Revisions for the thesis of Jonathan Feige

This document contains a list of revisions requested by Jonathan’s MS thesis committee and advisor. For each revision request, a description of specific changes made to address the request is provided. Items changed are indicated by section number in thesis. Requested revisions are numbered and are ordered by chapter and section of the thesis.

I would like to thank all committee members for their helpful comments.

1. Follow TAD template.
   * Thesis has been incorporated into TAD template and all guidelines have been followed including table of contents and list of figures / tables.
     + Template link:
     + <https://www.ohio.edu/sites/default/files/sites/graduate/files/TAD/2021.December%20Template%20Updates/Russ%20College%20of%20Engineering%20and%20Technology%20-%20APA%207%20Revised%2012-21.docx>
     + Format Guidelines:
     + https://www.ohio.edu/sites/default/files/sites/graduate/files/TAD/2020-Document-Format-Checklist\_3.doc.pdf
   * All figures and tables are left aligned
2. Follow NIH guidelines for a specific aims page.
   * Introduction modified to fit NIH specific aims
     + First paragraphs Chapter 1
3. Improve literature review / related work.  A reader should understand your position in the landscape of your research.
   * Expanded introduction and background section as follows:
     + Included connection Between CIMP and Chemotherapy.
       - Chapter 1.2
     + Expanded DNA Methylation section as follows:
       - Inclusion of DNMT’s their function and relationship to hypermethylation
         * Chapter 1.1
       - DMNT and their relationship to CIMP
         * Chapter 1.1
       - Hypermethylation relationship to mutations and cancer
         * Chapter 1.1
     + CpGs Island Methylator section
       - Added description and requirement of CpGs.
         * Chapter 1.2
       - The relationship between CpG’s and gene promoters
         * Chapter 1.2
       - More description on obtaining methylation data
         * Chapter 1.2
     + Research hypothesis and aims section
       - Added major objective and goal statement
         * Chapter 1.3
4. Lay introduction to help CS people understand the context
   * Emphasizing relationship between methylation and cancer to provide a better picture of the problem
     + Chapter 1.1
   * CpG island’s broken down into more basic topic
     + Chapter 1.2
   * The aims relate back to computer Science concepts
     + Chapter 1.3
5. Move new figures presented in thesis defense to document.
   * Presentation Figures have been implemented into the document
     + Figure 2
       - Figure was altered to include CIMPi group
       - Chapter 1.3
     + Figure 5
       - Figure was added to show all for statistical graphs in a single image
       - In the text the graphic is used to show that there is a strong correlation between CIMP and the mutations in cancer samples with high accuracy regardless of the classification model used.
       - Chapter 2.2
     + Figure 6
       - Figure was updated to match figure 5
       - Chapter 2.2
     + Figure 8
       - The figure depicts the intersection of three mutational selectors and the mutations at the intersection of all three selectors
       - The text uses this image to show prominent mutations that act as drivers for classification between mutations and CIMP
       - Chapter 4.1
     + Figure 9
       - This figure depicts the modified flow chart for the validation classification.
       - The text uses this image to show the new steps in the pipeline and describe the differences between the datasets.
6. Elaborated on the methods used to build the classification models
   * More repeatable methods
     + Chapter 2.2
   * Added python libraries for classifiers and cross fold validation
     + Chapter 2.2
7. Your references should refer back to original works, not the most recent work (that can go elsewhere in the related work).
   * MLP and SVM references have been adjusted to reflect the original citation.
   * MLP:
     + Altered to reflect original publication
     + F. Rosenblatt, *Recent work on Theoretical Models of Biological Memory*. Cornell University, 1967.
     + Chapter 2.1
   * SVM:
     + Altered to reflect original publication
     + C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
     + Chapter 2.1
   * Corrections:
     + Citation 2: S.-W. Jiang, J. Li, K. Podratz, and S. Dowdy, “Application of DNA methylation biomarkers for endometrial cancer management,” *Expert Review of Molecular Diagnostics*, vol. 8, no. 5, pp. 607–616, 2008.
     + Chapter 1
8. Motivate choices for mutation selectors, etc.
   * Justifying selection of 10 fold cross validation due to the performance of the classifiers that used it. (Compared between 3, 5, 10)
     + Chapter 2.1
   * Justifying using random forest over other classifiers due to the additional insights on mutations using gini importance, which is unique to the random forest
     + Chapter 2.1
     + Chapter 2.2
9. More thorough Association Rule mining discussion needs to be provided.
   * Added python library name
     + Chapter 4.1
   * Preformed additional association rule mining form figure 6, table 2, and table 3 mutations.
     + Chapter 4.1
10. Chapter 5 added to show aim 3
    * Showing mutations and their relevance to cancer
    * Showing association rule mining mutations and relationship to cancer
    * Showing Endometrial KEGG pathway
11. Similarly, at the end, in the conclusion, reiterate your distinct contributions to the field.
    * Future Steps and Conclusions were separated
      + Chapter 6
    * Expanding the next steps to include topics from presentation
      + Chapter 6
12. Moved discussion sections adjusted and moved to end discussion and next steps
    * Chapter 6
13. Add a "future work" / "next steps" section.  For ideas of where things should go next (or things you ran out of time to do).
    * Next steps section added as 5.1
      + Chapter 6
    * Conclusions section added as 5.2
      + Chapter 6