Peer review

1.I do not agree with your word “causal”; especially since you have chosen to explicitly differentiate it from the word “correlational”. What the model does - double machine learning notwithstanding - is no different from what most statistical models do when correcting for confounding variables, i.e. restriction, randomization, matching, stratification, and for multavariate models (logistic regression and ANCOVA). All these methods can do is report results as a predictor variable (not cause) after correcting for confounders. The semantics of causation in the biomedical sciences imply that mechanisms (such as signaling pathways) are understood. In these cases, a ‘chain-of-causation’ can also apply. I will need discussion on why you think your method finds causative factors - rather than predictive factors. Please explain with relevant references in the manuscript. I strongly suggest that you replace the word ‘causal’ and ‘causation’ with something like ‘predictor variables’ or ‘predicting factors’. I will accept those substitutions.

2.Additionally, as you state in the manuscript, you have used heuristics to differentiate between covariates and treatment variables. This itself implies a correlational genesis, rather than a mechanistic causal inference; as I have described in point 1.

3.Section 2.4.2 has some sentence repeatability. Please check. Please also explain more in the text how the multiplier of -0.022 was obtained. Does this coefficient have any units?

4.The paragraph above figure 4 lists some treatment variables and some confounding variables. However, you state earlier that the program used 249 features and that you used PCA to decrease these features to a lesser number. What was this lesser number and can you list the number and titles of features that were used after the PCA analysis in the manuscript. I am familiar with PCA using components - which are themselves comprised of ‘pieces’ of individual factors. How exactly did you then retain the features themselves with the most predictive power from the PCA analysis? Please provide a description of the method used to achieve this. Please also provide the original 249 features as a list in a supplementary file.

5.Please explain the legend of Figure 4 with actual regressions. For example, show the residual plot for the confounders against Riluzole, then the residual plot of the confounders against ALSRFS. If actual plots are not available, please show the progression of how the RF and/or XGBoost actually acomplished this. An example is needed so that the readers can readily replicate your research. Why were two regression models used for XT while only one was used for XY ?

6.Again, in figure 5, I cannot accept that the DL model uncovers causal relationships (as distinct from correlational relations). These are again, Granger predictive or forecasting variables and not necessarily causative to the progression of the disease. Investigators have produced copious lists of blood biomarkers as predictors of ALS progression; that does not mean that all of these are necessarily causative (see as just one example, https://onlinelibrary.wiley.com/doi/pdf/10.1155/2014/525097) Most of the supposed biomarkers you state are causative for ALS are actually generic electrolytes, metabolic waste products, pH stabilizers or related to fuel utilization and protein binding. I would hazard a guess and state that these ‘predictive’ set of biomarkers would predict many more diseases in terms of their progression rather than ALS alone.

7.you state “….Lower serum albumin levels are associated with a slower decline in ALS scores, as reflected by the negative treatment effect of -0.18 (95% CI: -0.29 to -0.08, p = 0.001). Given that albumin is a marker of nutritional status, this finding suggests that better nutritional health may slow

ALS progression…..” If albumin is protective, should not HIGHER albumin levels be associated with slower decline? See: https://doi.org/10.1001/jamaneurol.2014.1129, again notice how in this reference, the authors use the words “significantly associated with ALS outcome” and do not mention “cause”.

8.Table 1, the row header “gastrointestinal issues” is too broad. Either define what these ‘issues’ are specifically as a footnote to the table or leave the row out.
9.You state “…..The study also highlights the role of electrolyte imbalances in ALS. Elevated bicarbonate levels are associated with faster disease progression, with a treatment effect of 0.60 (95% CI: 0.42 to 0.77, $p < 0.001$), potentially indicating respiratory issues…..” This seems to be a consequence of ALS, not a predictive variable. Please define whether you consider the consequences of ALS progression also as predictor variables in the manuscript. Again, it is extremely difficult to deconvolute cause-from-consequence, as decades of wasted effort on the amyloid cascade hypothesis has demonstrated. Please therefore, remove all words such as ‘cause’ and ‘consequence’ and ‘effect’ and replace them per point 1. Similarly blood glucose utilization is reduced in ALS (see: https://doi.org/10.4103/1673-5374.335154) so that assigning this as a ‘cause’ of ALS is erroneous.

10.Please perform a thorough search of the literature so that you can align your positive or negative treatment effect for each of these predictive factors with the literature. See for example, the references I mention in points 7 and 9.

11.You state “……Electrolytes, particularly chloride, bicarbonate, and calcium, also play crucial roles in ALS progression……” Please cite this paper https://doi.org/10.3389/fnins.2018.00510 with discussion in the manuscript. Your paper does reinforce the intriguing possibility that rigorous continuous fine tuning of blood electrolytes may slow down ALS and this needs discussion in the manuscript with pertinent references.
In summary, I think this paper has significant merit but the assignment of causality is highly questionable. If this claim were changed along the lines suggested in points 1 and 2 throughout the manuscript, and the other concerns addressed appropriately, I would be happy to review this manuscript again.

