Using Deep Learning to Predict Diabetes Patients' Readmission to the Hospital

Ahmad Hammoudeh[a,d,*], Ghazi Al-Naymat[a], Ibrahim Ghannam[b], Nadim Obied[a,c]

[a]Princess Sumaya University for Technology, Amman 11941, Jordan
[b]RWTH Aachen University, Aachen 52074, Germany
[c]University of Jordan, Amman 11942, Jordan
[d]School of AI, Amman 11123, Jordan

Abstract

Hospital readmissions increase the healthcare costs and negatively influence hospitals' reputation. Predicting readmissions in early stages allows prompting great attention to patients with high risk of readmission, which leverages the healthcare system and saves healthcare expenditures. Machine learning helps in providing more accurate predictions than current practices. In this work, an approach that balances between data engineering and neural networks' ability to learning representations is proposed for predicting hospital readmission among diabetic patients. A combination of Convolutional neural networks and data engineering were found to outperform other machine learning algorithms when employed and evaluated against real life data.

***Keywords**:* deep learning, diabetes, predictive modelling, data mining ;

Introduction

Diabetes is a wide spread chronic disease that is accompanied with irregularities of blood glucose levels due to problems related to insulin. The number of people with diabetes in the world has risen from 108 million in 1980 to 422 million in 2014. The prevalence of diabetes is growing most rapidly in low- and middle-income countries [1]. In Jordan for example, the prevalence of Type 2 diabetes was around 17.1% in 2004 with 30% increase in a decade which is a dramatic increase [2].

Hospital readmission is expressed by the time that a patient takes before getting back to the hospital. Readmission is considered a quality measure of hospital performance as well as a mean to reduce healthcare costs. Hospitals are financially penalized when the permitted rate of 30-day readmissions is exceeded. The Medicare Payment Advisory Commission in the US estimated that 12% of readmissions can be avoided. Preventing 10% of readmissions would save Medicare in the US more than $1 billion [3]. For diabetes; the cost analysis estimates that $250 million can be saved across 98,000 diabetic patients by incorporating predictive modeling and prompting greater attention to those who were predicted to get readmitted [4].

Current practices to identify at-risk diabetic patients are subjective; a clinician will assess the patient and decide what the appropriate care plan for that individual is. Research has shown that these methods for determining readmission are slightly better than random guessing [5]. On the other side, machine learning plays a vital role in many predicting tasks. Hence, predicting hospital readmissions using machine learning sounds a worth implementing approach.

This work shows deep learning as an effective approach for predicting diabetic patients' readmissions. The results show that deep learning predicts hospital readmissions among diabetics better than other machine learning algorithms, such as Logistic Regression, Naïve Bayes, or Random Forest.

The organization of the paper is as follows: Section 2 discusses the related work to hospital readmissions among diabetic patients. Section 3 briefly introduces deep learning. Then the methodology and the experimental results are presented in section 4. In section 5, conclusions are drawn.

Related work

In general, Neural networks (NNs) are common in medical classification research. A review by Dreiseitl and Ohno- Machado [6] to some implemented models in medical classification tasks showed that logistic regression (LR) followed by NNs are the most popular classifiers in medicine. Authors in [6] reported that both LR and shallow NNs perform on about the same level more often than not. However, the superior popularity of LR was attributed to the interpretability of model parameters and the ease of use. An obstacle for neural networks is the black-box property that hinders the model interpretability.

Various published papers studied readmission rates of diabetic patients [7-10]. Some studies used machine learning models to predict the risk of all-cause readmissions among patients with diabetes. Bhuvan et al.[4] compared different classifiers that were applied to this problem (same dataset) such as Naïve Bayes, Bayes Network, Random Forest, Adaboost Trees, and shallow NNs. Bhuvan et al.[4] showed that the performance of Random Forest and shallow NNs outperforms other classifiers with a slight preponderance in Random Forest's favor. However, only a single hidden layer of neural network was used. The AUC of the published papers ranges between 0.5 and 0.7 [4, 11]. Ensemble models of various classifiers were proposed by Mingle in [11]. First patients were divided into 3 groups according to the age {less than 30 years, between 30 and 70, greater than 70 years}. A separate ensemble model was built for each group by combining multiple classifiers (random forest, different types of gradient boosted trees, and kernel SVM). The obtained results for the developed ensemble models are summarized in Table 1.

The proposed classifier in this paper is distinguished from other reported classifiers by following an approach that combines data engineering and deep NNs to achieve higher performance. The proposed approach balances between widening the boarders that separate different classes (by involving data engineering), and detecting tiny details that contribute to determining the correct classes by incorporating a larger feature space by incorporating all possible features and assigning the task of determining the importance of each feature to the Deep NNs.

