Diabetes and Pancreas Transplant Waitlist Outcomes

Billy Block, Ruben Jimenez, Nick Patrick

Department of Statistics, California Polytechnic State University—San Luis Obispo

Diabetes Risk Factors

Diabetes is a major public health issue in the United States, affecting over 11% of the population and contributing to over 100,000 deaths yearly. We explore the comparative performance of several classification models for predicting diabetes.

Important Features

Ordered Correlations Between Features and Diabetes
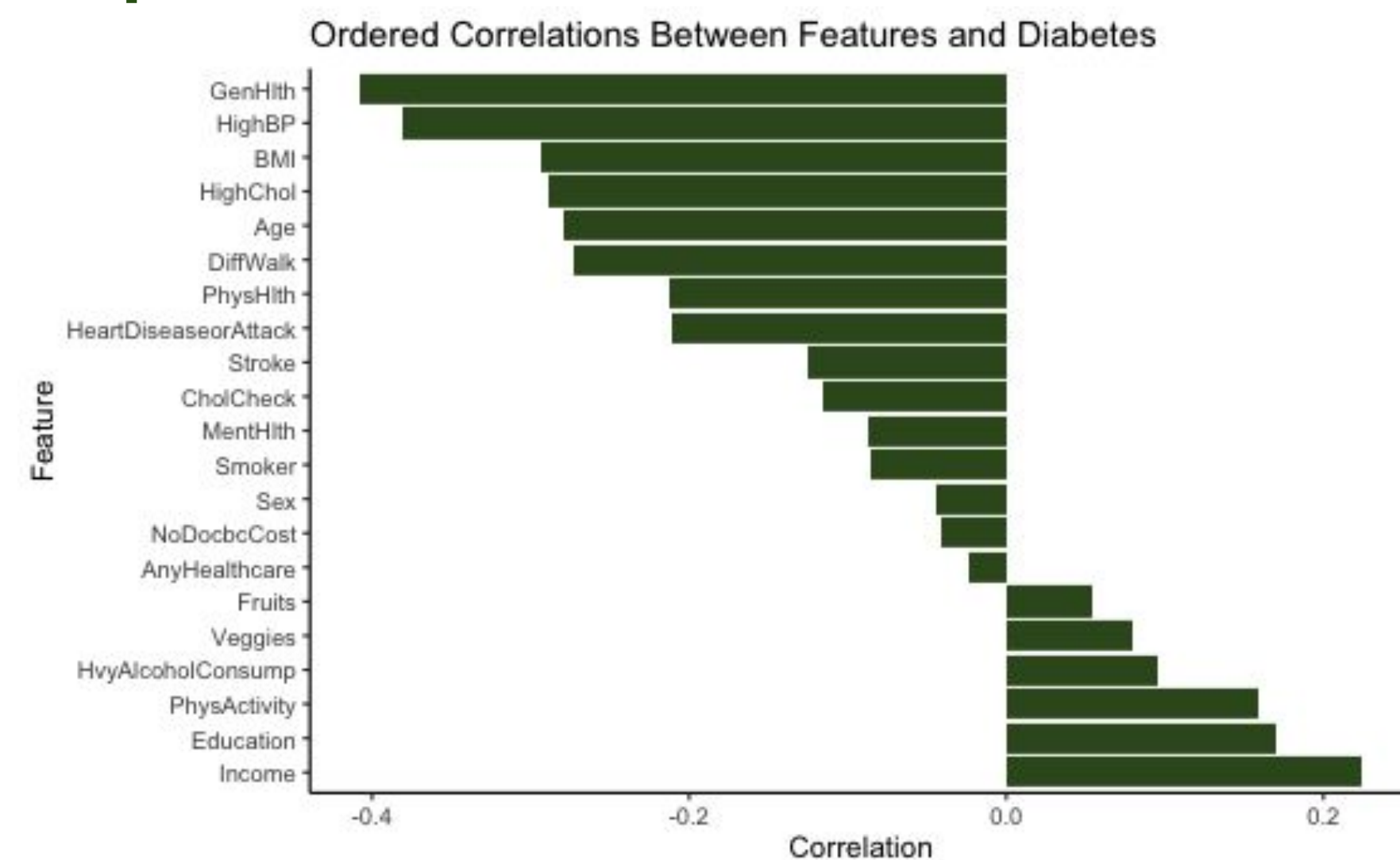
Both medical features (i.e. blood pressure and cholesterol) and socioeconomic features (i.e. income and education) are strongly correlated with diabetes status.

Model Performance

Random Forest

	Truth	
Prediction	Prediabetes/Diabetes	Healthy
Prediabetes/Diabetes	1176	2478
Healthy	194	6152

Recall = 0.86, Accuracy = 0.73

Random Forest performed best with a very low false negative rate, ensuring strong detection rates.

