Clinical heart failure data visualisation

Alessia Marcolini, Bence Fazekas and Harish Sekar
Group 59

1 Introduction

Worldwide more than 26 million people suffer from heart failure every year. When the clinicians are unable to identify the cause of heart failure, they use endomyocardial biopsy (EMB) which is the most common evaluation of the disease. However, there is high inter-rater variability in the interpretation of the EMB, due to human evaluation. To improve this process, there are ongoing research studies to support interpretation and quantification of Whole Slide Images (WSI) with Deep Convolutional Neural Networks [4]. We aim to support the exploration of the current models and data sets with different visualisation techniques.

1.1 Problem Description

In this article, we show the relations between the patient's demographics data (gender, age, ethnicity) and the outcome predicted by a deep-learning classifier, which aims at identifying patients with clinical heart failure using H&E-stained WSI [6]. We are aiming to explore the UMAP projection [3] of the features extracted from the classifier in an interactive scatter plot. Also, we are interested in the relation of UMAP-projected features and demographic data. For example, we can show how a particular type of ethnicity is clustered concerning a disease subtype. Additionally, we are showcasing differences and similarities between ethnicities, age groups and clinical outcomes.

2 Data Analysis

2.1 Domain Data Specification

For the purpose of this assignment, we first start with a dataset of tiles extracted from 209 WSI of the left ventricular tissue, each corresponding to a single patient. From each WSI, eleven non-overlapping images or regions of interest are extracted at 5× magnification. The whole collection of 2,299 tiles is publicly available on the Image Data Resource repository[1] (IDR number: idr0042). Slides are subdivided into images of heart failure (HF) (n = 94) and those of non-heart failure (non-HF) (n = 115). Slides in the former category are further grouped according disease subtypes: ischemic cardiomyopathy (n = 51); idiopathic dilated cardiomyopathy (n = 41); undocumented (n = 2). Patients with no heart failure are categorized into: normal cardiovascular function (n = 41); non-HF and no other pathology (n = 72); non-HF and other tissue pathology (n = 2).

The training dataset has been used to train a Convolutional Neural Network for feature extraction, i.e., to learn a vector representation of the data. Full experimental settings can be found in [1]. The training features are used to learn a 2-dimensional representation using the UMAP unsupervised multidimensional projection method [3], which is used to project both the training features (1824 samples × 979 features) and the test features (448 samples × 979 features) on a bi-dimensional space (`n_neighbors` = 40; `min_dist` = 0.01; `n_components` = 2; `metric` = "euclidean"; `random_state` = 42).

Ultimately, the set of data used for this Information Visualization assignment consists of original tiles, training and test UMAP embeddings, patient demographics and clinical information.

[1] `http://idr.openmicroscopy.org`

2.2 Data Abstraction: What

Our data is composed of images with their associated tabular metadata, organized as follows: Dataset Partition (binary: *Training/Test*) - Image Name (*Text*) - Diagnosis (binary: *chronic heart failure/not chronic heart failure*) - Disease Sub-type (categorical: see Section 2.1) - Sex (categorical: *Male/Female/Unknown*) - Ethnic or Racial Group (categorical: *African American/Caucasian/Hispanic/Unknown*) - Age (continuous: *Years*) - Patient Id (*categorical: Number*) - Clinical History (categorical: see Section 2.1).

For each of the 2,299 images, we have the corresponding demographic and clinical information, enriched with the UMAP embedding of the respective feature vector.

We filtered the dataset by selecting only those rows which contained a value for gender and age, thus removing 22 rows (corresponding to two patients) from the dataset.

Moreover, we have transformed the content of the Age column, as it contained the value as a string (for example: "69 years"). We removed the unnecessary text and converted the values to numbers.

Thereafter, we were able to use most of the original dataset (2,277/2,299 rows).

3 Task Analysis

3.1 Domain-Specific Tasks

We aim to discover relations between the patient's data and the outcome predicted by a deep learning classifier. Therefore, we are identifying eight different tasks.

