Assessment of Machine Learning Models in Patients with Diabetes

Andrew Yi, Sr. Business Intelligence Developer, Quality, Safety and Risk Prevention, Vanderbilt University Medical Center

Acknowledgements: The author gratefully thanks **Jason Robert Hessler**, **Neesha Choma**, **Jennifer M Slayton** at QSRP, VUMC, for administrative and technical support, **Henry Domenico** at Department of Biostatistics, Vanderbilt University School of Medicine, for guidance in statistical analysis, and **Advanced Computing Center for Research and Education at Vanderbilt (ACCRE)** for high-performance computing services.

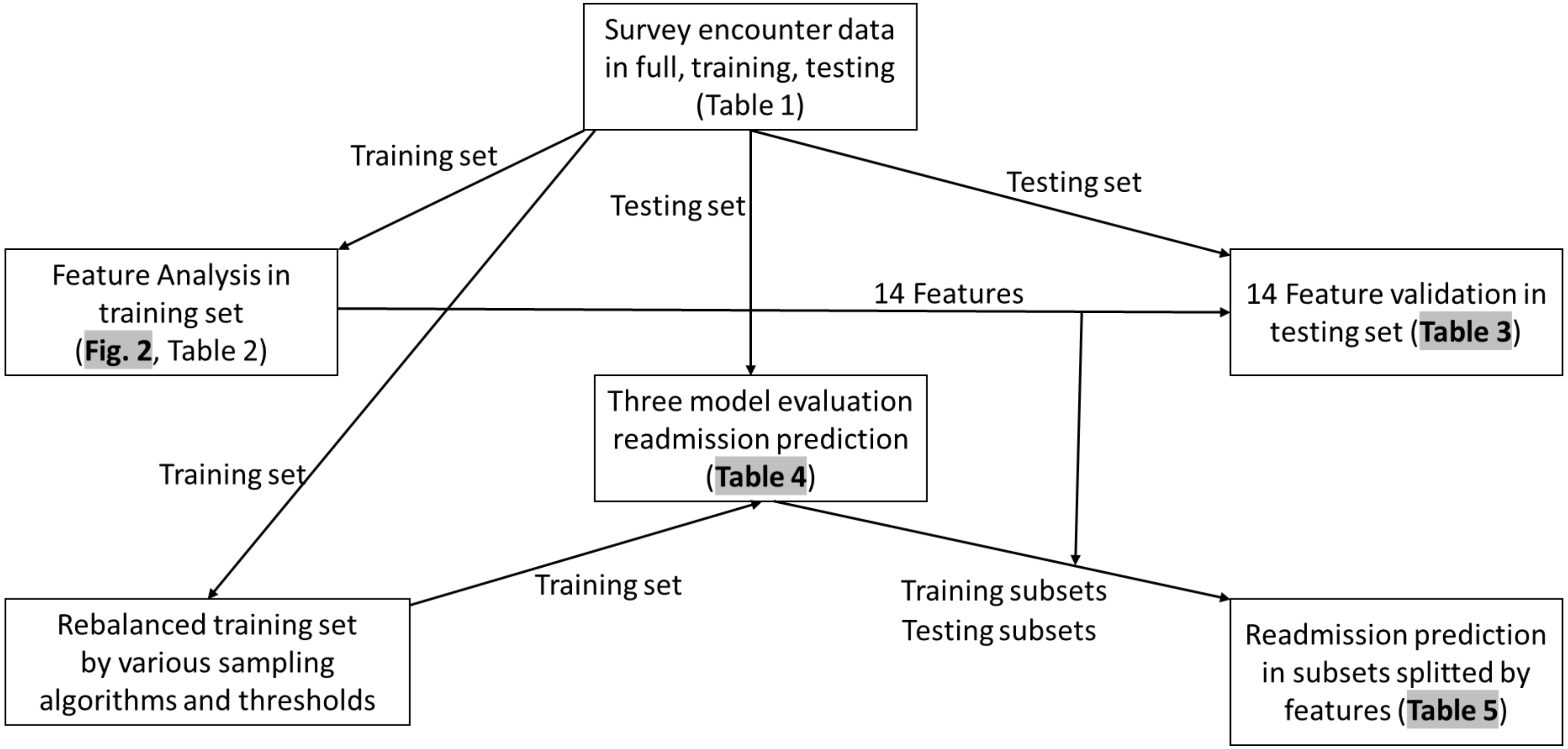
Sources of Funding: None

Background

- Diabetes is a widely spread (34.2 M, 10.5%) chronic disease, and repeating hospitalizations are associated with health care quality and cost
- In order to deploy targeted interventions for readmission reduction, it is critical to identify patients at greater risk and develop accurate predictive models
- Public diabetes dataset from 130 hospitals in US represents 10 years period (1999–2008)
- The dataset was downloaded from

  http://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008#

Study Methods and Design



Survey encounter data in full, training, testing (Table 1)

