Predictive Modeling for Telemedicine Service Demand

Nancy Hung and Agni Kumar
{nancy.hung1068, agnikumar.ea}@gmail.com

Amar Gupta
agupta@mit.edu

The authors are grateful to Robyn Baek of SOC Telemed for the valuable inputs, information, and other help provided to them. The views expressed in this paper are exclusively those of the authors unless expressly stated otherwise via citations or other means.

Massachusetts Institute of Technology

15 May 2018

Predictive Modeling for Telemedicine Service Demand

N. Hung[1, a)] and A. Kumar[1, b)]
Electrical Engineering and Computer Science Department, MIT

(Dated: 15 October 2018)

Telemedicine services range from real-time consultations, remote monitoring, to electronic medical record data transfers. This case study of data from private telemedicine provider, SOC Telemed, explores the matched needs of hospitals and service of telemedicine providers based on consult volume. A predictive model for demand forecasting can optimize telemedicine resources to improve patient care and help hospitals decide how much to invest in telemedicine services.

This study trained consultation data from current hospital telemedicine clients. The model analyzes characteristics, including hospital size (number of beds), per department size and annual volume, patient demographics, time of consultation, reason for consult, and more. Various regression techniques were used to demonstrate a strong correlation between these features and weekly demand with $r^2 = 0.7821$. Reason for Consult in the past week was the strongest predictor for the demand in the next week with $r^2 = 0.7899$.

Keywords: telemedicine, teleneurology, forecasting demand, predictive modeling

I. INTRODUCTION

Telemedicine is a more than 100 year-old concept that describes the process of using information technologies and telecommunications platforms to deliver healthcare services, which range from consultations, education, exchanging patient records data, research, and more[1]. In fact, telemedicine practices date back to the 1920s when ship-to-shore radios transmitted cardiac sounds[4]. Recently, telemedicine market growth was projected to be \$41.8 billion with $> 19\%$ growth from 2018-2022, yet currently only addresses 0.1% of global demand[5,6]. In this paper, we refer to telemedicine as the use of technology to deliver care remotely between patients and providers. Specifically, we are analyzing TeleNeurology, an application of telemedicine to treat patients who require a neurology specialist, such as for acute ischemic stroke, migraines, or epilepsy.

Consultation demand is the quantity and frequency of consult requests from hospitals. Hospitals across the nation are interested in predicting the number of future visits within a specific time (i.e. within one week or one month). This information is valuable because it impacts hospital cost assessment in terms of overcrowding, hospital understaffing, and insufficient bed availability[2].

TeleStroke and TeleNeurology are telecommunication-based services specific to treating patients who have stroke and neurological disorders in need of care from a neurologist. In a randomized blinded, and prospective trial, TeleStroke was more patient-specific, sensitive, and higher predictive values than traditional telephone-based consultations[19]. These are typically deployed in underserved facilities that lack 24/7 stroke expertise[2,13,15]. Previous studies have shown that TeleStroke services performed as reliably as bedside neurologists by comparing their assigned NIHSS scores, the National Institute of Health Stroke Scores, from neurologists in person and neurologists treating via telemedicine[4]. The study concludes that doctors practicing through telemedicine can be used to help emergency physicians administer tissue plasminogen activator (tPA), a drug that dissolves blood clots used to treat ischemic stroke[2,4,6].

TeleStroke has also been shown to be cost-effective, with Incremental Cost-Effectiveness Ratio of \$108,363/QALY in the 90-day horizon and \$2,449/QALY in the lifetime horizon[14]. Similar studies found that the cost-effectiveness of hub-and-spoke TeleStroke can increase the number of patient discharges[16]. Demand prediction could improve these services even further to predict how many clinicians expected to be in the hospital because clinicians on staff in hospital clients and telemedicine clinicians providers can be more efficiently staffed. However by lowering the barrier to get access to neurologists, there is a higher representation of Stroke Mimics, which describes patients having a diagnosis that has symptoms similar to stroke, such as seizures and migraines[17,20].

