Predictive Modeling for Pressure Ulcers from Intensive Care Unit Electronic Health Records

Pacharmon Kaewprag[1], Cheryl Newton[4], Brenda Vermillion[3,5], Sookyung Hyun[2,3],

Kun Huang[1,2], Raghu Machiraju[1]

[1]Department of Computer Science and Engineering, [2]Department of Biomedical Informatics,

[3]College of Nursing, The Ohio State University

[4]Department of Critical Care Nursing, [5]Department of Health Services Nursing Education,

The Ohio State University Wexner Medical Center

Abstract. *Our goal in this study is to find risk factors associated with Pressure Ulcers (PUs) and to develop predictive models of PU incidence. We focus on Intensive Care Unit (ICU) patients since patients admitted to ICU have shown higher incidence of PUs. The most common PU incidence assessment tool is the Braden scale, which sums up six subscale features. In an ICU setting it's known drawbacks include omission of important risk factors, use of subscale features not significantly associated with PU incidence, and yielding too many false positives. To improve on this, we extract medication and diagnosis features from patient EHRs. Studying Braden, medication, and diagnosis features and combinations thereof, we evaluate six types of predictive models and find that diagnosis features significantly improve the models' predictive power. The best models combine Braden and diagnosis. Finally, we report the top diagnosis features which compared to Braden improve AUC by 10%.*

Keywords. EHRs, intensive care unit, pressure ulcers, machine learning, predictive model

1 Introduction

Our goal in this study is to find risk factors associated with Pressure Ulcers (PUs) and to develop predictive models of PU incidence. A pressure ulcer is a localized injury to the skin and/or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear[1]. We focus on PU incidence in the Intensive Care Unit (ICU). Patients admitted to ICU have higher incidence of PUs than those admitted to the general hospital wards[2]. In the United States, the prevalence of patients afflicted with pressure ulcers across several ICUs ranged from 16.6% to 20.7% in 2009[2]. An estimated 2.5 million patients are treated each year in acute care settings at an additional cost of $11 billion per year mostly arising from PU incidence[3], which is possibly preventable.

The common model for PU assessment is called the Braden scale. An advantage of the Braden scale is that it uses only six simple features to compute a summed score indicating the risk of acquiring PU. This simplicity is also a disadvantage: It omits important risk factors shown to be significant with PU incidence in ICU setting[13, 17]. It has also been shown to have high sensitivity with low specificity, resulting in predicting too many patients as "at risk"[16, 17]. Many potentially significant risk factors of PU incidence not considered by the Braden scale can be found in Electronic Heath Record (EHR) systems. We posit that features extracted from patient records in these systems can significantly improve the quality of PU assessment scoring.

In particular, we study medication and diagnosis features extracted from EHRs. For medication features, we use the set of medications prescribed during the patient's ICU stay. For diagnosis features, we use the ICU discharge diagnoses as encoded by International Classification of Diseases (ICD)-9 codes[12]. There are many such medication and diagnosis features, so we first perform univariate analysis to identify the features of each type strongly associated with PUs. Without such a feature selection process, predictive models (discussed next) would consist of a huge number of features which, as we show in Section 4, can harm prediction performance due to the curse of dimensionality.

Predictive modeling methods provide a framework by which clinicians can predict the likelihood that a patient will be diagnosed with a disease in the future. Accurate predictive models can help clinicians recommending preventive care to the patients. We evaluated six types of predictive models for PU incidence using five sets of features: 1) Braden, 2) Medication, 3) Diagnosis, 4) Braden & Diagnosis, and 5) Braden & Medication & Diagnosis. We find that using diagnosis features significantly improves the models' predictive power. Finally, we report the top diagnosis features, which improve assessment quality over only Braden features (as measured by AUC) by 10%. The overall process of our study is shown in Figure 1.
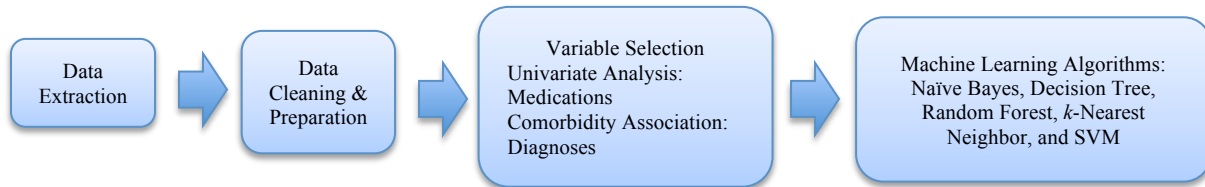

