

Predicting 30-Day Emergency Readmission Risk

Arkaitz Artetxe^{1,2(✉)}, Andoni Beristain^{1,2}, Manuel Graña²,
and Ariadna Besga³

¹ Vicomtech-IK4 Research Centre, Mikeletegi Pasealekua 57,
20009 San Sebastian, Spain

{aartetxe, aberistain}@vicomtech.org

² Computation Intelligence Group, Basque University (UPV/EHU),

P. Manuel Lardizabal 1, 20018 San Sebastian, Spain

³ Department of Internal Medicine,
Hospital Universitario de Alava, Vitoria, Spain

Abstract. *Objective:* Predicting Emergency Department (ED) readmissions is of great importance since it helps identifying patients requiring further post-discharge attention as well as reducing healthcare costs. It is becoming standard procedure to evaluate the risk of ED readmission within 30 days after discharge. *Methods.* Our dataset is stratified into four groups according to the Kaiser Permanente Risk Stratification Model. We deal with imbalanced data using different approaches for resampling. Feature selection is also addressed by a wrapper method which evaluates feature set importance by the performance of various classifiers trained on them. *Results.* We trained a model for each scenario and subpopulation, namely case management (CM), heart failure (HF), chronic obstructive pulmonary disease (COPD) and diabetes mellitus (DM). Using the full dataset we found that the best sensitivity is achieved by SVM using over-sampling methods (40.62 % sensitivity, 78.71 % specificity and 71.94 accuracy). *Conclusions.* Imbalance correction techniques allow to achieve better sensitivity performance, however the dataset has not enough positive cases, hindering the achievement of better prediction ability. The arbitrary definition of a threshold-based discretization for measurements which are inherently is an important drawback for the exploitation of the data, therefore a regression approach is considered as future work.

Keywords: Readmission risk · Imbalanced datasets · SVM · Classification

1 Introduction

The number of people aged over 65 is projected to grow from an estimated 524 million in 2010 to nearly 1.5 billion in 2050 worldwide [1]. This trend has a direct impact on the sustainability of health systems, in maintaining both public policies and the required budgets.

This growing population group represents an unprecedented challenge for health-care systems. In developed countries, older adults already account for 12 to 21 % of all ED visits and it is estimated that this will increase by around 34 % by 2030 [14].

Older patients have increasingly complex medical conditions in terms of their number of morbidities and other conditions, such as the number of medications they use, existence of geriatric syndromes, their degree of physical or mental disability, and the interplay of social factors influencing their condition [9]. Recent studies have shown that adults above 75 years of age have the highest rates of ED readmission, and the longest stays, demanding around 50 % more ancillary tests [15]. Notwithstanding the intense use of resources, these patients often leave the ED unsatisfied, and with poorer clinical outcomes, and higher rates of misdiagnosis and medication errors [16] compared to younger patients. Additionally, once they are discharged from the hospital, they have a high risk of adverse outcomes, such as functional worsening, ED readmission, hospitalization, death and institutionalization [17].

In this paper we present our recent work on ED readmission risk prediction. We utilize historic patient information, including demographic data, clinical characteristics or drug treatment information among others. Our work focuses on high risk patients (two higher strata) according to the Kaiser Permanente Risk Stratification Model [11]. This includes patients with prominence of specific organ disease (heart failure, chronic obstructive pulmonary disease and diabetes mellitus) and patients with high multi-morbidity. Predictive models are built for each of the stratified groups using different classifiers such as Support Vector Machine (SVM) and Random Forest. In order to deal with class imbalance and high dimensional feature space, different filtering techniques have been proposed during experimental approach.

The main contributions of this work are:

- We extend the work by Besga et al. [2] applying well-known machine learning techniques such as class balancing and feature selection in order to obtain better sensitivity.
- We compare two well established supervised classification algorithms, Random Forests and SVM, and analyze their performance in different scenarios.
- We make use of a wrapper feature selection method that maximizes the prediction ability while minimizes models' complexity.

The paper is organized as follows. In Sect. 2 we present some related works on predictive modelling for readmission risk estimation. In Sect. 3 we present the dataset as well as the methodological approach followed in order to build our models. Next, we describe the evaluation methodology and the experimental results. In Sect. 5 we discuss the conclusions and future work.

2 Related Work

Readmission risk modelling is a research topic that has been extensively studied in recent years. The main objective is usually to reduce readmission costs by identifying those patients with higher risk of coming back soon. Patients with higher risk can be followed-up after discharge, checking their health status by means of interventions such as phone calls, home visits or online monitoring, which are resource intensive. Predictive systems generally try to model the probability of unplanned readmission (or death) of a patient within a given time period.

