts of the text removed, parts of the text added) made to address this issue.

RESPONSE - R3

We agree that the AI solution should be supportive, and that the final decision should be made by the responsible clinician. We have indicated this point in the second paragraph of the conclusions.

NEW COMMENT - R3

R3. Thank you for addressing this issue. Were the more assertive parts of the text removed as well?

RESPONSE - R4

We have clarified the indicated sentence, stating that it is indeed a form of support for clinicians, who should be ultimately responsible for any decision (third paragraph of the second page: "A model with these characteristics could potentially enhance safety in decision-making processes, thereby helping to reduce subjectivity, always as a form of support for clinicians in decision-making"). We also point this out in the second paragraph of the conclusions ("It is important to consider that decisions with clinical impact, whether related to patient diagnosis and treatment or patient management, should always be made by the clinicians, who can benefit from the information provided by an intelligent system while using their knowledge and experience, as well as taking responsibility for the decisions they adopt.").

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

Comment 1.2):

R1. In chapter 2 (methods and materials), it is necessary to add, in the introductory section (2.1), the study context of the specific study-case. Mainly, please describe the origin and the features (basic population statistics) of used data, about which nothing at all is reported in the current version of the manuscript. Also, data dimension should be put here rather than in the discussion."

R2. I find it hard to believe that “no missing or inconsistent data were identified” across 22 years of data, dating back to 1999. Unfortunately, I see from data availability statement that authors are not allowed to disclose original data.

RESPONSE - R3

The problems we identified stemmed primarily from the use of natural language processing. We have included that limitation in the third paragraph of the discussion. We have also expanded the volume of the dataset, to include the most recent data.

NEW COMMENT - R3

R3. This does not address the raised issue.

RESPONSE - R4

We have incorporated more specific information about the context in section "3.1. Data Set":

This study uses the historical data from critical patients obtained from a reference clinical center. The data set used in this study covers the totality of patients treated in the critical care unit of a major clinical center located in a large urban area. This large data set includes a diverse patient population representative of the city's broad demographics. Specifically, data related to the clinical information of patients who were hospitalized in the critical care unit from the time prior to admission until discharge are used. Historical data for the 24,876 episodes occurring between 1 January 1999 and 31 December 2023 were available for this study.

NEW COMMENT - R4

R4. The requested information is:

- 1 - Statistics about the patients population in data (e.g. age, sex, etc.)
- 2 - Geographic area in which data are collected
- 3 - Geographic area in which the clinical center is located
- 4 - How is it possible that “no missing or inconsistent data were identified” across 22 years of data, dating back to 1999?

RESPONSE – R5

1. We have incorporated the breakdown by age and gender in section 3.1:

The patient population was categorized by sex and age as follows: 14,141 males, 9,392 females, and 1,343 unspecified. Regarding age, 316 patients were under 20 years; 3,359 patients were aged 20 to 40 years; 7,567 patients fell into the 40 to 60 years range; 11,416 patients were between 60 and 80 years; and 2,218 patients were over 80 years of age.

2. We have indicated the geographic area in section 3.1.
3. We have also indicated the geographical area of the clinical center in section 3.1:

The data were collected from patients located in the geographical area of Madrid, where the clinical center is situated.

4. This is indeed poorly explained on our part. The critical hospitalization episodes with inconsistent and complete data are so after filtering out those records with inconsistent or incomplete data in fields necessary for the elaboration of the model. We have corrected the text and incorporated the clarification.

We have fixed this in section 3.1:

For this study, historical data were available for episodes occurring between January 1, 1999 and December 31, 2023. The data were collected from patients located in the geographical area of Madrid, where the clinical center is situated.

The data were subjected to a comprehensive analysis followed by an exhaustive cleansing process, which was essential to ensure the integrity and accuracy of the dataset. This included a thorough verification of the fields which inclusion is critical to confirm both the completeness and consistency of the processing. Records failing to meet these criteria were excluded from the analysis. Consequently, this rigorous approach yielded a refined dataset comprising 24,876 episodes of critical hospitalizations, each characterized by complete and consistent data suitable for constructing predictive models.

