



GEORGETOWN UNIVERSITY

Prediction Fatal Aviation Accidents Using Deep Learning and Neural Networks

Qiaowei Jiang, Shiqi Ning, Guanzhi Wang, Xi Yang
Georgetown University Analytics Program

Abstract

This project researched on predicting fatal aviation accidents through various deep learning and neural network models. Multi-layer Perceptron Neural Network is used to predict the injury severity level (fatal or non-fatal). Dropout on the hidden layers of neural network is then researched and tested. The usage of dropout with proper dropout rate significantly ameliorate potential over-fitting problem. Super learner, which is a stacked ensemble model, is also applied to predict fatal aviation accident. However, super learner does not enhance prediction results and model robustness in this study.

Introduction and Problem Statement

Aviation transportation serves as a vital tool that provides people with mobility and access to various activities across the world. Even though aviation transportation has been said to be the safest way of traveling, according to US Department of Transportation (DOT) (2012), higher risk of safety can be posted due to various aspects:

- Human Behavior:** human error, fatigue, driver health, risky human behaviors such as drunk driving or being distracted by other means of communication
- Emerging and Automated Technologies:** fetal design of the public transportation tools that can cause fatality, and automatic pilot
- Weather-related Factors:** heavy rain and heavy fog.

This projects will thus identify the factors that cause the accidents, integrate them and find the weights of these factors by their importance in terms of severity level.

Related Work

Academic: New method of accident severity analysis and new quantitative method of predicting and assessing accident are proposed by Miao (2018). Mabrouk (2016) applied artificial intelligence techniques as well as the development of several approaches and tools to help with analysis and assessment of the safety for railway transport systems in France.

Government agency: NYPD uses the roll out of a system, ShotSpotter (SST) and Domain Awareness System (DAS) to promote their operation and investigating gunshot, applying tools such as Hunchlab, or ShotSpotter Mission. The predictive analysis (mass) implementation of identifying and mapping criminality profile of different regions has gradually becomes the new focus .

Description of Data Sources

Our data is collected from the National Transportation Safety Board (NTSB) aviation accident .The database can be queried by several filters and can be downloaded into XML format or delimited text format. The link to the database is attached:

https://www.nts.gov/_layouts/nts.gov/aviation/index.aspx

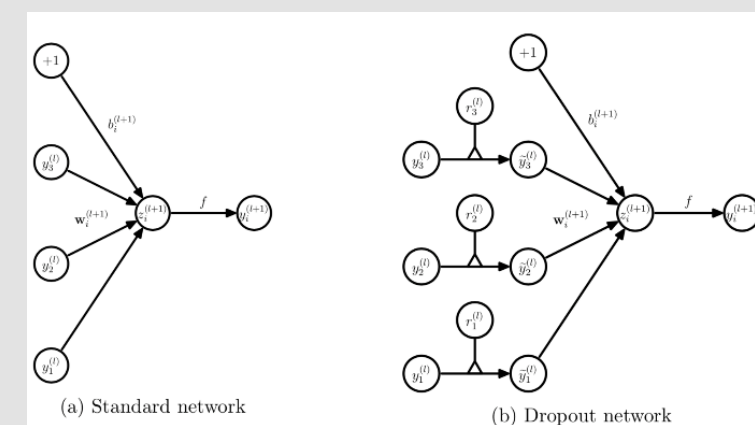
Data Description

Only features of interest are preserved, including:

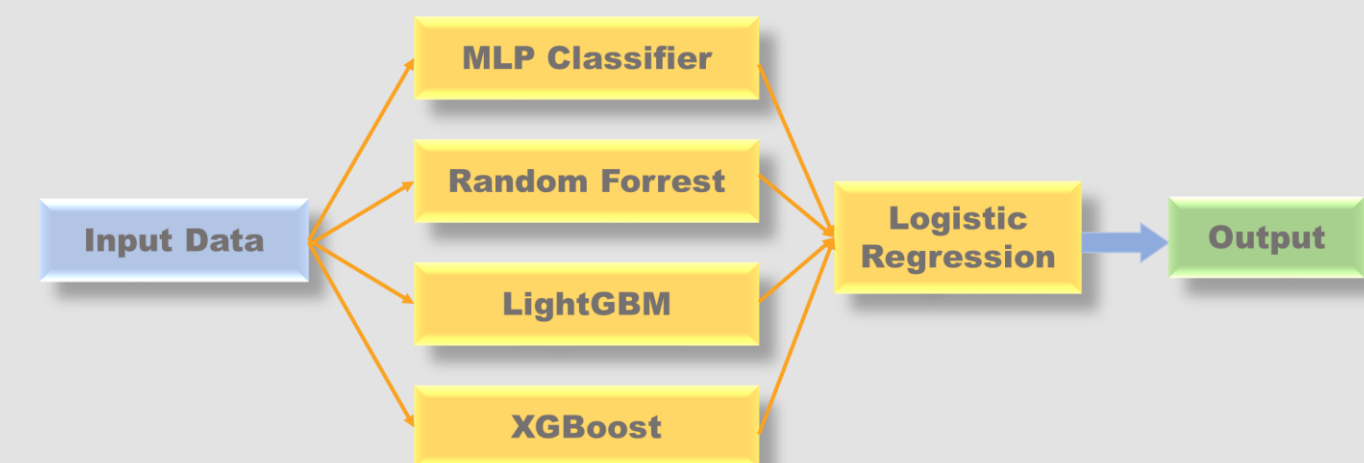
- 1 response:** Injury Severity [Fatal or Non-Fatal]
- 14 features:** Year, Month, Weekday, State, Investigation Type, Country, Amateur Built, Aircraft Category, Aircraft Damage, Make, Broad Phase of Flight, Purpose of Flight, FAR Description, and Weather Condition.

Description of Models

- Multi-Layer Neural Network Model with Dropout
 - Two hidden Layers are added, the first hidden layer is considered to use 20-64 neurons, and second hidden layer is considered to use 10-32 neurons.
 - In dropout networks, drop out layer is applied and 30%-70% dropout rate are tested.
 - Neural networks without dropout are also tested for comparison.



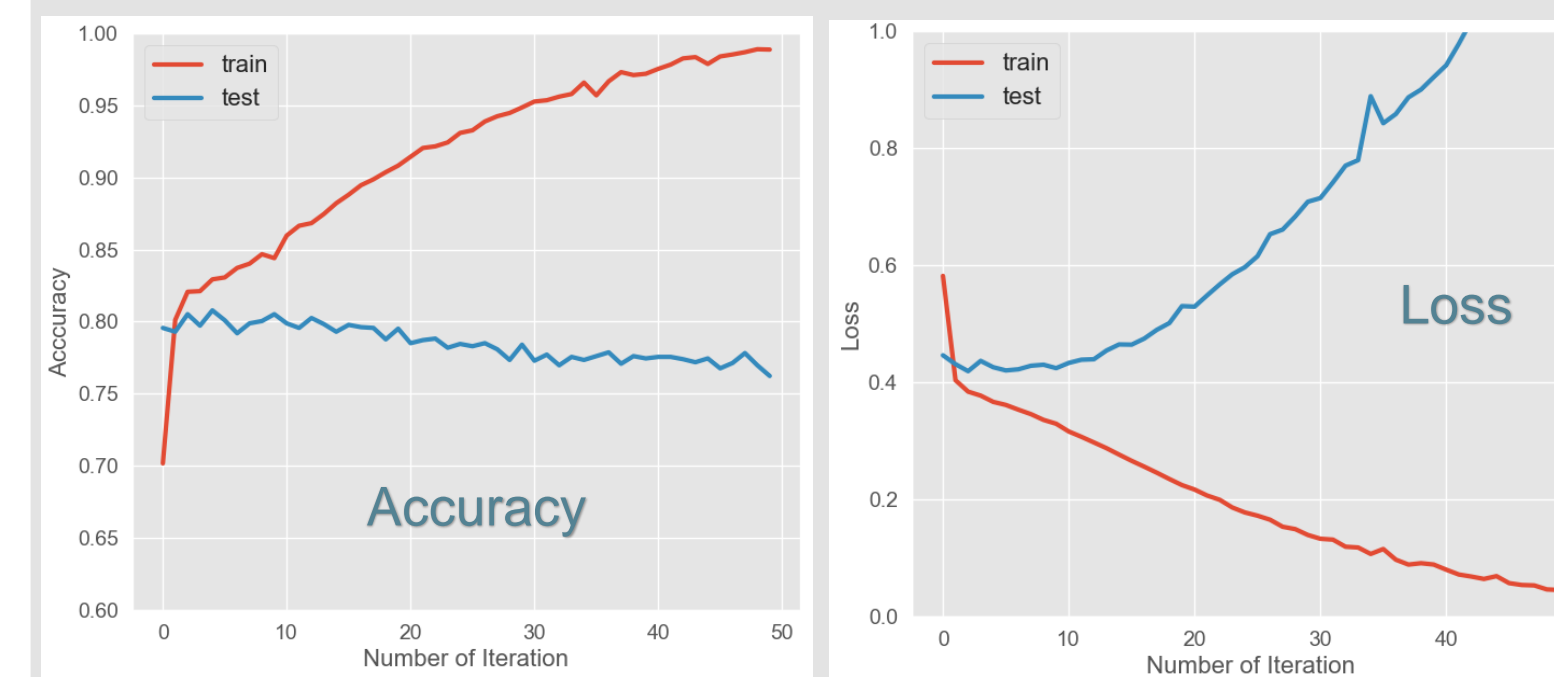
- Usage of Super Learner
 - Super learner model which is also known as stacking ensemble is used to enhance base machine learning model performance on regression or classification problem.
 - Super learner has been applied in biology domain including genetics, medicine, epidemiology, and healthcare .
 - Super learner algorithm is proposed by Polley and van der Laan (2010) as a powerful tool for machine learning.