Table 1. Evaluation measures of ensemble models developed by [16]

Evaluation measure\ Age	<30	30-70	>70
Accuracy	0.848	0.785	0.685
Precision	0.365	0.229	0.186
Recall	0.498	0.437	0.49
AUC	0.79	0.70	0.65

3. Deep Learning

Deep learning is simply a neural network (NN) of multiple layers that learns representations of data with multiple levels of abstraction. Different from conventional machine-learning techniques that of limited ability to enhance by getting exposed to more data, deep learning has the capacity to scale effectively even with raw data [12]. In this work two architectures for the neural networks are studied: the traditional artificial neural networks (ANNs), and the convolutional nets (CNNs). Since CNNs were found to perform better than ANN for the same layers, only CNN results are reported.

The traditional architecture of neural networks consists of basic units arranged in successive layers; each unit computes a weighted sum of the input and applies a non-linear function (e.g. rectified linear unit ReLU(z) = max(0 , z)). Fig. 1.a shows a simple example of the basic unit of NN. Each unit in a traditional ANN layer is connected to all units in the previous layer as shown in Fig. 1.b. As distinct from traditional ANN, units in convolutional layer in CNNs [13] are not fully connected to all features in the previous layer, and the units in a layer share the same weights (filter) as shown in Fig. 1.c. It is worth mentioning that neural networks learn by modifying internal parameters (weights) to reduce the error between the estimated output and the correct outputs.

Neural networks classifiers have the capacity to discriminate non-linearly separable classes and generalize far from the seen examples. Moreover, deep layers eliminate considerable amount of features engineering work that was essential for shallow classifiers; deep neural networks can learn the representation automatically. Adding more layers to the neural networks increases both selectivity and invariance of the representation, where a classifier can draw extremely complex functions of its inputs that are sensitive to tiny details, which distinguish one class from another, and simultaneously insensitive to large irrelevant variations between instances of the same class [13].

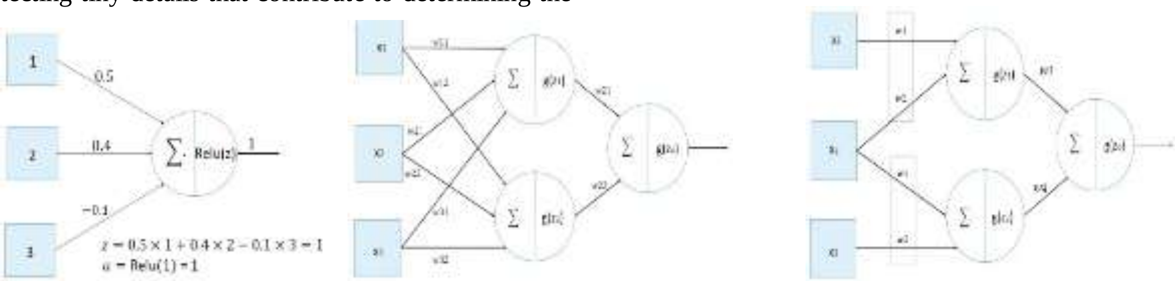

Fig. 1. a. Basic unit of NN.

Fig. 1. b. A simple diagram of traditional ANN.

Fig. 1. c. A simple diagram of a convolutional layer.

Experimental results

Developing an efficient classifier in this work comes in two folds: the first is developing the machine learning model and the second is developing the representation: Going into deeper neural networks is the approach to developing the model. On the other hand, the approach to develop the representation is by carrying out some data engineering techniques. This section discusses the experimental steps followed to preparing the data, building CNNs,and evaluating the model.

Dataset description

The dataset of this work [14, 15] consists of 100,000 medical records for 70,000 patients with diabetes collected from 130 hospitals in the USA over 10-years' period from 1999 to 2008. Medical records in the dataset include 50 attributes that are the risk factors, in addition to a label indicating the readmission status of a patient indicates whether a patient was readmitted to the hospital in 30 days or not. The dataset encounters satisfy the following conditions:

- It is an inpatient encounter (a hospital admission)
- It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.
- The length of stay was at least 1 day and at most 14 days.
- Laboratory tests were performed during the encounter.
- Medications were administered during the encounter.

4.1. Data Engineering

Deep learning approach allows a machine to learn from raw data. However, learning directly from raw data entailsa large number of training examples. When data are not sufficient for representation learning, data engineering turns out to be essential to overcome the shortage of data. In this work, we balance between deep learning capabilities, andthe size of the data.

Feature selection: the proposed approach in this work assumes feature selection as part of the deep learning model.By feeding the attributes to deep learning model, the NNs learn the influence of each attribute and assigns the weights accordingly. On the other side, conventional machine learning approaches only consider the attributes of the highest influence, which contributes to enhancing the interpretability and the simplicity of the model at the expense of the performance. Considering attributes of minimal contributions can lead to improvement in the performance. Hence, our approach is to consider all attributes unless any of the below conditions was met:

The attribute is completely irrelevant to patients' readmission such as encounter id and patient number.