Dr. Apte: I do not agree with your word “causal”; especially since you have chosen to explicitly differentiate it from the word “correlational”. What the model does - double machine learning notwithstanding - is no different from what most statistical models do when correcting for confounding variables, i.e. restriction, randomization, matching, stratification, and for multavariate models (logistic regression and ANCOVA). All these methods can do is report results as a predictor variable (not cause) after correcting for confounders. The semantics of causation in the biomedical sciences imply that mechanisms (such as signaling pathways) are understood. In these cases, a ‘chain-of-causation’ can also apply. I will need discussion on why you think your method finds causative factors - rather than predictive factors. Please explain with relevant references in the manuscript. I strongly suggest that you replace the word ‘causal’ and ‘causation’ with something like ‘predictor variables’ or ‘predicting factors’. I will accept those substitutions.

Berke: Added the following paragraphs to clarify the definition of Causal:
“In this manuscript, causality refers to the quantitative predictive value of a blood biomarker on the quantitative progression of Amyotrophic Lateral Sclerosis (ALS). The notion of causality enables the analysis of how a system would respond to an intervention, providing actionable insights into potential treatment effects by predicting treatment outcomes including efficacy and toxicity (Feuerriegel et al., 2024). This approach is supported by the framework of statistical causality, which emphasizes estimating causal effects and predicting treatment outcomes in a systematic manner (Pearl, 2009).

However, this notion of causality does not—and cannot—single-handedly differentiate between correlation and causation in the context of molecular biology, where causation implies a knowledge of the mechanism of action (Russo & Williamson, 2007). In molecular biology, causation involves a signaling pathway where molecular conformational or chemical changes lead to a pharmacological effect. The complexity of these pathways, along with pleiotropy—where one element can impact multiple physiological processes—and redundancies built into the system, means that causation typically encompasses a chain of causative events. This intricate network ensures system robustness but makes it challenging to identify a singular cause for an effect. In this context, causative events can be necessary or sufficient to produce the outcome, with the progression from cause to effect being mechanistically understood, such as through well-characterized molecular pathways.
Recent literature, including the "Evidence Based Medicine +" approach, emphasizes the importance of integrating clinical research with mechanistic evidence as described by the Russo and Williamson Thesis (Williamson, 2019). This combination provides a more comprehensive understanding of causality, moving beyond traditional evidence-based methods to include the explanatory power of mechanisms (Khan et al., 2012)."

Berke: I also expanded the introduction to provide readers with additional context on machine learning-based causal studies. This added background aims to clarify the principles of ML-based causal inference and its applications, helping readers understand how these methods are utilized in the current study. This added section precedes the discussion on the distinction between statistical causality and causation in molecular biology.
"In recent years, machine learning-based causal inference techniques have gained prominence in healthcare research, particularly for addressing complex, high-dimensional datasets. Sanchez et al. (2022) and Feuerriegel et al. (2024) underscore the role of Double Machine Learning (DML) and other causal inference methods in these contexts, illustrating how they yield robust insights that support decision-making in clinical settings. Since ALS progression is influenced by multiple, interrelated factors, causal inference methods are valuable for identifying biomarkers that may alter disease trajectory. While several causal inference methods exist, DML has emerged as a particularly powerful tool in healthcare research for its ability to adjust for confounding factors and isolate treatment effects. This approach is especially effective in high-dimensional datasets, where relationships between variables can be complex and multi-layered.
DML has demonstrated its versatility across various healthcare applications, enabling researchers to analyze high-dimensional data while adjusting for confounding influences. For instance, it has been used to evaluate the effects of the Women, Infants, and Children (WIC) program on infant health outcomes, allowing researchers to identify impact variations based on risk profiles (Peet et al., 2022). In oncology, DML has informed pharmacogenomic research by assessing epithelial-mesenchymal transition (EMT) signatures in cancer cell lines, thereby revealing specific drug susceptibilities and guiding precision treatment strategies (Ohnmacht et al., 2024).
These examples highlight DML's flexibility in healthcare settings, where complex datasets and accurate treatment effect estimation are essential. By applying DML to blood biomarker data, this study similarly aims to advance our understanding of ALS progression, offering a statistically rigorous foundation for targeted interventions."

Dr. Apte: Additionally, as you state in the manuscript, you have used heuristics to differentiate between covariates and treatment variables. This itself implies a correlational genesis, rather than a mechanistic causal inference; as I have described in point 1.

