

PAPER • OPEN ACCESS

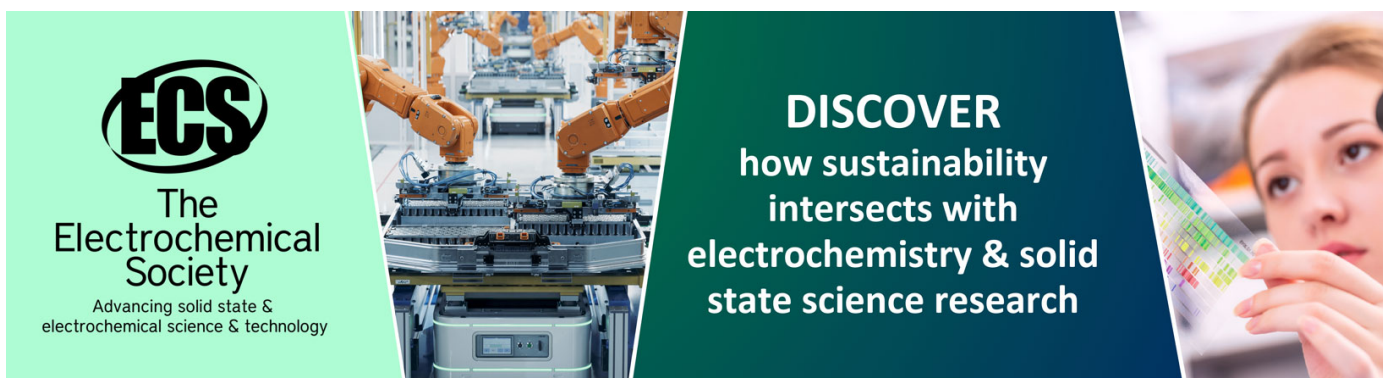
Predictive Modelling for Hospital Readmission Risk in the Philippines

To cite this article: J Landicho *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **864** 012061

View the [article online](#) for updates and enhancements.

You may also like

- [Chernobyl cleanup workers from Estonia: follow-up for cancer incidence and mortality](#)
Kaja Rahu, Anssi Auvinen, Timo Hakulinen et al.
- [The risk of childhood leukaemia following exposure to ionising radiation—a review](#)
Richard Wakeford
- [EPI-CT: design, challenges and epidemiological methods of an international study on cancer risk after paediatric and young adult CT](#)
Magda Bosch de Basea, Mark S Pearce, Ausrele Kesminiene et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Predictive Modelling for Hospital Readmission Risk in the Philippines

J Landicho^{1,2}, V Esichaikul², R Sasil³

¹Department of Information Technology, College of Information Technology and Computing, University of Science and Technology of Southern Philippines, Cagayan de Oro City, Philippines 9000

²Department of Information and Communication Technologies, School of Engineering and Technology, Asian Institute of Technology, Pathum Thani, Thailand 12120

³Department of Internal Medicine, Northern Mindanao Medical Center, Cagayan de Oro City, Philippines 9000

Email: juarlandicho@ustp.edu.ph

Abstract. Predictive models have been developed over the years to identify patients at risk of readmission. The goal of this study is to identify the risk factors associated to a patient's readmission within one year in the cohort study including acute myocardial infarction (AMI), Heart Failure (HF), Chronic Obstructive Pulmonary Disease (COPD) and Pneumonia (PN) in a reputed Philippine hospital. Four predictive models were used and evaluated using performance metrics. The study found Logistic Regression as the most performing model in most of the cohort studies. There are 6 to 8 variables significantly associated with the readmission of high-risk patients.