1. How are the UMAP projections related to the heart tissue images, for example, are there any recognizable features within close items?
   To answer this question we plot a two-dimensional UMAP projection of the features extracted from the classifier and show the image of the heart tissue.

2. Are there any visible clusters with respect not only to the outcome (HF/non-HF) but also to demographics data?
   To discover these relations we are linking the UMAP scatter plot to the other plots.

3. Are patients further clustered according to disease subtypes?
   To show this relation we link the UMAP to the plots which are considering the subtypes.

4. Is there a racial group/gender which is getting the heart transplant earlier/later?
   To accomplish this task we are showing the patients' gender and their age of heart transplant.

5. Is there an age group more propitious to get one specific subtype?
   To answer this question we are going to show the age of the patients and their related subtype.

6. Does the ethnicity have an impact on the heart failure rate?
   To give some insight to the experts we are showing the distribution of HF vs non-HF patients among the racial groups.

7. Does the gender have an impact on the heart failure rate?
   To provide this information we are showing the distribution of the patients' gender, grouped according to HF/non-HF.

8. Is there a racial group more propitious to get one specific subtype? For this question we are showing the distribution of the subtypes between the ethnicities.

3.2 Task Abstraction: Why

High-Level abstraction
On this level, we would like to provide tools to discover relations between the features of the data. For example, in Task 1 our goal is to show images which are considered to be "close" by the deep learning classifier due to visual similarity, possibly finding sub-clusters which could be exploited by clinicians in order to offer better-targeted treatment plans. Moreover, our tool could be used to communicate the results to a not-necessary expert commission to help their decision-making process. For example, Task 4 could be used to highlight differences between age or ethnicity groups which could be used to implement new treatment or prevention strategies.

Mid-Level abstraction
On mid-level, our tool provides opportunities to use several filtering utilities. The filter bar allows the user to locate different groups among the patients on each plot of the application. Moreover, the linked plots can be used to explore the correlation between the features.

Low-Level abstraction
On the low-level, the user can identify a patient's demographics data, their images and their location on the UMAP space. Furthermore, the experts can compare different groups to each other, e.g., Hispanic females to Caucasian males.

4 Visualization and Interaction Design: How

4.1 UMAP embeddings exploration

The main purpose of this plot is to find any clusters or correlations between various parameters that may or may not contribute to the classifier outcomes and then to the heart failure rate. A scatter plot showing the 2-dimensional projected UMAP embeddings is used to address the tasks 1, 2 and 3, as shown in Figure A1. The main asset of the scatter plot is that it allows the user to visualize a large number of points in a compact and readable method. The points are depicted with different markers to distinguish among training and test datasets and they are colour-coded with respect to the diagnosis type, i.e., HF/non-HF. This allows the user to quickly identify any clusters present in the data.

Interaction:
The filter bar allows the user to directly analyze the scatter plot considering only subsets of patients. The scatter plot is also linked with the other plots in the dashboard by selecting any point and the related data is highlighted in the other plots as well. Hovering over the points displays a tooltip that contains the coordinates of the point, the filename and the index of the tile. Moreover, the user can view the tile image of the selected point on the right side of the plot. The user can select a point with a tap, select multiple points with a lasso, and zoom in and out.

4.2 Comparison of age of heart transplant concerning the racial groups and gender

To address the question of Tasks 4 - 5, dot plots are used to get insights about which racial group/gender is getting the heart transplant earlier or later, as shown in Figures A2 - A3. The dots are colour-coded with respect to gender. This allows the user to understand and discover any clustering of disease subtypes for different racial groups. The user will be able to find any similarities or dissimilarities in age transplantation between the genders who have certain disease subtype for different races in all the three-dot plots.
Interaction:
The dot plots are linked to the UMAP scatter plot and to the other plots of the dashboard. This allows the user to select specific points with a tap or multiple points with the lasso, highlighting the same patients on the other plots.