ORIGINAL COMMENT

Comment 1.3):

R1. Methods and materials: on top of describing the goals of the various phases/tasks, it is also necessary to describe how they are carried out in the presented study-case, which is currently reported (at least in part) in the results. I understand that the definition of the workflow is a primary part of the methodology, but so is the actual implementation in the addressed stud-case, whereas the results are the quantitative/qualitative data (or information in general) that is generated through the application of what was implemented (output). Additionally, the baseline reference method for patients' assessment (which is compared to ML model) should be presented in details in this section."

R2. I can't find in the manuscript the content matching authors' response. Please indicate it punctually.

RESPONSE - R3

We have updated the structure of the methodology in order to make it clearer. We have also simplified and made the comparison more balanced by unifying the data source for the comparison of the two models.

The update of the dataset is shown in section 3.1, and we clarify the process of the methodology in the first paragraph of section 3.2.

NEW COMMENT - R3

R3. Despite the changes, this is still a description of a general framework, with minimal reference to the actual study-case implementation on which the study is proposed.

RESPONSE - R4

We have incorporated in the third paragraph of the introduction to section "3. Materials and Methods" the possibility of practical application of the model generated:

The model generated, if the metrics obtained are considered to be sufficiently satisfactory, could be used as a support for decision making by clinical professionals.

NEW COMMENT - R4

R4. The request is mainly (but not only) referred to Figure 1. The framework reported here is general and does not present what was actually implemented for the proposed study. In the new version, this is presented in the results section. However, it is still hard for the reader to build a clear idea of the structure of the implemented framework.

RESPONSE – R5

We have incorporated the application of the actual case study within the methodology (sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4.) We also included in the application of the case study the comparative analysis with the scoring system:

In the specified case study, a systematic identification of pertinent clinical criteria essential for the predictive models was conducted. This process included an extensive compilation of requisite information and the identification of relevant data sources that were utilized throughout the study.

...

In the context of the current case study, the existing data architecture was carefully adapted to facilitate access to and compatibility with clinical data, essential for automating the research processes. This adaptation targeted the comprehensive data processing demands of the developed model, ensuring that all architectural components were suitably aligned to support the investigative objectives.

...

For the case study under consideration, specific technologies were thoughtfully chosen. The design and implementation of robust ETLs were carried out to populate the information model. This model was structured to ensure the continuous availability of updated data, which is crucial for the functioning of machine learning systems.

...

In the case study under review, a series of advanced artificial intelligence techniques were utilized to develop predictive models. This application encompassed the selection of appropriate AI methodologies and their strategic implementation to achieve optimal prediction accuracy.

In addition, a comparative analysis is carried out with a reference method, the scoring system based on the PRLI criteria.

ORIGINAL COMMENT

Comment 1.5):

R1. The features selection process is not described in detail. By reading the manuscript, what I understand is that explicative variables are those reported in table 1 (which are all binary variables), is that correct? Were explicative variables somehow pre-processed and selected? For example, was collinearity between the variables assessed?"

R2. If I understand properly, the variables selection is based on a literature reference, and no further processing is applied. Also due to the fact that this reference is dated 2009, I think that a more rigorous approach would be necessary when it comes to implementing a ML model.

RESPONSE - R3

Indeed, we start from the Philip R. Lee Institute's criteria as the standard for carrying out the analysis. It is possible to locate other criteria, but the one selected is a solid criterion that allows us to realistically contrast both models.

The possibility of a comparative analysis between different criteria could be of interest, although it would exceed the scope of the manuscript.

NEW COMMENT - R3

R3. While the reference proposed by authors is undoubtedly valid, I still think that a more updated and data-driven approach is necessary. Also, the first part of the inquiry was not addressed at all.

RESPONSE - R4

In the second paragraph of the section 4.1, we include the justification for the use of PRLI criteria as a reference:

Concerning the data, we proposed using the PRLI criteria, used in several previous studies that consider data from multiple ICUs as a reference for the assessment of critically ill patients [19, 20]. These criteria are shown in Table 1.

In the third paragraph of the section 4.1, we clarify that it is possible to go deeper into the variables used, although it exceeds the proposed scope:

A deeper analysis of the variables used to assess the impact of each one of them or the collinearity could contribute to improve the model, however, it is beyond the scope of the research and is not essential for the achievement of the proposed objectives.