1. Introduction

Hospital readmission happened when a patient hospitalized again due to same or different diagnosis within a time frame after being discharged. Hospital readmission is either planned or unplanned treatment in the same (index) hospital or the different hospital within 7, 15, 30, 60, or 365 days [8]. In most cases, unplanned readmission is always associated with lower quality of care received during index admissions. There are many factors associated with readmission. Some are intrinsic, attributable to the reduced patient reserve due to disease progression and severity at each admission. Some may have related to the clinical planning and care coordination while patients are still in the hospital. Others are related to post-discharge care and other social factors [1]. The most common medical condition diagnosis that has higher readmission rates are acute myocardial infarction, congestive heart failure, pneumonia, chronic obstructive pulmonary disease, total hip arthroplasty and total knee arthroplasty [2].

Hospital readmission is not only degrading the quality of healthcare but also affects patient financial due to increasing medical expenses. Some developed countries like England and Denmark consider readmission applicable to almost all conditions. In the US, the readmission penalty program focuses on the diagnosis that is more expensive and which can be preventable [1,3]. Furthermore, in developing countries such as South Africa and Nigeria, hospital readmission problems are much pronounced due to limited resources and lack of funding to grapple [7].



One of the strategies in reducing unplanned hospital readmission rates is to apply classifiers and predictive models, including Logistic Regression and machine learning algorithms. Several reviews to date have developed readmission risk prediction models to identify high-risk patients, where it emphasized that not every predictive model works equally for each hospital. Hence, the applicable model depends on the setting, and the population studied, and the overall performance of reviewed models was still inconsistent [4,5]. With around 80% readmission occurring at the same hospital, it may be reasonable to assume that the models developed based on same-hospital readmission data are also able to predict all-hospital readmission [6]. Also, few studies have published regarding hospital readmissions in developing countries, including the Philippines [9, 10]. Furthermore, there is a lack of digitized data available in the Philippine hospital, and no national program penalized hospitals for higher readmission rates. This study aims to design and build a predictive model to identify the predictors and patients at risk of readmission within a one-year timeframe in a same-hospital setting.

2. Methodology

2.1. Data Collection and Definition

We used administrative data and clinical data obtained from Northern Mindanao Medical Center (NMMC), Cagayan de Oro City, Philippines approved from their Research Ethics Board. This study was a retrospective cohort study that aims to analyses patients hospitalized in a primary diagnosis based on the identified conditions within 1-year readmission. The initial dataset contains 2,234 patients who readmitted between June 2017 and December 2018. The inclusion criteria consisted of patients who were 18 years of age or older and had an index inpatient admission with a primary diagnosis of four conditions, including AMI, HF, PN, and COPD. These conditions extracted using the International Classification of Disease, Ninth Edition (ICD-9) codes to confirm diagnosis and discharge. For AMI, the codes used were I20.0, I21.9; for HF, I50.0, I50.9; for COPD, J44.9, J44.8; and for PN, J18.92, J18.93, J18.99, J69.0. Although it is possible for a patient to have multiple admission and readmission, this study only limits the population or sample to the first occurrence to ensure that a unique patient is analyzed. This study excludes records when a patient has missing records to specific attributes such as missing length of stay. It also considers excluding patients that lead to transfer to another acute care hospital and died during readmission. The final sample consisted of 200 AMI patients, 127 HF patients, 75 COPD patients, and 261 PN patients.

2.2. Data Preparation

It is essential to have appropriate preprocessing since the data acquired are inconsistent and incomplete. First, constant features or features that provide no information in the dataset removed to avoid error in models. Second, we applied multiple imputations with 5 imputations using predictive mean matching (PMM) for accounting missing data. Finally, combine filter and wrapper feature selection methods to identify relevant and best features in predicting hospital readmission. All analyses conducted using R statistical software.

2.3. Constructing Predictive Modelling

Before constructing a predictive model, each cohort dataset was separated into training (70%) and testing (30%) sets which is one of the common ratios used in medical research [11]. Logistic Regression, Support Vector Machine, Random Forest, and Neural Network were used to classify using the training set [12]. To further enhance the models' performance, tuning was conducted to find the best parameters. The result from each model was used to predict the testing set. The effectiveness of each model was evaluated using the area under the curve (AUC), accuracy, sensitivity, specificity, and precision with 10-fold cross-validation.

