

עבודה 3 דאטא סיינס

הערה- צימצמנו את גודל ה dataset ל-1000 רשומות

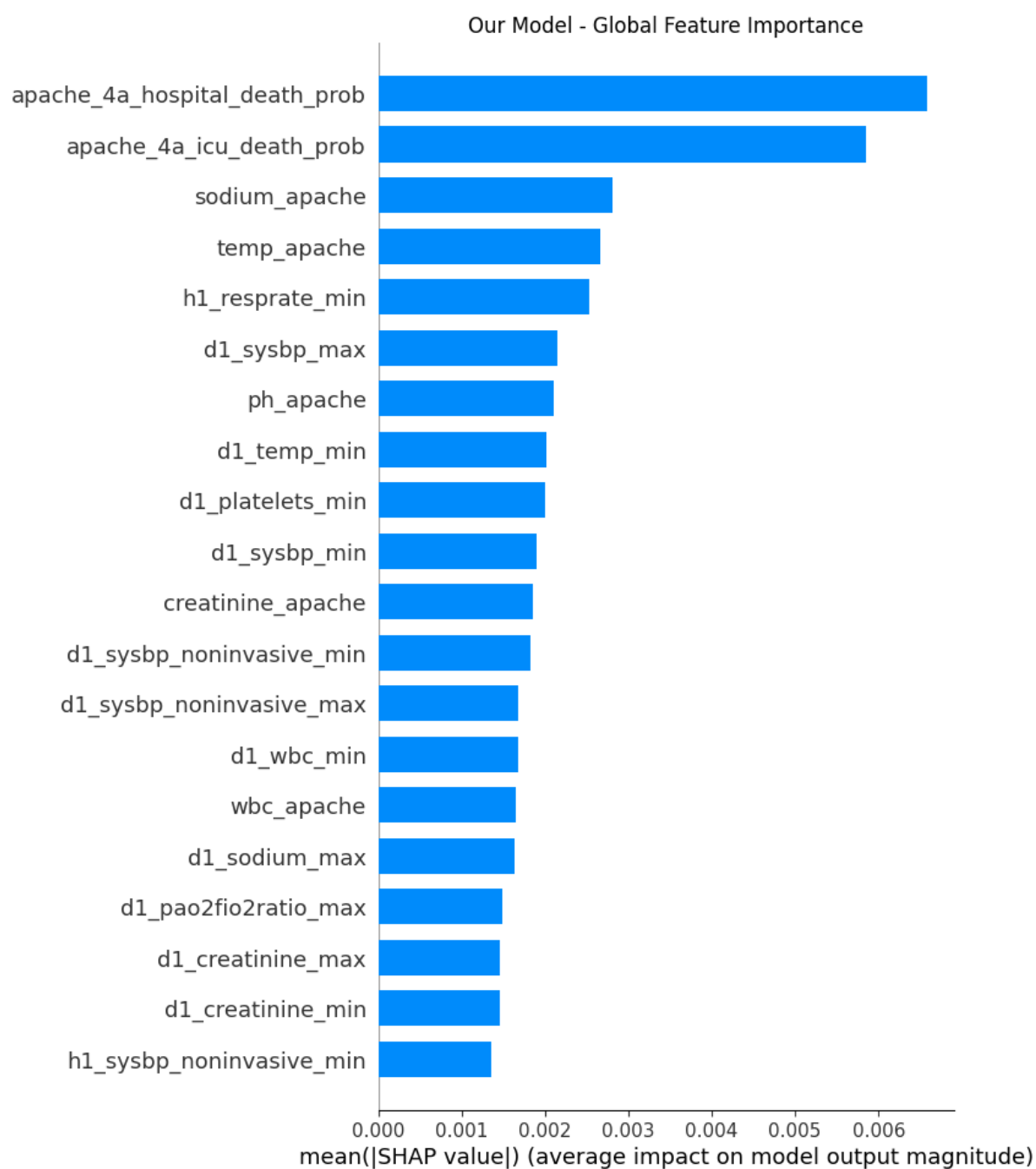
סעיף 4:

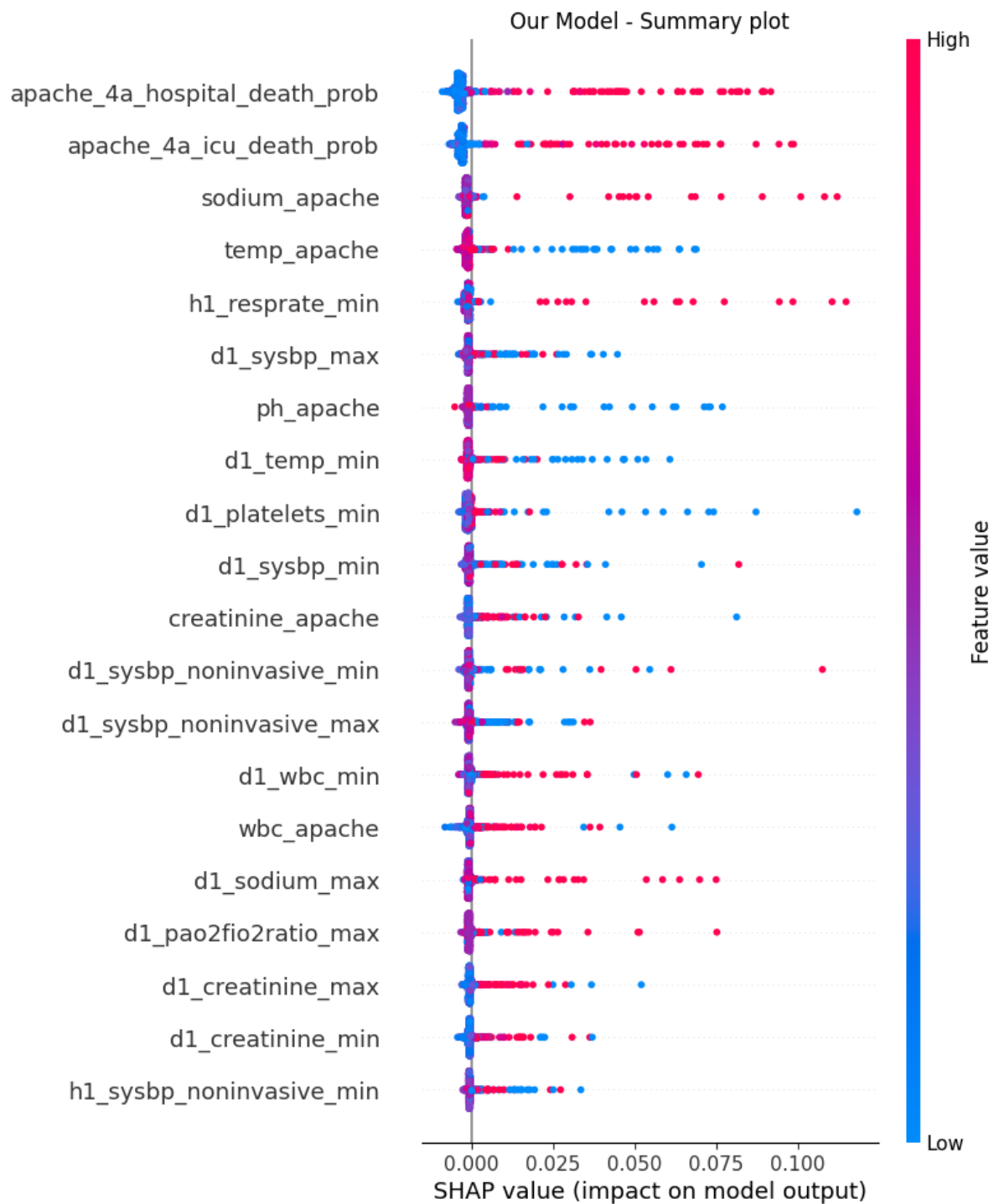
```
Baseline Model Accuracy - Train Set: 1.0  
Baseline Model Accuracy - Test Set: 0.945  
Our Model Accuracy - Train Set: 0.966125  
Our Model Accuracy - Test Set: 0.936
```

להלן ביצועי המודלים לפי מדד הדיוק. ניתן לראות שלפי מדד ה מודל הבייס ליין מראה בצועי דיוק גבוהים יותר. חשוב להדגיש כי מדד בלבד זה אינו די מספיק בשביל להסיק מסקנות בנוגע לשאלה "מי טוב יותר", וראינו כי בעבודה 1 שלמדד זה אין קורלציה מושלמת עם שאר מדדי הדיוק.

סעיף 5:

הפעלנו את שיטת SHAP (TreeShap) על המודל שלנו עם train test ולהלן התוצאות:

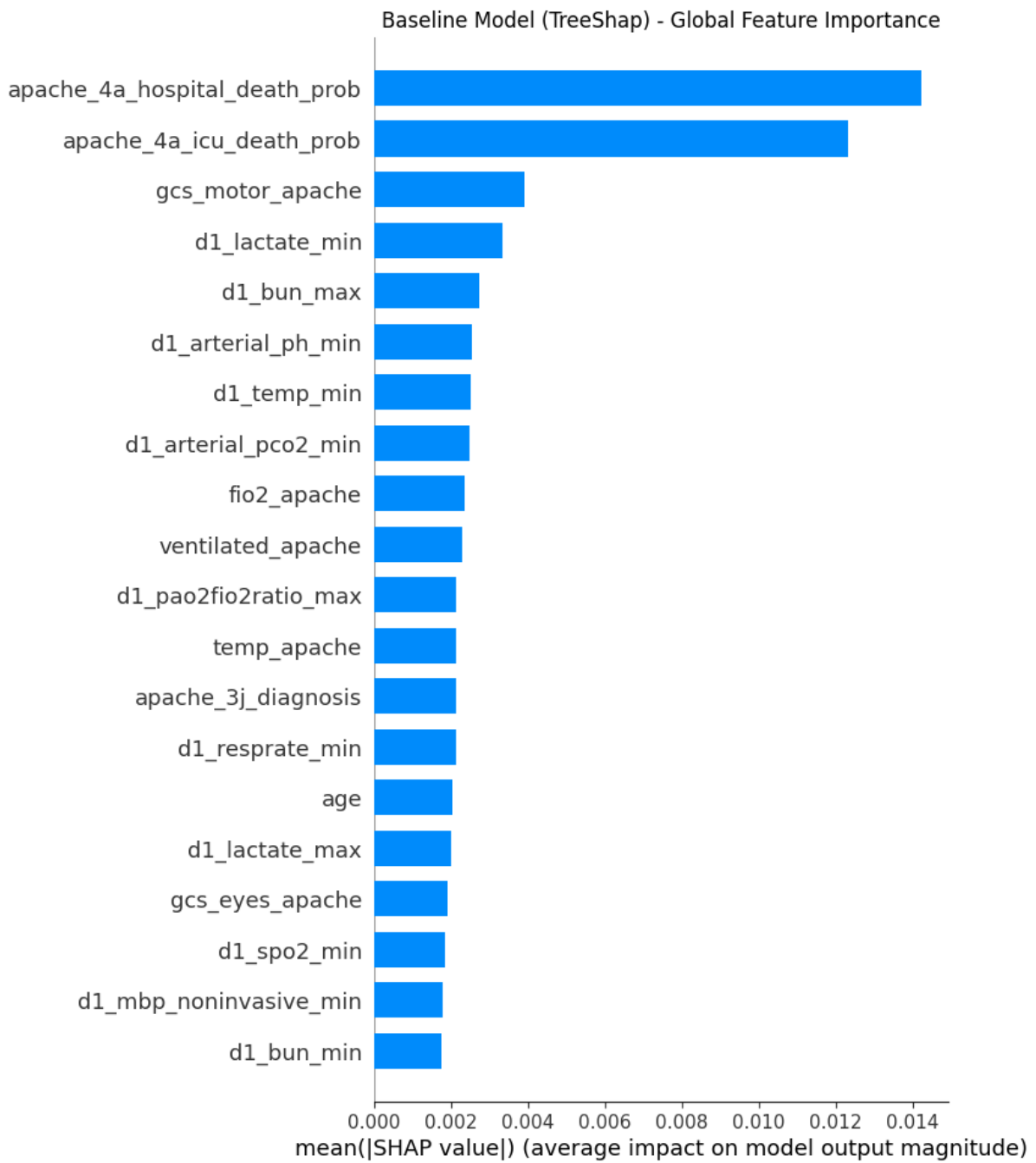


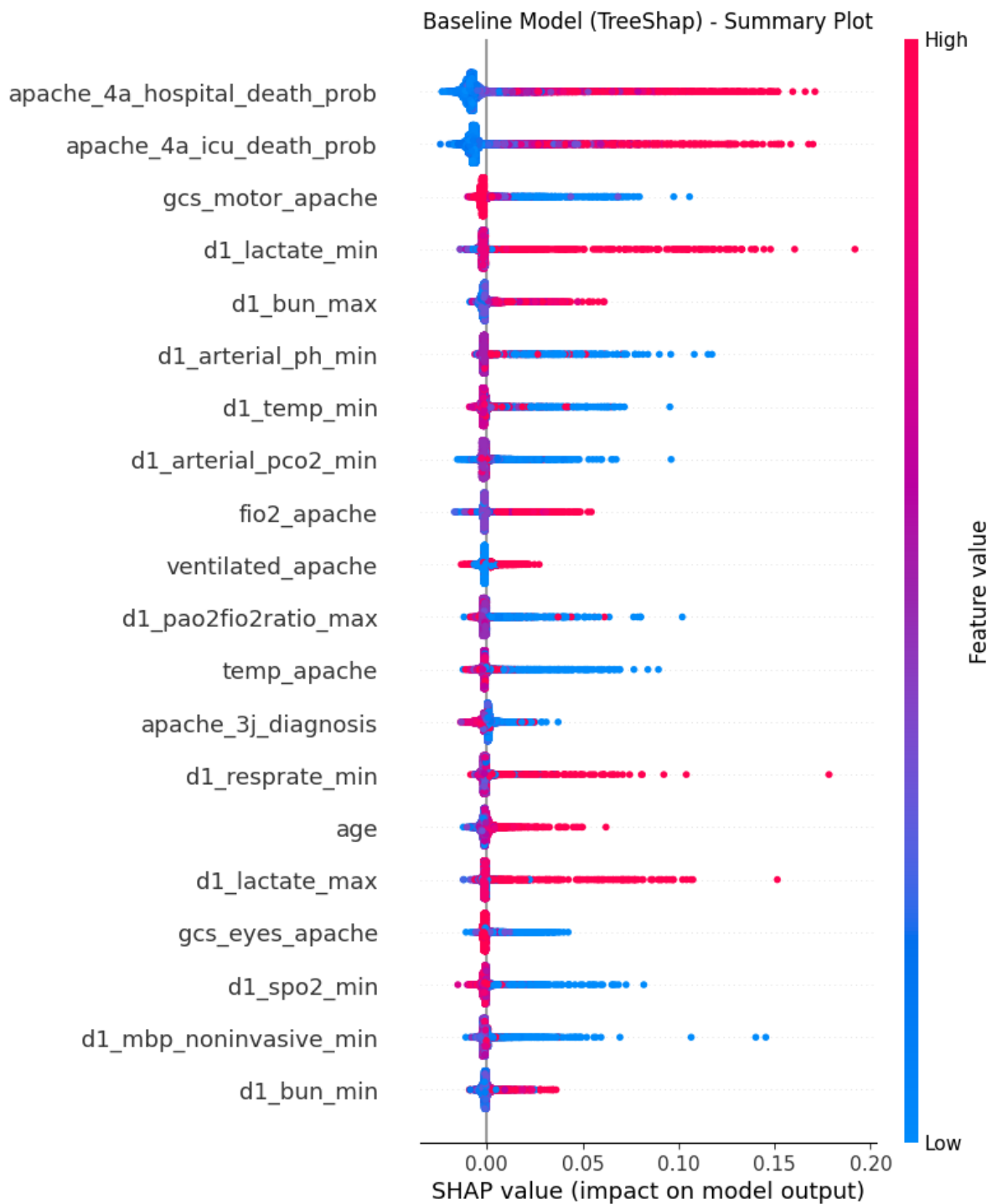


רואים לפי הגרף שיש שני פיצ'רים מרכזיים בעלי השפעה גדולה, ושאר הפיצ'רים בפער משמעותי מהם, ובעלי השפעה דומה ביניהם. ניתן לראות כי עבור שני הפיצ'רים בעלי ערכי ההשפעה המשמעותיים יותר, ככל שערכי הפיצ'רים גבוה יותר כך גם ערך SHAP שלהם (ולהפך).