In the fourth paragraph of the section 4.1, we have also included the specification that the data type is binary in all the criteria used:

The selected criteria indicated in Table 1 represent the relevant information for the assessment of patients using a reduced number of values, all of them of binary type and obtained in an automated way from heterogeneous data sources

Regarding the use of PRLI criteria, there is indeed the possibility of using other alternatives. We have incorporated this as a limitation in the fourth paragraph of the discussion:

Another limitation identified was the use of a single set of criteria (PRLI criteria). Although this is a robust set of criteria, alternative sets of criteria could be used for cross comparison.

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

Comment 1.6):

R1. The validation metrics considered can be consistently improved. For instance, mortality is a binary target variable, therefore I think it is recommendable to perform a ROC-based validation considering AUC as the performance metric. Also in the comparison between traditional and ML model, basing it on the accuracy alone is poorly significant."

R2. I appreciate the efforts made by authors to include the ROC analysis. However, it is not addressed in the methodology, and its description in the results is very poor: it is unclear what model was implemented for, and with what target outcome. Moreover, the obtained result (AUC 0.65) indicates a low performance, discouraging model's application. This is not addressed in the discussion.

RESPONSE

We have made a strong update of the entire artificial intelligence section. We describe the ROC AUC in section 3.2.4, and comment on the limited performance in the second paragraph of the discussion.

NEW COMMENT - R3

R3. I appreciate authors effort in improving this section, which is now consistently more exhaustive.

RESPONSE - R4

Thank you.

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

Comment 1.7):

R1. Methods and materials: << Therefore, it is not considered within the scope of the study to technically deepen the tools used, nor to carry out exhaustive improvements and optimizations.>> even in terms of hyper-parameters tuning? I do agree that there is no need to implement a new algorithm, but at least the basic settings should be (to some extent) optimized. For instance, with a ROC-based validation it would be possible to optimize the threshold of the labelling probability resulting from the ML block, thus improving the performance, as well as setting the dimensioning of the decision trees and random forest. Or again, the classification strategy for length of stay is arbitrary, while a testing protocol could prove which one can enable a higher output performance.

As a matter of fact, the performance is quite poor, especially in predicting LOS (where the reached performance is even lower than the traditional expert-based model).

R2. I don't see any changes made to address this issue. Please point it out clearly. **RESPONSE**

It is correct that the level of detail and optimization done was limited and could be improved. Therefore we have completely rebuilt the artificial intelligence part. We have extended the set of techniques to be used, and performed optimizations and improvements (mainly class balancing and hyperparameter optimization). It is detailed in point 4.4.1

NEW COMMENT

R3. As for the previous issue, I appreciate authors effort in elaborating on this aspect, which is now consistently improved.

RESPONSE - R4

Thank you.

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

Comment 1.8):

R1. I don't understand why the dataset for the traditional model is so limited. To the best of my understanding (although, as already stated in comment 3, its functioning must be explained better) the traditional model is based on a formula applied to the same data structure. Wasn't it possible to label them with the traditional 'manual' method?"

R2. I can't find the part of the manuscript where this issue is addressed. Please indicate it punctually.

RESPONSE

We have completely eliminated the use of the reduced dataset for the traditional scoring model, and we now keep the single large dataset for comparison. We consider it much more fair and balanced.

NEW COMMENT

R3. Thank you for the adjustment, which I deem have improved the validity of the analysis.

RESPONSE - R4

Thank you.

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

Comment 1.9):

R1. All figures and table captions should be (as far as possible) self-explaining. The current form is way too concise."

R2. Implemented changes are too limited to make the captions self-explanatory.

RESPONSE - R3

We have adjusted the names of the figures and tables to make them more descriptive. We have also incorporated additional tables and figures.

NEW COMMENT - R3

R3. I find the current captions still too concise.

RESPONSE - R4

We have increased the depth of the description in the captions of the figures and tables, and this is the result:

Figure 1. Methodological phases of the development of an intelligent model for length of stay and survival in a Critical Care Hospital Unit.

Figure 2. Architecture for information management that contemplates the integration between clinical information systems.

Figure 3. Diagram of the information architecture focused on data exploitation.

Figure 4. Figure 4. Information flows for data processing in the proposed model.