3. Results and Discussion

3.1. Patient and Clinical Characteristics

Table 1 shows the sociodemographic, clinical records, hospital utilization, and medical history for the 4 cohort study. The results showed there are 200 admissions for AMI, 127 admissions for HF, 75 admissions for COPD, and 261 admissions for Pneumonia. All of the cohort studies have at least 21% of hospital readmission rates. The mean (standard deviation) age was 62.12(11.16) for AMI, 54.02 (17.24) for HF, 66.73 (10.71) for COPD, and 57.43 (18.33) for PN, with the proportion of females 37%, 52.8%, 29.3% and 51% respectively. It is observed that there are 67.5% have a medical history of Hypertension admitted in AMI, 45.3% have a history of PTB treatment admitted in COPD, and the rest of cohort study patients have found no medical history during admission.

Table 1. Summary descriptive statistics of the Independent Candidate Variables for each cohort study.

Predictors	Acute Myocardial Infarction	Heart Failure	Chronic Obstructive Pulmonary Disease	Pneumonia
	(N = 200)	(N = 127)	(N = 75)	(N = 261)
Readmission, n (%)				
Yes	42 (21.0)	50 (39.37)	26 (34.67)	74 (28.35)
No	158 (79.0)	77 (60.63)	49 (65.33)	187 (71.65)
Socio Demographics				
Age, years, mean (SD)	62.12 (11.66)	54.02 (17.24)	66.73 (10.71)	57.43 (18.33)
<65	115 (57.5)	96 (75.6)	30 (40.0)	163 (62.4)
≥65	85 (42.5)	31 (24.4)	45 (60.0)	98 (37.5)
Gender, n (%)				
Male	126 (63.0)	60 (47.2)	53 (70.7)	128 (49.0)
Female	74 (37.0)	67 (52.8)	22 (29.3)	133 (51.0)
Marital Status, n (%)				
Single	27 (13.5)	25 (19.7)	5 (6.7)	56 (21.5)
Separated	1 (0.5)	-	-	1 (0.4)
Widow	33 (16.5)	23 (18.1)	18 (24.0)	44 (16.9)
Married	137 (68.5)	79 (62.2)	51 (68.0)	158 (60.5)
Missing	2 (1.0)	-	1 (1.3)	2 (0.8)
Social Services, n (%)				
C1	20 (10.0)	6 (4.7)	4 (8.2)	16 (6.1)
C2	5 (2.5)	4 (3.1)	-	8 (3.1)
C3	42 (21.0)	27 (21.3)	15 (20.0)	47 (18.0)
D	133 (66.5)	85 (66.9)	55 (73.3)	187 (71.7)
Missing	-	5 (3.9)	1 (1.3)	3 (1.1)
Clinical Records				
No. of Operations, mean (SD)	0.14 (0.45)	0.32 (0.64)	0.13 (0.44)	0.26 (0.52)
No. of Medications, mean (SD)	11.14 (3.76)	11.47 (4.60)	9.45 (4.11)	8.99 (4.03)
No. of Laboratory Test, mean (SD)	14.24 (9.05)	16.57 (9.47)	13.56 (7.92)	15.29 (8.50)
No of Comorbidities, mean (SD)	2.45 (1.89)	3.83 (1.73)	2.32 (1.24)	2.79 (1.93)

Table 1. Summary descriptive statistics of the Independent Candidate Variables for each cohort study (Continued...).