A. Background Literature

Many hospital departments are interested in forecasting demand for care for a variety of reasons. For example, emergency departments (ED) in the U.S. experi-

[a)]Electronic mail: nhung@mit.edu.
[b)]Electronic mail: agnik@mit.edu.

enced widespread overcrowding issues, which negatively impacts patient quality of care and hospital costs[7]. Predictive studies using Poisson models on EDs have shown with up to 90% confidence that the highest number visits during the week is Monday and increases from 7:00AM until it peaks at noon[2,18].

Another study forecasted hospital bed demand by analyzing prediction techniques: hourly historical average, seasonal autoregressive integrated moving average (ARIMA), and sinusoidal with an autoregression (AR)-structured error term. ARIMA performed the best and was able to estimate demand for bed count 4 to 12 hour in advance[7].

However, to our knowledge, there is no study that evaluates the cost-effectiveness of TeleNeurology implementation by forecasting demand for consultations across different clients. This is true because telemedicine services vary nationwide based on the demands of individual hospitals, which can be on available bed count, hospital type, value-based care or fee-for-service care, hospital staffing, etc. This study can help shed light on what factors affect return on investment for telemedicine services for both hospitals and providers.

B. Data and Methods

The dataset has 411 hospitals and 97,593 consultations since July 2015. The number of consultations were predominantly focused on stroke, followed by Transient Ischemic Attack (TIA), encephalopathy, seizures, and stroke that requires tPA Administration (tPA Stroke). Although less frequent, tPA Stroke is the highest clinical priority. The hospital clients bed count sizes ranged from 10 to 1432. Geographically, the data shows a heavier representation of hospitals in Southeast, which occupy around 35% of the total number of clients. They are in Figure 2, which shows the fields used to run our models.

Within Tele-Neurology, there is an even split between males and females who received consultations: around 53,000 females received consultations compared to 43,000 males, which is summarized by the pool of females dominating at 55% overall. As the number of consultations was dependent on Provider Diagnosis: an overwhelming majority of consultations were based on stroke, followed by TIA, then tPA stroke, which is much more serious and life-threatening.

Data Visualizations

The relationship between tPA administration and a variety of factors, including gender and age, is highly correlational. Although there was no correlational relationship between gender, there is a normal-like distribution that models the relationship between age and tPA (Figure 1). The data shows that tPA Administration accumulates mostly in New Jersey at 278 cases, followed by Pennsylvania at 165 cases (Figure 4).

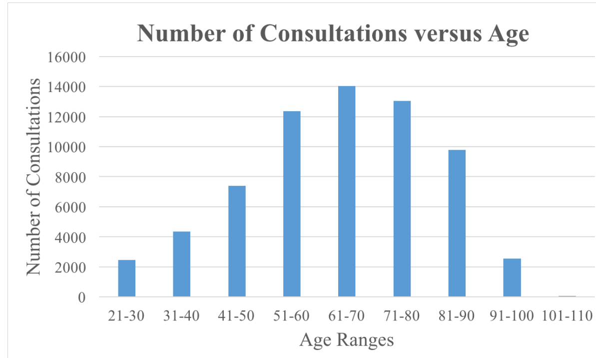

Figure 1. Shows the number of consultations based on the age of the patient.

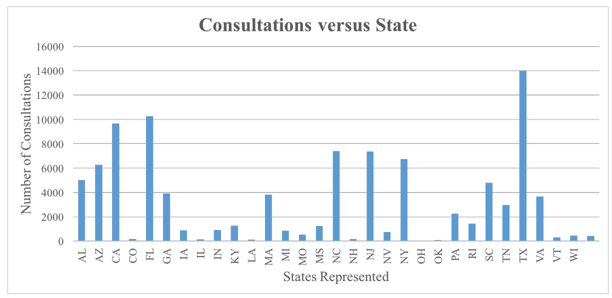

Figure 2. Demonstrates variety of hospitals represented in data.

Interestingly, New York, which had less than half the number of tPA administrations than North Caroline or New Jersey despite experiencing similar demand for TeleNeurology.