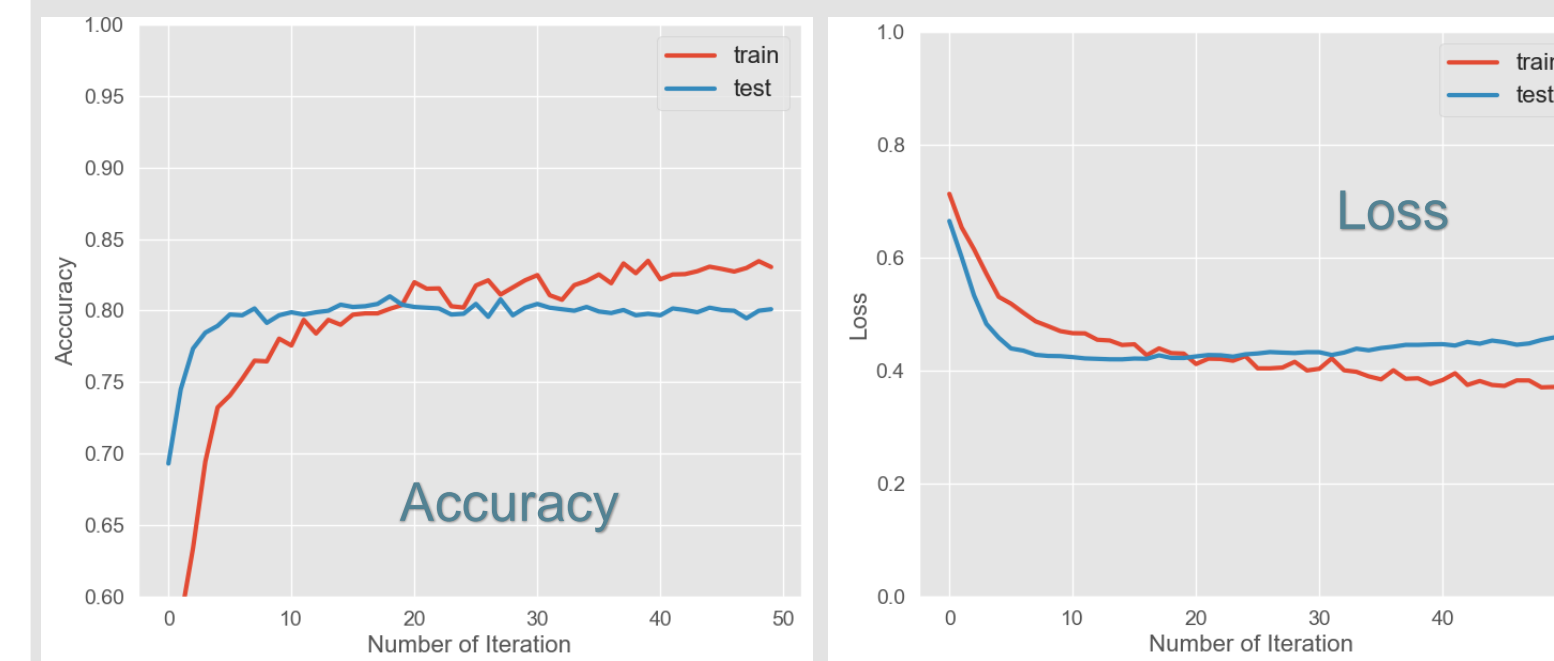
Analysis of Results

Accuracy or Loss Function versus Number of iterations

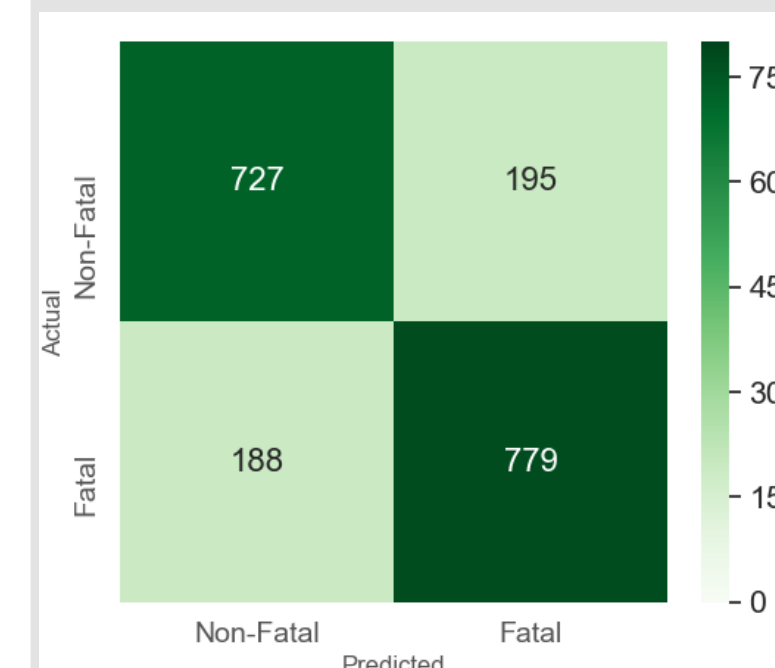
First Hidden Layer: 40 Neurons [Without Dropout]
