Proceedings of the 2019 IISE Annual Conference
H.E. Romeijn, A. Schaefer, R. Thomas, eds.

A Predictive Analytics Approach for Nursing Home Hurricane Evacuation

Nazmus Sakib
Department of Industrial and Management Systems Engineering
University of South Florida, Tampa, Florida 33620, USA

Kathryn Hyer, Debra Dobbs, Lindsay Peterson, Dylan J. Jester
School of Aging Studies
University of South Florida, Tampa, Florida 33620, USA

Nan Kong
Weldon School of Biomedical Engineering
Purdue University, West Lafayette, IN 47907, USA

Mingyang Li*
Department of Industrial and Management Systems Engineering
University of South Florida, Tampa, Florida 33620, USA

Abstract

Whether to evacuate a nursing home (NH) or shelter in place in response to the approaching hurricane is one of the most complex and difficult decisions encountered by nursing home administrators. A variety of factors may affect the evacuation decision, including storm and environmental conditions, nursing home characteristics, and the dwelling residents' health conditions. Successful prediction of evacuation decision is essential to proactively prepare and manage resources to meet the surge in nursing home evacuation demands. In current nursing home emergency preparedness literature, there is a lack of analytical models and studies for nursing home evacuation demand prediction. In this paper, we propose a predictive analytics framework by applying machine learning techniques, integrated with domain knowledge in NH evacuation research, to extract, identify and quantify the effects of relevant factors on NH evacuation from heterogeneous data sources. In particular, storm features are extracted from Geographic Information System (GIS) data to strengthen the prediction accuracy. To further illustrate the proposed work and demonstrate its practical validity, a real-world case study is given to investigate nursing home evacuation in response to recent Hurricane Irma in Florida. The prediction performance among different predictive models are also compared comprehensively.

Keywords
Nursing Home, Evacuation, Aging Population, Prediction, Data Analytics

1. Introduction
Natural disasters, such as hurricanes, have detrimental effects and even life-threatening adverse consequences on nursing home (NH) facilities and their elderly care recipients due to associated physical damage of healthcare properties, long-lasting power loss, widespread communication failures, and disruption of medical supplies. Evidence shows that both mortality and morbidity of NH residents significantly increase during hurricanes [1, 2]. Consequently, many NHs have to evacuate and relocate their residents to safer hosting facilities. However, as many of the NH residents suffer from diverse chronic diseases and multi-functional (e.g., physical, mental, and social) limitations, the disruption associated with evacuation and the trauma of moving can result in functional declines and depression [3, 4]. Therefore, whether to evacuate the NH or not becomes one of the most critical and complex decisions faced by NH administrators. Successful modeling and accurate prediction of appropriate NH evacuation becomes important since it will not only provide evidence-based decision support to NH administrators, but will also assist operational

decisions for NH providers as well as local emergency authorities to proactively plan and manage healthcare resources to meet with NH evacuation demand surge.

In the existing evacuation modeling literature, many studies focus on community-dwelling households and individuals from general population and model their evacuation decisions by identifying various influencing factors, such as demographics and socioeconomic status of individuals [5, 6]. Many of these studies are also based on surveys conducted under non-disaster conditions due to limited availability of evacuation data [7]. NHs are long-term care organizations and their dwelling residents are specific and vulnerable elderly population who are frail and less autonomous. NH evacuation is an organization-level decision, and various internal and external factors describing the organization characteristics need to be incorporated to understand and model it. The existing literature on NH evacuation mainly focus on conceptual and qualitative studies based on single source of data, such as NH survey data [8-11]. There is a lack of evidence-based analytical modeling approaches in improving the understanding and prediction of NH evacuation. Moreover, with the advancement of data acquisition and storage systems, rich Geographic Information System (GIS) data are collected for capturing storm characteristics during hurricane occurrences. There is a need to integrate such GIS data with conventional data sources to further enrich the understanding of NH evacuation and strengthen the prediction performance.

