Post-stroke Discharge Disposition Prediction using Deep Learning

Jin Cho*, Zhen Hu*, and Mina Sartipi*
*Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA
Email: cwp636@mocs.utc.edu, zhen-hu@utc.edu, and Mina-Sartipi@utc.edu

***Abstract*—Stroke is the fifth leading cause of death and a major cause of long-term disability in the United States. Because of its life-threatening consequences, the stroke system of care including acute treatment and post-acute rehabilitation is crucial. In order to support the stroke system of care in the age of big data, predictive analytics can be applied to forecast what will happen in the future. For example, we can predict post-stroke discharge disposition at an acute stroke admission, which will facilitate optimizing acute treatment and planning post-acute rehabilitation with the desired outcomes. In this preliminary study, we will explore deep learning for post-stroke discharge disposition prediction and evaluate prediction performance using hospital discharge data provided by Tennessee Department of Health. Deep learning is a powerful tool to deal with large-scale data and provides an end-to-end solution to multi-class classification including dimensionality reduction. Our preliminary results will demonstrate the effectiveness of deep learning and suggest the further exploration for performance improvement and clinical adoption.**

***Index Terms*—post-stroke, discharge disposition, prediction, deep learning**

I. INTRODUCTION

Stroke is the fifth leading cause of death and a major cause of long-term disability in the United States. Because of its life-threatening consequences, the stroke system of care including acute treatment and post-acute rehabilitation is crucial [1]. In the age of big data, predictive analytics can be explored to support the stroke system of care with the desired outcomes by forecasting what will happen in the future. For example, post-stroke discharge disposition can be predicted at an acute stroke admission [2], [3]. Based on such an early prediction, we can plan the stroke system of care and optimize the quality of care.

As large-scale data (i.e., electronic health record, mobile health data, and gene data) become available and accessible, big data analytics (including descriptive analytics, predictive analytics, and prescriptive analytics) by taking advantage of machine leanring/artificial intelligence for health informatics and healthcare has gained a lot of attention and efforts for both pre-clinical research and clinical practice. We can extract unknown information and discover actionable knowledge from massive data [4], [5], [6], [7], [8], which will guide us towards data-driven precision medicine and healthcare.

Deep learning is the latest breakthrough in machine learning/artificial intelligence. In theory, deep learning attempts to model high-level abstractions in massive data by using multiple processing layers with complex structures and non-linear operations [9], [10]. Deep learning also provides an end-to-end solution to dimensionality reduction of high-dimensional data. Handcrafted feature engineering is no longer needed, which will reduce the requirement of domain knowledge and bias of extracted features.

In this preliminary study, we will explore deep learning for post-stroke discharge disposition prediction. We will utilize hospital discharge data provided by Tennessee Department of Health to evaluate prediction performance. Different from our previous study that only considered two discharge disposition statues [11], [3], we will take into account more than two discharge disposition statuses in this paper. Thus, the prediction problem under investigation here is a multi-class classification instead of binary classification. Even though the multi-class classification problem needs more discriminative capability, it can provide more medical implications.

II. DATA AND METHODS

A. Data

We used hospital discharge data provided by the Tennessee Department of Health. Our investigation was focused on patients with the principal diagnosis of stroke (i.e ischemic and hemorrhagic strokes), hospitalized from 2010 to 2014. The raw data included 139,706 records with the following information for our research:

- Demographics: Sex, Age, and Race
- Clinical Characteristics: Principal Diagnosis Code, Admitting Diagnosis Code, and Other Diagnosis Codes
- Insurance: Primary and Secondary Payer Classes
- Source of Admission
- Discharge Disposition Status.

Data preprocessing was applied on raw data to build a clean dataset for predictive analytics. We excluded 12,125 records due to invalid or missing data. We exploited one-hot encoding to convert categorical data into binary feature format. Overall, the final dataset included 127,581 records with 341 features which include 21, 200, and 100 features corresponding to principal diagnosis codes, other diagnosis codes, and admitting diagnosis codes, respectively. The 16 discharge disposition statuses were considered:

1) home or self care (routine discharge)
2) another short term general hospital for inpatient care
3) a skilled nursing facility (SNF)
4) an intermediate care facility (ICF)

5) a Designated Cancer Center or Children's Hospital
6) home under care of organized home health service organization
7) Admitted as an inpatient to this hospital
8) Still a patient or expected to return for outpatient services
9) a Federal Health Care Facility
10) a hospital-based swing bed within this institution
11) another rehabilitation facility including rehabilitation distinct parts units of a hospital
12) a Medicare-certified long-term care hospital (LTCH)
13) a nursing facility certified under Medicaid but not certified under Medicare
14) a psychiatric hospital or psychiatric distinct part unit of a hospital
15) a Critical Access Hospital (CAH)
16) Another Type of Healthcare Institution.