Predictors	Acute Myocardial Infarction	Heart Failure	Chronic Obstructive Pulmonary Disease	Pneumonia
Initial Lab Test				
Creatinine, mean (SD)	1.82 (3.95)	1.92 (1.94)	1.44 (2.17)	1.82 (3.31)
Hematocrit, mean (SD)	37.78 (6.47)	38.04 (8.87)	40.86 (9.89)	34.75 (11.03)
Hemoglobin, mean (SD)	13.20 (2.99)	12.94 (3.80)	13.84 (4.14)	11.73 (4.03)
Blood Urea Nitrogen, mean (SD)	38.7 (28.87)	52.55 (39.79)	37.98 (21.15)	43.73 (37.61)
Potassium, mean (SD)	4.80 (10.29)	4.17 (0.86)	3.99 (0.70)	3.80 (0.68)
Sodium, mean (SD)	4.80 (10.29)	135.46 (7.00)	135.42 (6.95)	132.72 (19.17)
White Blood Cells, mean (SD)	11.55 (4.15)	118.67 (1100.97)	11.74 (5.22)	98.85 (888.97)
Hospital Utilization				
Length of Stay, mean (SD)	7.33 (4.19)	9.87 (6.11)	8.81 (6.90)	10.98 (23.01)
No. of Previous Admission, mean (SD)	0.32 (0.93)	0.75 (1.49)	0.71 (1.26)	0.80 (0.45)
No. of Previous Outpatient, mean (SD)	1.74 (4.49)	3.03 (5.91)	3.15 (9.37)	3.24 (7.59)
Utilization of Emergency room, mean (SD)	1.94 (2.15)	2.71 (2.89)	3.73 (5.16)	3.24 (4.84)
Medical History				
Hypertension, n (%)	Yes	135 (67.5)	52 (40.9)	27 (36.0)
	No	55 (27.5)	65 (51.2)	44 (58.7)
	Not Ask / No Answer	10 (5.0)	10 (7.9)	4 (5.3)
Diabetes Mellitus, n (%)	Yes	44 (22.)	19 (15.0)	3 (4.0)
	No	142 (71.0)	93 (73.2)	62 (82.7)
	Not Ask / No Answer	14 (7.0)	15 (11.8)	10 (13.3)
Bronchial Asthma, n (%)	Yes	7 (3.5)	6 (4.7)	18 (24.0)
	No	168 (84.0)	97 (76.4)	43 (57.3)
	Not Ask / No Answer	25 (12.5)	24 (18.9)	14 (18.7)
Arthritis, n (%)	Yes	14 (7.0)	4 (3.1)	4 (5.3)
	No	125 (62.5)	73 (57.5)	51 (68.0)
	Not Ask / No Answer	61 (30.5)	50 (39.4)	20 (26.7)

3.2. Predictors of 1-year Readmission

Predictors were determined by using both forward and backward selection method. Initially, there are 90 variables collected in each cohort study, and it was then reduced to 54 variables (Table 2). A range

of 6 to 8 variables is considered to have a significant effect on the 1-year readmission in four cohort studies.

Table 2. Predictors selected by the filter and wrapper-based machine learning techniques for predicting 1-year readmission.

Cohort Study	Selected variables
Acute Myocardial Infarction	Number of Previous Outpatient, Utilization of Emergency room, White Blood Cell, Family History of Hypertension and Heart Disease, Family History of Bronchial Asthma, and Number of Comorbidities
Heart Failure	Pulse Rate, Number of Previous Outpatient, History of Bronchial Asthma, Height, History of Arthritis, Initial Diagnosis and Temperature
Chronic Obstructive Pulmonary Disease	Initial Diagnose, Utilization of Emergency room, Weight, Family History of Bronchial Asthma, Type of Social Service, Red Blood Cell, History of Bronchial Asthma and Age
Pneumonia	History of Previous Transfusion, History of Past Surgery, Number of Previous Admission, Sodium, Creatinine, Initial Diagnosis, Weight and Temperature

3.3. Performance of Predictive Models

Table 3 summarizes the performance of the best models for each disease cohort. As shown, in terms of AUC and other performance metrics. Logistic Regression is the best performing model for predicting readmission in HF and COPD. SVM show better performance in predicting AMI provides high result in specificity and precision in most of the cohort studies. Neural Network is the best performing model for Pneumonia, and it can perform better in less hidden layers. It is observed that Random Forest was the least impressive model in all disease cohorts.