To address the aforementioned research need, we propose a predictive analytics framework by extracting and incorporating multi-factorial NH features, such as NH facility, resident, and storm characteristics, to improve NH evacuation prediction based on integration of heterogeneous data sources, including real evacuation observations, NH survey data, and GIS data. In particular, we extract and integrate facility-specific environmental features from storm GIS information to strengthen the model prediction accuracy. To further improve the overall prediction performance in real-world context, we consider different machine learning methods, such as linear classification and its regularization-based variants, non-linear classification and ensemble learning methods, to comprehensively evaluate their prediction performances in a real case study based on the recent hurricane Irma.

2. Methodology

2.1 Overview

To accurately predict NH evacuation, the proposed predictive analytics framework consists of several sub-modules, as shown in Figure 1. Technical details are described in the subsequent sections.

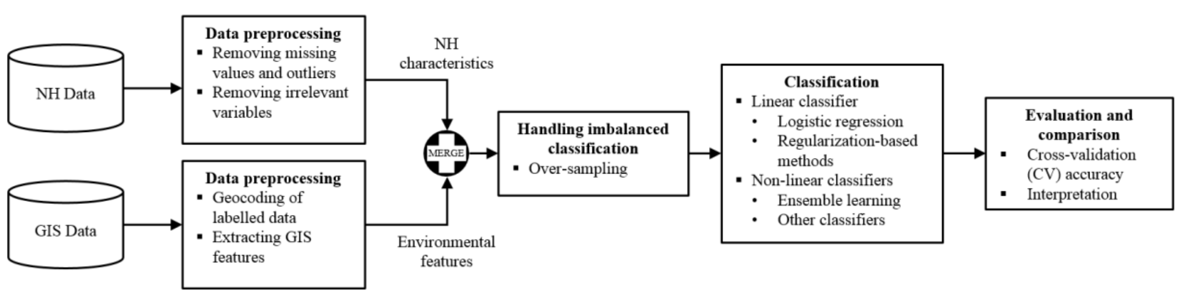
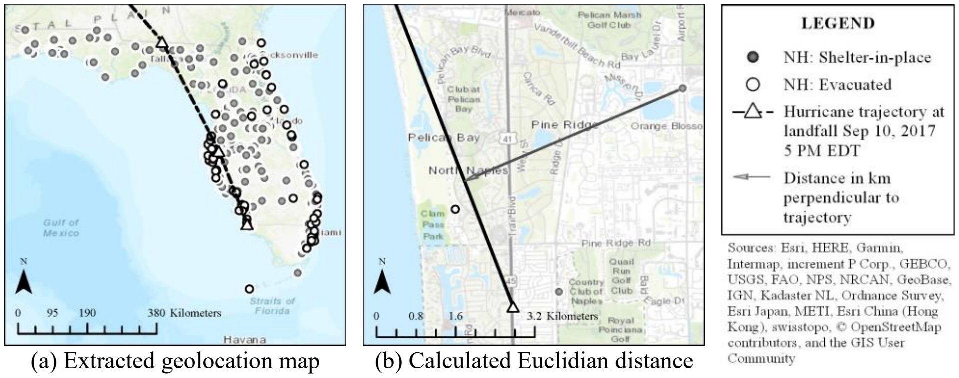


Figure 1: Overview of the predictive analytics framework

2.2 Data Preprocessing and Feature Extraction

NH evacuation observations are first obtained from the Florida Agency for Health Care Administration (AHCA). The NH addresses are geo-processed to extract and obtain the geolocation, i.e., latitude and longitude, of each NH. NHs and their dwelling residents' characteristics are extracted from NH inspection survey data available from the Centers for Medicare and Medicaid Services (CMS) in collaboration with the Florida Policy Exchange Center on Aging. To further incorporate GIS features to capture storm threat and its effect on NH evacuation, the projected hurricane track is extracted from the storm GIS data available at the National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA) [12] and the Euclidian distance between each NH geolocation and the projected hurricane track is calculated. If an NH is evacuated, the distance between the NH location and the projected storm trajectory on the day of evacuation is calculated. If NH is shelter-in-place, the distance between the NH location and the storm trajectory when it made landfall at Marco Island, FL, is calculated. Figure 2(a) shows the geolocation map of all NHs extracted and Figure 2(b) describes the distance calculation, which represents the proximity between NH and the projected hurricane track.