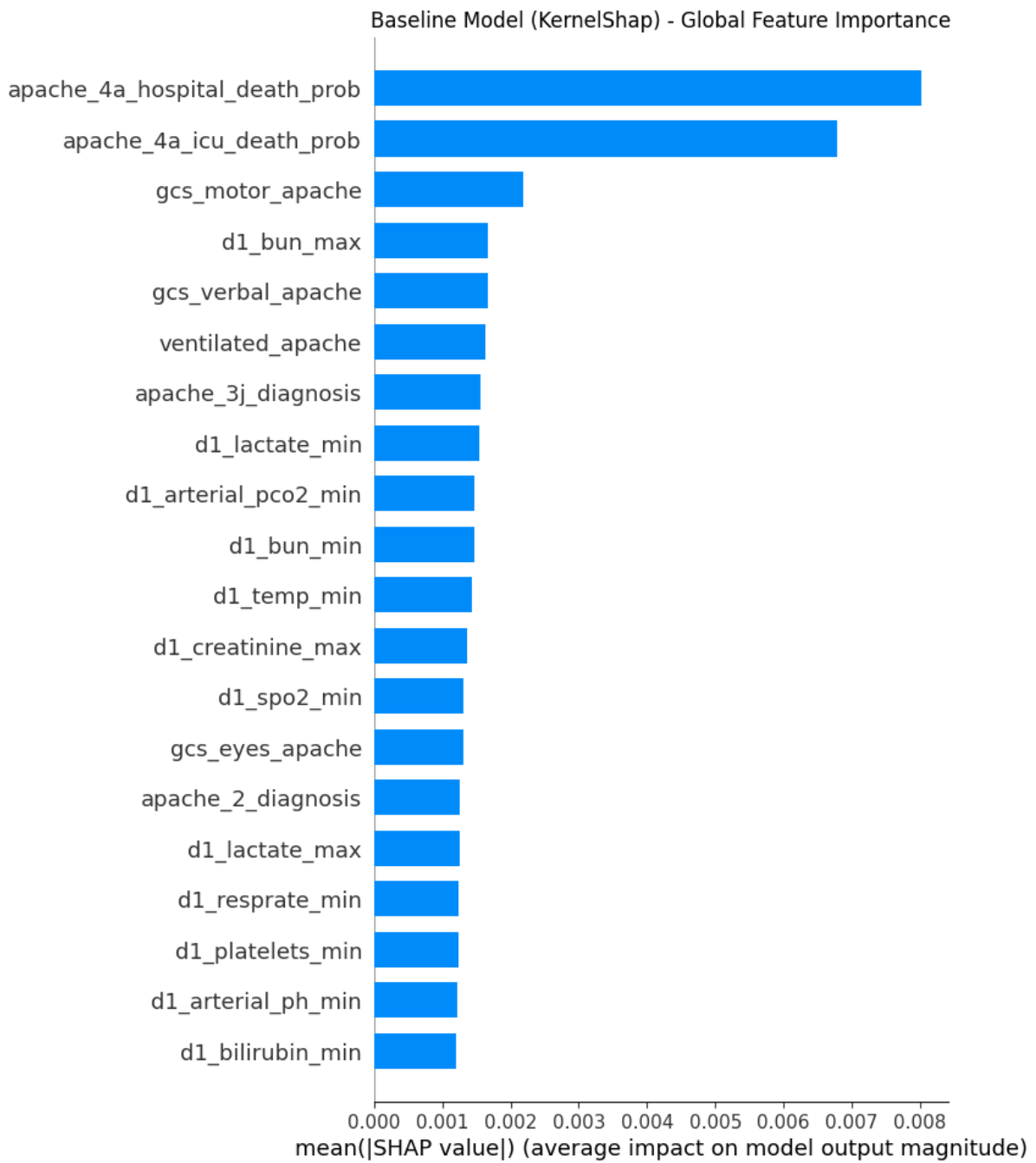
סעיף 6:

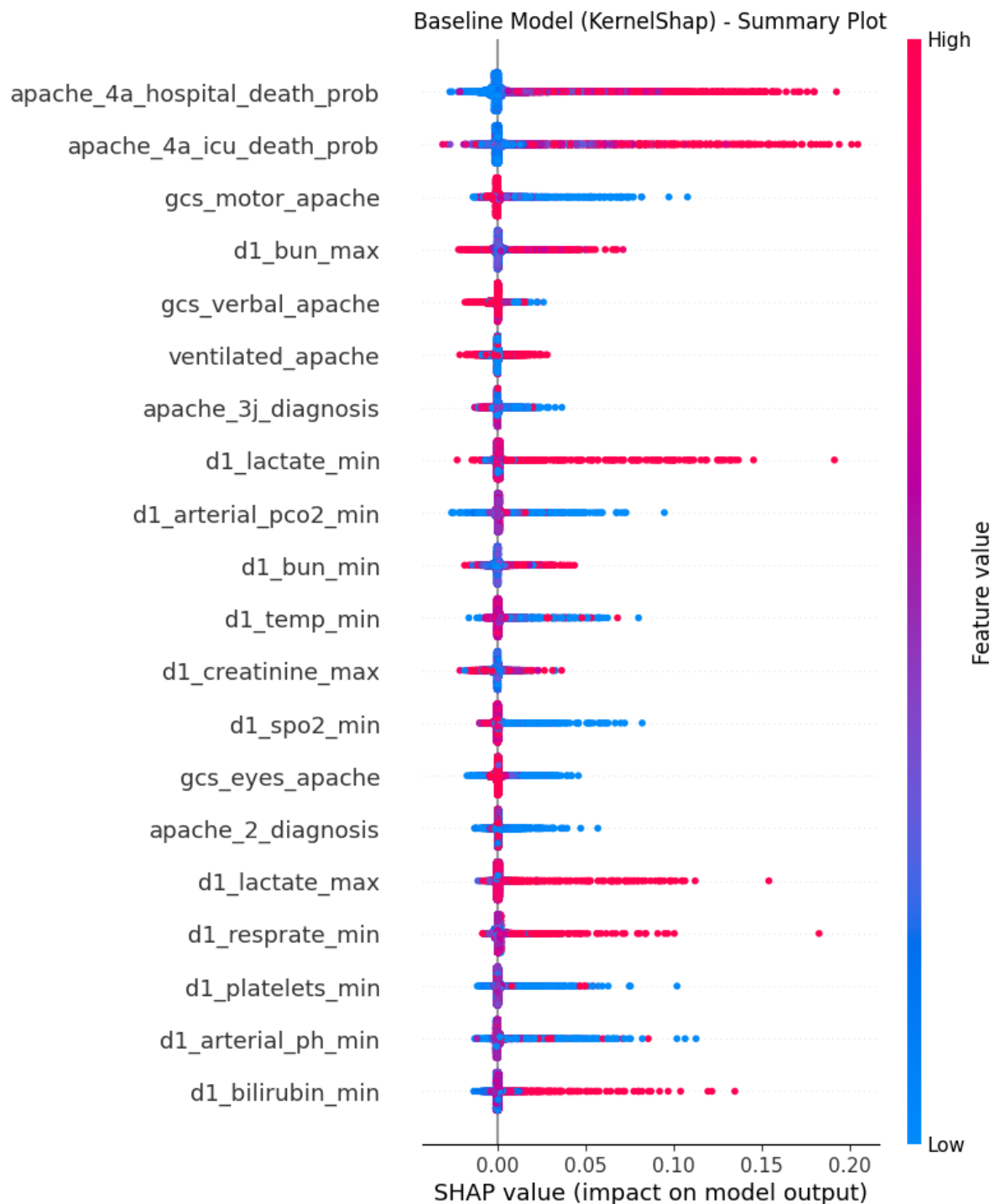
הפעלנו את שיטת SHAP בשיטת TreeShap על מודל הבייס ליין עם test train ולהלן התוצאות:





הפעלנו את שיטת SHAP בשיטת KernalShap על מודל הבייס ליין עם test train ולהלן התוצאות:



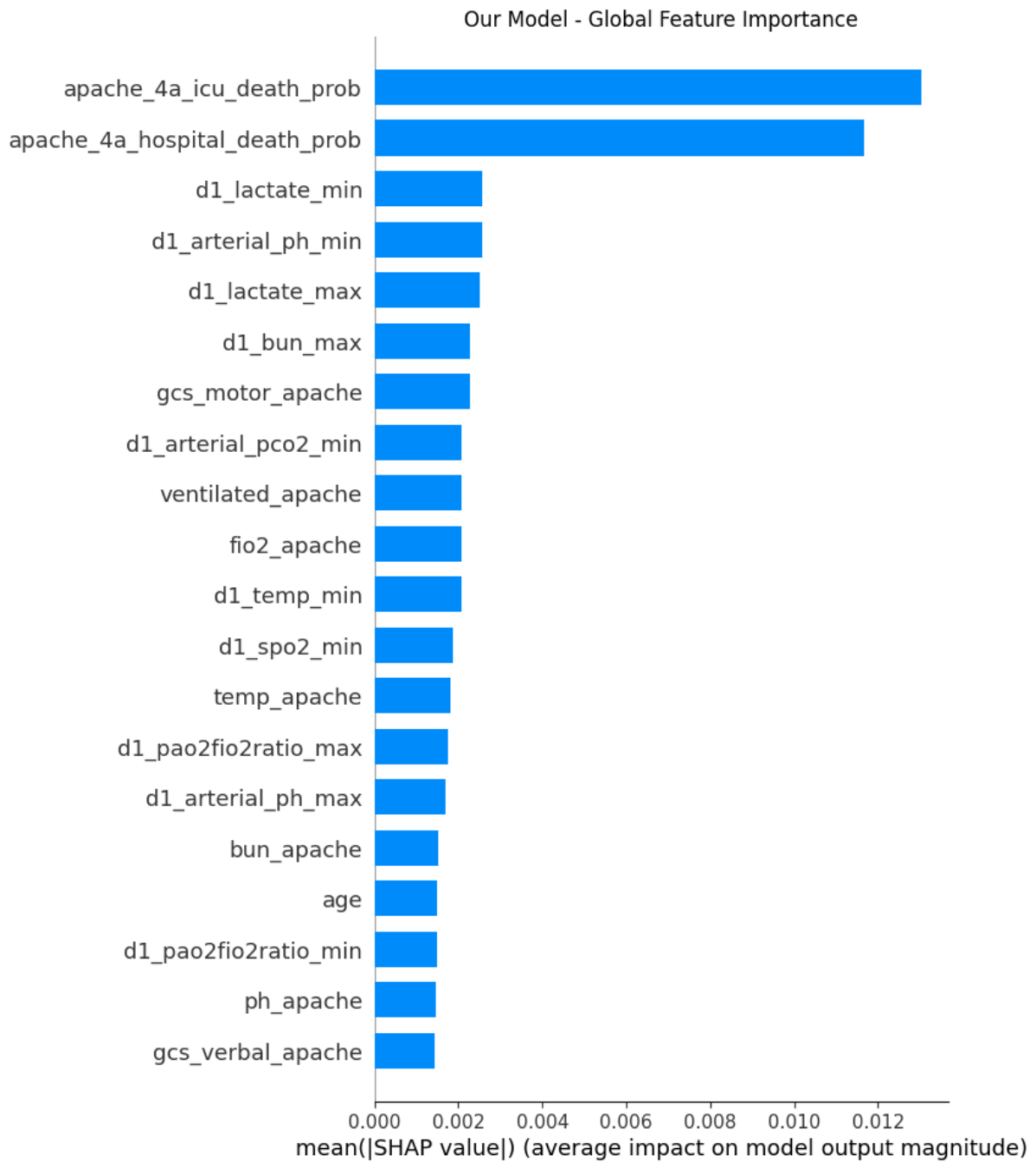


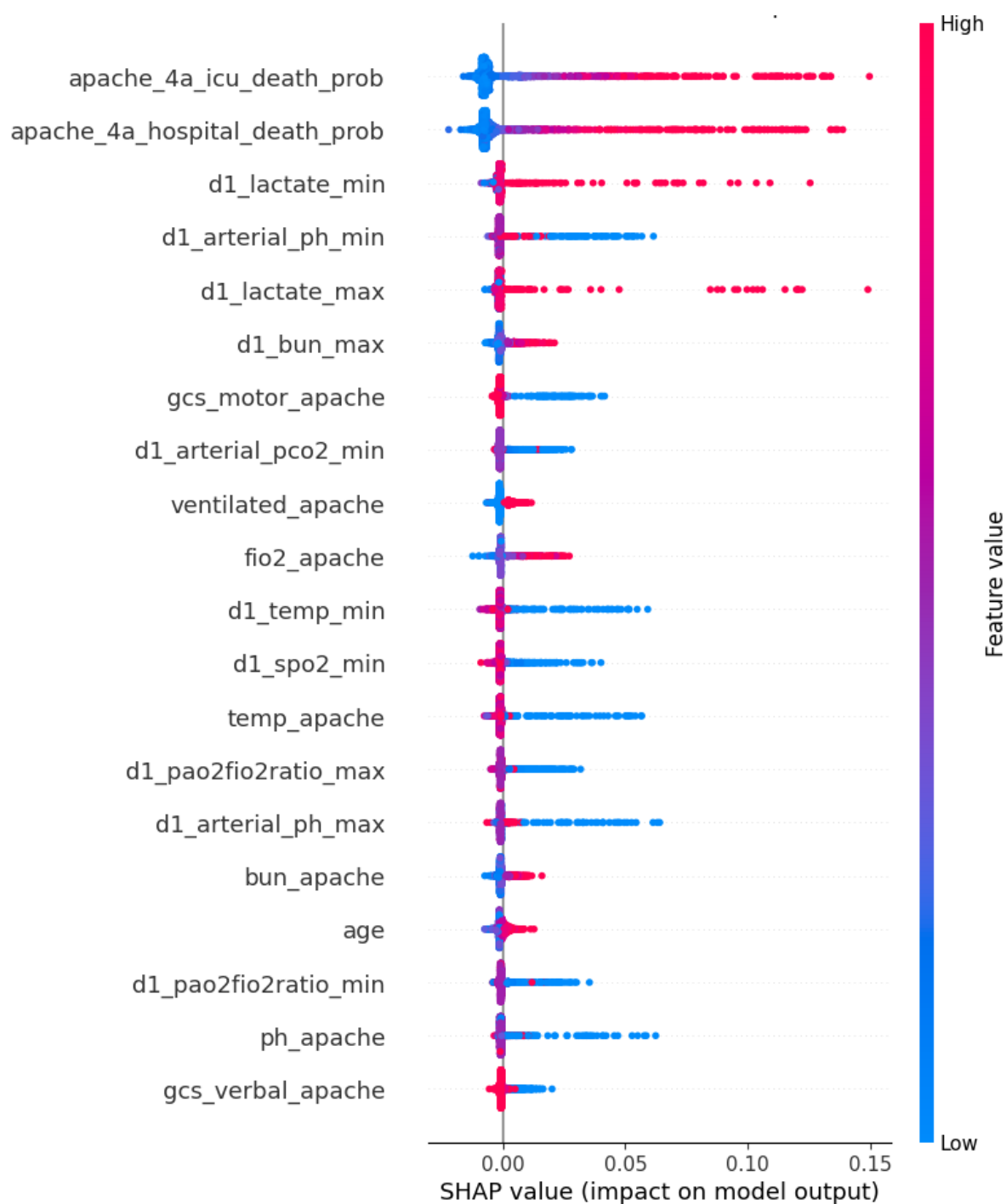
גם כאן רואים לפי הגרף שיש שני פיצ'רים מרכזיים בעלי השפעה גדולה, ושאר הפיצ'רים בפער משמעותי מהם, ובעלי השפעה דומה ביניהם. ניתן לראות כי עבור שני הפיצ'רים בעלי ערכי ההשפעה המשמעותיים יותר, ככל שערכי הפיצ'רים גבוה יותר כך גם ערך SHAP שלהם (ולהפך).

אם נשווה בין תוצאות SHAP בין המודלים עם השיטות השונות, נוכל לראות תוצאות דומות, וכי כל התוצאות מצאו כי ישנם שני פיצ'רים משמעותיים יותר מהשאר, אך בכל אחת מהתוצאות יש דירוג שונה בין השניים המובילים. בנוסף יש הבדלים בין הפיצ'רים הזניחים יותר כאלה שקיימים במודל ראשון ולא קיימים בשני, אך משם שהם זניחים יותר מלכתחילה לא נוכל להצביע על כך כהבדל משמעותי.