Using polynomial regression fitting, tPA administration and the months of the year follow a cubic relationship with r^2 value 0.757.

C. Results

OLS linear regression and polynomial regression models are used to find strong correlations between features and demand. Ridge regression and the lasso were also tested on the dataset to find the best fit model. Ridge regression is used for fine-tuning the complexity of the model with a regularization term, while lasso regression

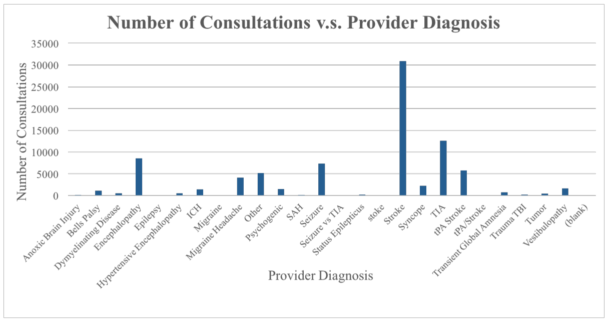

Figure 3. Number of consultations categorized by condition entered by physician.

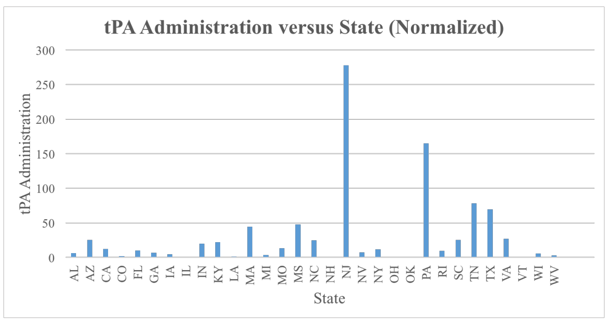

Figure 4. Number of consultations tPA Administration normalized by state.

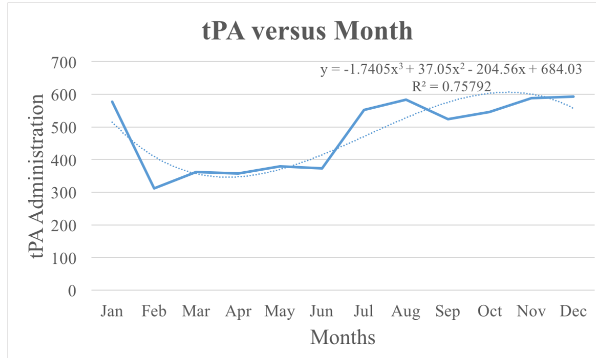

Figure 5. tPA administration varies polynomially based on time of year.

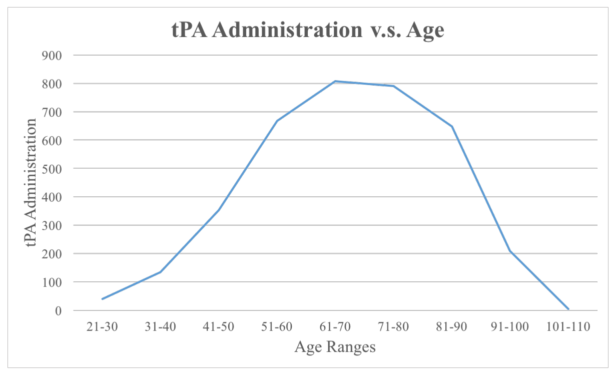

Figure 6. tPA administration varies based on age range.

suits sparse data by forcing the sum of the absolute value of the regression coefficients to be less than a fixed value, causing some certain coefficients to be set to zero and reducing the problem to a simpler model that does not include those coefficients.[12]

Given below are plots depicting the weekly and monthly telemedicine service demand aggregated across all partnering hospitals, over a period of approximately a year.

The weekly demand (Figure 7) appears to be centered around 850-950 patients utilizing telemedicine services. These counts oscillate around this center on an approximately 1.5-week basis, and demand is noted to spike slightly around the holidays and new year periods.