Berke: I added the following text to the Data Analysis section to explain:

“

2.3.1 Feature Engineering:

The initial phase began with feature engineering, expanding the Pro-ACT dataset from 147 columns to 415 columns by constructing additional variables relevant to ALS progression. These newly constructed features capture the time series aspects of the biomarkers, such as measurements at different time points and changes in values over specific periods.

2.3.2 Data Preprocessing:

*Following the feature engineering, univariate analysis was conducted to explore each feature independently, examining distributions, central tendencies, and variability. This step provided crucial insights into the data, enabling the identification of outliers and ensuring the quality of features for subsequent analysis. Data cleaning activities are employed such as removing features with high percentages of missing values, imputing missing data, and handling outliers. These processes ensured the dataset was of high quality and ready for subsequent analysis**.** In this step, features with more than 30% missing values were removed, resulting in a feature list of 249 columns.*

2.3.3 Feature Selection and Dimensionality Reduction:

There are nuanced perspectives on the need for dimensionality reduction in Double Machine Learning (DML). Chernozhukov et al. (2018), in the original paper introducing DML, highlight its capability to handle high-dimensional data effectively without requiring dimensionality reduction. Nonetheless, some studies suggest that reducing the feature set can improve model performance and interpretability, particularly in cases with complex, high-dimensional datasets where noise or redundancy might impact results. Carefully selecting variables and tuning algorithms can significantly enhance the accuracy and robustness of causal effect estimation (Fuhr et al., 2024; Angrist and Pischke, 2009; Cinelli et al., 2022). Therefore, in this analysis, feature selection was conducted progressively, using PCA, bivariate analysis, stepwise regression, Random Forest, and XGBoost.

In this study, absolute PCA loadings were used to identify the most important features. For each Principal Component, features with the highest absolute loadings were retained, as these loadings indicate the features that contribute most to the variance captured by each component. This approach preserves individual features with the strongest predictive power, allowing the model to remain interpretable—unlike using principal components directly, which could obscure relationships between original features and the outcome (Chahboun & Maaroufi, 2021; Wuttichaikitcharoen & Babel, 2014). Selecting features based on PCA loadings is less common than using the components themselves but is especially beneficial when interpretability is a priority, as is often the case in healthcare applications (Habehh & Gohel, 2021; Lipton, 2018). By retaining features with the highest PCA loadings, the resulting model maintains the biological interpretability of individual biomarkers, which is crucial in clinical settings.

Following PCA, bivariate analysis was performed to investigate relationships between the outcome variable and each potential covariate and treatment variable. This was followed by stepwise regression and Random Forest analysis, which further refined the feature set to approximately 75 columns, reducing noise and potential collinearity. Finally, XGBoost was applied to evaluate feature importance among these 75 features, resulting in a final selection of around 34 features for DML analysis. This structured approach balanced model performance with interpretability, ensuring that the retained features provided diverse, biologically relevant information, minimized redundancy, and reduced collinearity—a key factor in improving stability and performance.

2.3.4 Confounders and Covariates:

In terms of the DML framework, all features in this final list were used as both covariates and treatments. This approach is rooted in DML’s robustness with high-dimensional data and its capacity to handle complex inter-variable relationships (Chernozhukov et al., 2018). Using all selected features as covariates and treatments ensures that confounders are accounted for, minimizing bias in the

estimation of causal effects. This aligns with a key premise of DML: its ability to manage high-dimensional datasets and correct for confounders to provide unbiased causal estimates. By including all relevant features in the model, we reduce the risk of missing confounders that might otherwise distort the results."

Dr. Apte: Section 2.4.2 has some sentence repeatability. Please check.

Berke: Deleted the following paragraph:
"The final phase of the methodology involved interpreting the results through a Mixed-Effects Linear Regression Framework. This framework was utilized to analyze the outcomes, extending insights into the broader implications for patient life span and quantifying the treatment effects on survival time. By incorporating mixed-effects models, the study accounts for both fixed and random effects, offering a more nuanced understanding of how treatments impact ALS progression across different patient subgroups. "

Dr. Apte: Please also explain more in the text how the multiplier of -0.022 was obtained. Does this coefficient have any units?

Berke: It is unitless. I added a paragraph to the paper regarding the coefficient:
"In the mixed-effects model, the coefficient for the variable "days" is -0.022, which represents the slope of the ALSFRS score over time. This value is unitless because it reflects a rate of change in ALSFRS score per day, quantifying the average decline in functional status without requiring additional units. Specifically, it indicates that, on average, the ALSFRS score decreases by 0.022 points per day, showing a gradual decline in function over time as ALS progresses."

Dr. Apte: The paragraph above figure 4 lists some treatment variables and some confounding variables. However, you state earlier that the program used 249 features and that you used PCA to decrease these features to a lesser number. What was this lesser number and can you list the number and titles of features that were used after the PCA analysis in the manuscript. Please also provide the original 249 features as a list in a supplementary file.