Second Hidden Layer: 20 Neurons [Without Dropout]



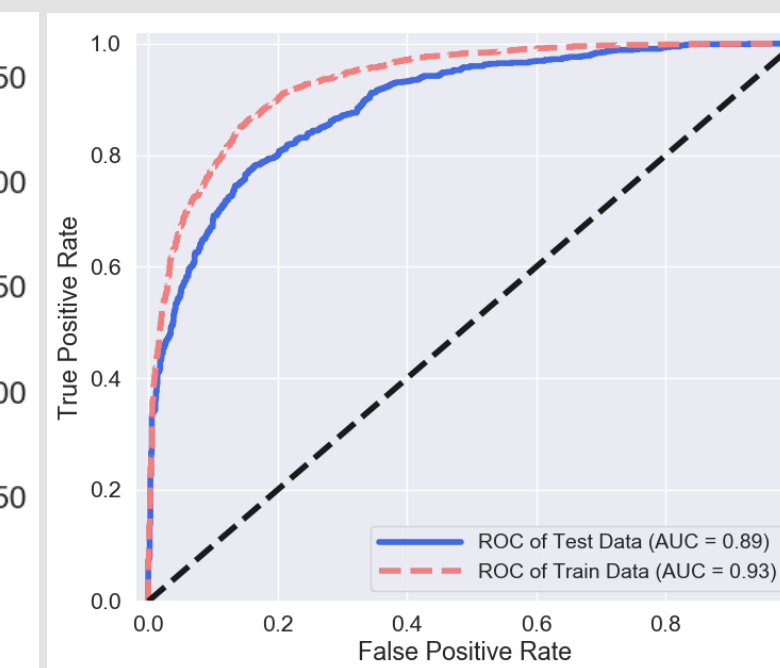
First Hidden Layer: 40 Neurons [Dropout = 0.7]
Second Hidden Layer: 20 Neurons [Dropout = 0.7]