B. Predictive Analytics

1) Logistic Regression: Logistic regression has been widely used in many classification problems due to its simpleness and ability to predict with probability estimation [3]. For our study, logistic regression was used as a baseline method.

2) Deep Learning: Stacked autoencoders, one of the popular deep learning architectures, were exploited for discharge disposition prediction. In such a deep learning architecture, an autoencoder is the basic building block to encode the input for dimensionality reduction [12], [8]. The main parameters we can choose to manipulate autoencoders are the number of neurons (p1), the weight for the L_2 regularization (p2), the weight for sparsity regularization (p3), sparsity proportion (p4), and maximum number of training epochs (p5) [13]. In our preliminary study, stacked autoencoders included one or two encoder layers and one softmax layer. The input and output dimensions of stacked autoencoders were 341 and 16, respectively. In order to maximize the performance of stacked autoencoders, both unsupervised pre-training and supervised fine-tuning were applied.

III. RESULTS

For performance evaluation, we selected first 101,223 records (2010-2013) as the training set and the remaining 26,358 records (2014) as the testing set. We considered predictive models built from logistic regression (LR), stacked autoencoders with one encoder layer (SA1), and stacked autoencoders with two encoder layers (SA2). The parameters of stacked autoencoders were shown in Table I.

TABLE I: Parameters of stacked autoencoders.

Model	p1	p2	p3	p4	p5
SA1	200	0.004	4	0.15	400
SA2 – Encoder 1	200	0.004	4	0.15	400
SA2 – Encoder 2	100	0.002	4	0.1	200

The performances for the testing set in terms of prediction accuracy were shown in Table II. Logistic regression achieved 68.0% accuracy. The performances of stacked autoencoders were slightly better and the prediction accuracy can go up to 69.0%.

TABLE II: Performance comparison.

Model	Accuracy
LR	0.680
SA1	0.686
SA2	0.690

IV. CONCLUSION

In this preliminary study, we demonstrated post-stroke discharge disposition prediction using deep learning. The prediction problem under investigation here is a multi-class classification, with consideration of 16 discharge disposition statuses. The prediction performance verified the effectiveness of deep learning. Due to data imbalance among different discharge disposition statuses, there is still a room for performance improvement. We will continue this line of work by addressing the data imbalance problem and investigating other deep learning architectures.

ACKNOWLEDGMENT

Research reported in this paper was supported by National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award number R15EB015700. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also partially funded by Tennessee Higher Education Commission.

REFERENCES

[1] M. L. Clark and T. Gropen, "Advances in the stroke system of care," *Current treatment options in cardiovascular medicine*, vol. 17, no. 1, pp. 1–13, 2015.

[2] M. Mees, J. Klein, L. Yperzeele, P. Vanacker, and P. Cras, "Predicting discharge destination after stroke: A systematic review," *Clinical neurology and neurosurgery*, vol. 142, pp. 15–21, 2016.

[3] J. Cho, Z. Hu, N. Fell, G. Heath, R. Qayyum, and M. Sartipi, "Hospital discharge disposition of stroke patients in the state of tennessee," 2016, submitted to Southern Medical Journal.

[4] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, vol. 13, no. 6, pp. 395–405, 2012.

[5] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 1, 2014.

[6] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G.-Z. Yang, "Big data for health," *IEEE journal of biomedical and health informatics*, vol. 19, no. 4, pp. 1193–1208, 2015.

[7] J. S. Rumsfeld, K. E. Joynt, and T. M. Maddox, "Big data analytics to improve cardiovascular care: promise and challenges," *Nature Reviews Cardiology*, 2016.

[8] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. A. Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, 2016.

[9] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[11] J. Cho, M. Sartipi, and Z. Hu, "Big data analytics for smart health," in *ACM Mid-Southeast Conference*, 2016.

[12] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[13] https://www.mathworks.com/help/nnet/ref/trainautoencoder.html, accessed: 2017-01-18.