4.3 Chronic and Non-Chronic patient data distributed over ethnicity and gender

A stacked bar plot was chosen to answer Tasks 6 and 7. The plot represents the number of patients who have been diagnosed with HF vs non-HF, distributed over their ethnicity groups. This allows the user to discover and analyze whether a particular ethnic group as a whole has a direct impact on the heart failure rate. To get a better understanding of this, a sub-classification on the gender has also been incorporated in the plot, specifically addressing the question of Task 7. It allows the user to gather insights on whether the genders contribute equally to the failure rate or if one of the genders is more likely to be susceptible compared to the other.

Interaction:
The user can select the points from the UMAP scatter plot and it shows the number of Chronic and Non-Chronic patients in the stacked bar plot accordingly. Filtering buttons are used for each of the categories described above. The user can select the desired gender, diagnosis and ethnicity groups and also the combination of all of these to analyze the impact of heart failure rate. It allows users to check if a racial group or gender is more vulnerable to the risk of being affected by heart failure or not.

4.4 Disease sub-type data distributed over different ethnic groups

To understand how different ethnic groups are diagnosed with certain disease subtypes and therefore accomplish Task 8, we can compare them quantitatively. For this purpose, a clustered bar chart representing the disease subtypes for the different racial groups is employed. This visualisation allows the user to identify and compare the number of patients tagged to the disease subtypes between the races by directly observing the height of the bar charts. Therefore, the user can make inferences about the racial groups who are more propitious to get one specific disease subtype.

Interaction:
The bar charts are linked to all other plots of the dashboard. This allows the user to select specific ethnicity groups along with the gender and diagnosis types by clicking on the filter buttons. This is in turn reflected onto the bar charts by providing ways to compare between the disease subtypes distributed among the ethnicity groups. The comparison can also be made by selecting the points from the UMAP scatter plot.

5 Realization

The assignment has been implemented in Python using the Bokeh visualisation library [2]. We choose Bokeh because it is a very flexible and powerful tool. It is an open-source project which can produce interactive and linked plots. Moreover, Python and Bokeh are cross-platform, therefore the application can run on Windows, Linux and macOS without any modification. Moreover, the functionalities of the library can be extended by using JavaScript within Python.

During the implementation, we only had to make one compromise. For Task 4 and Task 5, we first thought that a range chart with a heatmap or a box-and-whisker plot would be the most appropriate way to address them. However, later on, we realized that these plots would not be able to produce clear results regarding the age distributions, and given the small size of the data, we could directly show the raw data points. Therefore we solved these tasks with dot plots, which show the age distribution correctly and it is not sensitive to outliers (see Figure A2).

6 Use Cases

Each method described in Section 4 was applied as a visualization tool for the analysis of the HF dataset. Additionally, a screencast of the functionalities of the application has been recorded to support the claims of this paper. The consumers that might be interested in this visualization tool are (i) clinicians who want to explore factors that might influence the heart failure rate; (ii) machine / deep learning practitioners who

[2] https://bokeh.org

want to validate their trained models with an unsupervised analysis of the extracted features (i.e., with a UMAP projection) and understand if the model is affected by batch effects or information which was not explicitly used as input of the model in the first place (e.g., demographics data).

To accomplish Task 1, we can interact (e.g., select points with a tap or a lasso, zoom in) with the scatter plot showing the UMAP embeddings and observe the image of the heart tissue corresponding to the selected point, as described in Section 4.1. The plot presents two main clusters corresponding to HF and non-HF tiles, as expected with a well-trained deep learning model used to extract the features. Moreover, a small cluster can be found far from the others, containing only the tiles of five non-HF patients (patients IDs: 36172, 36173, 36174, 36175, 36176). After careful examination, we understood that the reason for this additional separation of these patients was due to the lighter tissue staining intensity, as shown in Figure 1. Differences in tissue staining procedures are recognized as a critical issue for multi-cohort studies involving deep learning [2, 5], confirming our claim.