Figure 1. Our workflow for developing predictive model of PU among ICU patients using EHR data

2 Methods

Data Source: Data are 7717 patient records from three adult ICUs at The Ohio State University Wexner Medical Center (OSUWMC). An Information Warehouse (IW) compiles EHR data from an Essentris© documentation system, administrative system (ADT), laboratory system, computerized provider order entry (CPOE), and medication system. Patients (age ≥ 18) admitted to ICUs between the years 2007 to 2010 comprise the sample. EHR data elements pertinent to patient demographics, diagnoses, and medications are retrieved from the IW. Patients who have contracted PU are identified by reviewing discharge diagnoses marked with ICD-9 codes. For instance, if a patient has an ICD-9 code, 707.07 (Pressure ulcer, Heel), the patient is included in the PU group. On the other hand, if a patient does not have any of the ICD-9 codes representing PUs, the patient is then included in the non-PU group. Institutional Review Board (IRB) approval is obtained for data extraction. Data are de-identified by the IW staff as the honest broker. Patient demographics are summarized using descriptive statistics as shown in Table 1.

Data Cleaning and Preparation: First, patients who have a PU at the time of admission are excluded. In addition, patients whose ICU stay is shorter than 72 hours are excluded since PUs generally develop after 72 hours of admission[11]. Second, if a patient has multiple hospitalizations (for any reason) during the study period, only the first hospitalization record is included. If a patient has more than one ICU admission record during the hospitalization, only the first ICU admission record is included in the analysis. This is because our objective is to find risk factors of patients who have the first incidence of PUs during ICU stay. Patients who have PUs at the time of admission may have previously been exposed to unknown risk factors of which we have no data, and of which clinicians have no control. This patient selection process is consistent with our previous study[4].

The Braden scale contains 6 subscales: sensory perception, moisture, activity, mobility, nutrition, and friction & shear. Our Braden features include these, since most of the subscales have significant association with PU incidence, as well as a summed Braden scale for consistency with previous work[13, 14].

Medications that are used for the patients in PU and non-PU groups during the ICU stay are listed. This list is reviewed by a research team, which is comprised of a registered nurse, two ICU clinical nurse specialists, and a dietician. Through a manual review, medications are grouped into salient categories, for instance, Meperidine and Nalbuphine are classified into the Analgesia category. The vasoactive category contains vasodilators e.g., sildenafil and vasoconstrictors e.g., dopamine. Medication categories are coded as dichotomous variables.

Discharge diagnostic ICD-9 codes are used to identify patients experiencing maladies during their sojourn at ICU. The diagnostic data are coded with ICD-9 codes that are 5 digits long and are extracted from the EHR system. The first three digits indicate a main disease and the last two provide additional information about the disease. ICD-9 codes are collapsed into 3 digits in order to analyze the main diseases. Most of the 707 ICD-9 codes are considered PUs except 707.1 (Ulcer of lower limbs), 707.8 (Chronic ulcer of other specified sites), and 707.9 (Chronic ulcer of unspecified site). Those codes are grouped as 707-notPU.

Variable Selection: Univariate analysis is carried out to determine what medication categories are highly associated with PUs. Chi-square statistics are applied to medication categories with frequencies greater than 20. Otherwise, Fisher's Exact Test (FET) is used. The variable is considered significant if its *p-value* is less than 0.1. Hence, the medication categories that are significantly associated with PUs are retained.

