COVID-19:
Which Factors Impact Mortality?

By Nick Wawee, Connor MacMillan, Kshitij Saxena, & Grant Ferrell

COVID-19:

Which Factors Impact Mortality?

Nick Wawee, Connor MacMillan,
Kshitij Saxena, Grant Ferrell

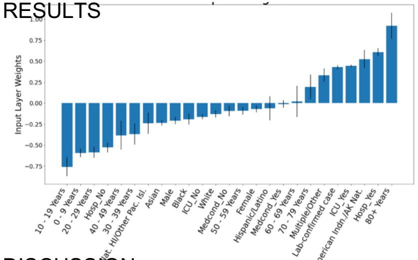
INTRO

- COVID-19 is a deadly pandemic. Which demographic factors increase that deadliness
- Factors that increase mortality should show up in fatal cases

METHODS

1. Use the CDC Public Use data
2. Focus only on relevant inputs. Final data shape: (328,977 x 8)
3. Find correlations
4. Use Logistic Regression with 5 K-Folds.
5. Use GridSearchCV and MLPClassifier with 5 K-Folds
6. Compare each model's metrics
7. Use weights and coefficients to find each feature's impact

RESULTS

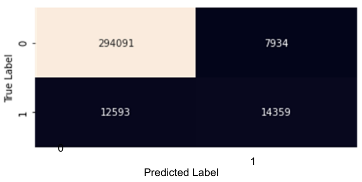


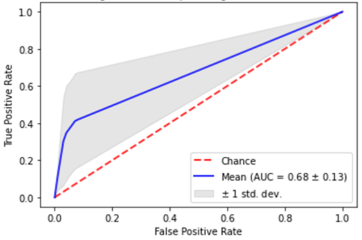
DISCUSSION

Both models performed similarly, but the MLP was more consistent. The closer the weight is to 0, the less important to mortality it is. Being 80+ has the highest impact, and some races/ethnicities are more vulnerable.

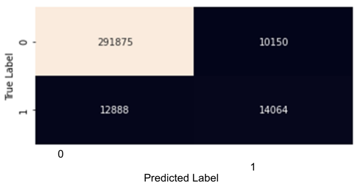
Age has a **higher impact** on the mortality rate of COVID-19 than **pre-existing medical conditions**