The monthly demand plot (Figure 8) below confirms the above observation, where dips and peaks are observed either once or twice a month (the sudden downward-sloping lines at the immediate left and right of the plots are due to the provided data beginning and ending in the middle of a month).

An OLS linear regression and polynomial regression were conducted on the dataset (Figure 9), where the generated trends are depicted with a dotted green line and solid red line respectively, overlayed on a scatterplot of demand aggregated weekly.

The following results were outputted for the OLS regression, with the slope of the line being 0.202485, the y-intercept as 845.707773, and the correlation coefficient as 0.0357. The low correlation coefficient outputted prompted running a model with additional input variables. After experimenting with various orders of fitted polynomials, training, testing, and cross-validation were executed to generate accuracy scores, which were no longer highly variable from run to run due to an augmented independent variables set in the system.

The following features were used to train the new regression models: Time Zone, Visit Initiated, State, Sex, Age, Service Line, Reason for Consult, Provider Diagnosis, Hospital Type, Bed Count, Stroke Center, Advanced Comprehensive Stroke Center, and Total ER Visits to predict telemedicine consultations. Running linear regression on these features, most of the datapoints matched our predicted value where $r^2 = 0.7821$ (Figure 10). Applying lasso regression on the same features, the $r^2 = 0.7695$, which was less optimal than OLS Linear Regression. Ridge regression on all inputs yielded

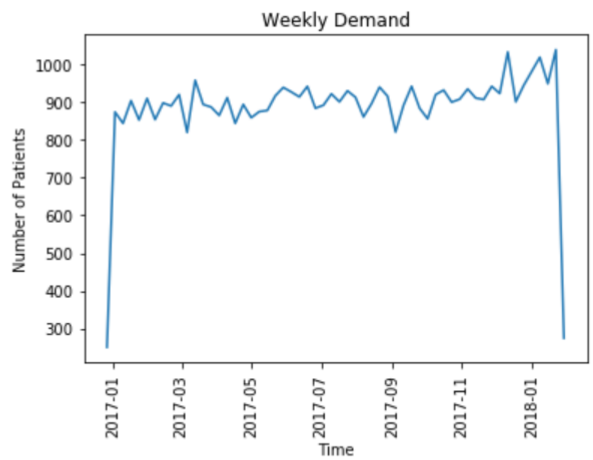

Figure 7. Weekly telemedicine service demand

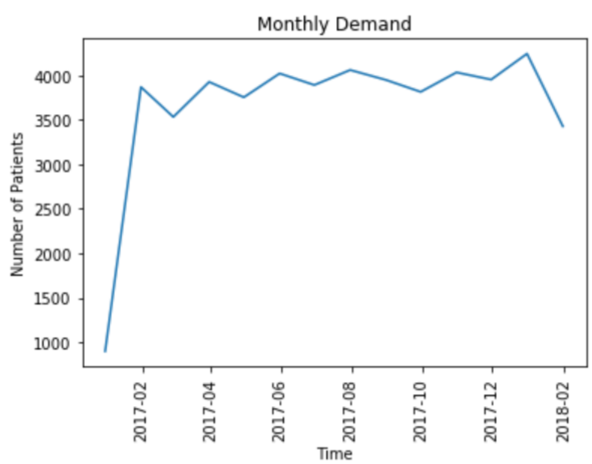

Figure 8. Monthly telemedicine service demand

slightly better results, $r^2 = 0.7829$ (Figure 10), and performed the best overall. This was expected because ridge regression penalizes incorrectly trained labeled predictions.