In a recent work, Kansagara et al. [9] presented a systematic review of risk prediction models for hospital readmission. Many of the analyzed models target certain subpopulation with specific conditions or diseases such as Acute Myocardial Infarction (AMI) or heart failure (HF) while others embrace general population.

One of the most popular models that focus on general populations is LACE [3]. The LACE index is based on a model that predicts the risk of death or urgent readmission (within 30 days) after leaving the hospital. The algorithm used to build the model is commonly used in the literature (logistic regression analysis) and, according to the published results, the model has a high discriminative ability. The model uses information of 48 variables collected from 4812 patients from several Canadian hospitals.

A variant called LACE + [4] is an extension of the previous model that makes use of variables drawn from administrative data.

A similar approach is followed by Health Quality Ontario (HQP) with their system called HARP (Hospital Admission Risk Prediction) [10]. The system aims to determine the risk of patients in short and long term future hospitalizations. HARP defines two periods of 30 days and 15 months for which the model infers the probability of hospitalization, relying on several variables. From an initial set of variables of 4 different categories (demographic, feature community, disease and condition and meetings with the hospital system) the system identifies two sets of variables, a complex and a simpler one, with the most predictive variables. Using these sets of variables and a dataset containing approximately 382,000 episodes, two models for one month and 15 months are implemented. The models were developed using multivariate regression analysis. According to the committee of experts involved in the development of HARP, the most important metric was the sensitivity (i.e. the ability to detect hospitalizations). Regarding this metric, claimed results suggest that both simple and complex models achieve high sensitivity, although the complex model gets better results. The authors of this work suggest that the simple model could be a good substitute when certain hospitalization data is not available (e.g. to perform stratification outside the hospital).

A recent work by Yu et al. [5] presents an institution-specific readmission risk prediction framework. The idea beneath this approach is that most of the readmission prediction models have not sufficient accuracy due to differences between the patient characteristics of different hospitals. In this work an experimental study is performed, where a classification method (SVM) is applied as well as regression (Cox) analysis.

In [2] Besga et al. analyzed patients who attended Emergency Department of the Araba university Hospital (AUH) during June 2014. We exploit this dataset improving their results with further experiments.

3 Materials and Methods

The dataset, presented by Besga et al. in [2], is composed of 360 patients divided into four groups, namely: case management (CM), patients with chronic obstructive pulmonary disease (COPD), heart failure (HF) and Diabetes Mellitus (DM). For each patient a set of 97 variables were collected, divided into four main groups: (i) Sociodemographic

data and baseline status, (ii) Personal history, (iii) Reasons for consultation/Diagnoses made at ED and (iv) Regular medications and other treatments. Dataset contains missing values.

In order to build our model following a binary classification approach, the target variable was set to *readmitted/not readmitted*. Those patients returning to ED within 30 days after being discharged are considered readmitted (value = 1), otherwise are seen as not readmitted (value = 0).

It is noteworthy that one patient returning the first day and another returning the 30th are both considered as *readmitted*. On the other hand, a patient returning the 31th day is considered as *not readmitted*, while in practice underwent a readmission. We believe that having the number of days passed before readmission would have been much more meaningful for identification and would have permitted even identifying a more accurate prediction, including the predicted time for readmission.

All the tests were conducted using 10-fold cross-validation. The evaluation metrics that we have used are: sensitivity, specificity and accuracy. In order to avoid any random number generation bias, we have conducted 10 independent executions with different random generating seeds and averaged the results obtained.

Table 1. Distribution of variables by category

Variable	No. (%) of variables n = 96
Sociodemographic and baseline status	4 (4.2)
Personal history	43 (44.8)
Reasons for consultation	16 (16.7)
Regular medications	33 (34.3)

According to the data shown in Table 1 our dataset has a high dimensional feature space. In this scenario we have carried out some feature selection techniques. The goal is to find a feature subset that would reduce the complexity of the model, so that it would be easier to interpret by physicians, while improving the prediction performance and reducing overfitting.

We are going to use the following approaches: filter methods and wrapper methods. Filter algorithms are general preprocessing algorithms that do not assume the use of a specific classification method. Wrapper algorithms, in the other hand, “wrap” the feature selection around a specific classifier and select a subset of features based on the classifier’s accuracy using cross-validation [18]. Wrapper methods evaluate subsets of variables, that is, unlike filter methods, do not compute the worth of a single feature but the whole subset of features.

