The paper examines the tensile properties of Carbon Fiber Reinforced Cementitious Matrix (FRCM) composites using a LASSO regression model. The study is based on 107 uniaxial tensile test samples.

Impact of Data Size The dataset consists of 107 observations, which is relatively small for machine learning applications. The authors acknowledge this limitation, stating that it constrains the use of more complex models. The use of LASSO regression, which is well-suited for smaller datasets due to its feature selection capability, is justified in this context. However, a larger dataset could improve the model's generalizability and reduce the risk of overfitting. Additionally, as discussed in the Outliers section, removing features may lead to suboptimal model performance. One way to mitigate this risk is by increasing the number of observations to compensate.

Handling of Missing Data The study explicitly addresses missing values, employing imputation for missing independent variables. While mode imputation is a simple and commonly used method, it may not always be the best approach, particularly if the missing data mechanism is not random; especially since the paper does not go into detail about why the data is missing as this could provide some input about the material behavior that is important for the model.

Treatment of Outliers The authors use Cook's distance to identify and remove influential outliers from the dataset. This method is effective in ensuring that extreme values do not unduly influence the regression model. The study does not explore the impact of outlier removal on the results, nor does it discuss potential biases introduced by eliminating these observations. A sensitivity analysis on outlier treatment could strengthen the findings. For example, running the models with and without the outliers to compare performance can increase confidence in the results of the paper, seeing that with a limited dataset, removing these outliers can impact performance.

Evaluation of Results The study finds that coating of fibers, support systems, and loading speed significantly impact tensile strength. The regression models achieve a reasonable fit (R squared values ranging from 0.42 to 0.92), suggesting that despite the dataset's limited size, meaningful results are obtained. However, the lower R squared values for some models indicates that additional variables or a larger dataset might improve the model's accuracy. It is unknown whether or not the process of handling missing data and outliers would increase or decrease the R squared values and should be something that the authors reevaluate.

Conclusion The paper utilizes LASSO regression to analyze FRCM tensile properties. While the handling of missing data and outliers is sound, the study would benefit from a larger dataset and further exploration of imputation and outlier treatment methods to ensure the decisions taken in this experiment were the best. Overall, the findings align with expectations, but additional validation with more data would improve confidence in the paper.

I agree with the findings of the paper, in its use of LASSO regression. However, the study could benefit from further validation with a larger dataset to confirm the reliability of the conclusions. The authors acknowledge that data size limits the complexity of models used, which I agree

with. Additionally, while their approach to handling missing data and outliers is reasonable, a deeper exploration of these factors could improve the confidence of the findings.