Running the weekly prediction linear regression model with just Provider Diagnosis, the model had $r^2 = 0.7433$ (Figure 14). Using the same model with Service Line in place of Provider Diagnosis, the accuracy increased to $r^2 = 0.7698$. Weekly prediction with only Reason for Consult performed the best even compared to training all the features together with $r^2 = 0.7899$ (Figure 14). Interestingly, weekly forecasting with Reason for Consult and Provider Diagnosis bumped the accuracy to $r^2 = 0.8057$. It is expected that Reason for Consult and Provider Diagnosis would yield strong results, but it was surprising that the Service Line was also an equally

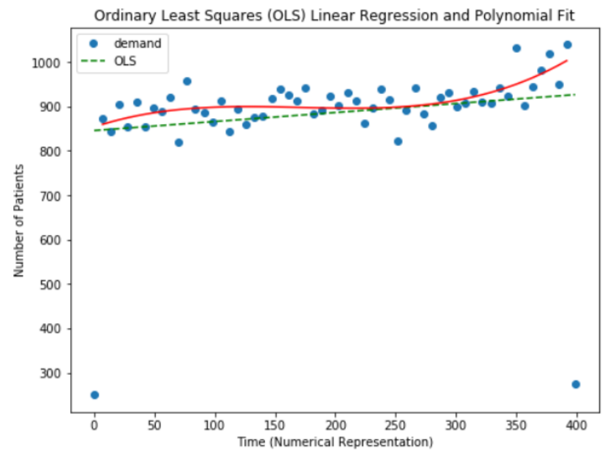

Figure 9. Linear and polynomial fits on weekly aggregated demand

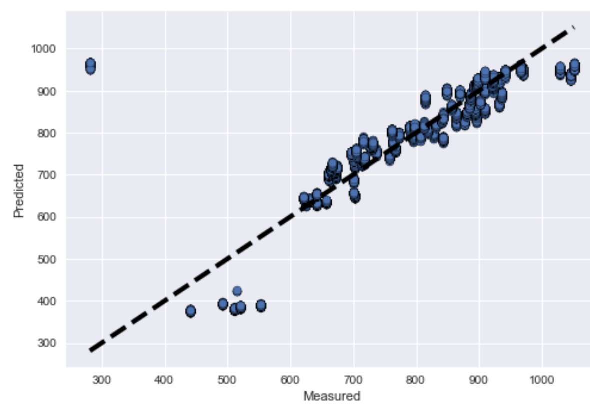

Figure 10. Visualization of relation between measured and predicted demand.

good predicting factor.

The sparse nature of this dataset enabled a slight change in the target variable to cause huge variances in the calculated weights. The plots below, one for each of ridge regression (Figure 11) and the lasso (Figure 12), set a certain regularization (alpha) to reduce this variation. When alpha is very large, the regularization effect dominates the squared loss function, causing the coefficients tend to zero. At the end of the path, as alpha tends toward zero and the solution toward the ordinary least squares, coefficients exhibit large oscillations. In outputting accuracy statistics for the models built for this study, we set an alpha enabling a maximal score.

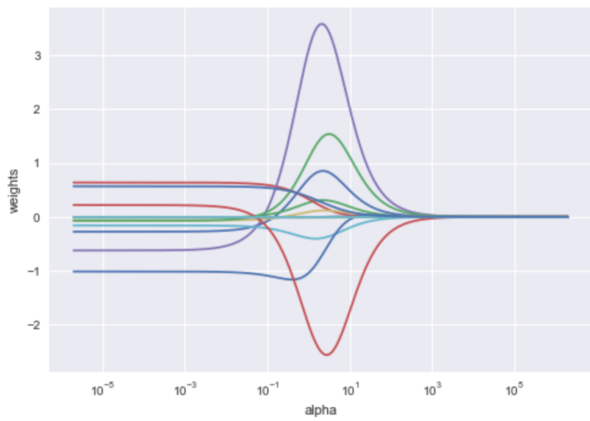

Figure 11. Ridge coefficients as a function of the regularization.

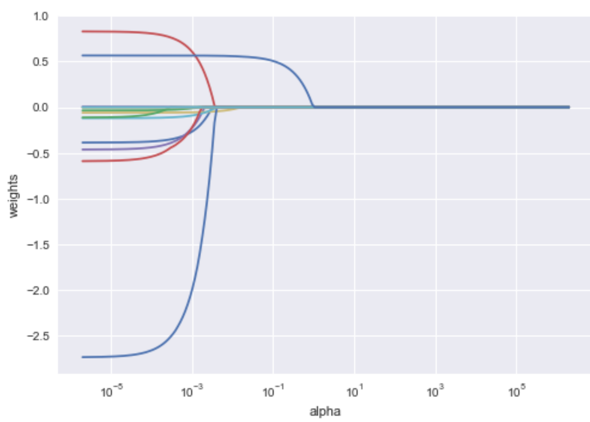

Figure 12. Lasso coefficients as a function of the regularization.