(a) Extracted geolocation map (b) Calculated Euclidian distance

Figure 2: Geo-processing of NH location and facility-specific GIS feature extraction

2.3 Classification

After data preprocessing and feature extraction, the binary response variable is coded as class "1" if a NH is evacuated and "0", otherwise. Since the majority class "0" accounts for almost 90%, oversampling of minority class is considered to resolve the classification imbalance issue. Given the balanced data set, $\mathbf{D} = \{y_i, \mathbf{x}_i\}_{i=1}^{n}$, where $y_i = \{0,1\}$ is labeled data, $\mathbf{x}_i$ is a vector of covariates (i.e., features extracted in Section 2.2) for each NH, and n is the sample size, different machine learning methods are considered to construct the predictive models and used to perform the binary classification. Specifically, linear classifier, such as logistic regression (LR), and its regularization-based variants are first applied. They aim to minimize the overall loss function, $l(\boldsymbol{\theta}_{LR}) = \sum_{i=1}^{n}\{-y_i\boldsymbol{\theta}_{LR}^T\boldsymbol{x}_i + \log[1+\exp(\boldsymbol{\theta}_{LR}^T\boldsymbol{x}_i)]\} + \lambda_1\left|\left|\boldsymbol{\theta}_{LR}\right|\right|_2^2 + \lambda_2\left|\left|\boldsymbol{\theta}_{LR}\right|\right|_1$, where the first term is the negative log-likelihood function of LR with model parameters $\boldsymbol{\theta}_{LR}$ and the last two terms are L_2 and L_1 penalty terms, respectively. When $\lambda_1 = \lambda_2 = 0$, the model becomes LR; when $\lambda_1 = 0$ and $\lambda_2 > 0$, it becomes regularized LR with ridge penalty; when $\lambda_1 > 0$ and $\lambda_2 = 0$, it becomes regularized LR with LASSO penalty. The regularization-based variants provide less data over-fitting by shrinking or excluding irrelevant covariates and restrict the unnecessary growth of the model complexity.

To further capture the potential non-linear relationship between input and output variables, different non-linear classification methods are considered, namely Naïve Bayes, K-nearest neighbor (KNN), decision tree, Support Vector Machines (SVM) and Artificial Neural Network (ANN) for the classification problem [13, 14]. Different methods have their own model structures and prediction schemes. For instance, the Naïve Bayes method utilizes Bayes' rule to predict the binary class (e.g., $C_k, k = 0, 1$) based on the highest posterior probability given the data, i.e., $\hat{y} = \underset{k\in\{0,1\}}{\operatorname{argmax}} \Pr(C_k) \prod_{i=1}^{n} \Pr(\boldsymbol{x}_i|C_k)$. KNN performs the prediction based on the majority vote of its K nearest neighbors, i.e., $\hat{y} = \underset{k\in\{0,1\}}{\operatorname{argmax}} \frac{1}{K}\sum_{N_x} \mathrm{I}(y_i = k)$, where N_x is an index set of K nearest neighboring observations for input variables $\boldsymbol{x}$ and $\mathrm{I}(\cdot)$ is an indicator function. The decision tree method represents the predictive model as an upside-down tree structure by recursively splitting nodes and creating branches. SVM and ANN allow construction of highly non-linear classification models by either formulating the classification problem as an optimization model or capturing the input-and-output relationship with a multi-layer network structure, respectively. To overcome the potential high variance of developing a single predictive model, ensemble learning methods, such as random forest and boosting trees, are also considered. They tend to reduce the variance by generating a large number of predictive models (e.g., simple decision trees) in either parallel or serial structure.

2.4 Prediction Performance Evaluation

To evaluate the performance of the predictive models discussed and developed in Section 2.3 and avoid over-fitting of the training data, 10-fold cross-validation (CV) is considered. The CV accuracy is a good surrogate estimate of test accuracy [13]. Thus, the classification accuracy based on CV, i.e., $Acc = \frac{1}{10\cdot n_m}\sum_{m=1}^{10}\sum_{i=1}^{n_m} \mathrm{I}(y_{i,m} = \hat{y}_{i,m})$, is utilized as a metric to evaluate and compare prediction performance among different predictive models, where $y_{i,m}$ and $\hat{y}_{i,m}$ are observed and predicted values in the m-th validation set with sample size n_m, respectively.