Limitations

With the CDC's BRFSS from which our diabetes data was collected, where surveys are collected via phone call and answers self-reported, potential sources of bias include but are not limited to response bias and sampling bias with in regards to geographic location availability, bias against non-dominant language speakers, and intersectional factors. With the UNOS data from which our patient mortality data was collected, possible inequities again include sampling bias in regards to geographic location due to availability of organs, socioeconomic availability that manifests as multi-listing across several health centers, access to supplemental treatments. Additionally, issues specific to clinical data arise, such as limited demographic and quality-of-life metrics, as well as general public availability of such data.

Patient Mortality on the Pancreas Waitlist

We hypothesized that diabetes status is crucial to understanding health outcomes for patients on the pancreas waitlist. While diabetic patients had higher mortality rates, we find that diabetes is not an important feature for predicting mortality.

Important Features

Deaths By Diabetes Status
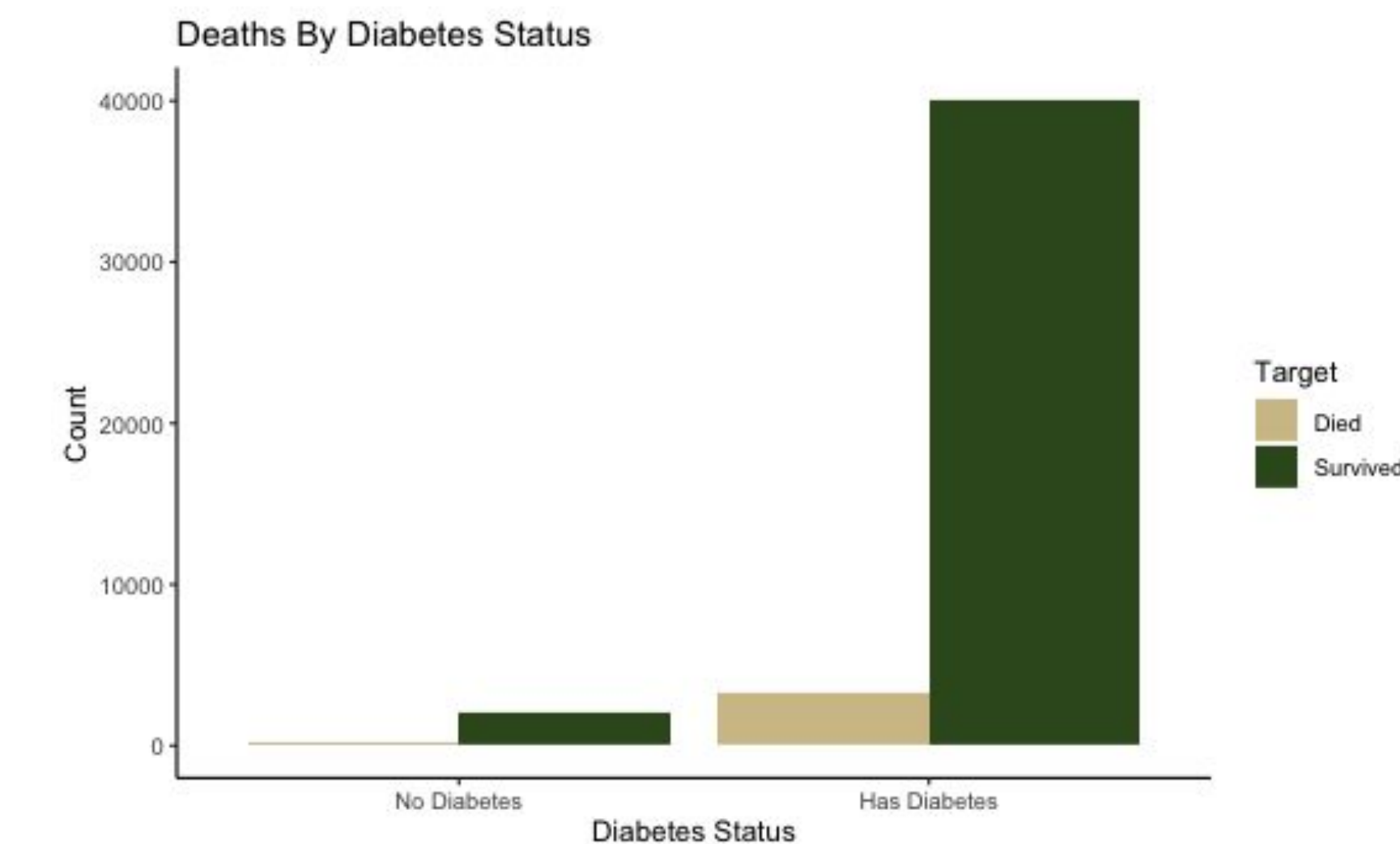
Patients without diabetes typically have more dangerous conditions, like cancer.

