**JEI Editor’s Letter**

**Manuscript:** Machine Learning and Data Shapley Data Selection to Predict Cirrhosis Stages

**Authors:** Ethan Daniel Taylor, Joseph Chai, Joyce Zheng, Juliana Maria Gaitan, Paulos Waiyaki, Tara Maria Salli, David Jose Florez Rodriguez

**Scientific Reviewer 1:** A graduate student at the University of Connecticut conducting research on polymer nanocomposites.

**Scientific Reviewer 2:** A graduate student at Texas A&M University whose research focuses on cellular treatments for neurodegenerative diseases.

**Scientific Reviewer 3:** A scientist at MilliporeSigma focused on cell and general workflow tools.

**Scientific Reviewer 4:** A graduate research assistant at the University of Hawaii in electrical engineering.

**Decision: Accept pending Presentation Changes**

**General Comments from the JEI Editors and Scientific Reviewers**

This section contains general comments written by both the JEI editors and the Reviewers. These comments do not contain any required/recommended revisions but instead are meant to convey the Editor’s/Reviewer’s opinion on the manuscript as a whole.

**JEI Editor Team**

Great job on this work! You demonstrated a great understanding of machine learning concepts which is a valuable skill to develop this early in your career! Good job applying this to an important problem and presenting some nice findings towards predicting cirrhosis stage.

**Scientific Reviewer 1**

Great job on this manuscript! You present some nice data which could be strengthened by further explanation.

**Scientific Reviewer 2**

Interesting paper! I made some recommendations to help with presenting your work more clearly and in the scientific writing style.

**Scientific Reviewer 3**

Congratulations on such a wonderful collaboration and technically challenging project. There was a strong effort placed on such technical model work. Everyone should be commended for such in-depth understanding of AI driven models. The discussion could be stronger, and some sections should be adjusted for a broader audience, with these improvements, this would make a great journal article for JEI.

**Scientific Reviewer 4**

Really impressive work here. Likely the most polish scientific paper I've reviewed in this journal. This was a pleasure to read. Coming from high school students, the level of research is above and beyond your ages. I'm glad to see such a large group for a research paper. Most JEI papers have one and maybe two student authors. Science is often done in groups and from what you've written, you seem to have learn a lot with the topic as well as working together. Machine Learning for medical diagnosis is a booming research area right now. As you guys have likely learned, accurate neural networks require very large datasets. Your dataset was relatively small, but your results were still very impressive. It is very easy to just present numbers and results for machine learning research, but it's important to provide insights and takeaways with every result. This will greatly strengthen the paper.

**Presentation Comments**

You presented some really exciting results and did a great job rationalizing them! You should be proud of this work! Spending a little bit of time on adjusting the presentation of these results will really help your reader understand and appreciate what you did.

**Required Changes**

We believe that these revisions must be made to publish in JEI. If any changes are impossible, please include an explanation in your cover letter.

**Manuscript-Wide:**

Conventionally ReLU is written in this manner rather than Relu.

**Summary:**

Great job on this summary! This abstract has all the key components that should be included but is a bit wordy. It would be helpful if you could look at where there is extraneous detail or how many sentences have coordinating conjunctions (for example: and) as this may help you spot overly wordy sentences.

If you are going to include Not a Number (NaN) in the summary it should be fully spelled before abbreviation can be used. You did define it later in the paper but it should be defined the first time you use it.

**Introduction:**

Very good intro! You do a great job introducing the motivation/application of this project! As JEI caters to a broad audience and range of topics it is likely that your readers may not have machine learning experience. It would be really helpful if you discussed some basic machine learning concepts in the introduction so every reader can understand and appreciate your results.

Additionally, your introduction should include the rationalization behind why you chose the specific AI models and how they compare to potential others that have been used to tackle the same issue of identifying cirrhosis. More detail on this would be awesome!  
  
**Results:**

Make sure you’re presenting any data you talk about in the discussion in the results section first. The results section is a presentation of all of your findings. The discussion section is where you elaborate and postulate why you might have gotten this data and how you can improve your models for next time. [1] (David) Please add the data you mention in your discussion section to your results section as well.

It would be helpful if you clarified why the accuracy of the other 180 [2](Ethan) models was not shown in the manuscript.

In the results there is a reference to figure \_. Was the intention to make a figure or a table[3]?(paulos)

Figures 3 and 4 should be introduced in the results section, not the discussion. [4](Ethan)  
  
**Discussion:**

You do a nice job of providing reasoning behind some of your results! This section would be strengthened if you implemented this structure/considered these questions:   
  
1. Summarize main findings, hypothesis outcome etc[4](paulos)  
2. Compare your results to key literature, accounting for  
differences, usually due to methodologic strengths and limitations[5] (Joseph)  
3. Demonstrate validity of study and findings. Identify and key  
limitations that could have biased study and show how you were  
able to mitigate their effects[6](paulos)  
4.Implications, significance, and generalizability of your findings?[7] (Joseph)  
5. What is the take-away message?[8] (Joseph)

It would be helpful for your reader to understand your results if you further explained some findings such us why MedNan, categorical, Relu model with four layers is the best, and why MedNan performed better than ZeroNan. [9](Ethan)

Your work suggests that when we have a large dataset (n larger than 200) the data selection you used is very difficult and complex. It would be good to discuss this as a limitation as this could be a restriction in applying this to other data sets. [10](Ethan)

**Methods:**  
Would tweak your section on “Why data selection” and “What is Data Shapely”. The methods section is like a recipe book where you list out what you did so that readers can reproduce your work if they want. You can potentially split these sections where the explanation goes into the introduction and how you used it goes into the methods section.[11](paulos)

This would be a good place to define the Relu, Tanh, and Sigmoid functions.[12](paulos)

**Recommended Changes**

We have also compiled a list of recommended changes. These are not required for publication, but we strongly encourage you to consider them. These revisions will further improve your manuscript and show you examples of good scientific writing.

**Results:**Removing the footnotes and placing them in the methods, discussion, or describe them when they appear in the text would be helpful for the flow of this work.[13] (Joseph) **Discussion:**A section on how you could improve your models for the next experiment/what would you do next would be awesome.[14]

**Figure Comments**

This section contains comments on the presentation of the data in figures and tables.

**Required Changes**

We believe that these revisions must be made to publish in JEI. If any changes are impossible, please include an explanation in your cover letter.

Lengthening Figure 3 would make it easier to see what is true and what your algorithms predicted. You may also change the type of plot so this is more clear. [15](paulos)

Similarly, a flowchart explaining the different models and algorithms would also really help your reader understand what’s going on.[16]David

Please number and write a description for very figure after figure 4 as well. [17](paulos)

A correlation constant highlighting the statistical similarities between the predicted and true data for the regression data would be helpful to include. There cannot be a statistical significance per say, but what about a correlation constant? Any statistical analysis should be included in the figures.[18]David

Similarly a statistical correlation with the bar charts comparing the different models would be helpful to include and should be referenced in text. [19]

Overall, make sure to explain what your figures are showing both in text and in the captions.

**Recommended Changes**

We have compiled a list of recommended revisions that would help further improve the figures. These recommended revisions are not required for publication, but we strongly encourage you to consider them. These revisions will further improve the data visualization and aesthetics of your figures.

You represent percentage two ways on their y-axes which may be confusing for your reader. It would be helpful to choose just one. [20](paulos)