3. Real case study

3.1 Data Description

To illustrate the proposed analytical framework and demonstrate its effectiveness in an application context, a real-world case study of NH evacuation in response to Hurricane Irma is provided. As one of the strongest hurricanes in the Atlantic basin, Hurricane Irma tracked northwest through the Caribbean and eventually hit Florida as a Category 4 storm during September 2017. It nearly affected the entire state of Florida and more than 10% of the NHs in the state evacuated. To develop the predictive model for NH evacuation, heterogeneous data sources described in Section 2 are collected and integrated. Oversampling technique is also considered to handle the classification imbalance issue by augmenting the original data set with balanced classes.

3.2 Data Analysis and Results

3.2.1 Predictive Model Performance Comparison

Based on the preprocessed data, different classification methods, ranging from linear classification to non-linear classification, are adopted to develop predictive models by establishing functional relationship between evacuation response variable and various input variables extracted from heterogeneous data sources. Numerical summary of prediction performance among different models are presented in Table 1. Both training and 10-fold CV-accuracy are reported. A higher CV-accuracy indicates better prediction accuracy of the predictive model.

Several findings are shown in Table 1. First, all predictive models incorporating the GIS feature exhibit better prediction accuracy (e.g., higher CV-accuracy) than those without incorporating the GIS feature. It justifies the benefits of integrating rich GIS information with conventional NH survey data in improving the prediction of NH evacuation decision. The prediction performance improvement is more significant for linear classification method, such as LR, as well as its regularized variants. As the classification model becomes more non-linear, the prediction improvement resulting from the GIS feature becomes smaller. Second, non-linear classification methods exhibit better prediction performance than linear classification methods, which implies that the underlying relationship between NH evacuation decision and different input variables is non-linear. Non-linear classifiers with more flexible modeling capabilities, such as SVM and ANN, are more accurate than simpler non-linear classifiers, such as KNN and Naïve Bayes. Ensemble learning methods, such as random forest and boosting tree, further improve the prediction performance over ANN and SVM, since their model averaging techniques mitigate the potential overfitting of non-linear models. Third, based on comparison between training and CV accuracies, non-linear classifiers tend to have more over-fitting than linear classifiers due to their more flexible modeling structures. Also, compared to LR, its regularization-based variants, such as LR with L1/L2 norms, exhibit less overfitting since less relevant input variables are shirked with more complex models penalized during the model learning process. Figure 3 further visualizes the prediction performance comparison among different predictive models. Random forest with incorporating the GIS feature exhibits the best prediction performance.

Table 1: Predictive model performance comparison

Model Name	Model Description	Model Type	Training Accuracy (%)			CV Accuracy (%)		
			Without GIS Data	With GIS Data	Difference	Without GIS Data	With GIS Data	Difference
L1-Log	Logistic Regression (LR)	Linear	77.06	79.92	2.86	70.85	75.97	5.12
L2-Lasso	Lasso LR	Linear	76.97	79.75	2.78	71.01	76.72	5.71
L3-Ridge	Ridge LR	Linear	76.30	79.92	3.62	71.18	76.55	5.37
NL1-RF	Random Forest	Non-linear	93.98	95.14	1.16	96.13	97.4	1.27
NL2-DT	Decision Tree	Non-linear	95.29	95.71	0.42	86.63	89.92	3.29
NL3-GBT	Gradient Boosted Trees	Non-linear	99.16	99.33	0.17	94.54	95.04	0.5
NL4-SVM	Support Vector Machines (SVM)	Non-linear	95.88	96.47	0.59	90.34	93.28	2.94
NL5-KNN	K-Nearest Neighbor (KNN)	Non-linear	83.28	84.37	1.09	77.82	79.75	1.93
NL6-NB	Naïve Bayes	Non-linear	85.71	87.23	1.52	81.27	83.20	1.93
NL7-ANN	Artificial Neural Network (ANN)	Non-linear	99.08	99.58	0.50	91.34	93.28	1.94