To quantify the strength of the association between PU and each diagnosis, χ^2-statistics are employed. The premise behind the χ^2-test is to examine relatedness of two events by measuring the deviation between observed

and expected values. Only discharge diagnoses are included since they are clinically more meaningful than admission diagnoses. After the comorbidity association is created, we remove weakly comorbid conditions in the same way as we dealt with medication variables, through a statistical significance test. This process yields a subset of diagnoses highly associated with PUs to be used as variables for machine learning algorithms.

Machine Learning: We apply six diverse machine learning algorithms on results from the univariate analysis to build predictive models using 10-fold cross validation: logistic regression (LR), support vector machine (SVM), decision tree (DT), random forest (RF), k-nearest neighbor (kNN), and Naïve Bayes (NB). These methods were accessed through the WEKA software suite.

LR is a statistical method that predicts the probability of dichotomous outcomes from one or more independent variables and has been widely used in medical studies[13, 15-17]. SVM[5] has the ability to learn a highly nonlinear decision boundary and outperforms other methods on some datasets[6, 7]. DT generates descriptive models and has been widely used in clinical applications where interpretation is desired. RF offers robustness against overfitting and outliers, and has been found to give good performance in several applications[8, 9]. kNN is an intuitive method which classifies a data point based on majority votes of its neighbors. NB is a probabilistic classifier, which has higher bias than LR thus converges faster when there is little training data[18].

We select machine learning algorithm hyperparameters according to recommended practices in the literature[10]. For instance, for SVM grid search for the best regularization (c) and kernel function (g) parameters is carried out on training data. Both libSVM and SVMlight are investigated. For the DT method, parameter selection through cross-validation is used to find the best confidence factor (c) for pruning. For kNN the number of neighbors k is varied from 5 to 100. For RF the number of trees is varied from 10 to 250.

3 Results

Patient Demographics: A total of 7,717 ICU patients are included in the analysis. The number of patients in PU group is 590, while the number of patients in non-PU group is 7,127. Patient demographics are summarized in Table 1. Of the patients, 57.4% are male and 82.2% are demographically classified as white. The mean age of the patients is 57.7 years and the mean length of ICU stay is 10.1 days.

Table 1. Demographics of ICU Patients (N=7717)

Variable		Total	PU Group (N=590)	Non-PU Group (N=7127)	Statistic	*p* value
Gender, freq (%)	Male	4426	378 (64.1%)	4048 (56.8%)	x^2=11.9	<.000
	Female	3291	212 (35.9%)	3079 (43.2%)		
Race/Ethnicity, freq (%)	White	6345	469 (79.5%)	5876 (82.4%)	x^2=3.15	.076
	Non-white	1372	121 (20.5%)	1251 (17.6%)		
Age (years), mean (SD)		57.7 (15.9)	59.0 (15.5)	57.6 (16)	t=4.52	.034
Length of ICU stay (days), mean (SD)		10.1 (10)	13.4 (14.3)	9.8 (9.6)	t=70.56	<.000

Table 1 shows that gender and length of ICU stay are very statistically significant of PU development. However, demographic data are clinically obvious PU risk factors for clinicians. The clinicians are already attuned to the relationship between length of ICU stay or hospitalization and PU incidence. Consequently, we are looking for non-obvious relationships that could be related to PUs such as medications and diagnoses.

Medication variable selection: In total, 828 unique medications are administered to the patients in our study. Medications are grouped into 72 categories by clinical experts based on the clinical effects and pharmacology of the medications. For example, **Electrolytes:** calcium acetate, glucose, potassium chloride. **Analgesia:** hydromorphone, meperidine, morphine, nalbuphine. **Sedation, continuous:** lorazepam, midazolam, pentobarbital, propafol. **Neuromuscular Blockage:** pancuronium, rocuronium, succinylcholine, vecuronium. **Antibiotics/Antifungal/Antiviral:** imipenem, isoniazid, itraconazole, lamivudine (epivir), linezolid. **NSAIDS:** ibuprofen, naproxen. Categories whose frequency is less than 10 are removed, since they are not considered significant for the univariate analysis. Additionally, Electrolytes, IV fluid, Research drugs, and Miscellaneous categories are removed since they do not appear to be clinically meaningful; thus, 49 categories are used for univariate analysis. Only 18 medication categories are found to be significantly associated with PUs at significance level 90% (i.e., p-value < 0.1).