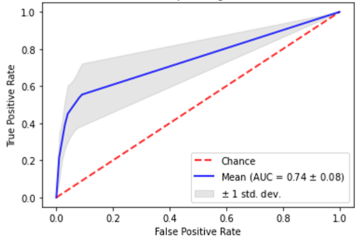
Logistic Regression





Multi-Layer Perceptron Classifier

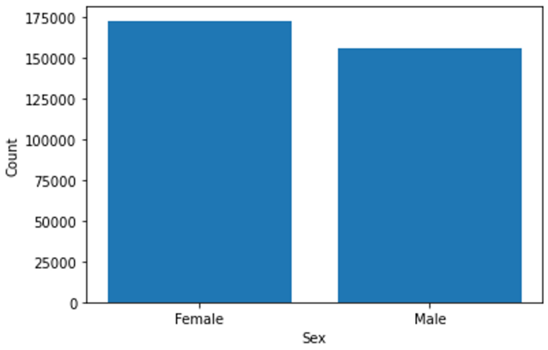


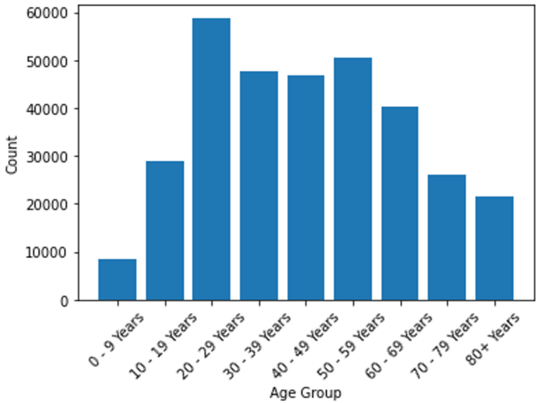


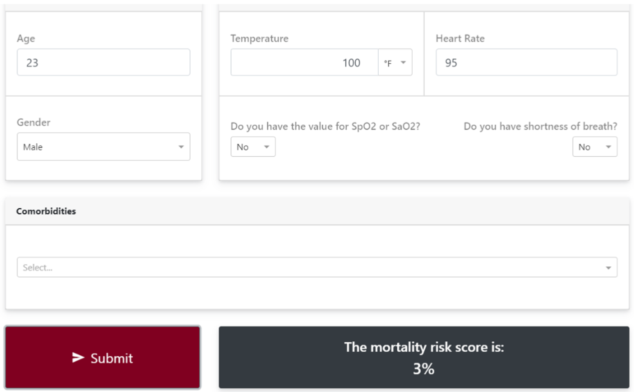
BGSU. DEPARTMENT OF Computer Science
BOWLING GREEN STATE UNIVERSITY

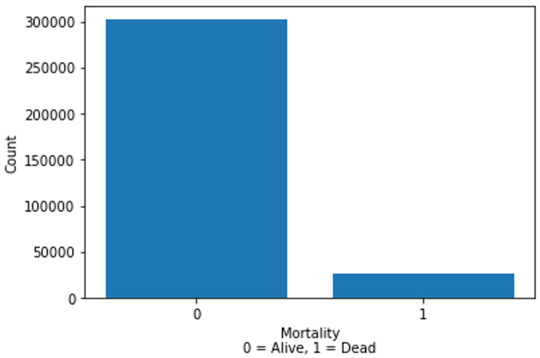
Introduction

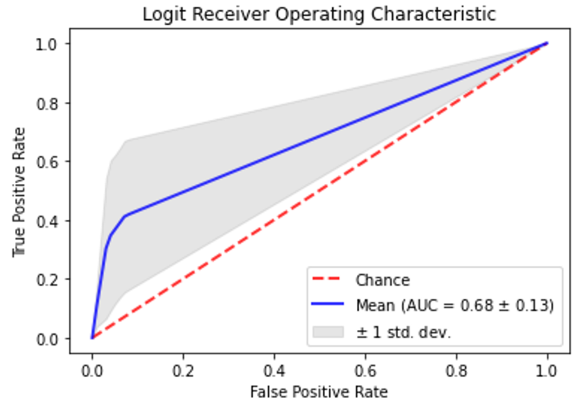
- Covid-19 Case Dataset from the CDC
- We removed rows with missing and unknown values
- What Raises Mortality Rate?
- Mortality Calculator?
- ethnicity, sex, med-condition, and age