Berke: Created list of features and provided as supplementary file
ALS_Study_Data Field.xlsx

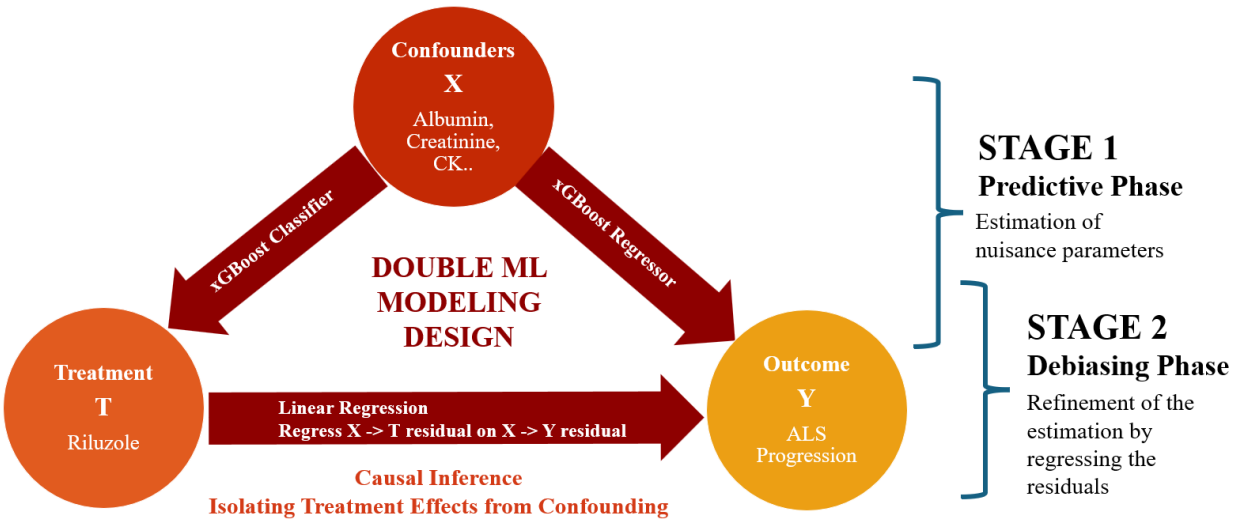
Dr. Apte: I am familiar with PCA using components - which are themselves comprised of 'pieces' of individual factors. How exactly did you then retain the features themselves with the most predictive power from the PCA analysis? Please provide a description of the method used to achieve this.

Berke: Added more context on using PCA approach as feature selection:" *In this study, absolute PCA loadings were used to determine the most important features, with the top features within each Principal Component selected to represent the most diverse set of features. PCA loadings measure the importance of each original feature in contributing to the formation of the principal components. Specifically, features with the highest absolute loadings were retained, as these loadings indicate which features contribute most to the variance captured by each Principal Component. This approach ensures that individual features with the most predictive power are preserved, allowing the model to remain interpretable, unlike using the principal components directly, which would obscure the relationship between original features and the outcome (Chahboun & Maaroufi, 2021; Wuttichaikitcharoen & Babel, 2014). At the conclusion of the data analysis step, the feature set was streamlined to approximately 100 features. This selection process ensured that the retained features*

provided diverse information across the data, while minimizing redundancy. Importantly, this method also helped minimize collinearity, which is a common issue in model building when predictor variables are highly correlated. Reducing the number of features using PCA helped to diminish potential correlations among predictors, thereby enhancing the stability and performance of the resulting non-tree-based machine learning models. While tree-based models (such as decision trees and random forests) are generally robust to multicollinearity, non-tree-based models—such as linear regression or support vector machines—can be adversely affected by it. Therefore, reducing collinearity is particularly valuable for these types of models to improve stability and performance.The approach of selecting features based on PCA loadings qis less common compared to directly using principal components but is particularly useful when interpretability is a priority. This is especially critical in healthcare applications, where understanding how specific features impact predictions is essential for decision-making and establishing trust in the model's outcomes (Habehh & Gohel, 2021; Lipton, 2018). By retaining features with the highest PCA loadings, the resulting model maintains the biological interpretability of individual biomarkers, which is crucial in clinical settings"

Dr. Apte: Please explain the legend of Figure 4 with actual regressions. For example, show the residual plot for the confounders against Riluzole, then the residual plot of the confounders against ALSRFS. If actual plots are not available, please show the progression of how the RF and/or XGBoost actually acomplished this. An example is needed so that the readers can readily replicate your research. Why were two regression models used for XT while only one was used for XY ?

Berke: I updated the visual and the caption and removed the Random Forest Regressor. The example in the picture is regarding the Riluzole usage and I used XGBoost Classifier as Riluzole usage is binary. For treatments that are continuous, I used Random Forest Regressor.