Comorbidity association: The number of main discharge diagnoses after collapsing the ICD-9 codes to three digits totaled 861 diagnoses. We construct the comorbidity association in the same manner as the medication variables are selected by removing diagnoses whose frequency is less than 10. Retained conditions are qualified by χ^2 statistic greater than 20 (i.e., significance level $\alpha < 0.001$), resulting in 61 comorbid conditions highly associated with PUs.

Machine Learning: We report the performance of predictive models of PU incidence in Tables 2 thru Table 6, which show five performance measures of six different machine learning algorithms on features that include Braden,

medication, and diagnosis features, and combinations thereof. The performance measures are sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), and area under ROC curve (AUC). We focus on AUC performance measure since it considers both true positive rate (sensitivity) and false positive rate (1-specificity). For SVM, we report the performance of libSVM which gave comparable to slightly better performance than SVMlight. For kNN, we used $k = 7$ and for RF, we used 150 trees which performed best.

As seen in Table 3, predictive models using medication features perform poorly, with the best model NB having only AUC 0.62. By contrast, Table 2 shows that Braden scale alone gives 0.73 AUC with its best model, LR. Table 4 shows that when using diagnosis features, LR and NB give the highest AUC of 0.80. Using the combination of three types of features, shown in Table 5, LR performed best with AUC 0.827 followed closely by NB and RF with AUC 0.82. Table 6 shows that Braden and diagnosis features outperform using all three types of features with AUC 0.830 from LR. Medication features perform poorly individually (Table 3) and worse jointly (Table 5) than when they are omitted from the model (Table 6). Diagnosis features (Table 4), on the other hand, outperform Braden features individually (Table 2) and give the best performing model in Table 6.

Since adding diagnosis features to Braden gives the best results, we present the most significant diagnosis features in Table 7, which along with the other results is discussed in more detail in the next section.

Table 2. Braden features

	SENS	SPEC	PPV	NPV	AUC
LR	0.007	0.999	0.190	0.924	**0.731**
NB	0.297	0.939	0.288	0.941	0.727
DT	0.046	0.993	0.368	0.926	0.712
kNN	0.030	0.997	0.441	0.925	0.707
RF	0.050	0.991	0.328	0.927	0.688
SVM	0.623	0.633	0.124	0.953	0.628

Table 3. Medication features

	SENS	SPEC	PPV	NPV	AUC
LR	0.001	1.000	0.033	0.923	0.615
NB	0.002	0.999	0.120	0.923	**0.617**
DT	0.017	0.990	0.132	0.924	0.562
kNN	0.002	0.999	0.031	0.923	0.529
RF	0.022	0.986	0.113	0.924	0.530
SVM	0.449	0.630	0.091	0.933	0.540

Table 4. Diagnosis features

	SENS	SPEC	PPV	NPV	AUC
LR	0.103	0.993	0.576	0.930	**0.801**
NB	0.477	0.869	0.232	0.953	**0.800**
DT	0.191	0.936	0.201	0.933	0.569
kNN	0.008	0.999	0.553	0.924	0.713
RF	0.066	0.992	0.398	0.928	0.779
SVM	0.711	0.713	0.172	0.968	0.712

Table 5. Braden, medication, and diagnosis features

	SENS	SPEC	PPV	NPV	AUC
LR	0.167	0.988	0.541	0.935	**0.827**
NB	0.642	0.809	0.218	0.965	0.815
DT	0.243	0.934	0.234	0.937	0.579
kNN	0.011	0.999	0.477	0.924	0.684
RF	0.088	0.994	0.541	0.929	0.817
SVM	0.743	0.724	0.183	0.972	0.734