- **Filter method:** We have used Correlation-based Feature Selection (CBFS) method since it evaluates the usefulness of individual features for predicting the class along with the level of inter-correlation among them [19]. In this work we have used the implementation provided by Weka [8].

- **Wrapper method:** We have selected SVM as the specific classification algorithm and Area Under the Curve (AUC) as evaluation measure. Since an exhaustive search is impractical due to space dimensionality, we used heuristics, following a greedy stepwise approach. In this work we have used the implementation provided by Weka.

3.1 Support Vector Machine

Support vector machines (SVM) are supervised learning models which have been widely used in bioinformatics research and many other fields since their introduction in 1995 [7]. It is often defined as a non-probabilistic binary linear classifier, as it assigns new cases into one of two possible classes. In the readmission prediction problem, the model would predict whether a new case (the patient) will be readmitted within 30 days.

This algorithm is based on the idea that input vectors are non-linearly mapped into a very high dimensional space. In this new feature space it constructs a hyperplane which separates instances of both classes. Since there exist many decision hyperplanes that might classify the data, SVM tries to find the maximum-margin hyperplane, i.e. the one that represents the largest separation (margin) between the two classes.

In this work we have used the libSVM¹ implementation of the algorithm, which is the common implementation used for experimentation, and can be easily integrated to weka [8] using a wrapper. We have used a radial basis kernel function: $\exp(-\gamma*|u-v|^2)$ where $\gamma = 1/\text{num_features}$ and $C = 1$.

3.2 Random Forest

Random Forest [6] is a classifier consisting of multiple decision trees trained using randomly selected feature subspaces. This method builds multiple decision trees at training phase. In order to predict the class of a new instance, it is put down to each of these trees. Each tree gives a prediction (votes) and the class having most votes over all the trees of the forest will be selected. The algorithm uses the bagging method, i.e. each tree is trained using a random subset (with replacement) of the original dataset. In addition, each split uses a random subset of features.

One of the advantages of random forests is that generally they generalize better than decision trees, which tend to overfitting and naturally perform some feature selection. They can also be run on large datasets and can handle thousands of attributes without attribute deletion. In this work we have used Weka's implementation of the algorithm.

¹ <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

4 Results

In this section we analyze the prediction performance of different models on the emergency department short-time readmission dataset presented in [2]. As shown in Table 2 we have considered besides the original four subpopulations a fifth dataset that encompasses all of them.

Table 2. Comparative information about the subpopulations of the dataset

	Readmission within 30 days, no. (%) of patients		
	Overall no. of patients	No	Yes
	$n = 360$	n = 296 (82.2)	n = 64 (17.7)
Case management	94 (26.1)	73 (77.7)	21 (22.3)
Heart failure	70 (19.4)	62 (88.6)	8 (11.4)
Chronic obstructive pulmonary disease	80 (22.2)	64 (80)	16 (20)
Diabetes mellitus	116 (32.2)	97 (83.6)	19 (16.4)

4.1 Class Balancing

In readmission prediction analysis like in any other supervised classification problem, imbalanced class distribution leads to important performance evaluation issues and problems to achieve desired results. The underlying problem with imbalanced datasets is that classification algorithms are often biased towards the majority class and hence, there is a higher misclassification rate of the minority class instances (which are usually the most interesting ones from the practical point of view) [13].

As shown in Table 3, class imbalance is causing an accuracy paradox. If we just

Table 3. Confusion matrix of SVM on the diabetes mellitus dataset

		Predicted	
		Readmitted	Not readmitted
Actual	Readmitted	97	0
	Not readmitted	19	0

look at the accuracy of the model we get an 83.62 % although SVM just behaves as suing only the greater *a priori* probability to make the classification decision.

Resampling. There are several methods that can be used in order to tackle the class imbalance problem. Building a more balanced dataset is one of the most intuitive approaches. In our experiment we have used under-sampling as a preliminary approach and continued with an over-sampling using synthetic samples.

Under-sampling with random subsample. Given that there is a low number of samples for the minority-class, which is also the most relevant for classification, we can anticipate that reducing the amount of samples for the majority-class to be comparable to the minority-class and avoid the class imbalance will lead to a model with poor generalization capability.

Focusing on the diabetes mellitus subpopulation dataset, it is composed of 97 instances belonging to the *not-readmitted* class and only 19 of the *readmitted* class. An experiment consisting of subsampling the dataset to a distribution of 1:1.5 between the minority and majority classes, and then applying a Random Forest classifier shows the following results in Table 4.