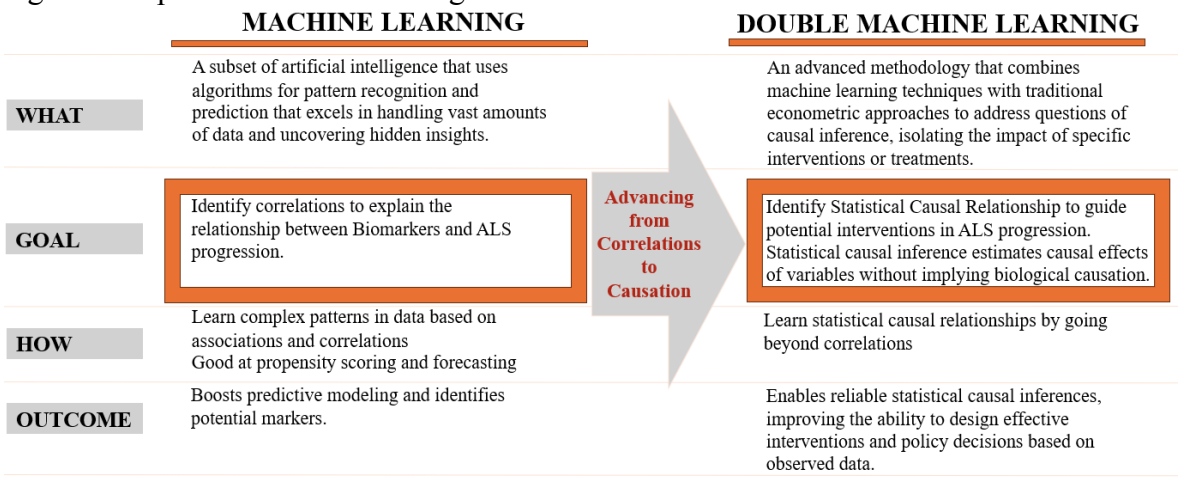


***"Figure 4:** Double Machine Learning (DML) framework applied to ALS research with Riluzole usage as the binary treatment variable. The figure illustrates the two-stage DML process to isolate the treatment effect on ALS progression. In Stage 1 (Predictive Phase), confounders such as Albumin, Creatinine, and Creatine Kinase (CK) are regressed on both the binary treatment variable (Riluzole usage) using an XGBoost Classifier and on the outcome (ALS progression) using an XGBoost Regressor. In Stage 2 (Debiasing Phase), the residuals from the predictive phase are then regressed to refine the estimation of the treatment effect of Riluzole usage on ALS progression, as measured by*

changes in ALSFRS scores. This approach helps to isolate the treatment effect from confounding influences, improving the accuracy and reliability of causal inference."

Dr. Apte: Again, in figure 5, I cannot accept that the DL model uncovers causal relationships (as distinct from correlational relations). These are again, Granger predictive or forecasting variables and not necessarily causative to the progression of the disease. Investigators have produced copious lists of blood biomarkers as predictors of ALS progression; that does not mean that all of these are necessarily causative (see as just one example, https://onlinelibrary.wiley.com/doi/pdf/10.1155/2014/525097) Most of the supposed biomarkers you state are causative for ALS are actually generic electrolytes, metabolic waste products, pH stabilizers or related to fuel utilization and protein binding. I would hazard a guess and state that these 'predictive' set of biomarkers would predict many more diseases in terms of their progression rather than ALS alone.

Berke: I changed the caption and content of Figure 5:



***"Figure 5.** Comparison of Machine Learning (ML) and Double Machine Learning (DML) approaches in the study of ALS progression. Traditional ML methods focus on identifying correlations between biomarkers and ALS progression, uncovering patterns that can suggest potential markers. However, these correlations do not imply causation. DML, by combining machine learning with econometric techniques, moves beyond correlation to enable statistical causal inference. This statistical approach isolates causal relationships, allowing for more robust predictions on how interventions may impact ALS progression. Thus, DML strengthens the causal insights derived from observed data, informing more targeted and effective treatment strategies."*

Dr. Apte: you state "….Lower serum albumin levels are associated with a slower decline in ALS scores, as reflected by the negative treatment effect of -0.18 (95% CI: -0.29 to -0.08, p = 0.001). Given that albumin is a marker of nutritional status, this finding suggests that better nutritional health may slow ALS progression….." If albumin is protective, should not HIGHER albumin levels be associated with slower decline? see: https://doi.org/10.1001/jamaneurol.2014.1129, again notice how in this reference, the authors use the words "significantly associated with ALS outcome" and do not mention "cause".