Table 6. Braden and diagnosis features

	SENS	SPEC	PPV	NPV	AUC
LR	0.160	0.990	0.556	0.934	**0.830**
NB	0.628	0.821	0.226	0.964	**0.815**
DT	0.232	0.938	0.238	0.936	0.588
kNN	0.023	0.999	0.737	0.925	0.670
RF	0.109	0.991	0.515	0.931	0.806
SVM	0.744	0.727	0.185	0.972	0.736

Table 7. Top 10 discharge diagnoses' comorbidity with PU

x^2	CD-9	Disease description
524.193	344	Other paralytic syndromes
487.9 1	995	Certain unclassified adverse effects
476.778	038	Septicemia
444.992	730	Osteomyelitis/periostitis/bone infections
308.461	785	Cardiovascular system symptoms
232.930	482	Bacterial pneumonia
211.980	599	Disorders of urethra and urinary tract
198.863	518	Other diseases of lung
168.651	112	Candidiasis
155.338	263	Other protein-calorie malnutrition

4 Discussion

Incorporating diagnosis along with Braden features as shown in Table 6 yields the best predictive model for PU incidence. LR and NB model variants scored 83% and 82% AUC, respectively. LR, the only linear model, performs best indicating that a linear separating boundary is effective for our data. NB is known to underperform LR when training size is large[18] (we have N=7717) and when significant features are collinear (our Braden total is with subfeatures). DT and RF can only make axis-parallel "cuts" and do not find a good non-axis parallel boundary. kNN with small k lacks enough global knowledge, and when increasing k to give more, loses PPV i.e., classifies all as negative as PU patients are the minority by a factor of approximately 1:12. SVM proves too difficult to tune; grid search does not find as good a separating boundary as LR. PPV and NPV are influenced by the ratio of PU patients in the dataset[19], resulting in low PPV and high NPV since PU patients are in the minority. All models have high SPEC and low SENS. Consequently, they are more appropriate for ruling out PU incidence, and are likely to properly classify healthy patients as not having a PU. This makes them candidates as a second-level test[19] for patients already identified as at risk by Braden scale, which is known to have the complimentary characteristics of low SPEC and high SENS[16,17].

Table 7 lists the top 10 discharge diagnoses from the comorbidity association ranked by x^2-statistic. Most of the diagnoses in Table 7 are associated with bed-rest, which is well known to be associated with PU incidence.

The codes 995, 038, 730, and 112 are consistent with Bours et al.'s findings[14] as to the primary causes of admission to ICU for PU patients. Empirically, our clinician collaborators have indicated that most PU patients in their unit suffer from specific adverse effects of Sepsis (995.91) and Systemic Inflammatory Response Syndrome (SIRS, 995.9). They also noted that with lung-related maladies such as pneumonia, patients sit up at a high angle, which places pressure in a small area, increasing the chance of PUs.

We found medications alone (Table 3) to be poor predictors of predict PU risk, consistent with findings of Kaitani et al.[15]; our models constructed with them underperform the others. Additionally, incorporating features which are not relevant to predicting PU can cause the model to produce a worse result. This is demonstrated by the 1% to 2% AUC performance difference between the worse model including Braden, medication, and diagnosis features (Table 5) compared to the better model with only Braden and diagnosis features (Table 6). Furthermore, we experimented on predictive models without univariate analysis feature selection, that is, all 72 medication and all 861 diagnosis features. For medication with all features, DT decreased by 4% AUC compared to with univariate analysis, while other methods increased by 1% - 4%. For all models with diagnosis, LR decreased by 13% to 16% AUC and SVM decreased by 5% to 6% while other models changed by -1% to 3%. Because these models with many more features delivered generally poorer classification performance due to the curse of dimensionality, and exhibited longer computational runtime (5 minutes to 2 hours for diagnoses), we used feature selection to reduce the number of features as described earlier in Sections 2 and 3.