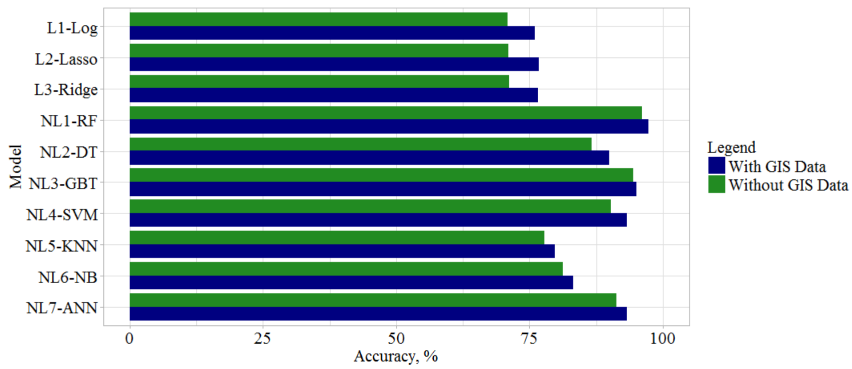


Figure 3: Comparison of CV-accuracy among different predictive models

3.2.2 Model Interpretation of Linear Predictive Model

From prediction performance perspective, linear predictive models, such as LR, exhibits less satisfactory prediction performance than non-linear models, such as random forest. However, LR provides much better model interpretation by identifying several significant factors and quantifying their influence of NH evacuation decision. Such model interpretation is important because it provides easy-to-interpret empirical evidence to enrich the scientific understanding of NH evacuation decision making. Table 2 summarizes several significant factors identified based on significance level of 0.05.

Table 2: Significant factors identified by the LR Model

Covariate	$\hat{\beta}$ (Std. Error)$^{\text{p-value}}$	Covariate	$\hat{\beta}$ (Std. Error)$^{\text{p-value}}$
Distance at decision, km	-0.013 (0.001)***	Count of residents under IV Therapy	-0.177 (0.052)***
Ownership Indicator: For Profit - Partnership	-1.023 (0.457)*	Count of residents under Respiratory Treatment	0.051 (0.008)***
Ownership Indicator: Government - Hospital district	2.567 (0.975)**	Count of residents under Tracheostomy Care	-0.149 (0.044)***
Ownership Indicator: Non-profit - Church	1.891 (0.418)***	Count of residents under Special Care with Injections	-0.02 (0.008)*
Ownership Indicator: Non-profit - Corporation	0.622 (0.18)***	Count of residents under Rehabilitative Services	-0.019 (0.006)**
Count of bedfast residents	-0.041 (0.016)*	Count of residents under Antianxiety Medications	-0.026 (0.008)***
Count of chair-bound residents	0.016 (0.005)***	Count of residents under Pain Management Program	0.009 (0.003)**
Count of residents with depression	-0.012 (0.004)***	Count of residents under Advanced Directives program	0.007 (0.003)**
Count of residents with pressure ulcers on admission	0.061 (0.024)**	Staff Count: Administrative Full-Time	-0.057 (0.019)**
Count of residents under Skin Integrity Preventive Care program	-0.009 (0.003)**	Staff Count: Medical Director Full-Time	1.53 (0.674)*
Count of residents under Radiation Therapy	-1.288 (0.366)***	Staff Count: PT Assistant Full-Time	-0.502 (0.086)***
Count of residents under Chemotherapy	0.2 (0.092)*	Staff Count: Speech Pathologist Full-Time	0.403 (0.143)**

Notes: *1) * p<0.05; ** <0.01; *** p<0.001,*
2) 95% Confidence Intervals for each parameter estimate are calculated by $\hat{\kappa}_j \pm 1.96 \times SE(\hat{\kappa}_j)$, where $\hat{\kappa}_j$ is the respective estimated covariate coefficient.

As shown in Table 2, when the estimated coefficient has a positive sign, it indicates that a larger covariate value tends to increase the probability of evacuation; and vice versa. For instance, the estimated coefficient for the GIS feature is negative, indicating that a NH facility farther away from the projected hurricane track is less likely to evacuate compared to a facility which is closer to the projected track. Similarly, for-profit NHs are less likely to evacuate compared to non-profit ones since evacuation involves transporting residents, staff and health resources to safer places, which is more costly. In addition to organization factor's influence, dwelling NH residents' characteristics also affect its evacuation decision making. For instance, if a NH has more bedfast and/or depressed residents, the probability of evacuation tends to decrease since these residents are particularly vulnerable to evacuation while shelter-in-place tend to be more protective for them in order to avoid any unnecessary physical injuries and/or mental discord exacerbation due to evacuation. Staffing also plays a role in evacuation decision. For instance, an NH with more administrative full-time staff is less likely to evacuate since it indirectly reflects that the NH facility is bigger and potentially more self-sufficient during storm with more resources.

