Predicting length of stay in hospitalized patients using SSL algorithms

Ioannis E. Livieris
Department of Computer Engineering & Informatics
Technological Educational Institute of Western Greece
Antirrio, Greece GR 263-34
livieris@teiwest.gr

Ioannis F. Dimopoulos
Department of Business Administration (LAIQDA Lab)
Technological Educational Institute of Peloponnese
Kalamata, Greece GR 241-00.
ydimop@teikal.gr

Theodore Kotsilieris
Department of Business Administration (LAIQDA Lab)
Technological Educational Institute of Peloponnese
Kalamata, Greece GR 241-00.
t.kotsilieris@teipel.gr

Panagiotis Pintelas
Department of Mathematics
University of Patras
Patras, Greece GR 265-00
ppintelas@gmail.com

ABSTRACT

Length of stay in hospitalized patients is acknowledged as a critical factor for healthcare policy planning that consequently affects the available human, technical and financial resources as well as facilities occupation. Over recent years, data mining and machine learning led to the development of several efficient and accurate models for predicting of how long a patient will stay in the hospital and support healthcare policy planning. As an alternative to traditional classification methods, semi-supervised learning algorithms have become a hot topic of significant research which exhibit remarkable performance over labeled data but lack the ability to be applied on large amounts of unlabeled data. In this work, we evaluate the performance of semi-supervised methods in predicting the length of stay of hospitalized patients. Our reported experimental results illustrate that a good predictive accuracy can be achieved using few labeled data in comparison to well known supervised learning algorithms.

CCS CONCEPTS

• **Theory of computation** → **Machine learning theory**; **Semi-supervised learning**; • **Applied computing** → *Health care information systems*;

KEYWORDS

Length of stay, data mining, semi-supervised learning, self-training, co-training, tri-training, classification.

ACM Reference format:
Ioannis E. Livieris, Ioannis F. Dimopoulos, Theodore Kotsilieris, and Panagiotis Pintelas. 2018. Predicting length of stay in hospitalized patients using SSL algorithms. In *Proceedings of 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, Thessaloniki, Greece, June 20–22, 2018 (DSAI 2018),* 7 pages.
DOI: 10.1145/3218585.3218588

978-1-4503-6467-6/18/06…$15.00
DOI: 10.1145/3218585.3218588

1 INTRODUCTION

Nowadays, the healthcare system integrates many service units which consist of a number of interacting departments and healthcare units such as outpatient department, emergency department, operating theatre, intensive care unit and inpatient wards. The main objective of hospital managers is the establishment of an appropriate healthcare planning and organization by allocating facility, equipment and manpower resources necessary for hospital operation according to the patients' needs while minimizing the cost of healthcare. Therefore, several techniques have been developed for scheduling elective admissions, predicting bed needs and measuring bed utilization. The major component in these techniques is the accurate prediction of how long a patient will stay in the hospital and the understanding of the factors that influence its stay.

Length of Stay (LoS) is usually defined as the duration of a patient hospitalization and it is calculated as the difference between the timestamp of a patient discharge and the timestamp of its admission. It is generally acknowledged as an indicative marker of inpatient hospitalization costs and resource utilization. Since hospitals have severely limited beds to hold inpatients and as most of them are facing substantial financial pressure, it is extremely important to find ways to reduce healthcare costs [15]. Due to the growing number of hospitalized patients, predicting the average LoS has become increasingly important for both resource planning and effective admission scheduling. Clinicians generally assume that LoS of individual patients is unpredictable and the accuracy of the prediction is heavily depended on experience. It is noted that many hospitals have no ability to predict and measure future admission requests [11]. Hence, the accurate prediction of LoS has become increasingly important not only for the healthcare systems but also for the patients. More comprehensively, this "knowledge discovery" can assist hospital managers for rehabilitation planning, resource allocation and healthcare units administration (e.g. patient admission/treatment/discharge, bed management, staff scheduling). Furthermore, with the accurate estimation of LoS, patients will be provided with better medical services in a hospital. Nevertheless, the development of an accurate model for the prediction of LoS is a very attractive and challenging task.

During the last decades, hospitals have managed to accumulate a large volume of data which enable researchers to measure and compare clinical performance and utilize these results to support or critique policy decisions. Machine learning techniques offer a first step and a helping hand in extracting useful information from these data and gaining insights into the prediction of LoS and on the major factors and elements which affect the duration of a patient hospitalization. To this end, several studies have attempted to predict LoS with heterogeneous methods [2, 29, 30, 34] and extract the factors affecting LoS among various types of patients [3, 14, 16]. The majority proportion of these studies examines the efficiency of supervised methods utilizing only labelled data to determine an accurate prediction model. Nevertheless, in many real-world classification problems, labelled instances are often difficult, expensive, or time consuming to obtain, since they require the efforts of empirical research and in contrast unlabeled data are easier to obtain, require less effort of experienced human annotators. As an alternative to traditional classification methods, semi-supervised learning algorithms constitute the appropriate technique to exploit medical data, since there is often a lack of labeled data, while unlabeled is vast.