There are limitations pertaining to our study. First, the data are from a single institution; thus, interpretation of the findings are limited. Second, our IW lacks APACHE II severity scores, hence we considered PU incidence as dichotomous. Finally, the predictive power for pressure ulcer incidence in this study is only based upon Braden scale, discharge diagnosis, and medication. In the future we will investigate other patient data such as demographics, procedures, and laboratory settings to determine which contribute meaningfully to a risk assessment model.

In conclusion, in this study we have identified PU risk factors and evaluated predictive models of PU incidence in ICU setting. Beyond the simple features used in the traditional PU assessment, we extracted medication and diagnosis features from patient EHRs and constructed predictive models. Studying baseline Braden, medication, and diagnosis features and combinations thereof, we evaluated six types of predictive models and found that diagnosis features significantly improve the models' predictive power. In our best model, Braden and diagnosis features perform the best and improve AUC performance by 10%. Of the predictive model types, LR and NB performed best throughout. Lastly, we investigated diagnosis features in detail and report the 10 most comorbid with PU. Most of them relate to patient infection, immobility and imperceptions. Both predictive models and risk factors can assist clinicians in administering preventive care to patients.

References

1 European Pressure Ulcer Advisory Panel and National Pressure Ulcer Advisory Panel. Prevention and treatment of pressure ulcers: quick reference guide 2009. http://www.npuap.org/ (accessed 1/29/2013).
2 VanGilder C, Amlung S, Harrison P, Meyer S. Results of the 2008-2009 international pressure ulcer prevalence survey and a 3-year, acute care, unit-specific analysis. J OWM. 2009 Nov 1;55(11):39–45.
3 Reddy M, et al. Treatment of pressure ulcers: a systematic review. JAMA. 2008 Dec 10;300(22):2647–62.
4 Hyun S, Vermillion B, Newton C, Fall M, Li X, Kaewprag P, Moffatt-Bruce S, Lenz ER. Predictive validity of the Braden scale for patients in intensive care units. AJCC. 2014 Jun 3;22(6):514-20.
5 Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995 Sep;20(3):273–97.
6 Tan PN, Steinbach M, Kumar V. Introduction to data mining. 1st ed. Addison Wesley; 2005 May 12.
7 Tang Y, Zhang YQ, Chawla NV, Krasser S. SVMs modeling for highly imbalanced classification. IEEE SMC. 2008 Dec 9;20(1):281–8.
8 Kawaler E, et al. Learning to predict post-hospitalization VTE risk from EHR data. AMIA Annu Sympp Proc. 2012 Nov 3;2012:436–45.
9 Mani S, et al. Type 2 diabetes risk forecasting from EMR data using machine learning. AMIA Annu Symp Proc. 2012 Nov 3;2012:606–15.
10 Chapelle O, Vapnik V, et al. Choosing multiple parameters for support vector machines. Machine Learning. 2002;46(1-3):131–159.
11 Carlson EV, Kemp MG, Shott S. Predicting the risk of pressure ulcers in critically ill patients. AJCC. 1999 Jul;8(4):262–9.
12 Searchable online version of the 2009 ICD-9-CM. http://icd9.chrisendres.com/ (accessed 7/31/2014).
13 Cox J. Predictors of pressure ulcers in adult critical care patients. AJCC. 2011 Sep;20(5):364-75.
14 Bours GJ, De Laat E, Halfens RJ, Lubbers M. Prevalence, risk factors and prevention of pressure ulcers in Dutch intensive care units. Results of a cross-sectional survey. Intensive Care Med. 2001 Oct;27(10):1599-605.
15 Kaitani T, et al. Risk factors related to the development of pressure ulcers in the critical care setting. JCN. 2010;19(3-4):414–21.
16 Frankel H, et al. Risk factors for pressure ulcer development in a best practice surgical intensive care unit. Am Sur. 2007;73(12):1215-7.
17 Slowikowski GC, Funk M. Factors associated with pressure ulcers in patients in a surgical intensive care unit. J WOCN. 2010;37(6):619–26.
18 Ng YA, Jordan MI. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. NIPS 2001.
19 Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. BJA: CEACCP. 2008;8(6):221-3.