Training set

Testing set

Testing set

Feature Analysis in training set (**Fig. 2**, Table 2)

14 Features

14 Feature validation in testing set (**Table 3**)

Three model evaluation readmission prediction (**Table 4**)

Training set

Training set

Training subsets
Testing subsets

Rebalanced training set by various sampling algorithms and thresholds

Readmission prediction in subsets splitted by features (**Table 5**)

- After **Preprocess**, the dataset had 69,990 encounter records and 40 data variables which was then randomly split in a 7:3 ratio into a training and a testing subsets
- **Outcome**: 30-day readmission (9%, imbalanced class distribution) and 39 potential predictors
- **Feature Analysis** was done in the training set by Logistic Regression Model (LR) and Validation in testing set
- Numerous **Sampling methods** and three **machine learning models** were examined using LR, Artificial Neural Network (ANN), and EasyEnsemble (EE)
- **Evaluation Metrics** included F1 statistics, Sensitivity, Positive Predictive Value (PPV)

The Most Influential Features

Figure 2. Identification of 14 most influential data features based on LR Model

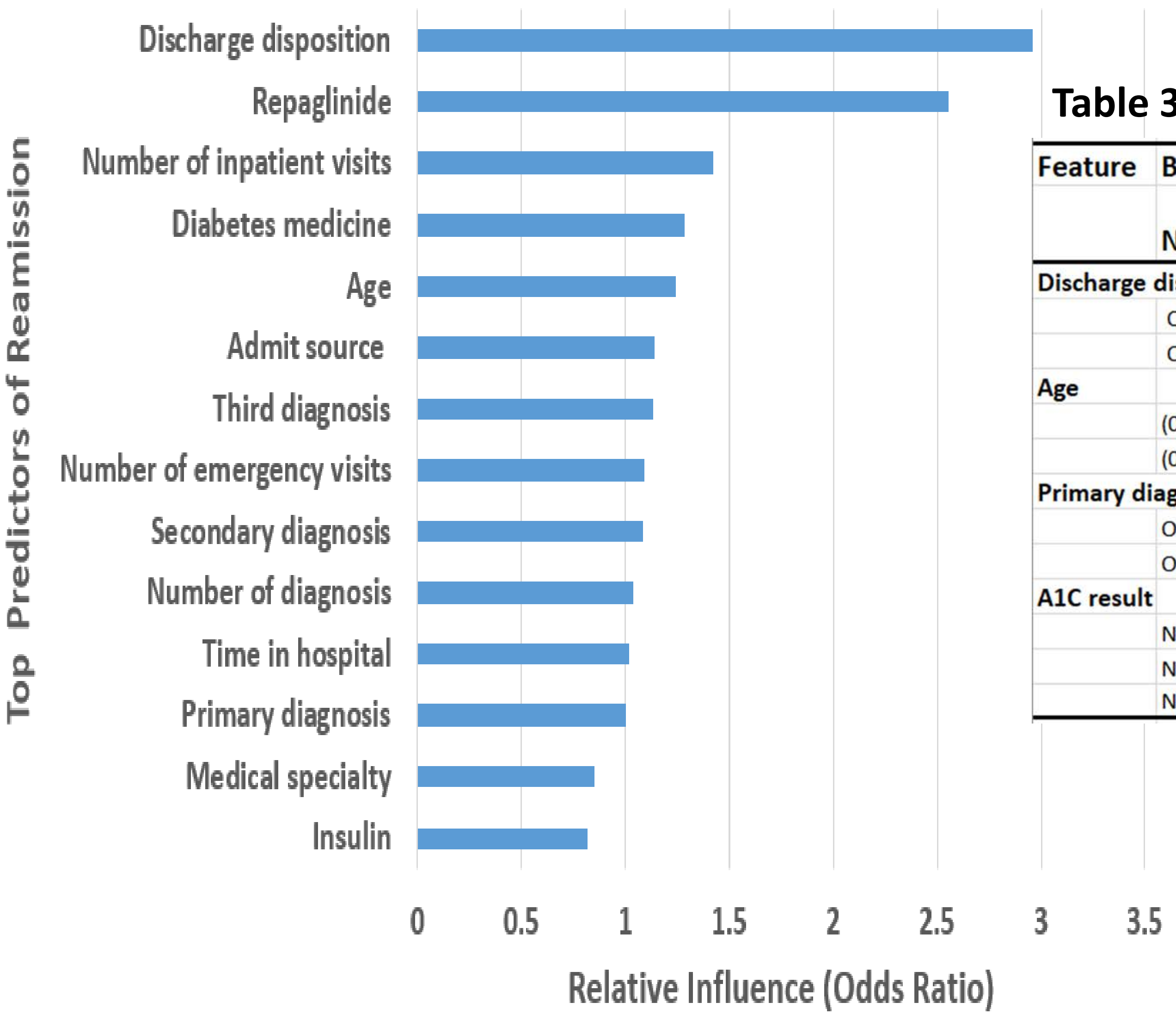
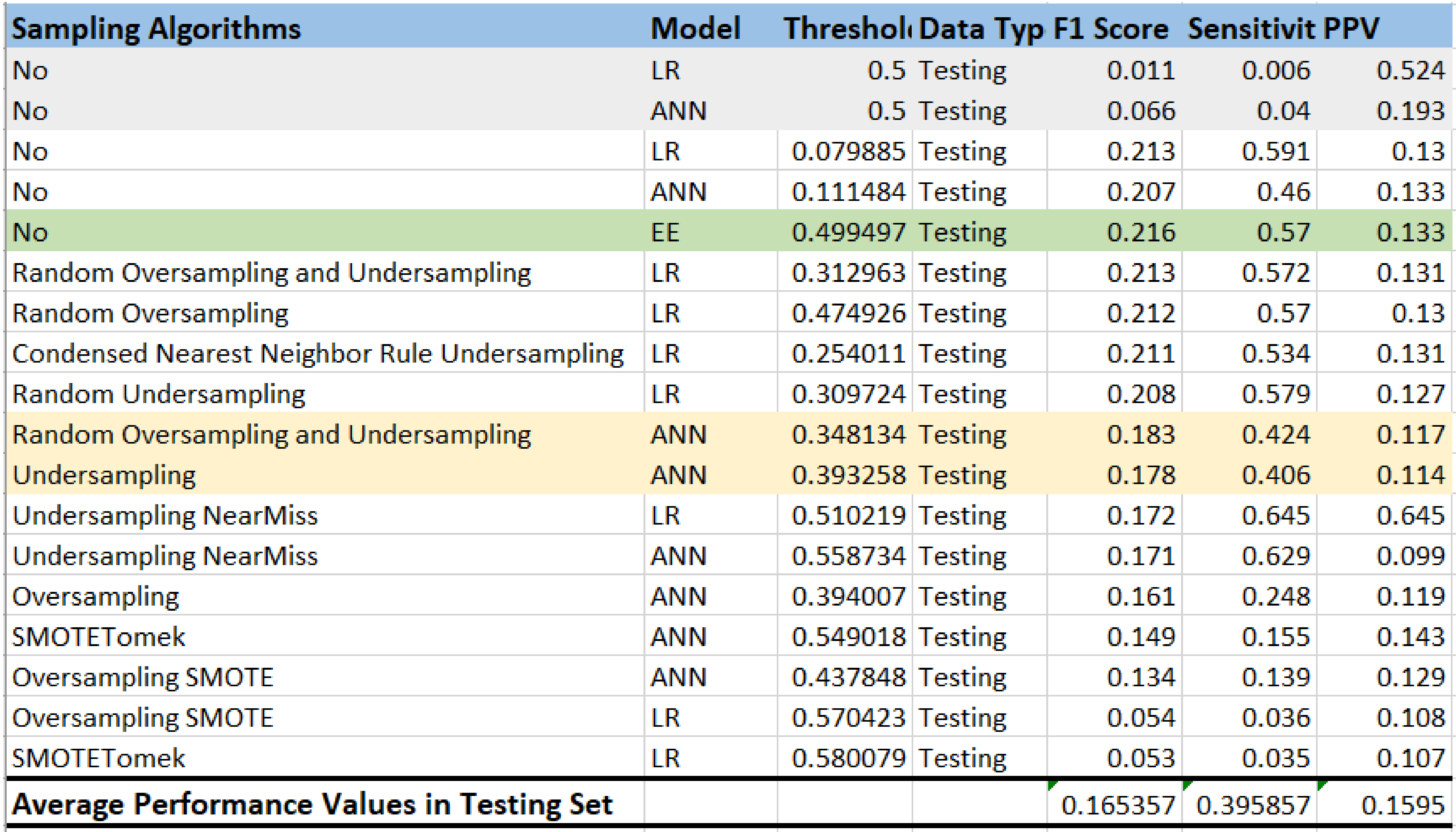