Table 4. Comparison of performance evaluation metrics for RF over original and under-sampled versions of diabetes mellitus dataset

Dataset	Accuracy	Sensitivity	Specificity
Original	84.48	10.52	98.96
Under-sampled	61.7	31.57	82.14

As seen in Table 4, although the classification sensitivity has increased, it is still low (31.57 %) despite the sacrifice of both accuracy and specificity performance. Taken into account the low number of instances contained in our dataset, we don't consider under-sampling an effective approach.

Oversampling with SMOTE. We used Synthetic Minority Over-sampling Technique (SMOTE) [20] for oversampling the minority class. In order to avoid overfitting, we applied SMOTE (percentage of new instances equal to 200) at each fold of the 10-fold cross validation. If oversampling is done before 10-fold cross-validation, it is very likely that some of the newly created instances and the original ones are both in the training and testing sets, thus causing performance metrics being optimistic.

Our approach is to test the performance of two classifiers, namely SVM and Random Forests, using the over-sampled dataset, in order to compare it with the results obtained using the original imbalanced dataset. The choice of these two classifiers was based on the fact that both SVM and RF have been widely used in the literature, achieving good results [6]. On one hand, SVM has the advantage of been able to deal with data which is difficult to directly separate in the feature space, while on the other Random Forest has the advantage of the embedded feature selection process, which is helpful in high dimensional feature spaces. The experiment will be carried out by generating a model for each of the subpopulations on each of the specified scenarios. Table 5 shows the results of our experiment.

Results show that class-balanced dataset achieved better sensitivity than the original dataset. Nevertheless, both accuracy and specificity achieve worse results. It is worth noting that while performance is similar for both classifiers using the original dataset, SVM performs much better (in terms of sensitivity) when using the over-sampled version. At last, we observe that sensitivity improvement is rather small and it is obtained mainly at the expense of worsening both sensitivity and accuracy.

Table 5. Performance comparison using SVM and RF classifiers on original and over-sampled datasets

		Original			Over-sampled		
		Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy
Case management	SVM	1	0.42	0.87	0.98	0.42	0.86
	RF	1	0.42	0.87	1	0.42	0.87
Heart failure	SVM	1	0	0.88	0.90	0.12	0.81
	RF	1	0	0.88	1	0	0.88
COPD	SVM	1	0	0.80	0.81	0.37	0.72
	RF	1	0.37	0.87	1	0.43	0.88
Diabetes mellitus	SVM	1	0	0.83	0.88	0.15	0.76
	RF	1	0.10	0.85	0.96	0.10	0.82
All	SVM	1	0.21	0.86	0.78	0.40	0.71
	RF	1	0.28	0.87	0.99	0.28	0.86

4.2 Feature Selection

Our dataset has a high dimensional feature space. With the use of feature selection algorithms we want to find a feature subset that would reduce the complexity of the model (so that it would be easier to interpret by the physicians) while improving the prediction performance and reducing overfitting. For that purpose we are using a filter method, with Correlation-based Feature Selection [19] as metric and a wrapper method, with SVM as the specific classifier, both presented in Sect. 3.

The experiment consists in training a SVM and a RF classifier using the original feature set and the generated feature subsets. The performance of the classifiers will be compared in terms of sensitivity, specificity and accuracy for each of the subpopulations.

It's worth noting that the feature selection must be done using cross-validation. If full training set is utilized during attribute selection process, the generalization ability of the model can be compromised.

Table 6. Performance comparison of both feature selection methods

		Filter (CBFS)			Wrapper (SVM)		
		Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy
Case management	SVM	0.97	0.33	0.82	0.94	0.23	0.78
	RF	0.86	0.42	0.76	0.89	0.38	0.77
Heart failure	SVM	0.96	0.12	0.87	0.90	0.12	0.81
	RF	0.96	0.25	0.88	0.98	0	0.87
COPD	SVM	0.96	0.18	0.81	0.95	0.37	0.83
	RF	0.89	0.18	0.75	0.92	0.37	0.81
Diabetes mellitus	SVM	0.98	0	0.82	0.98	0.05	0.83
	RF	0.98	0	0.82	0.98	0.05	0.83
All	SVM	0.98	0.06	0.82	0.97	0.14	0.83
	RF	0.94	0.15	0.80	0.95	0.18	0.81

In Table 6 the results of the experiment are shown. According to these results, although in some cases the sensibility has been increased, overall the results are not as promising as expected. Actually, even though models are much simpler than the original model (i.e. the one using full feature set), the prediction performance has been reduced. Moreover, both feature selection methods have performed similarly, even if selected feature subsets differs considerably.

