Group X Progress Report:
My Group's Project Name

First Author, Second Author, Third Author
{macid1,macid2,macid3}@mcmaster.ca

1 Introduction

Here, write a brief introduction to the problem you are solving. This can be adapted from your problem description and motivation from the original proposal. This should be around 0.25-0.5 pages.

2 Related Work

Here, talk about the related work you encountered for your approach. Cite at least 5 references. Refer to item 2. No one has done exactly your task? Write about the most similar thing you can find. This should be around 0.25-0.5 pages.

3 Dataset

You should write about your dataset here, following the guidelines regarding item 1. This section may be 0.5-1 pages. Depending on your specific dataset, you may want to include subsections for the preprocessing, annotation, etc.

4 Features

Describe any features you used for your model, or how your data was input to your model. Are you doing feature engineering or feature selection? Are you learning embeddings? Is it all part of one neural network? Refer to item 2. This may range from 0.25 pages to 0.5 pages.

5 Implementation

Describe your model and implementation here. Refer to item 4. This may take around a page.

6 Results and Evaluation

In our evaluation stage, we use stratified K-fold cross-validation. The reason for this choice is that our dataset contains only 918 patient records, and using a fixed train/validation/test split may lead to overfitting due to the small number of data points.

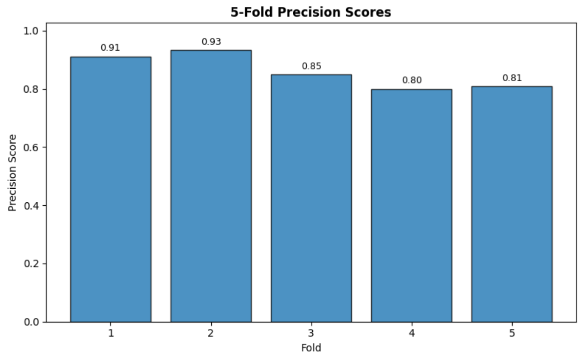

Figure 1: 5-fold cross validation demonstrating precision (Type II error) of each fold

```
===== K-Fold Cross Validation Results =====
Accuracy : 0.8616
Recall   : 0.8979
Precision: 0.8610
F1-score : 0.8773
===========================================
```

Figure 2: Summarized metrics across 5 folds

The metrics used in this project include accuracy, recall, precision, F1-score, and AUC–ROC. Our rationale for choosing these metrics is as follows. First, in disease prediction, precision (closely related to Type II error) is especially important. Hospitals cannot afford to misclassify a patient with the disease as healthy. Therefore, minimizing false negatives is crucial. In addition, we use the AUC–ROC curve because it is an effective way to evaluate the overall performance of a binary classifier.

7 Feedback and Plans

Our TA (Alex Abrehforoush) gave us two main pieces of feedback. First, in data preprocessing, our team converted several nominal categories into numbers using label encoding. For example, the four chest-pain types were encoded as 0, 1, 2, and

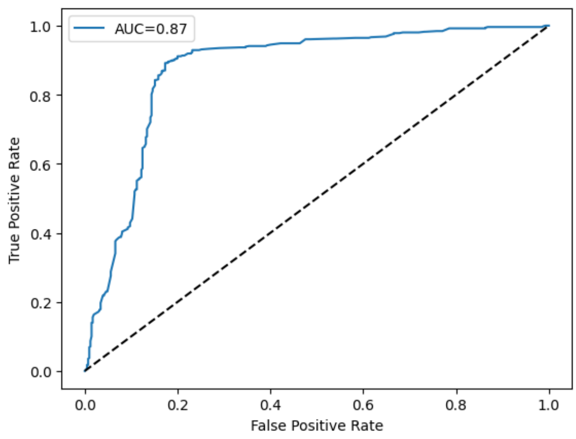

Figure 3: AUC–ROC curves across 5 folds

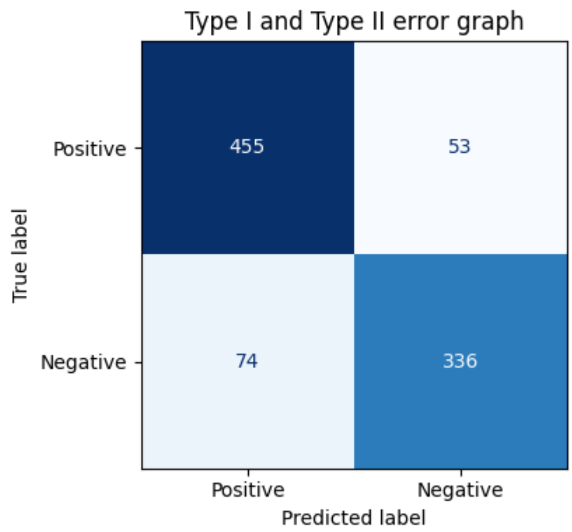

Figure 4: Confusion matrices across 5 folds

3. The TA suggested using one-hot encoding instead to examine whether the model performance improves.

The second suggestion concerns model performance: since the RBF-SVM model achieved only 86

Following the TA's suggestions, we will first implement one-hot encoding for the nominal categories and compare the performance. Afterward, our group will create a new GitHub branch named neural-network to experiment with a different model architecture.