Figure 5. Receiver Operating Characteristic (ROC) curve for the outcome variables of mortality and length of stay obtained in the intelligent model.

Figure 6. Comparison of the performance metrics of the scoring model and the intelligent model.

Table 1. Description and typology of the Philip R. Lee Institute criteria for evaluating patients who are candidates for admission to the ICU.

Table 2. Specifications of the extraction, transformation and loading processes (ETLs) for the proposed model.

Table 3. Metrics of the machine learning algorithms for initial approximation for mortality.

Table 4. Metrics of the machine learning algorithms for initial approximation for length of stay.

Table 5. Metrics of the optimized machine learning algorithms for mortality (class balancing and hyperparameter optimization techniques).

Table 6. Metrics of the optimized machine learning algorithms for length of stay (class balancing and hyperparameter optimization techniques).

Table 7. Metrics obtained using model stacking techniques optimized with class balancing and hyperparameter optimization strategies.

Table 8. Scoring model metrics for the outcome variables of mortality and length of stay obtained in the intelligent mode.

NEW COMMENT - R4

R4. CHECKED - typo in figure 4 (repeated).

RESPONSE – R5

We have corrected it, thank you.

ORIGINAL COMMENT

Comment 1.10):

R1. A limitations section is necessary in the discussion. Among them, please also address how the use of a specific custom NLP model, thus not replicable/extendible, can impact the model development and consequent performance."

R2. The limitations section should be beneficial for the readers, clarifying what are the boundaries of applicability of the presented research. To achieve this, a much more careful assessment should be performed, addressing the punctual issues encountered and clarifying their impact on the applicability, along with some suggestions for possible solutions. The limitations section presented in the current manuscript is minimal, and seems to be put there only to respond to reviewers concerns.

RESPONSE - R3

We have incorporated the limitations in the third paragraph of the discussion, I have included comments on how this point could be improved in the future.

NEW COMMENT - R3

R3. The limitations are now more clearly outlined.

RESPONSE - R4

Thank you.

NEW COMMENT - R4

R4. CHECKED

ORIGINAL COMMENT

2.1) The discussion section is very short and does not provide any insight on the obtained results or their applicability.

RESPONSE - R3

We have reorganized the information and expanded the discussion section.

NEW COMMENT - R3

R3. Despite the inclusion of a limitations section, the discussion is still insufficient, not providing any insight on the obtained results, their relevance, or their applicability. Also, no comparison with stateof-art is provided.

RESPONSE - R4

We have gone deeper into the implications and possibilities of the generated model, presenting it in the third paragraph of the discussion:

"The applicability of the model in clinical settings suggests a significant advance in medical decision support capabilities. This model, which uses advanced machine learning techniques to predict length of stay and survival rates in critical care units, provides healthcare professionals with a tool to improve resource management and patient care. When comparing the efficacy of our model to traditional scoring systems, it is observed to offer superior precision in predicting critical clinical outcomes. This improved precision can be crucial in optimizing hospital resources and care planning, especially in highdemand and resource-limited settings."

We have increased the depth of the limitations section, in the fourth and sixth paragraphs of the discussion, by suggesting future work by going deeper from multi-center data.

"Although the model shows marked improvement in several key performance metrics, it is necessary to recognize the limitations inherent in the single-center study design and the homogeneity of the data set. These factors may limit applicability in different clinical environments. Further studies incorporating multicenter data are essential to validate and refine the predictive model, ensuring that it can work effectively in a broader spectrum of clinical environments"

NEW COMMENT - R4

R4. CHECKED

NEW ADDITIONAL COMMENTS

3.1) Authors state *"In this section it is presented the reason for starting the work presented in this paper, based in the existence of a data set that was created and maintained along many years with the aim of using it in future research but without specific objective or method set."* Does this mean that the goal of the whole study is to put at use a dataset that was previously collected with no actual goal? This seems approximative from a scientific point of view.

RESPONSE - R4

The phrasing we used is not very clear, and may indeed give the impression that the reason for the research is to use a previous data set that had no purpose.

However, this is not the case: the data were maintained for research purposes, although the specific purpose of generating a predictive model for critically ill patients was not envisaged at the time.