D. Conclusion and Future Work

This paper aims to reasonably predict future consult demand to optimize hospital staffing and analyze return on investment for both parties. This study demonstrates that there is a strong correlation between month/week of a telemedicine consultation request and the predicted number of consultations. We have found that certain combinations of input features, including hospital characteristics and certain consult characteristics (namely Provider Diagnosis, Service Line, and Reason for Consult), yield forecasted telemedicine consult demand with around 78% accuracy overall.

Ongoing work includes improving the accuracy of the current predictive model by adding a classification by service line and the reason for consult and provider diagnosis component to it, in order to further break down the provided forecasted demand outputs. We will continue to work around dataset limitations, which include the dataset consisting entirely of data from one telemedicine provider and its clients, for a single practice, only over a one-year time period. Ideally, we would like more information on the degrees to which telemedicine was utilized in consults over various hospitals, patient demographics data (to analyze which hospitals get greater telemedicine service demand and if this is correlated with patient demographics), and notes of which partnering hospitals specialize in which service lines. Noting that the spikes and dips in demand over the year are not dramatic, we look to build in an anomaly detection component into the model, to search specifically for which points and microtrends most affect telemedicine service demand at certain points over time.

E. Acknowledgements

Many thanks to Dr. Amar Gupta for his guidance on helping us obtain data and introducing us to the challenges in telemedicine through this MIT course. Thank you also to Robyn Baek, VP of Analytics from SOC Telemed for partnering with us to let us analyze your datasets and for answering all our questions regarding the project.

F. References

[1] Moore, M. (1999). The evolution of telemedicine. *Future generation computer systems*, 15(2), 245-254.

[2] Schwamm, L. H., Rosenthal, E. S., Hirshberg, A., Schaefer, P. W., Little, E. A., Kvedar, J. C., ... & Levine, S. R. (2004). Virtual TeleStroke support for the emergency department evaluation of acute stroke. *Academic Emergency Medicine*, 11(11), 1193-1197.

[3] American Stroke Association. (n.d.). About Stroke. Retrieved from http://www.strokeassociation.org/STROKEORG/AboutStroke /Treatment/Stroke-Treatment_UCM_492017_SubHomePage.jsp

[4] Hess, D. C., Wang, S., Hamilton, W., Lee, S., Pardue, C., Waller, J. L., ... & Adams, R. J. (2005). REACH: clinical feasibility of a rural telestroke network. *Stroke*, 36(9), 2018-2020.

[5] Wootton, R. *Telemedicine support for the developing world.*. Journal of telemedicine and telecare, 14(3), 109-114, 2008.

[6] American Hospital Association. Telehealth: Helping Hospitals Deliver Cost-Effective Care. *Issue Brief*, 1-7, 2016.

[7] Schweigler, L. M., Desmond, J. S., McCarthy, M. L., Bukowski, K. J., Ionides, E. L., & Younger, J. G. (2009). Forecasting models of emergency department crowding. *Academic Emergency Medicine*, 16(4), 301-308.

[8] Schouten, P. Better patient forecasts and schedule optimization improve patient care and curb staffing costs, 2014.