Berke: This was a typo. Considering the negative treatment effect finding in the analysis. I updated the text in the paper. *: "Higher serum albumin levels are associated with slower ALS progression. "* Thank you for the link to the paper. The insights of my study aligns with Chio's as they also found higher albumin levels were associated with better survival outcomes, hence likely slower ALS progression.

I referenced Chio's article you shared the link for in the discussion section:
"The role of albumin in ALS has been explored in several studies, including Sun et al. (2020), Hertel et al. (2022), Chiò et al. (2014), and Gentile et al. (2023), with varying conclusions. Hertel et al. (2022) found a correlation between albumin levels and ALSFRS scores but did not establish a significant link to disease progression. In contrast, Sun et al. (2020) and Gentile et al. (2023) reported that lower albumin levels were associated with reduced survival in ALS patients. Additionally, Chiò et al. (2014) showed higher albumin is linked to better survival outcome. This study provides further evidence that higher albumin levels are associated with slower ALS progression, offering a more comprehensive understanding of albumin's protective role in the disease."

Dr. Apte: Table 1, the row header "gastrointestinal issues" is too broad. Either define what these 'issues' are specifically as a footnote to the table or leave the row out.

Berke: Added the footnote for Gastrointestinal Issues:
"Patients admitted to the hospital or with health records indicating any of the following conditions are marked as having gastrointestinal issues: abdominal hernias, anal and rectal conditions, benign and malignant neoplasms, dental and gingival conditions, diverticular disorders, exocrine pancreas conditions, gastrointestinal hemorrhages, inflammatory conditions, motility and defecation disorders, stenosis and obstruction, ulceration and perforation, vascular conditions, malabsorption disorders, oral soft tissue conditions, peritoneal and retroperitoneal issues, salivary gland conditions, and tongue conditions."*

Dr. Apte: You state "…..The study also highlights the role of electrolyte imbalances in ALS. Elevated bicarbonate levels are associated with faster disease progression, with a treatment effect of 0.60 (95% CI: 0.42 to 0.77, $p < 0.001$), potentially indicating respiratory issues….." This seems to be a consequence of ALS, not a predictive variable. Please define whether you consider the consequences of ALS progression also as predictor variables in the manuscript. Again, it is extremely difficult to deconvolute cause-from-consequence, as decades of wasted effort on the amyloid cascade hypothesis has demonstrated. Please therefore, remove all words such as 'cause' and 'consequence' and 'effect' and replace them per point 1. Similarly blood glucose utilization is reduced in ALS (see: https://doi.org/10.4103/1673-5374.335154) so that assigning this as a 'cause' of ALS is erroneous.

Berke: I replaced the statement and noted that there might be circular relationship here:
~~*"The study also highlights the role of electrolyte imbalances in ALS. Elevated bicarbonate levels are associated with faster disease progression, with a treatment effect of 0.60 (95% CI: 0.42 to 0.77, $p < 0.001$). potentially indicating respiratory issues*~~.
The study highlights the role of electrolyte imbalances in ALS, showing that elevated bicarbonate levels are associated with faster disease progression (treatment effect of 0.60, 95% CI: 0.42 to 0.77, $p < 0.001$). This association may reflect a complex, circular relationship: as ALS progresses, respiratory function may decline, leading to elevated bicarbonate levels, which in turn could accelerate disease progression. While the DML approach strengthens our ability to control for confounders, it cannot fully disentangle this potential feedback loop. Further investigation, including controlled studies, is needed to clarify the directional influence in this association."

Dr. Apte: Please perform a thorough search of the literature so that you can align your positive or negative treatment effect for each of these predictive factors with the literature. See for example, the references I mention in points 7 and 9.

Berke: I referenced the papers in point 7 (discussed in point 7 above) and 9. Below in red is where Lerskiatiphanich paper is referenced from Point 9:
"Higher blood glucose levels, which Sun et al. (2020) associated with increased mortality, and elevated cerebrospinal fluid glucose levels, correlated with faster ALS progression as identified by Gray et al. (2015) and Wuolikainen et al. (2016). This study confirms the causal role of glucose in ALS, specifically demonstrating that higher blood glucose levels causally contribute to faster ALS progression. These findings are consistent with those of Ahangaran et al. (2022), who also identified a causal association between glucose levels and ALS progression.
Additionally, Lerskiatiphanich et al. (2022) highlight systemic disruptions in glucose metabolism in ALS, noting increased glucose uptake and glycogen accumulation, alongside glucagon resistance, in the presence of ALS, further supporting the link between impaired glucose homeostasis and ALS progression."

Dr Apte: You state "……Electrolytes, particularly chloride, bicarbonate, and calcium, also play crucial roles in ALS progression……" Please cite this paper https://doi.org/10.3389/fnins.2018.00510 with discussion in the manuscript. Your paper does reinforce the intriguing possibility that rigorous continuous fine tuning of blood electrolytes may slow down ALS and this needs discussion in the manuscript with pertinent references.