Mortality Rate By Dialysis Status
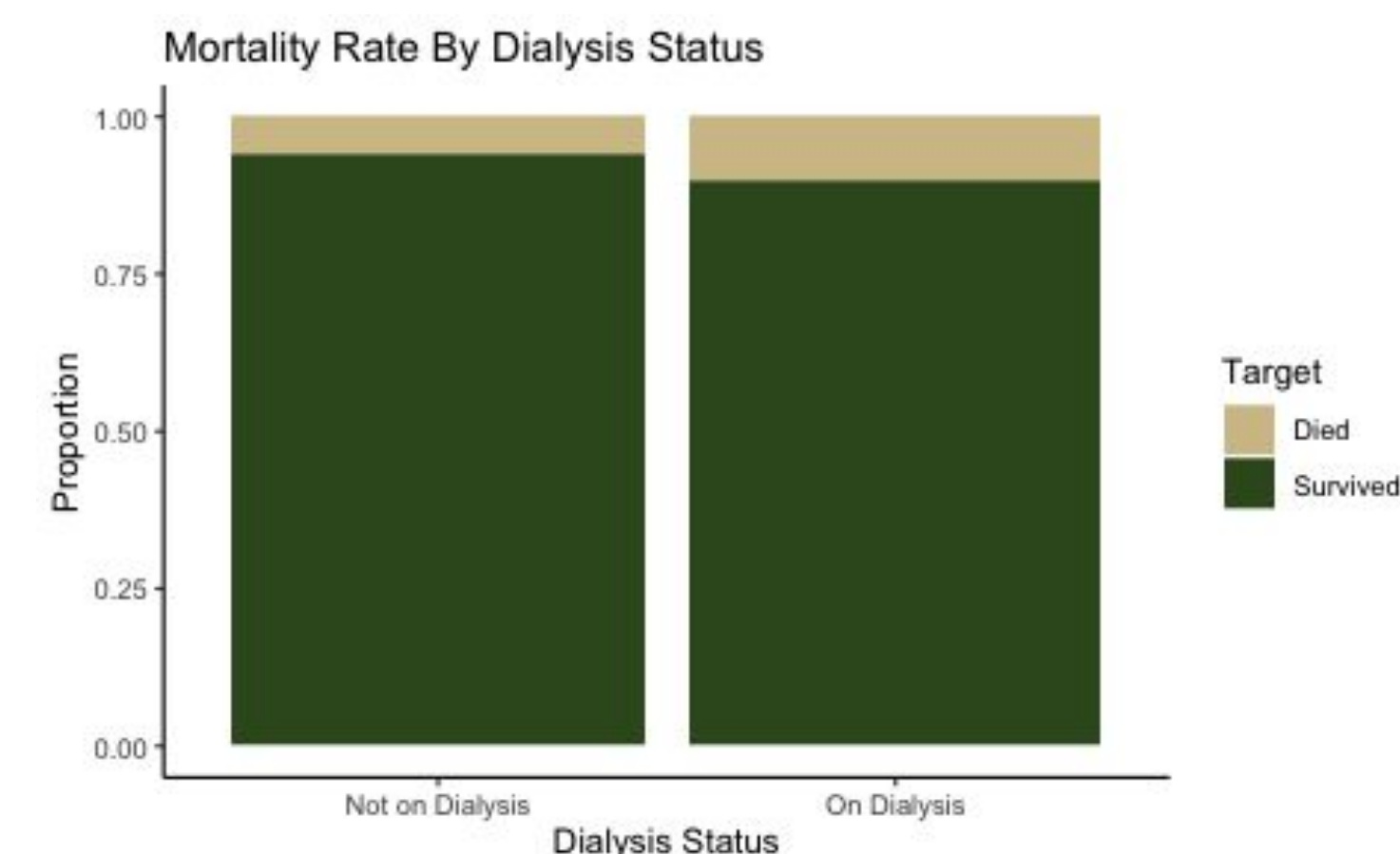
Patients on dialysis saw a higher mortality rate than patients not on dialysis.

Model Performance

Random Forest

	Truth	
Prediction	Death	Survive
Death	238	1300
Survive	490	7124

Recall = 0.32, Accuracy = 0.80

Feature Importances

Feature	Importance
On Dialysis	.021
CPRA (donor match %)	.020
Height	.017
Weight	.016
…	…
Diabetes status	.004

Diabetes was the least important feature for predicting mortality.

Conclusions

Simple parametric models lacked the complexity to accurately predict these health outcomes, requiring more powerful methods like Random Forest. Our approaches were aimed to minimize false negative rates. Simple measurements like weight and BMI, as well as socioeconomic factors like income and education level, were weakly to moderately predictive of diabetes and mortality.