With regard to Tasks 2 - 3, by interacting with the scatter plot and the dot plots, and by further filtering the data using the shared checkboxes, it appears the embedding is clustered with respect only to the classification outcome, as there is no clear indication that demographic characteristics (i.e., gender, ethnicity, and age) and/or disease subtypes (for HF patients) are influencing the classifier in the ability to distinguish HF and non-HF tiles.

Tasks 4 and 5 can be performed by observing the dot plots, showing ethnic groups and age of transplant, colour-coded by gender, as described in Section 4.2. Overall, it seems that Caucasian males are getting the transplant earlier (from 18 years old) and evenly during their lifetime, compared to the other ethnicities, whereas African American males are consistently being transplanted between the age of 52 to 68. In this dataset, all Hispanic patients are non-HF, therefore they are excluded from this discussion. The female patients in this cohort are mostly in the Unknown Racial Group category, so it is not possible to compare genders and ethnicities. Regarding disease subtypes, younger patients getting the transplant are most likely to be affected by cardiomyopathy, instead of ischemic cardiomyopathy. For older patients, there is no substantial difference among the subtypes and age of transplant.

To address tasks 6 and 7, we can examine the stacked bar plot presented in Section 4.3. The plot shows that there are more Caucasian and African American patients with non-HF than with HF. Hispanic patients are all non-HF. However, given this limited and unbalanced dataset, we cannot prove or disprove that belonging to a particular racial group has an impact on the heart failure rate. As mentioned earlier, most of the female patients with HF belong to the Unknown Racial Group category. However, with respect to only Caucasians, there is a clear difference between the ratio of females with HF and the ratio of females without HF: in the non-HF group, the number of females and males is balanced, while for HF group the males account for most of the patients. Even after considering this, again we cannot conclude anything about the relation between the gender and the heart failure rate.

Task 8 can be completed by studying the clustered bar chart described in Section 4.4. For both African American and Caucasian groups, there are consistently more patients affected by ischemic cardiomyopathy than by cardiomyopathy. However, African American patients are the only ones showing a great difference among subtypes, as one-third of them is affected by cardiomyopathy, while two-thirds by ischemic cardiomyopathy.

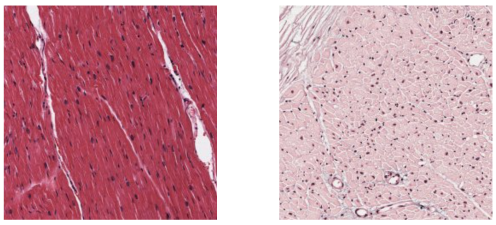

Fig. 1: *Comparison of a heart tissue tile with a proper staining intensity (A) to a tile with lighter staining intensity (B).*

7 Discussion and Conclusion

Our visualisation can be used to answer most of the questions presented in Section 3, with the exception of questions of tasks 6 and 7, as explained in Section 6. The UMAP scatter plot shows the embeddings for every image. We would be able to visualize if any of the ethnic groups would form clusters. We linked all the other plots to the dashboard to make it even more interactive as well as give meaningful insights.

However, during the development phase, we encountered certain problems. We initially planned to incorporate a range chart with a heatmap showing the distribution of age groups along with the box-and-whisker plot to understand who gets the heart transplant earlier or later. Moreover, after careful analysis of the dataset, we realised that we do not have enough data points for a meaningful heatmap. Therefore, we chose dot plots to show these relations, because they are not sensitive to outliers and they work best for small data.

A stacked back-to-back bar chart is provided to compare the HF vs non-HF patients over ethnicities and genders. However, since the dataset is unbalanced with respect to these characteristics, the bar chart was unable to provide certain meaningful insights because the dataset does not contain patients who are diagnosed with chronic heart failure belonging to the Hispanic racial group, moreover, the dataset is unbalanced regarding the ethnic groups. Though the plot could show meaningful data for different ethnicity groups, the missing data would have been crucial in this study for paving a way for some interesting results.