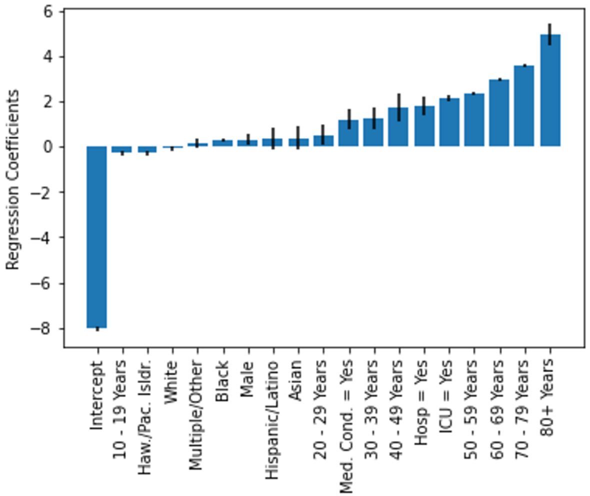




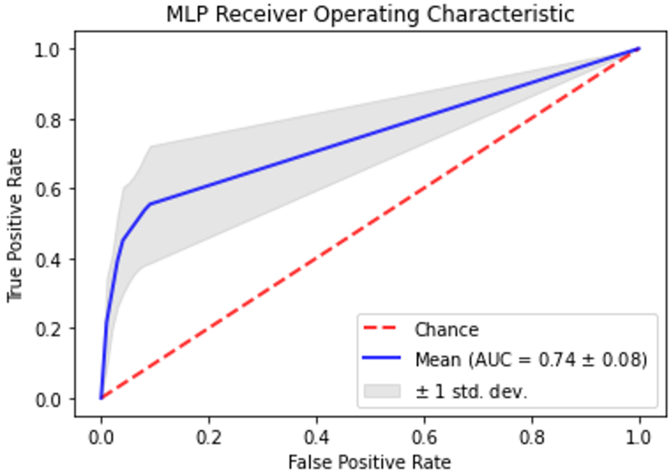




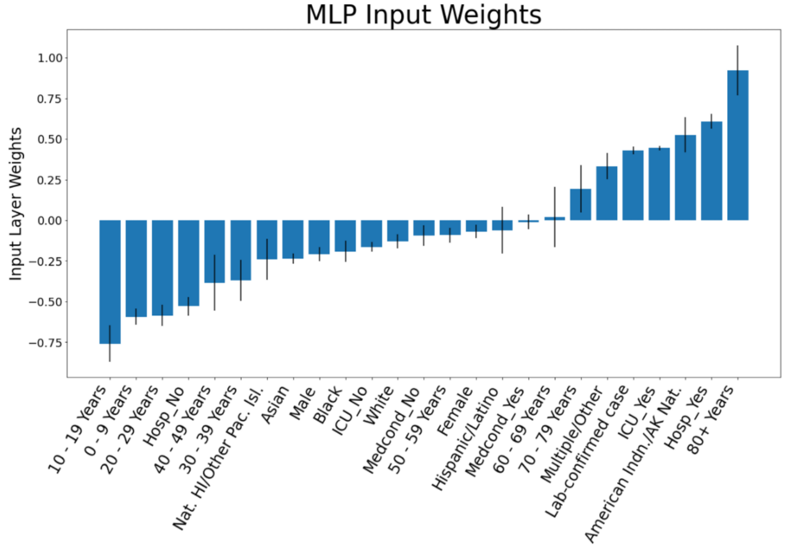


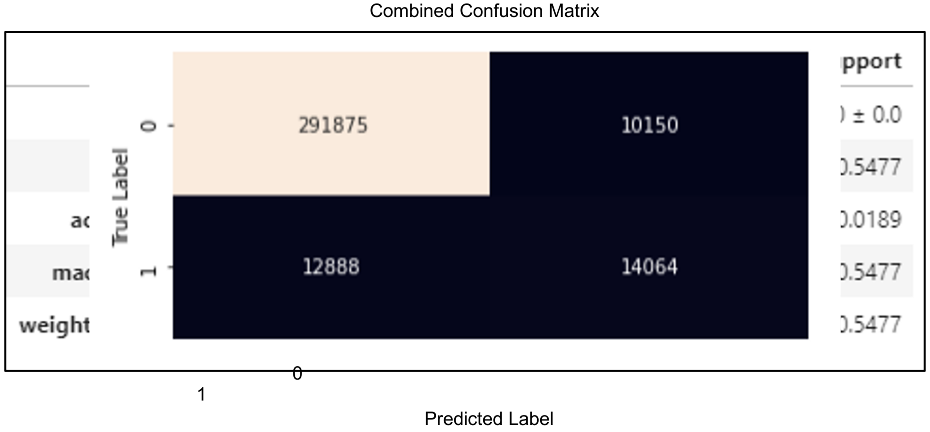


- **Logistic Regression**
  - Probabilities calculated using logit model ⇒ dead if p >= 0.5
  - K=5 Folds, ~330,000 records
  - Key Classification Metrics:
    - F1 = 0.683 ± 0.132,
    - AUC = 0.676 ± 0.149
  - Odds of mortality changes:
    - 0-9 yrs ⇒ 80+ years : **143**
    - ICU? ⇒ **8.59**
    - Hospitalized? ⇒ **6.21**
    - Medical Condition? ⇒ **3.36**
    - Race Comparisons
      - American Indian/Alaskan ⇒ Asian: **1.47**
    - Female ⇒ Male: **1.39**