Table 3. Validation of selected data features in the testing data set

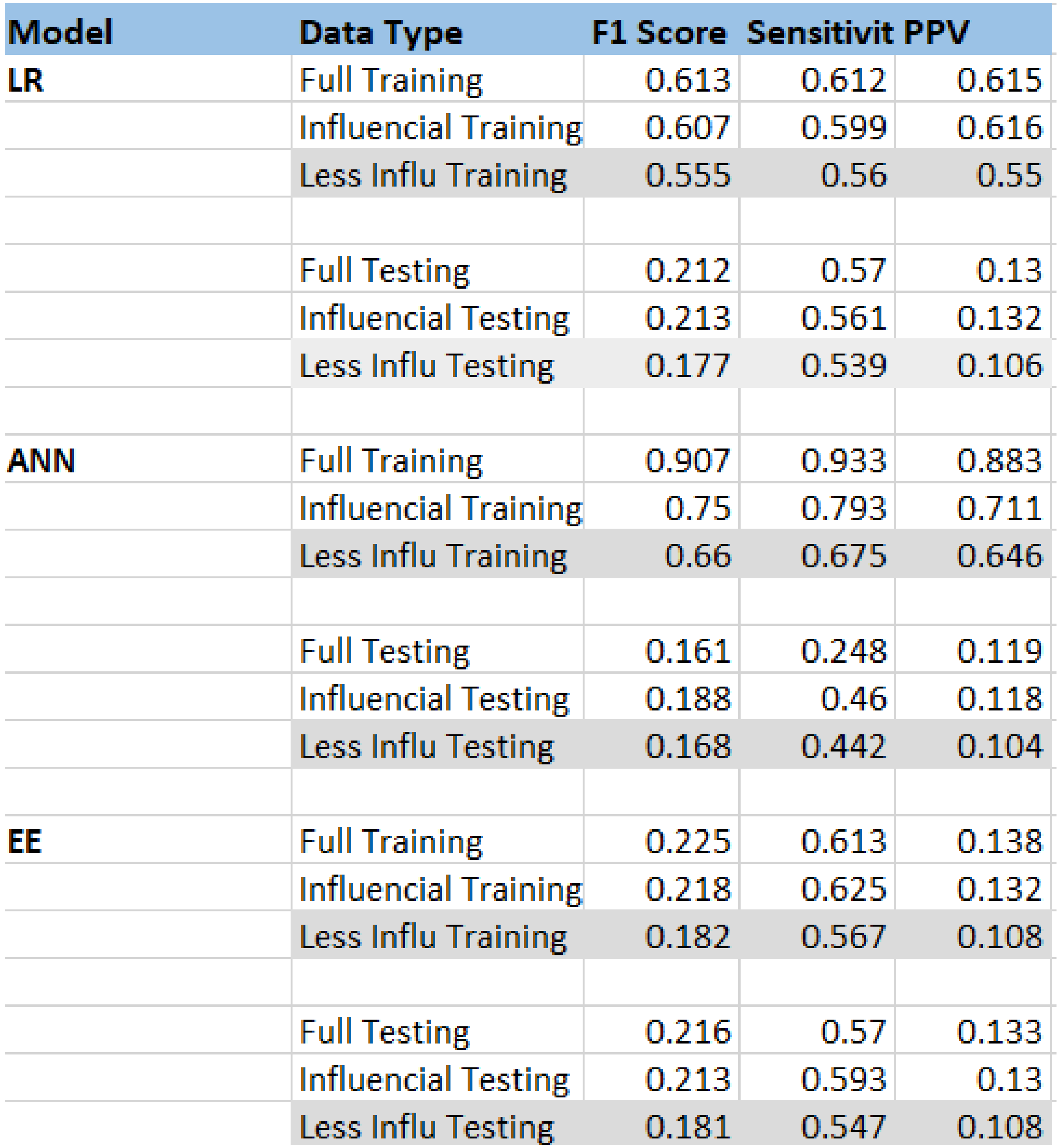
Feature	Base control group			Testing group			Statistics test	
	Name	Size	REA rate	Name	Size	REA rate	Odds Ratio	P value
Discharge disposition								
	Other locations	16371	7.70%	transfer to special care facility	430	27.67%	4.6	< 2.2e-16
	Other locations	16371	7.70%	transfer to acute care facility	4196	13.18%	1.8	< 2.2e-16
Age								
	(0-30)	537	6.33%	(31-60)	6603	7.27%	1.2	0.487
	(0-30)	537	6.33%	(61-100)	13857	10.24%	1.7	0.002138
Primary diagnosis								
	Other diagnosis	3648	9.70%	Circulatory diseases	6367	10.04%	1.0	0.6024
	Other diagnosis	3648	9.70%	Respiratory diseases	2904	6.99%	0.7	8.57E-05
A1C result								
	None	17108	9.35%	Norm	1164	9.02%	1.0	0.7547
	None	17108	9.35%	>7	826	8.11%	0.9	0.244
	None	17108	9.35%	>8	1899	8.53%	0.9	0.26