The clustered bar chart is used to compare the disease subtypes in accordance with the ethnicity groups. This chart is also linked to the other plots so that we could analyse the data populated from the other plots such as the number of patients who belong to HF/non-HF groups and also the age at which the heart transplantation was performed, linked to the disease subtypes providing useful information. This chart best suited our visualisation since we can have the three disease subtypes grouped together.

The dashboard presents a visual tool to inspect multiple perspectives of the dataset via filtering. The dashboard proposes the use of UMAP scatter plots, dot plots, stacked bar plots and clustered bar plots, helping to answer the tasks-related questions.

As we mentioned before, our dataset is unbalanced and regarding the demographics data we are unable to make clear conclusions about the impact of the race. In the following paragraph, we are writing down our insights based on our dataset, however further studies could confirm or refute these findings. Nevertheless, our tool could be used to explore larger, more comprehensive datasets. With the use of this application and this dataset, it can be concluded that the rate of heart transplant for African American males is more predominant and Caucasian males are getting affected by the disease earlier and are evenly distributed over the course of time. Younger patients are more affected by cardiomyopathy compared to ischemic cardiomyopathy and there is no differentiation for older patients. Besides, it was found that Caucasian and African American groups contributed more to the number of non-HF in comparison to the number of HF patients. The Unknown Racial Group category is mostly composed of HF females. African American and Caucasian groups are mostly affected by cardiomyopathy and they form a substantial portion of patients being affected to both cardiomyopathy and ischemic cardiomyopathy.

REFERENCES

[1] N. Bussola, A. Marcolini, V. Maggio, G. Jurman, and C. Furlanello. Ai slipping on tiles: data leakage in digital pathology, 2020.

[2] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. De Souza, A. Baidoshvili, G. Litjens, B. Van Ginneken, I. Nagtegaal, and J. Van Der Laak. The importance of stain normalization in colorectal tissue classification with convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 160–163. IEEE, 2017.

[3] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.

[4] J. J. Nirschl, A. Janowczyk, E. G. Peyster, R. Frank, K. B. Margulies, M. D. Feldman, and A. Madabhushi. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of h&e tissue. *PloS one*, 13(4):e0192726, 2018.

[5] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, and J. van der Laak. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical image analysis*, 58:101544, 2019.

[6] M. D. Zarella, D. Bowman, F. Aeffner, N. Farahani, A. Xthona, S. F. Absar, A. Parwani, M. Bui, and D. J. Hartman. A practical guide to whole slide imaging: A white paper from the digital pathology association. *Archives of Pathology & Laboratory Medicine*, 143(2):222–234, Feb. 2019.

A FIGURES

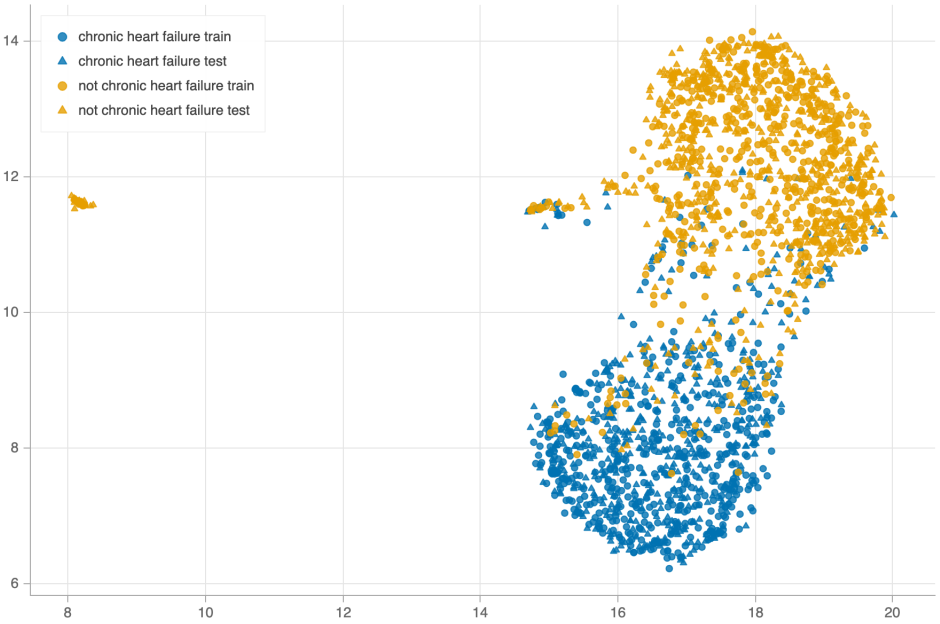

Fig. A1: *Scatter plot showing the UMAP embeddings of the features extracted from the deep learning classifier.*