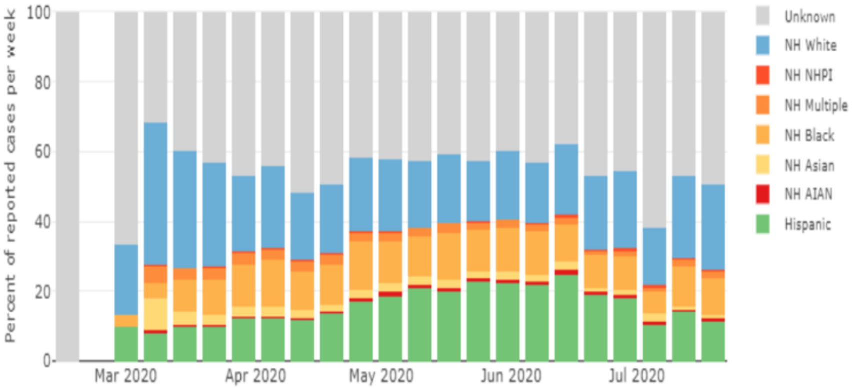
- **MLP Classifier**
  - GridSearchCV used to find optimal parameters
  - Stratified K-Fold for Cross-Validation
  - Input Layer Weights for Feature Impact



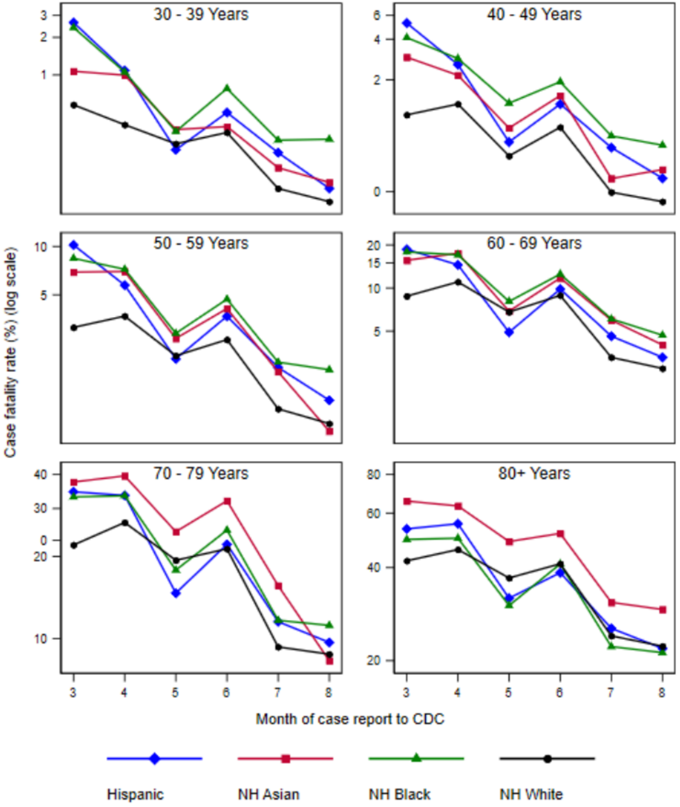


Background

- Many articles also analyze the same dataset from CDC, results are mostly consistent with ours
- 80+ year old had highest case fatality
- Contrary results to many studies for different ethinc groups



[2] Distribution of Covid-19 cases in the U.S.