Berke: I added the red text in the discussion section where Electrolytes are discussed and referenced the paper you shared the link for

"*Electrolytes, particularly chloride, bicarbonate, and calcium, also play crucial roles in ALS progression. While previous studies (e.g., Stambler et al., 1998; Qureshi et al., 2008; Ong et al., 2017; Manera et al., 2023) identified correlations between these electrolytes and ALS outcomes, this study leverages DML to show that elevated bicarbonate and calcium levels accelerate ALS progression, while higher chloride levels are protective. Mineral balance, particularly calcium and phosphorus, further emerges as a critical factor in ALS progression. Previous research (Katzeff et al., 2020; Gordon and Lerner, 2019) explored their roles, but the DML framework applied here provides stronger evidence that higher levels of these minerals drive faster disease progression. These findings underscore the importance of monitoring these biomarkers in ALS patients and exploring them as potential therapeutic targets.*
These findings underscore the importance of monitoring blood electrolytes and mineral levels in ALS patients, as they may serve as valuable therapeutic targets. This perspective is supported by Sirabella et al. (2018), who emphasize that disruptions in ionic homeostasis—including critical ions such as calcium, sodium, and iron—play a significant role in ALS pathophysiology. Their work highlights the potential of targeting ionic channels and transporters to restore balance as a therapeutic approach. Taken together, these insights suggest that a strategy of carefully regulating electrolyte levels, particularly bicarbonate, calcium, and chloride, could offer promising avenues for managing ALS progression."

Dr Apte: In summary, I think this paper has significant merit but the assignment of causality is highly questionable. If this claim were changed along the lines suggested in points 1 and 2 throughout the manuscript, and the other concerns addressed appropriately, I would be happy to review this manuscript again.

Other conversations with Dr. Apte:
Dear author,
I will be happy with an introductory paragraph that reads something along the lines below

"Causality, as described in this manuscript refers to the quantitative predictive value of a blood biomarker on the quantitative progression of disease (4). In this definition, the notion of causality thus enables analysis of how a system would respond to an intervention (5); thereby providing actionable insight on predicting the treatment outcome. This definition therefore does not – and cannot – differentiate between correlation and causation in terms of molecular biology; where causation necessarily implies a knowledge of the mechanism of action; typically along a signaling pathway; where a series of molecular conformational or chemical changes produce a pharmacological effect. The complexity of these pathways and the pleiotropy and redundancies built into biological systems means that causation can encompass a chain-of-causative-events; rather than a one-step mechanism. In this context, causative events may be necessary and/or sufficient to produce the effect. Either way, the progression of cause to effect is mechanistically understood."

Reference 4: The notion of causality thus enables analysis of **how a system would respond to an intervention**.
Reference 5: Causal machine learning (ML) offers flexible, data-driven methods **for predicting treatment outcomes** including efficacy and toxicity, thereby supporting the assessment and safety of drugs. In addition, please address all the other points in the review.
Best,

Dear Dr. Apte,
Thank you for your detailed feedback on my manuscript. I appreciate your emphasis on differentiating correlation and causation in the context of biomedical research and your concerns regarding the use of causal language. I will compile all the answers to your questions and submit the revised manuscript by November 3rd.
As I work on addressing your feedback, I would appreciate the opportunity to discuss my perspective and seek clarification on Feedback 1 and Feedback 2 to ensure that my revisions align with your expectations.

Feedback 1: Addressing the Use of "Causal" Language
You noted that causality in the biomedical sciences often implies an understanding of underlying biological mechanisms. While I acknowledge this perspective, I believe it is important to differentiate between biological causality and statistical causality. In my study, the use of Double Machine Learning (DML) aims to address confounding and estimate causal effects within a statistical framework. As described by Chernozhukov et al. (1), DML leverages machine learning techniques provide unbiased causal estimates even in the presence of high-dimensional data and complex relationships. DML has gained popularity in the industry over the last couple of years as well (2, 3).
There are also publications that refer to promising Causal ML techniques in healthcare (4, 5), and a comparison of different causal ML models (6).
Correcting for confounders using methods like traditional regressions helps reduce bias and focus on the effect of a predictor. However, this doesn't automatically establish causality. These methods can improve the accuracy of results by reducing confounding effects, but demonstrating something as truly causal requires more than just statistical adjustments—it also depends on study design and key assumptions. (publication 6 touches this)
Advanced methods like Double Machine Learning (DML), Propensity Score Matching, Quasi-control, or Instrumental Variables aim to go beyond simple associations by leveraging additional assumptions or external information to establish causality. However, I want to acknowledge that these methods also come with their own set of assumptions.