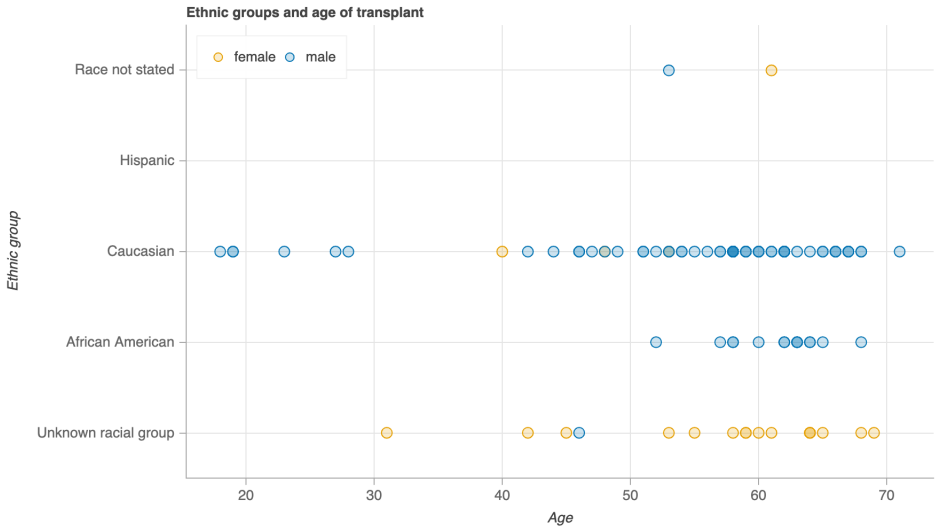

Fig. A2: *Dot plot showing the ethnicity group and the patients' age of heart transplant.*

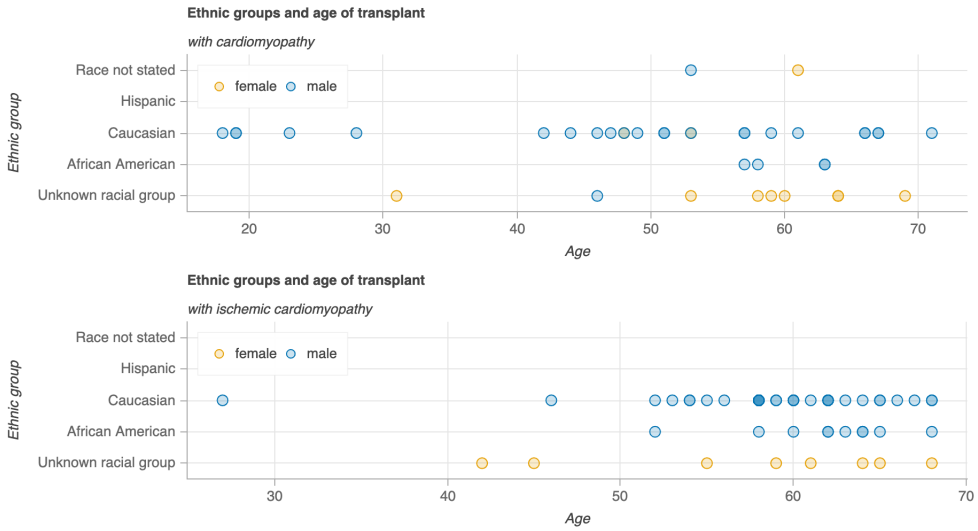

Fig. A3: *Dot plots showing the ethnicity group and the patients' age of heart transplant, divided by disease subtype.*