4. Conclusion

In this paper, a predictive analytics framework for better prediction of whether to evacuate or shelter in place in response to hurricane is presented. Facility-specific GIS feature is integrated with multi-factorial NH characteristics to improve the model prediction accuracy. Different machine learning methods, including both linear and non-linear classification methods, are investigated and compared from both model prediction and interpretation perspectives. Non-linear predictive model based on random forest exhibits the best prediction performance, while linear predictive model, such as LR, has better model interpretation. As future work, the proposed predictive analytics framework will be further extended with prescriptive analytics techniques to realize more proactive and cost-effective managerial decision-making for NH evacuations.

Acknowledgements

This work was supported in part by National Institutes of Health under Grant R01AG060581, and in part by National Science Foundation under Grants 1825761 and 1825725.

References

1. Dosa, D., Hyer, K., Thomas, K., Swaminathan, S., Feng, Z., Brown, L., and Mor, V., 2012, “To evacuate or shelter in place: implications of universal hurricane evacuation policies on nursing home residents,” Journal of the American Medical Directors Association, 13(2), 190-e1.
2. Dosa, D., Feng, Z., Hyer, K., Brown, L. M., Thomas, K., and Mor, V., 2010, “Effects of Hurricane Katrina on nursing facility resident mortality, hospitalization, and functional decline,” Disaster medicine and public health preparedness, 4(S1), S28-S32.
3. Thomas, K. S., Dosa, D., Hyer, K., Brown, L. M., Swaminathan, S., Feng, Z., and Mor, V., 2012, “Effect of forced transitions on the most functionally impaired nursing home residents,” Journal of the American Geriatrics Society, 60(10), 1895-1900.
4. Brown, L., Rothman, M., and Norris, F., 2007, “Issues in mental health care for older adults after disasters,” Generations, 31(4), 21-26.
5. Whitehead, J. C., Edwards, B., Van Willigen, M., Maiolo, J. R., Wilson, K., and Smith, K. T., 2000, “Heading for higher ground: factors affecting real and hypothetical hurricane evacuation behavior,” Global Environmental Change Part B: Environmental Hazards, 2(4), 133-142.
6. Wolshon, B., Urbina, E., Wilmot, C., and Levitan, M., 2005, “Review of policies and practices for hurricane evacuation. I: Transportation planning, preparedness, and response,” Natural hazards review, 6(3), 129-142.
7. Yazici, M. A., and Ozbay, K., 2008, “Evacuation modelling in the United States: Does the demand model choice matter?” Transport Reviews, 28(6), 757-779.
8. Dosa, D. M., Grossman, N., Wetle, T., and Mor, V., 2007, “To evacuate or not to evacuate: lessons learned from Louisiana nursing home administrators following Hurricanes Katrina and Rita,” Journal of the American Medical Directors Association, 8(3), 142-149.
9. Hyer, K., Polivka-West, L., and Brown, L., 2007, “Nursing homes and assisted living facilities: Planning and decision making for sheltering in place or evacuation,” Generations, 31(4), 29-33.
10. Dosa, D. M., Hyer, K., Brown, L. M., Artenstein, A. W., Polivka-West, L., and Mor, V., 2008, “The controversy inherent in managing frail nursing home residents during complex hurricane emergencies,” Journal of the American Medical Directors Association, 9(8), 599-604.
11. Blanchard, G., and Dosa, D., 2009, “A comparison of the nursing home evacuation experience between Hurricanes Katrina (2005) and Gustav (2008),” Journal of the American Medical Directors Association, 10(9), 639-643.
12. National Hurricane Center (NHC), National Oceanic and Atmospheric Administration (NOAA), 2018, Accessed at: https://www.nhc.noaa.gov/
13. Hastie, T., Tibshirani, R., and Friedman, J., 2009, The elements of statistical learning, 2nd Edition, Springer, New York.
14. Bishop, C. M., 2006, Pattern recognition and machine learning, Springer-Verlag, New York.