I also acknowledge that randomized experiments are the gold standard for establishing causality, which I noted in my manuscript. However, achieving a fully randomized study is especially challenging in low-prevalence diseases like ALS. Studies often have fewer than 100 patients, making it difficult to run a randomized experiment and control for all individual differences. Therefore, my recommendation in the manuscript conclusion is to conduct randomized clinical trials to test the insights derived from my DML research. This approach would allow researchers to narrow down the features to study, aiding both the setup of a randomized experiment and the attainment of statistical significance in a shorter time. Here is what I noted in my manuscript: *"While the DML approach provides promising insights into ALS progression, further validation through clinical trials is critical. The causal relationships identified here are based on retrospective, observational data, and future studies must validate these findings prospectively in more diverse ALS populations."*
After reviewing your feedback, I revisited the papers I had referenced in the field of medicine and examined how similar methodologies are used in these studies (7,8, 9, 10).

Based on this, I plan to address your feedback with the following revisions:I will revise the manuscript to clarify that DML is used to estimate causal effects statistically, rather than establishing biological causation.I will modify the language to state that "biomarker A is strongly correlated with ALS progression based on Average Treatment Effect estimates from the DML causal inference model," emphasizing that the findings represent statistical associations with a focus on the treatment effect.
Feedback 2: Differentiating Covariates and Treatments
In the manuscript, I did not use heuristics to differentiate between covariates and treatment variables. Instead, I conducted a bivariate analysis to investigate the relationships between the outcome variable (ALS progression) and the features available in my dataset. The feature list derived from the PCA and bivariate analysis formed the foundation of the Double ML model. In terms of covariates and treatments, I used all the features as both confounders and covariates. The rationale for this is rooted in DML's robustness with high-dimensional data and its ability to handle complex relationships between variables. This approach aligns with one of the key premises of the DML algorithm, as noted by Chernozhukov et al. (1), which is the algorithm's capacity to handle high-dimensional datasets while effectively correcting for confounders to provide unbiased causal estimates. By doing this, I aimed to avoid missing confounders that might bias the results.

Based on this, I plan to address your feedback with the following revisions: Refine the Explanation of Feature Selection: I will enhance the manuscript to more clearly describe the rationale behind using bivariate analysis and PCA for feature selection and how these features were used as both covariates and treatments in the model.
Emphasize the Methodological Assumptions: I will include a discussion of the assumptions made in this approach and provide additional justification for including all selected features as potential confounders, in line with the literature on DML's robustness.
Sincerely,
Berke Yilmaz

References in the text above:
Double Machine Learning is introduced by Chernozhukov et al. with the following paper. The premise of the paper is DML is robust and work with large amount of features.
1 https://doi.org/10.1111/ectj.12097
Double Machine Learning based causal inference methods are widely used in Economics, as well as industry in companies like Amazon and Microsoft. As a matter of fact, the EconML Phyton package I used for my research is built and shared by Microsoft Research.

2 https://assets.amazon.science/88/6c/67945ccf40e7af9e63f3ed99d539/double-machine-learning-at-scale-to-predict-causal-impact-of-customer-actions.pdf
3 https://learn.microsoft.com/en-us/azure/machine-learning/concept-causal-inference?view=azureml-api-2
The following papers refer to Double ML technique and other new Causal inference ML techniques and its role and importance in healthcare research
4 https://royalsocietypublishing.org/doi/10.1098/rsos.220638
5 https://www.nature.com/articles/s41591-024-02902-1
(arxiv version of the paper: https://arxiv.org/pdf/2410.08770v1)
The following paper compares some causal inference machine learning models and OLS.
6 https://pmc.ncbi.nlm.nih.gov/articles/PMC6863230/
Examples of usage of DML and other Causal ML techniques in ALS:
7 Wang, C. (2021). The Application of Double Machine Learning Onto Genomics Data Associated with Amyotrophic Lateral Sclerosis. In Massachusetts Institute of Technology, *Massachusetts Institute of Technology*. https://dspace.mit.edu/bitstream/handle/1721.1/139444/Wang-cwang506-meng-eecs-2021-thesis.pdf?sequence=1&isAllowed=y
8 Ahangaran, M., Peimankar, A., Wiil, U. K., & Chio, A. (2022). Impact of laboratory biomarkers on ALS disease progression: a probabilistic causation approach. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *11*(1). https://doi.org/10.1007/s13721-022-00386-y
Examples of usage of DML in Healthcare research:
9 Variation in the infant health effects of the women, infants, and children program by predicted risk using novel machine learning methods https://onlinelibrary.wiley.com/doi/full/10.1002/hec.4617
10 The pharmacogenomic assessment of molecular epithelial-mesenchymal transition signatures reveals drug susceptibilities in cancer cell lines
https://www.biorxiv.org/content/10.1101/2024.01.16.575190v1.full.pdf

---

Thank you for addressing my comments. Accepted.