[9] Mayo Clinic. (2016, April 20). Chronic traumatic encephalopathy. Retrieved May 15, 2018, from https://www.mayoclinic.org/diseases-

conditions/chronic-traumatic-encephalopathy/symptoms-causes/syc-20370921

[10]Mayo Clinic. (2018, March 03). Seizures. Retrieved May 15, 2018, from https://www.mayoclinic.org/diseases-conditions/seizure/symptoms-causes/syc-20365711

[11]Mayo Clinic. (2016, April 20). Chronic traumatic encephalopathy. Retrieved May 15, 2018, from https://www.mayoclinic.org/diseases-conditions/chronic-traumatic-encephalopathy/symptoms-causes/syc-20370921

[12]VanderPlas, J. (n.d.). In Depth: Linear Regression. Retrieved from https://jakevdp.github.io/PythonDataScienceHandbook/05.06-linear-regression.html

[13]Adams, H. P., Del Zoppo, G., Alberts, M. J., Bhatt, D. L., Brass, L., Furlan, A., ... & Lyden, P. D. (2007). Guidelines for the early management of adults with ischemic stroke. *Circulation*, 115(20), e478-e534.

[14]Nelson, R. E., Saltzman, G. M., Skalabrin, E. J., Demaerschalk, B. M., & Majersik, J. J. (2011). The cost-effectiveness of telestroke in the treatment of acute ischemic stroke. *Neurology*, 77(17), 1590-1598.

[15]Singh, R., Mathiassen, L., & Mishra, A. (2015). Organizational Path Constitution in Technological Innovation: Evidence from Rural Telehealth. *Mis Quarterly*, 39(3).

[16]Switzer, J. A., Demaerschalk, B. M., Xie, J., Fan, L., Villa, K. F., & Wu, E. Q. (2012). Cost-effectiveness of hub-and-spoke telestroke networks for the management of acute ischemic stroke from the hospitals perspectives. *Circulation: Cardiovascular Quality and Outcomes*, CIRCOUTCOMES-112.

[17]Ali, S. F., Viswanathan, A., Singhal, A. B., Rost, N. S., Forducey, P. G., Davis, L. W., ... & Schwamm, L. H. (2014). The TeleStroke mimic (TM)-score: a prediction rule for identifying stroke mimics evaluated in a Telestroke Network. *Journal of the American Heart Association*, 3(3), e000838.

[18]McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., & Kelen, G. D. (2008). The challenge of predicting demand for emergency department services. *Academic Emergency Medicine*, 15(4), 337-346.

[19]Capampangan, D. J., Wellik, K. E., Bobrow, B. J., Aguilar, M. I., Ingall, T. J., Kiernan, T. E., ... & Demaerschalk, B. M. (2009). Telemedicine versus telephone for remote emergency stroke consultations: a critically appraised topic. *The neurologist*, 15(3), 163-166.

[20]Ali, S. F., Hubert, G. J., Switzer, J. A., Majersik, J. J., Backhaus, R., Shepard, L. W., ... & Schwamm, L. H. (2018). Validating the TeleStroke Mimic Score: A Prediction Rule for Identifying Stroke Mimics Evaluated Over Telestroke Networks. *Stroke*, 49(3), 688-692.

G. Appendix

Sample fields used in models and regression result tables are shown on the next page.

Field	Field Definition	Example Categories	Definition of Categories
Reason for Consult	Reason for initiating telemedicine consultation	TIA	Transient Ischemic Attack
		Acute ischemic stroke	Requires brain imaging for diagnosis within 60 min in order to intervene
		Migraine	Throbbing/pulsing head pain
		Tremor	involuntary and rhythmic shaking[9]
Provider Diagnosis	Clinical diagnosis from telemedicine consultation	Stroke/tPA Stroke	Stroke that requires tPA administration
		Seizure	Sudden, uncontrolled electrical disturbance in the brain[10]
		Encephalopathy	Brain degeneration likely caused by repeated head traumas[11]
		Migraine	Throbbing/pulsing head pain
Service Line	Immediacy for consultation scheduling	Neuro Routine	Middle priority
		Neuro Emergency	Highest priority
		Neuro Routine – Scheduled	Lowest priority

Figure 13. Key fields, example fields, and descriptions used in data analysis.

Regression	Results	Rank of Performance
Linear (OLS)	$r^2 = 0.7821$	2
Ridge	$r^2 = 0.7829$	1
Lasso	$r^2 = 0.7695$	3

Figure 14. Results for linear, ridge, and lasso regression models.