In any case, the accumulated data set contains a large number of detailed clinical variables that allow a thorough and rigorous analysis, essential for the predictive model, to be carried out. The use of existing data is a common practice in research, especially in fields where data collection is costly or complicated. Thus, this approach favors an efficient use of resources already invested.

We have corrected the wording to clarify this point in the first paragraph of section 3, resulting as follows:

"In this section, we first introduce the dataset. The research is based on a comprehensive dataset that has been collected and maintained over several years in a clinical setting. Although these data were not initially gathered with a specific research objective in mind, their depth and breadth provide a remarkable opportunity for developing advanced predictive models that can significantly impact the care of critical patients."

NEW COMMENT - R4

R4. CHECKED

3.2) Authors state *"Although the restriction to a single center implies a more limited data set, it also increases the homogeneity and reliability of the data analysis."* I disagree with this statement. A more homogeneous data set is more likely subject to overfitting, and this reduces the reliability of any AI-based model when applied to different data.

RESPONSE - R4

You are right, and we appreciate your comment. Indeed, restricting the dataset to a single center does increase the risk of overfitting. While we address this risk using various techniques, we acknowledge that the second paragraph of section '3.2. Methodology' was inaccurate and has been removed. This change provides us with an opportunity to enhance the description and clarification in the third paragraph of section 5, where we discuss the limitations:"

"During the development of the study, methodological limitations were identified due to the dataset being restricted to a single center, which makes it susceptible to overfitting. To enhance the model's generalizability and reduce the risk of overfitting, specific measures were implemented, including the use of resampling techniques, the generation of synthetic samples, and hyperparameter optimization through grid search. Another limitation is the use of natural language processing techniques, which inevitably lead to erroneous interpretations of some terms (e.g., abbreviations). More rigorous data collection by the center is required to increase the reliability of the model, especially the use of standardized nomenclatures."

NEW COMMENT - R4

R4. CHECKED

3.3) Authors report *"The process that is carried out is as follows: A first approximation is made using various machine learning techniques. Optimizations are then performed on each of these models, obtaining results that allow discerning those with the best metrics. Among these models, a stacking model is proposed to obtain the most solid model for the prediction of both outcome variables."* What do authors mean by 'stacking model'? Is it the series of different computational algorithms, the parallel, or else? This requires elaboration to state that the final prediction could benefit from this approach.

Additionally, this disagrees with the previous statement *"It is not necessary to obtain the optimal predictive model, but rather to verify an improvement using machine learning models in relation to score systems. Therefore, it is not considered within the scope of the study to technically fine-tune the tools used, or to carry out exhaustive improvements and optimizations."*

RESPONSE - R4

In the manuscript, the term "stacking model" refers to an ensemble learning technique where predictions from multiple machine learning models are combined to form a new model. This is achieved by using a meta-classifier that learns how to best combine the predictions from the base models, which are typically trained in parallel. The objective is to leverage the strengths of various predictive models to achieve better accuracy and generalizability than any individual model alone.

Regarding the apparent contradiction you highlighted, we acknowledge that our phrasing may have caused some confusion. When we stated that "it is not necessary to obtain the optimal predictive model," we intended to emphasize that the primary goal of our study was to demonstrate a clear improvement in predictive performance over traditional scoring systems using machine learning techniques, rather than achieving the absolute best possible model through fine-tuning. However, we do utilize techniques such as hyperparameter optimization and model stacking to ensure that our approach is robust and yields practical improvements in prediction accuracy.

To clarify these points, we have incorporated the following additions:

Fifth paragraph of point "3.2.4 Artificial Intelligence":

"The stacking model is an ensemble learning technique. This approach combines the predictions of multiple base machine learning models using a meta-classifier. The base models are trained in parallel, and the meta-classifier learns to optimally combine their predictions to improve the final predictive accuracy and generalizability over any single model"

Fourth paragraph of point "3.2.4 Artificial Intelligence":

The primary objective is to demonstrate significant enhancements over traditional scoring systems, rather than achieving the definitive optimization of the model. To this goal, tuning techniques such as hyperparameter optimization and model stacking are employed to enhance model performance. While it is not intended to exhaustively fine-tune each parameter, the application of these methods is essential to ensure that the machine learning models are accurate enough to represent a practical and meaningful improvement.

NEW COMMENT - R4

R4. CHECKED