The attribute has large portion of missing values such as weight (~98% missing values), Payer code (around 40%), Medical specialty (around 50%).

The attribute has purity >99.95%: purity is the highest percentage of records which have the same value for that attribute. For example, attributes that have a single value for all record are useless.

For features of highest influence on diabetics' readmission, the reader is referred to Bhuvan et al. in [4]. The main interest of this work is improving the performance of the classifier at the expense of the interpretability (black-boxproperty of NNs)

Feature creation: To compensate for the lost information from dropped records, 2 additional features were created before eliminating the useless attributes. The first feature is the number of medications and the second feature is the number of changes in the medications. Both features are extracted from the drug attributes.

Feature transformation: Since NNs assume numerical inputs, Categorical data were transformed into binary encoding or One-hot encoding. And since model stability and parameter estimate convergence are influenced when multi-scaled variables are used. Z-norm standardization is applied by subtracting the arithmetic mean ($\bar{x}$) of an attribute and dividing over the standard deviation (σ) as shown in equation (1).

$$z = \frac{x - \bar{x}}{\sigma} \quad (1)$$

- Imbalanced data: The problem of imbalanced data is one of the obstacles for many Machine Learning (ML) algorithms, it arises when the data are dominated by a majority class and a minority class is rarely detected. As a result, the classifier performance on the minority may be insufficient when compared to the majority. For example, a dumb classifier that always predicts the majority class can achieve high accuracy.

  There are two main methods to deal with imbalanced data [16]: 1- Under-sampling methods that balances the classing by eliminating great portion of the majority class. 2- Over-sampling methods such as Synthetic Minority

  Oversampling Technique (SMOTE) [17] which increases a number of new minority class instances by interpolation. SMOTE creates new instances rather than replicating the existing instances. In this work around 10%of the patients were readmitted in less than a month. To overcome imbalanced data, SMOTE was chosen due to the fact that it increases the sample size. More data benefits deep learning models since the performance of deep NNs improves with more data. For the impact of over sampling in enhancing the performance of NNs, the reader is referred to Hensman and Masko in [18].
- Handling duplicate records: Out of multiple records for the same patient, a single record is kept and the remainingwere deleted which led to reducing the number of records to ~70,000. The first record is chosen because it has the highest probability of readmission which helps in balancing the data. For example, all last encounters are labeled by the minority class (not readmitted) which would aggravate the problem of Imbalanced data.

4.2. Deep Learning Models

Convolutional neural networks model is developed. Since the classifier distinguishes between two classes (0 as not readmitted and 1 as readmitted), Sigmoid activation function was chosen for the output layer. However, ReLU was chosen for all layers other than the output layer because it leads to efficient computation and fewer vanishing gradient problems. The selected optimization algorithm was Adam [19].

In order to make sure that the model generalizes and does not over-fit, the data were split into two parts: 20% for testing and 80% for training and development. The latter is divided repeatedly into 80% for training and 20% for cross validation. The test set is hold out to evaluate the performance of the model. Early stopping technique was used to avoid overfitting, when the validation error increases for a specified number of iterations, the training is stopped. Since the deep learning model did not see the test set before the final evaluation, the evaluation of the testing set determines whether the model generalize well or not. For model evaluation measures such as accuracy, AUC, and F1, the reader is referred to [20].

Machine Learning models were built in Python using the Scikit-learn (http://scikit-learn.org/), Tensorflow (https://www.tensorflow.org/), and Keras (https://keras.io/).

Fig. 2 Compares the obtained results of CNN to the best results in the literature, which were partially obtained by Mingle [11] (Results were summarized in Table 1). Deep neural networks have resulted in ~8% improvement in accuracy, ~45% improvement in F1-score, and ~14% improvement in AUC. The comparison shows a predominance of deep learning approach.

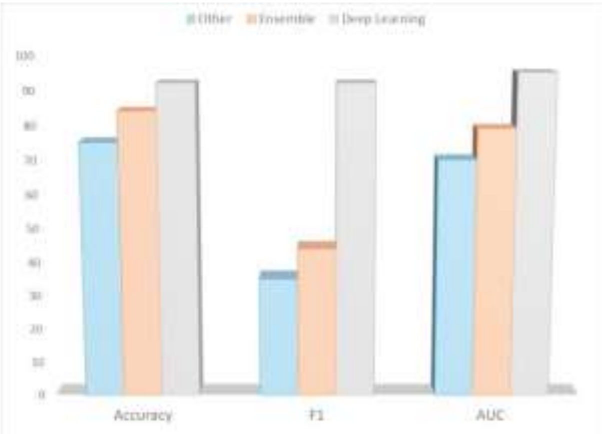

Fig. 2. Evaluation measures of CNN, traditional ANN, Mingle's ensemble model, and other reported machine learning algorithms (CNNhitting ~95% AUC, and ~92%

accuracy and
F1-score).