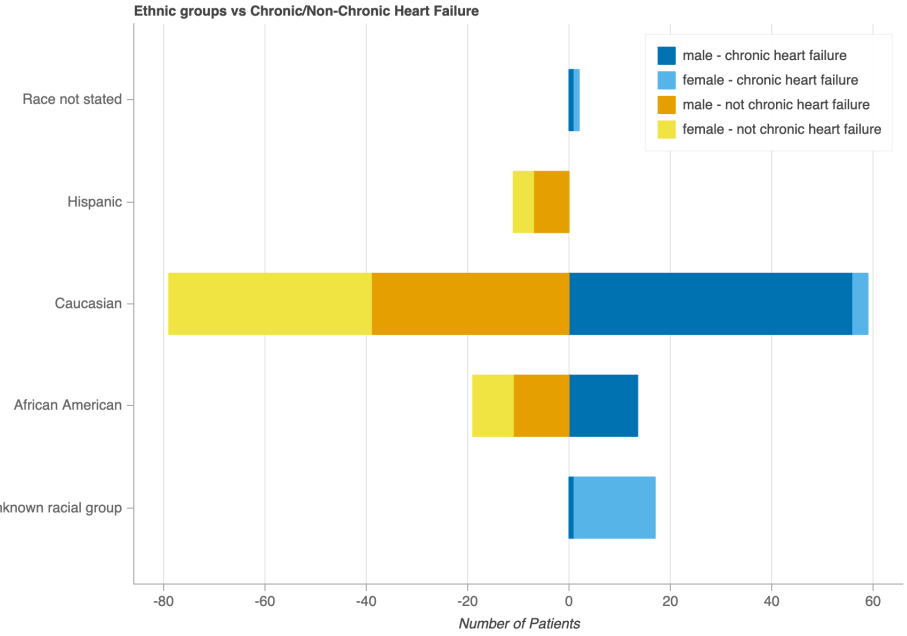

Fig. A4: *Stacked bar chart showing the number of patients belonging to HF vs non-HF groups within the ethnicity groups.*

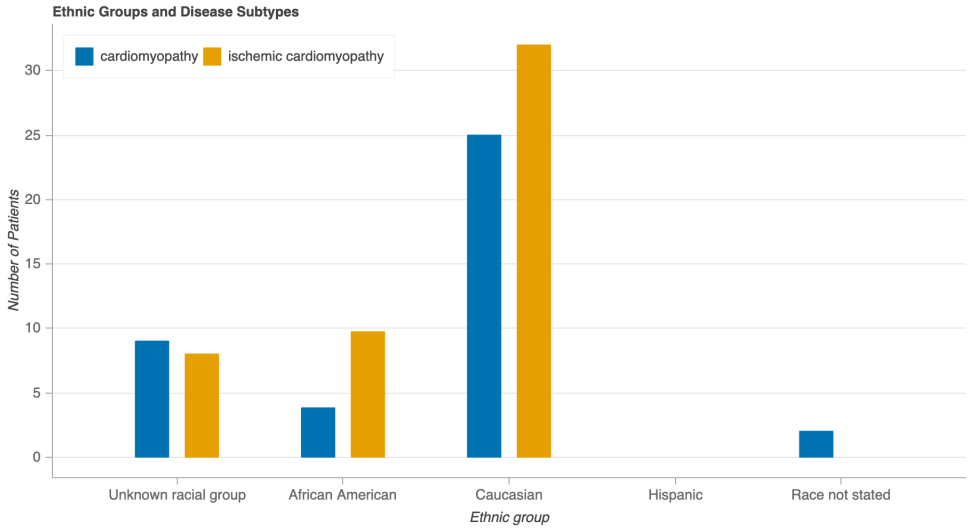

Fig. A5: *Clustered bar chart showing the distribution of the disease subtypes within the ethnicity groups.*