Confusion Matrix



ROC Curve

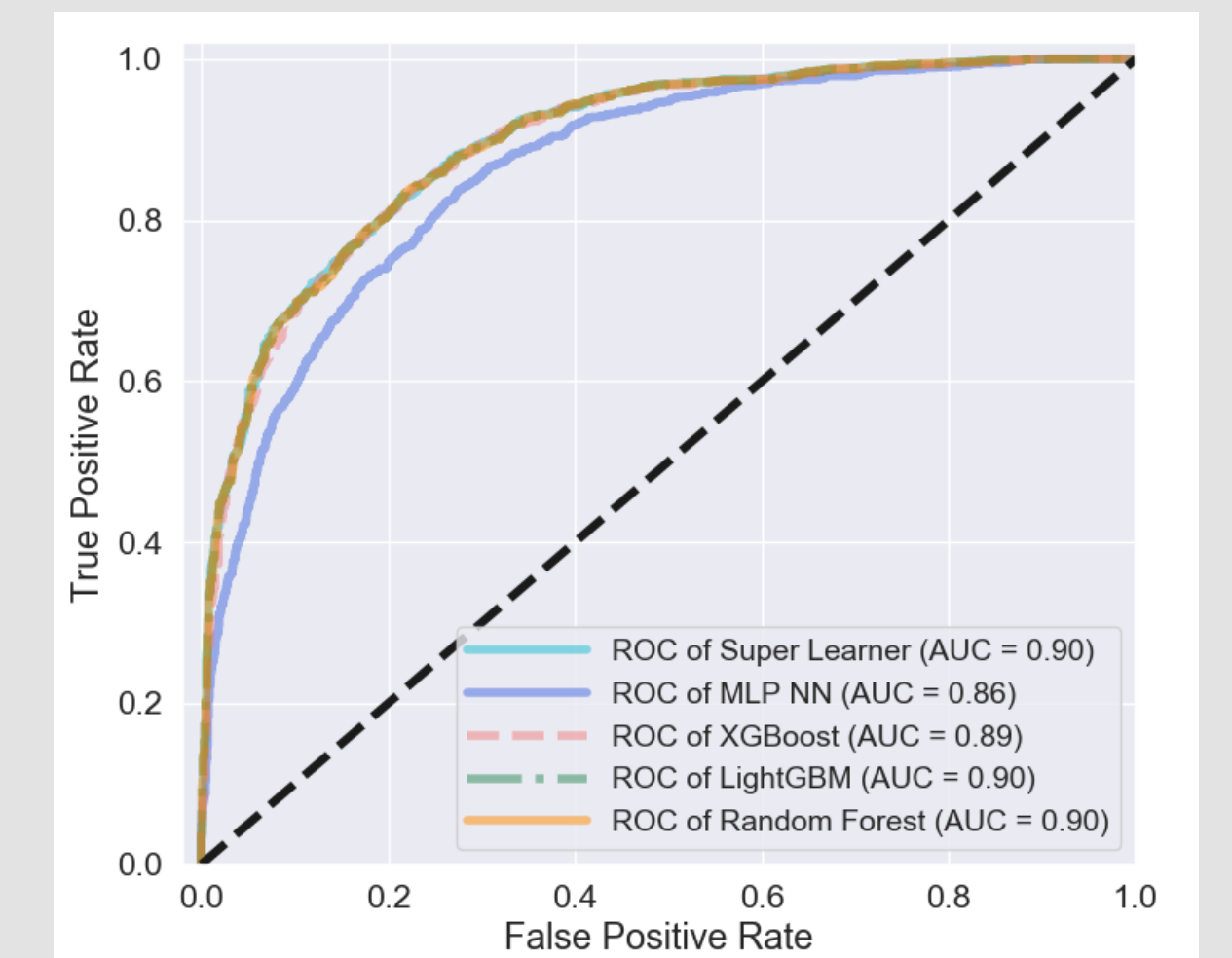


Test Dataset Prediction Summary

	precision	recall	f1-score	support
0	0.79	0.79	0.79	922
1	0.80	0.81	0.80	967
accuracy			0.80	1889
macro avg	0.80	0.80	0.80	1889
weighted avg	0.80	0.80	0.80	1889

Analysis of Results

Comparison of Super Learner, MLP Neural Nets, and other Classifiers



Conclusion

- Dropout (neural networks) significantly helps improve overfitting problem in our case. From the plots shown, the difference between train loss and test loss set as well as the difference between train accuracy and test accuracy are reduced when proper dropout rate is applied.
- Dropout algorithm guarantees the overall accuracy level of the neural network model. Also, dropout allows training with the more important hidden nodes.
- Super learner, which is another “deep” learning algorithm we applied using other classifiers as base learners, does not help with building more robust model in our study. However, super learner may have a broader future in research.

References

- Miao, Z. (2018). Transportation Safety Analytics with Statistical Machine Learning. Retrieved from <https://repository.arizona.edu/handle/10150/631360>
- Machine learning from experience feedback on accidents in transport - IEEE Conference Publication. (2016). Retrieved from <https://ieeexplore.ieee.org/document/7939874>
- Tami Lin, Malgorzata Rejniak. (2018). Smarter New York City: How City Agencies Innovate. Columbia University Press
- New York Police Department.(2016). The way moving forward. NYPD Publication
- Chantele Dubois.(2017). The future of AI and predictive Policing. Retrieved from <https://www.allaboutcircuits.com/news/the-future-of-ai-and-predictive-policing/>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.
- Polley, E.C. and van der Laan, M.J. (2010). Super Learner in Prediction. U.C. Berkeley Division of Biostatistics Working Paper Series: Working Paper 266.