5 Conclusions and Future Work

This paper has presented a work on the prediction of 30-day readmission risk in Emergency Department. Several contributions have been presented regarding the enhancement of predictor's performance, with special focus on sensitivity, i.e. predictive power of the critical class of readmitting patients. First, we have conducted an experiment that shows the performance variations produced by class-balancing techniques. Second, we analyze different feature selection methods and metrics and evaluate their performance. Two classification algorithms have been used (SVM and Random Forest) in order to evaluate the different approaches.

According to the results of our analysis, we conclude that although class balancing improves sensitivity results, the dataset seems not to have enough minority-class instances. In addition, setting 30 days as the arbitrary threshold for assigning the binary class label may cause situations such as labelling a patient readmitted the 30th day as "readmitted" and another readmitted the 31st day as "not readmitted". This imposes a clear limitation to any generated model, since actually both patients should be treated as similar (in terms of readmission).

Future work will include addressing the problem with a regression approach, instead of supervised classification. Thus we want to avoid the mentioned arbitrary labelling problem. With a regression analysis approach we plan to predict not only the readmission risk but also the approximate readmission window (i.e. the time interval from hospital discharge and readmission).

We also plan to increase the size of the dataset, including more instances of the minority class. Extending the samples of the readmission class we expect to achieve better predictions and ultimately generate a better-generalizing model.

References

1. World Health Organization: Global health and ageing. World Health Organization, Geneva, Switzerland (2011)
2. Besga, A., Ayerdi, B., Alcalde, G., et al.: Risk factors for emergency department short time readmission in stratified population. *BioMed Res. Int.* 2015, 7 pages (2015). Article ID 685067, doi:[10.1155/2015/685067](https://doi.org/10.1155/2015/685067)
3. Van Walraven, C., et al.: Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Can. Med. Assoc. J.* **182**(6), 551–557 (2010)

4. Van Walraven, C., Wong, J., Forster, A.: LACE+ index: extension of a validated index to predict early death or urgent readmission after hospital discharge using administrative data. *Open Med.* **6**(3), 80–89 (2012)
5. Yu, S., Farooq, F., van Esbroeck, A., Fung, G., Anand, V., Krishnapuram, B.: Predicting readmission risk with institution-specific prediction models. *Artif. Intell. Med.* **65**(2), 89–96 (2015)
6. Ho, T.K.: Random decision forests. In: 1995 Proceedings of the Third International Conference on Document Analysis and Recognition, pp. 278–282. IEEE (1995)
7. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *SIGKDD Explor.* **11**(1), 10–18 (2009)
9. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., Kripalani, S.: Risk prediction models for hospital readmission: a systematic review. *JAMA* **306**(15), 1688–1698 (2011)
10. Health Quality Ontario - Early Identification of People At-Risk of Hospitalization. ISBN 978-1-4606-2908-6 (PDF) Queen’s Printer for Ontario (2013). Accessed 09 Mar 2016. Enlace: https://secure.cihi.ca/free_products/HARP_reportv_En.pdf
11. Feachem, R.G., Dixon, J., Berwick, D.M., Enthoven, A.C., Sekhri, N.K., White, K.L.: Getting more for their dollar: a comparison of the NHS with California’s Kaiser Permanente. *BMJ* **324**(7330), 135–143 (2002)
12. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
13. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **250**, 113–141 (2013)
14. Carpenter, C.R., Heard, K., Wilber, S., Ginde, A.A., Stiffler, K., Gerson, L.W., et al.: Research priorities for high-quality geriatric emergency care: medication management, screening, and prevention and functional assessment. *Acad. Emerg. Med.* **18**(6), 644–654 (2011)
15. Lopez-Aguila, S., Contel, J.C., Farre, J., Campuzano, J.L., Rajmil, L.: Predictive model for emergency hospital admission and 6-month readmission. *Am. J. Manage. Care* **17**(9), e348–e357 (2011)
16. Han, J.H., Zimmerman, E.E., Cutler, N., Schnelle, J., Morandi, A., Dittus, R.S., et al.: Delirium in older emergency department patients: recognition, risk factors, and psychomotor subtypes. *Acad. Emerg. Med.* **16**(3), 193–200 (2009)
17. New guidelines for geriatric EDs: guidance focused on boosting environment, care processes. *ED Manage* **26**(5), 49–53 (2014)
18. Phuong, T.M., Lin, Z., Altman, R.B.: Choosing SNPs using feature selection. In: 2005 IEEE Computational Systems Bioinformatics Conference (CSB 2005), pp. 301–309. IEEE (2005)
19. Hall, M.A.: Correlation-based feature selection for machine learning (Doctoral dissertation, The University of Waikato) (1999)
20. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)