Table 3. Comparison of different predictive models.

Model	Acute Myocardial Infarction	Heart Failure	Chronic Obstructive Pulmonary Disease	Pneumonia
Logistic Regression				
AUC	0.5787	0.7353	0.6919	0.5948
Accuracy	0.7714	0.7632	0.7391	0.7711
Sensitivity	0.2000	0.4706	0.5714	0.2381
Specificity	0.9574	1.000	0.8125	0.9516
Precision	0.6000	1.000	0.5714	0.6250
Support Vector Machine				
AUC	0.6000	0.6765	0.6205	0.5000
Accuracy	0.8065	0.7105	0.6957	0.7831
Sensitivity	0.2000	0.3579	0.5714	0.1429
Specificity	1.000	1.000	0.7500	1.000
Precision	1.000	1.000	0.5000	1.000

Table 3. Comparison of different predictive models (Continued...).

Model	Acute Myocardial Infarction	Heart Failure	Chronic Obstructive Pulmonary Disease	Pneumonia
Random Forest				
AUC	0.5149	0.5524	0.6607	0.4750
Accuracy	0.6774	0.5789	0.6957	0.6627
Sensitivity	0.2000	0.2353	0.5714	0.0952
Specificity	0.8298	0.8095	0.7500	0.8548
Precision	0.2727	0.5000	0.5000	0.1818
Neural Network				
AUC	0.5333	0.7059	0.6696	0.6187
Accuracy	0.7742	0.7368	0.6522	0.7831
Sensitivity	0.667	0.418	0.7143	0.2857
Specificity	1.000	1.000	0.6250	0.9516
Precision	1.000	1.000	0.4545	0.6666

4. Conclusion and Future Works

In this study, we developed and compared four predictive models. All models were evaluated based on several performance metrics. It found that Logistic regression performs better when predicting readmissions. It is determined which predictors have significant associated with 1-year readmission using selection technique. The use of the predictive model can be a great tool in providing insights to design disease-specific interventions and decrease the readmission of high-risk patients in the Philippine hospital. In future studies, several critical features in medical records like health history, lifestyle, and social factors need to be collected. Explore other statistical methods or machine learning algorithms that could provide better results.

References

- [1] Tabak Y P, Sun X, Nunez C M, Gupta V and Johannes R S 2017 *Med. Care.* **55** pp 267-275
- [2] Centers for Medicare & Medicaid Services. Outcome Measures. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/OutcomeMeasures.html> Accessed March 6, 2018
- [3] Kristensen S, Bechb M and Quentin W 2015 *Health Policy* **119** pp 264-273
- [4] Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M and Kripalani S 2011 *JAMA* **306** 1688–98
- [5] Zhou H, Della P R, Roberts P, Goh L and Dhaliwal S 2016 *BMJ Open* **6** pp 1-25
- [6] Nasir K, Lin Z, Bueno H, Normand S, Drye E, Keenan P and Krumholz H 2010 *Med Care* **48** 477–481
- [7] Amoah D and Mwanri L 2016 *Aus J Pub Health Epidemiology* **3** p 1049
- [8] Schjodt I, Larsen P, Johnsen S P, Stromberg A and Logstrup B 2017 *JBIC Database of Systematic Rev Implementation Rep* **15** pp 1561-74
- [9] Alyahya M, Hijazi H, Alshraideh H, Alsharman M, Abdi R and Harvey H 2016 *Int J Integr Care* **16** p 12
- [10] Duggal R, Shukla S, Chandra S, Shukla B and Khatri S 2016 *Int J Diabetes Dev Ctries* **14** pp 519-528
- [11] Lui H and Cosea M 2017 *Granul Comput* **2** pp 357 – 386
- [12] Garcia-Arce A, Rico F and Zayas-Castro J 2017 *Journal of Healthcare Quality* pp 1-10