Conclusion

Hospital readmissions raise health care costs and negatively influence hospitals' reputation. Hence, predicting hospital readmissions among diabetics is of great interest. This paper presented deep learning as an effective approach in predicting hospital readmissions among diabetic patients. A combination of Convolutional neural networks and data engineering were found to outperform other machine learning algorithms when employed and evaluated againstreal life data.

Acknowledgements

Acknowledgement is given to the anonymous reviewers for providing valuable comments that contributed to enhancing the paper.

References

World Health Organisation. (2016). *Global Report on Diabetes*.
Ajlouni, Kamel, Yousef S. Khader, Anwar Batieha, Haitham Ajlouni, and Mohammed El-Khateeb. (2008). "An increase in prevalence of diabetes mellitus in Jordan over 10 years." *Journal of Diabetes and its Complications*, ***22*****(5)**: 317–324.
Medicare Payment Advisory Commission. (2007). *Report to the Congress promoting greater efficiency in Medicare*. Washington, DC.
Bhuvan, Malladihalli S., Ankit Kumar, Adil Zafar, and Vinith Kishore. (2016). "*Identifying diabetic patients with high risk of readmission*."
Allaudeen, Nazima, Jeffrey L. Schnipper, E. John Orav, Robert M. Wachter, and Arpana R. Vidyarthi. (2011). "Inability of Providers to
Predict Unplanned Readmissions." *Journal of General Internal Medicine* **26(7)**:771–76.
Dreiseitl, Stephan and Lucila Ohno-Machado. (2002). "Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review." *Journal of Biomedical Informatics* **35(5)**:352–59.
Silverstein, Marc D, Huanying Qin, S Quay Mercer, Jaclyn Fong and Ziad Haydar. (2008). "Risk Factors for 30-Day Hospital Readmission in
Patients 65 Years of Age." *Baylor University Medical Center Proceedings* **21(4)**:363–72.
Jiang, H.Joanna, Daniel Stryer, Bernard Friedman, and Roxanne Andrews. (2003). "Multiple Hospitalizations for Patients with Diabetes."
Diabetes Care **26(5)**:1421–26.
Strack Beata, DeShazo Jonathan, Gennings Chris, Olmo Juan, Ventura Sebastian, Cios Krzysztof, and John N. Clore. (2014). "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records." *BioMed Research International*.
Yifan, Xing and Jai Sharma. (2016). *Diabetes Patient Readmission Prediction Using Big Data Analytic Tools*.
Mingle, Damian. (2017). "Predicting Diabetic Readmission Rates: Moving Beyond Hba1c." *Current Trends in Biomedical Engineering & Biosciences* **7(3)**:555707. 007.
Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* **521(7553)**:436–444.
Lecun, Yann. (1989). *Generalization and Network Design Strategies*. Retrieved (http://masters.donntu.edu.ua/2012/fknt/umiarov/library/lecun.pdf%5Cnpapers3://publication/uuid/2C3E3B3C-DE43-4109-A816- 7220A5C08617).
Strack Beata, DeShazo Jonathan, Gennings Chris, Olmo Juan, Ventura Sebastian, Cios Krzysztof, and John N. Clore. (2014). "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records." *BioMed Research International*.
Strack Beata, DeShazo Jonathan, Gennings Chris, Olmo Juan, Ventura Sebastian, Cios Krzysztof, and John N. Clore. (2014). "Diabetes 130-US hospitals for years 1999-2008 Data Set." *UCI Machine Learning Repository*.
Jeatrakul, Piyasak, Kok Wai Wong, and Chun Che Fung. (2010). "Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm." in Wong K.W., Mendis B.S.U., Bouzerdoum A. (eds) *Neural Information Processing. Models and Applications*. ICONIP 2010. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-17534-3_19
Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W.Philip Kegelmeyer. (2002). "SMOTE: Synthetic Minority Over-sampling
Technique." *Journal of Artificial Intelligence Research* **16(1)**:321–357.
Hensman, Paulina and David Masko. (2015). "The Impact of Imbalanced Training Data for Convolutional Neural Networks." KTH RoyalInstitute of Technology.
Kingma, Diederik P. and Jimmy Lei Ba. (2015). "Adam: A Method for Stochastic Optimization." *International Conference on Learning Representations 2015*.
Zheng, Alice. (2015). "Evaluating Machine Learning Models: A Beginner's Guide to Key Concepts and Pitfalls." *O'Reilly Media, Inc.*,Sebastopol, CA.