8 Template Notes

You can remove this section or comment it out, as it only contains instructions for how to use this template. You may use subsections in your document as you find appropriate.

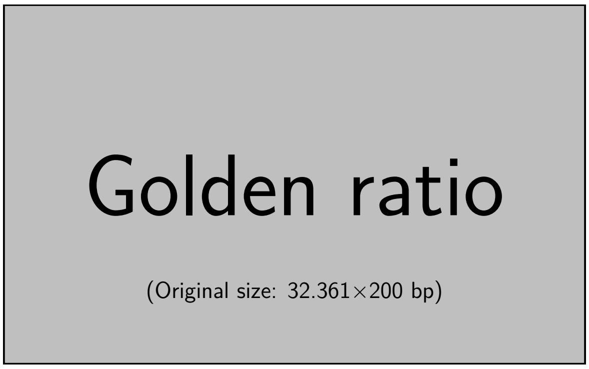

Figure 5: A figure with a caption that runs for more than one line. Example image is usually available through the `mwe` package without even mentioning it in the preamble.

8.1 Tables and figures

See Table 1 for an example of a table and its caption. See Figure 5 for an example of a figure and its caption.

8.2 Citations

Table 1 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get "author (year)" citations, like this citation to a paper by Gusfield (1997). You can use the command `\citep` (cite in parentheses) to get "(author, year)" citations (Gusfield, 1997). You can use the command `\citealp` (alternative cite without parentheses) to get "author, year" citations, which is useful for using citations within parentheses (e.g. Gusfield, 1997).

8.3 References

Many websites where you can find academic papers also allow you to export a bib file for citation or bib formatted entry. Copy this into the `custom.bib` and you will be able to cite the paper in the LaTeX. You can remove the example entries.

8.4 Equations

An example equation is shown below:

$$A = \pi r^2 \tag{1}$$

Labels for equation numbers, sections, subsections, figures and tables are all defined with the `\label{label}` command and cross references to them are made with the `\ref{label}` command. This an example cross-reference to Equation 1. You can also write equations inline, like this: $A = \pi r^2$.

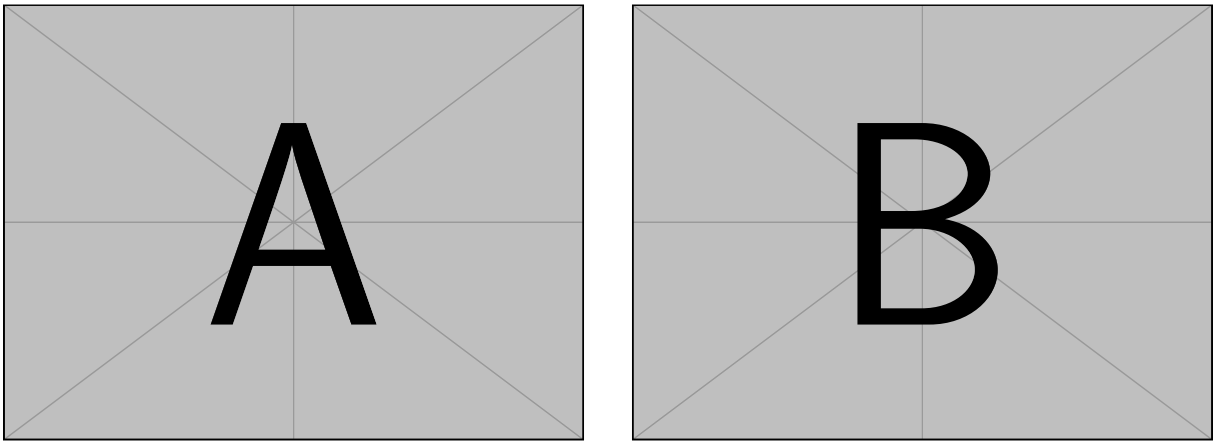

Figure 6: A minimal working example to demonstrate how to place two images side-by-side.

Output	**natbib command**	**ACL only command**
(Gusfield, 1997)	`\citep`	
Gusfield, 1997	`\citealp`	
Gusfield (1997)	`\citet`	
(1997)	`\citeyearpar`	
Gusfield's (1997)		`\citeposs`

Table 1: Citation commands supported by the style file.

Team Contributions

Write in this section a few sentences describing the contributions of each team member. What did each member work on? Refer to item 7.

References

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.

Mohammad Sadegh Rasooli and Joel R. Tetreault. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. Version 2.