סעיף 8

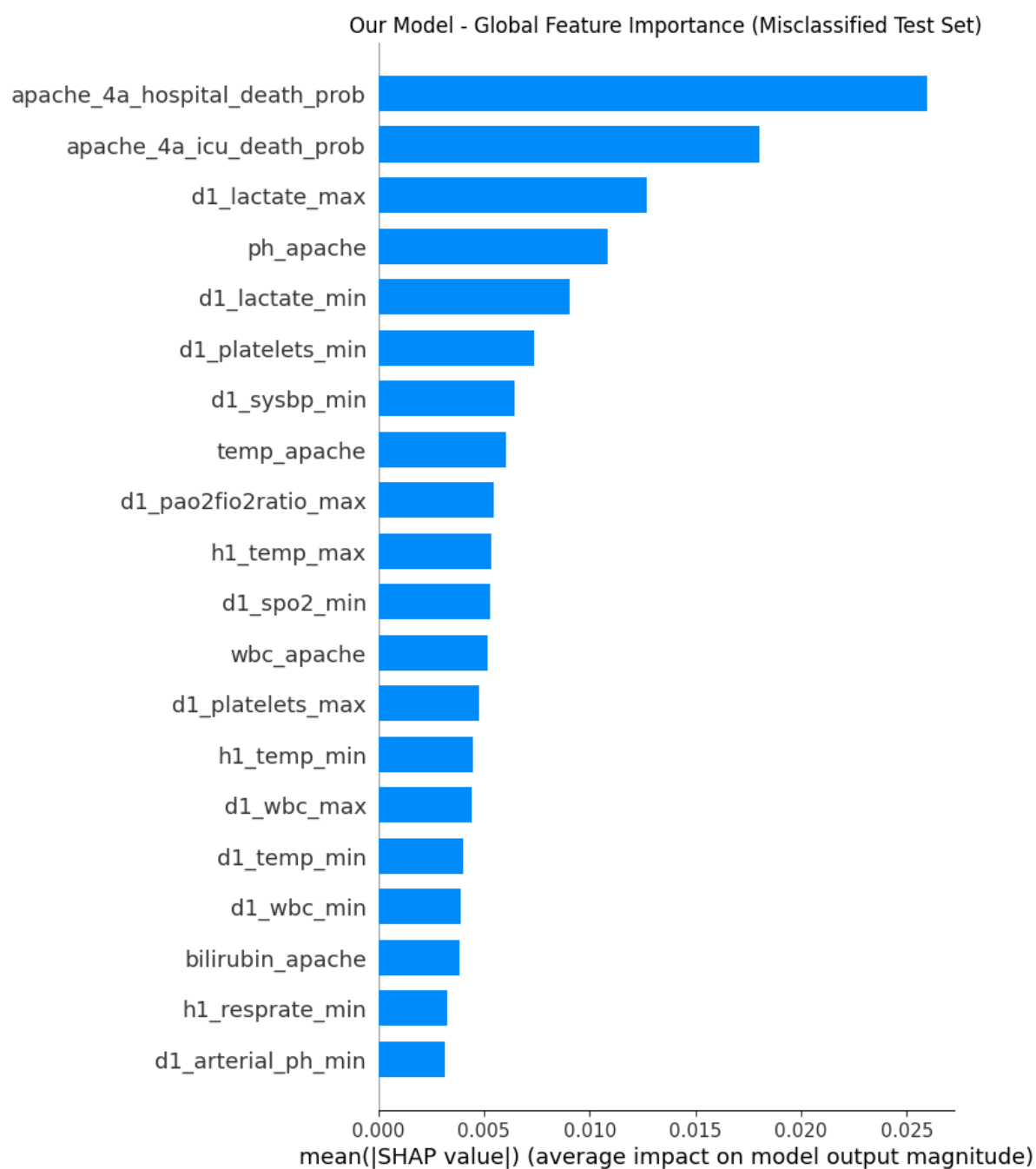
הפעלנו את שיטת SHAP (TreeShap) על המודל שלנו עם test set ולהלן התוצאות:

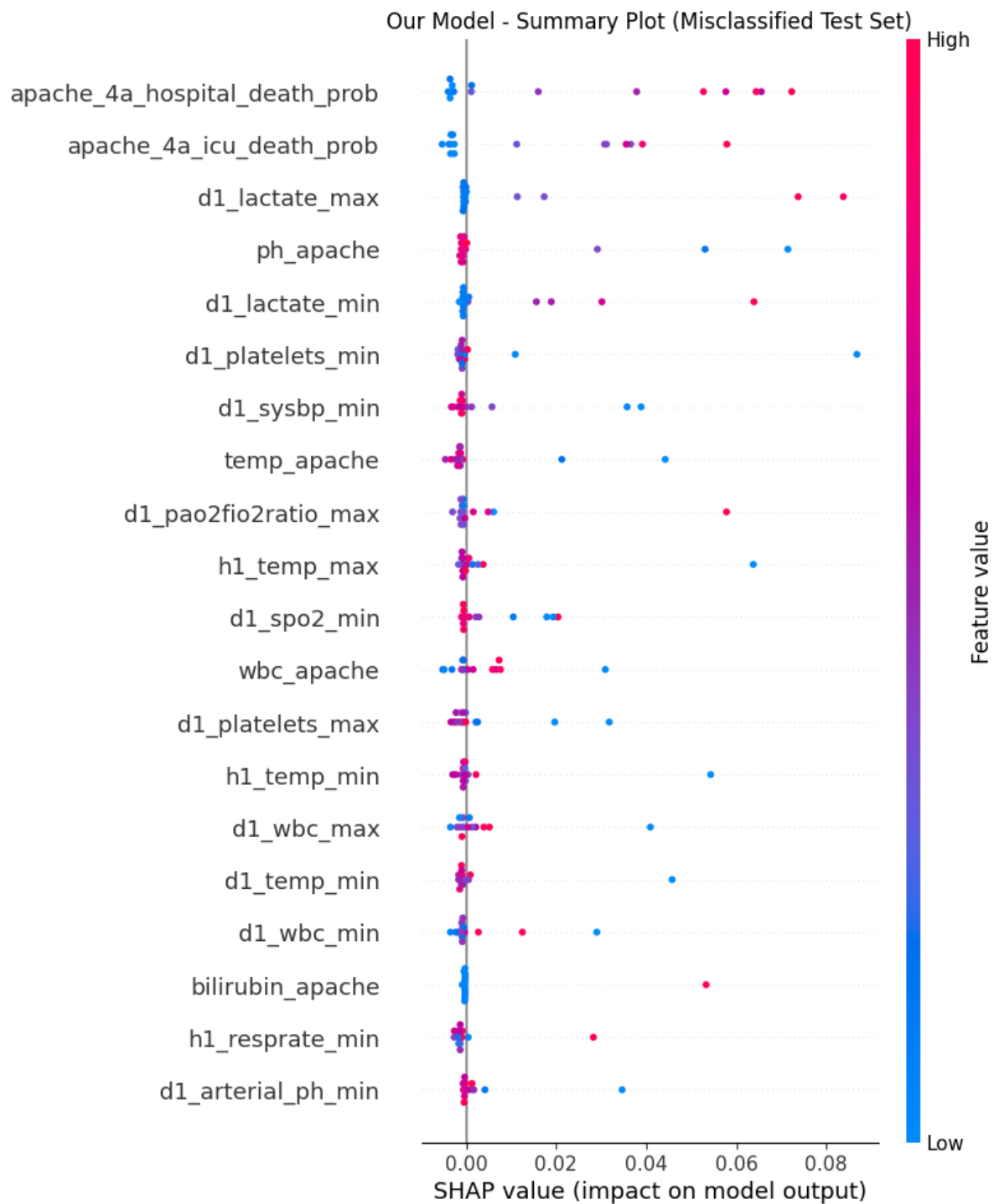




בהשוואה להרצה עם train setn בסעיף 5, ניתן לראות כי התוצאות עקביות. אם נסתכל על גרף הsummary אז נראה כי התפזרות הנקודות בקובץ האימון קטנה יותר מאשר בקובץ המבחן וזה להשערתנו מאחר וכמות הרשומות במבחן קטנה יותר ולכן התפזרות הנקודות בטווח הנל תיצור שונות גדולה.

סעיף 9:



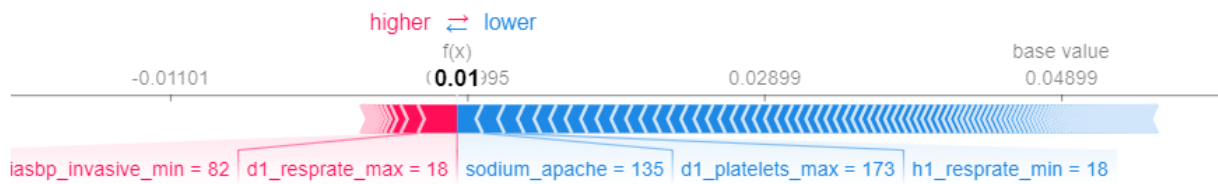


ניתן לראות כי עדיין נמצאו שני פיצ'רים משמעותיים וזהים לסעיפים הקודמים, אך הפער בין ציונם גדל. זאת ועוד, ההבדל בין הפיצ'רים המשמעותיים לשאר הפיצ'רים פחות מובהק וכן שאר הפיצ'רים בחלקם שונים מאשר תוצאת הSHAP על כל הסט הטסט.

סעיף 10

להלן הדוגמאות לרשומות שסווגו נכון:

דוגמא 1:



דוגמא 2:



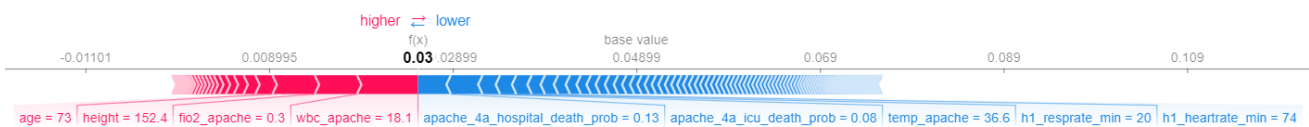
ניתן לראות כי בשתי הדוגמאות האלו הסיווג היה נכון והינו 0. בדוגמא הראשונה היו פיצ'רים שניסו "לדחוף" את הניבוי לכיוון 1 אך משום שרוב הפיצ'רים "דחפו" את הניסוי לכיוון האפס, כך נובה הסיווג 0. לעומת זאת בדוגמא השניה כל הפיצ'רים "דחפו" את הניבוי לכיוון 0, כלומר היו תמימי דעים.

להלן הדוגמאות שסווגו לא נכון:

דוגמא 3



דוגמא 4



כבר בשתי הדוגמאות הללו ניתן לראות כי "הביטחון" של המודל בנוגע לניבוי שהוא עומד לבצע פחות גובהה ביחס לרמת הביטחון בדוגמאות של הסיווג הנכון, בייחוד בדוגמא השלישית. בנוסף, הניבוי השגוי בדוגמא 3 נראה יותר הגיוני מאחר והערך קרוב לחצי בהשוואה לשאר הערכים בפער משמעותי ולכן ההסתברות שהמודל יטעה עולה. לעומת זאת בדוגמא 4 רמת הוודאות של המודל די קרובה לביטחון מלא ולכן הטעות שלו פחות צפויה ומובנת.