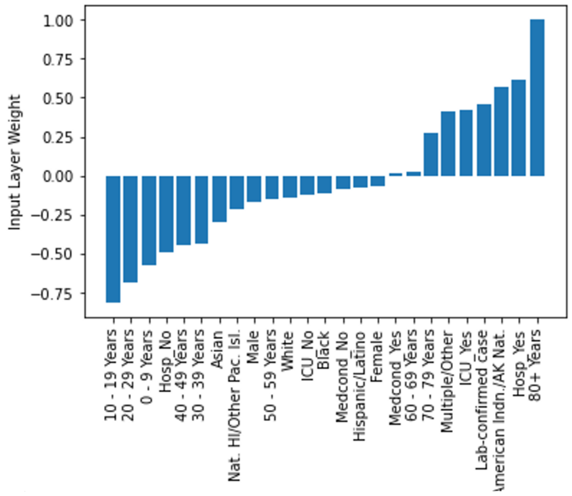


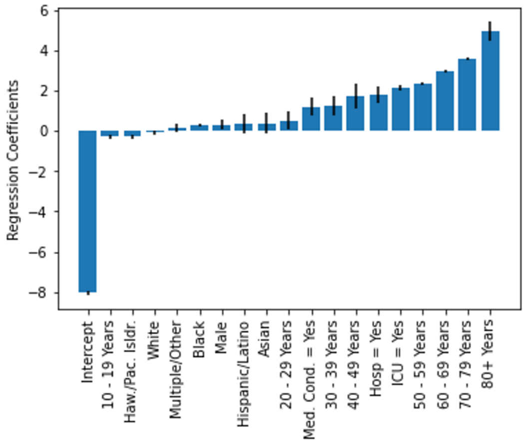
[1] Age adjusted data showing case fatality for certain age groups

Conclusion

- Case surveillance data from CDC was analyzed
  - Data sized from 1,048,576 to just 328,977 rows

- Logistic regression and MLP Classifier considered but they didn't agree with every single feature
- -Age is the biggest factor in for predicting low and high chance of death
- -Another interesting find is that males have lower mortality than females in MLP model, but higher mortality in logit
- - Being admitted to the hospital and ICU both increased your chances of dying.





Any Questions?

References

[1] M. S. Fabic, Y. Choi, D. Bishai, “Deaths among COVID Cases in the United States: Racial and Ethnic Disparities Persist,” *medRxiv*, 2020.

[2] M. S. Fabic and Y. Choi, “Trends in age distribution of COVID-19 cases, hospitalizations, and deaths by race in the United States”, 13-Aug-2020. [Online]. Available: osf.io/preprints/socarxiv/7edgu.

[3] B. T. Rumain, M. Schneiderman, A. Geliebter, “Prevalence of Covid-19 in Adolescents and Youth Compared with Older Adults in States Experiencing Surges”, medRxiv, Oct. 2020.

[4] Centers for Disease Control and Prevention, "COVID-19 Case Surveillance Public Use DataCase Surveillance," 11 November 2020. [Online]. Available: https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf. [Accessed 16 November 2020].

[5] Statsmodels, [Online]. Available: https://www.statsmodels.org/stable/index.html. [Accessed 16 November 2020].

[6] scikit-learn, [Online]. Available: https://scikit-learn.org/stable/. [Accessed 16 November 2020].

[7] A. Patricio, R. S. Costa and R. Henriques, "COVID-19 in Portugal: predictability of hospitalization, ICU and respiratory-assistance needs," COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv, vol. Preprint, pp. https://www.medrxiv.org/content/10.1101/2020.09.29.20203141v1.article-info, 2020

[8] A. S. Yadaw, Y.-c. Li, S. Bose, R. Iyengar, S. Bunyavanich and S. Pandey, "Clinical features of COVID-19 mortality: development and validation of a clinical prediction model," Lancet Digital Health, vol. 2, p. e516–25, 2020.

[9] Noam Barda, Dan Riesel, Amichay Akriv, et al., "Developing a COVID-19 mortality risk prediction model when individual-level data are not available," Nature Communications, Vols. 11, no. 4439, 2020.

[10] R. Bhatia, J. Klausner, “Estimating individual risks of COVID-19-associated hospitalization and death using publicly available data”, Cold Spring harbor Laboratory Press, 2020.