Table 4. Assessment of Three Machine Learning Models



Sampling Algorithms	Model	Threshol	Data Typ	F1 Score	Sensitivit	PPV
No	LR	0.5	Testing	0.011	0.006	0.524
No	ANN	0.5	Testing	0.066	0.04	0.193
No	LR	0.079885	Testing	0.213	0.591	0.13
No	ANN	0.111484	Testing	0.207	0.46	0.133
No	EE	0.499497	Testing	0.216	0.57	0.133
Random Oversampling and Undersampling	LR	0.312963	Testing	0.213	0.572	0.131
Random Oversampling	LR	0.474926	Testing	0.212	0.57	0.13
Condensed Nearest Neighbor Rule Undersampling	LR	0.254011	Testing	0.211	0.534	0.131
Random Undersampling	LR	0.309724	Testing	0.208	0.579	0.127
Random Oversampling and Undersampling	ANN	0.348134	Testing	0.183	0.424	0.117
Undersampling	ANN	0.393258	Testing	0.178	0.406	0.114
Undersampling NearMiss	LR	0.510219	Testing	0.172	0.645	0.645
Undersampling NearMiss	ANN	0.558734	Testing	0.171	0.629	0.099
Oversampling	ANN	0.394007	Testing	0.161	0.248	0.119
SMOTETomek	ANN	0.549018	Testing	0.149	0.155	0.143
Oversampling SMOTE	ANN	0.437848	Testing	0.134	0.139	0.129
Oversampling SMOTE	LR	0.570423	Testing	0.054	0.036	0.108
SMOTETomek	LR	0.580079	Testing	0.053	0.035	0.107
Average Performance Values in Testing Set				0.165357	0.395857	0.1595

Table 5. Readmission Prediction with Selected Features



Model	Data Type	F1 Score	Sensitivit	PPV
LR	Full Training	0.613	0.612	0.615
	Influencial Training	0.607	0.599	0.616
	Less Influ Training	0.555	0.56	0.55
	Full Testing	0.212	0.57	0.13
	Influencial Testing	0.213	0.561	0.132
	Less Influ Testing	0.177	0.539	0.106
ANN	Full Training	0.907	0.933	0.883
	Influencial Training	0.75	0.793	0.711
	Less Influ Training	0.66	0.675	0.646
	Full Testing	0.161	0.248	0.119
	Influencial Testing	0.188	0.46	0.118
	Less Influ Testing	0.168	0.442	0.104
EE	Full Training	0.225	0.613	0.138
	Influencial Training	0.218	0.625	0.132
	Less Influ Training	0.182	0.567	0.108
	Full Testing	0.216	0.57	0.133
	Influencial Testing	0.213	0.593	0.13
	Less Influ Testing	0.181	0.547	0.108

Conclusions

- Identified fourteen most influential data features
  - with three machine learning models
  - traditional models (LR and EE) performed better in predicting readmission than ANN
- Continuous improvement relies on
  - better prepared data source and more clinical variables
  - optimizing models