In this work, we examine the effectiveness of semi-supervised methods for the prediction of LoS in hospitalized patients. To the best of our knowledge, no study exists dealing with the implementation of semi-supervised methods for predicting the expected LoS. Our objective is to provide an accurate prediction model considering demographic, clinical, and geographical factors which can be assessed at the time of admission for predicting LoS of acute care hospitalization for patients. Our preliminary numerical experiments illustrate that the classification accuracy can be significantly improved, utilizing a few labeled and many unlabeled data for developing reliable prediction models.

The remainder of this paper is organized as follows: Section 2 presents a survey of recent studies concerning the application of data mining in the prediction of LoS. Section 3 presents a brief discussion of the semi-supervised learning algorithms and Section 4 presents a detailed description of the data collection and data preparation used in our study. Finally, Section 5 presents experimental results, while Section 6 sketches concluding remarks and future work directions.

2 RELATED WORK

Hospitals are daily faced with a significant uncertainty that is the LoS of a hospitalized patient. As future admission requests appears to be a more complicated problem within an effective and long-term healthcare system planning, accurate prediction of in-hospital stay duration would allow, in the short-term, for efficient human resources allocation and facilities occupancy. LoS prediction is a substantial problem that attracted research community's attention since the '60s [10, 19] by employing statistical methods.

Since then, several scientific fields have risen, providing mathematical and computing classification and prediction techniques. Following the evolution of machine learning and data mining, research efforts focused on employing relevant algorithms in the field of LoS prediction. In [31], Walczak et al. studied the applicability of neural networks for LoS and injury severity for pediatric trauma and pancreatitis patients. Initially, they used pediatric trauma cases and as far as LoS is concerned, the neural networks with a single hidden layer of 25 nodes achieved the best accuracy. The same research methodology was applied for patients with acute pancreatitis. Three LoS specificity scores were adopted, namely: short (less or equal to one week), medium (one to two weeks) and long (more than two weeks). Under the assumption that one-day overlap among the categories is allowed, the accuracy obtained was roughly 52%.

Hachesu [11] applied data mining techniques to extract useful knowledge and draw an accurate model to predict the LoS of heart patients. The data in their study consist of patients who had suffered coronary artery disease admitted to a cardiovascular center. Based on their experimental results, the authors stated that a LoS greater than 10 days was associated with comorbidity and diastolic blood pressure features and there was a significant tendency for LoS to be longer in patients with lung or respiratory disorders and high blood pressure. Moreover, their proposed ensemble algorithm exhibited the best performance than any individual algorithm, presenting 98.2% of successful classification.

Panchami and Radhika [22] proposed a novel approach for predicting whether the LoS of a hospitalized patient is greater than one week. Their approach identified groups of similar hospital claims from the dataset utilizing a density based clustering approach called DBSCAN which uses these groups as the training set to classify the LoS of patients with high accuracy. An advantage of the DBSCAN clustering is that it eliminates the noise points from the dataset. Their experiments presented that prediction models based on support vector machine gives the best performance and the training set created by DBSCAN approach provides the best performance.

Luigi et al. [18] proposed the Growing Neural Gas model to predict the LoS of hospitalized patients which is based on an ensemble algorithm. The dataset used in their study consisted of 274962 instances of hospital admissions and the accuracy of their ensemble algorithm obtained reached 96.36% which significantly outperformed the classical classification algorithms.

Nouaouri et al. [21] proposed an inpatient LoS prediction approach based on evidential data mining. Their methodology handles the uncertainty, imprecision and missing data within a database. Moreover, they introduce the Evidential Length Of Stay prediction algorithm which allows the accurate prediction of a new patient the length of stay. Their experimental results confirmed the suitability of their approach by testing and experimenting the proposed algorithm on a healthcare datasets containing 270 patients.

Tsai et al. [28] performed a two stages LoS prediction, the predischarge and the preadmission ones. The predischarge stage uses all the available data of in-hospital patients, while the preadmission one uses only the data available before a patient's admission. The prediction results of predischarge patients were utilized to evaluate the LoS prediction performance at the preadmission stage. The sample consisted of 2377 of cardiovascular disease patients with one of the three primary diagnoses: Coronary Atherosclerosis

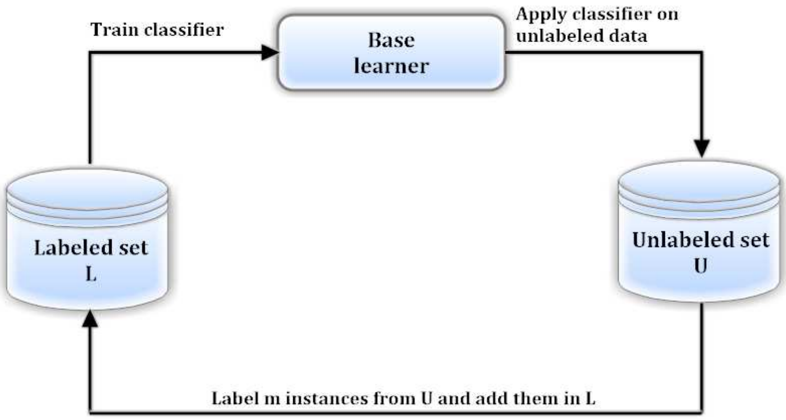

Figure 1: Semi-supervised learning framework

(CAS), Heart Failure (HF) and Acute Myocardial Infarction (AMI). Their proposed classification model was able to predict correctly for 88.07% to 89.95% CAS patients at the predischarge stage, respectively and for 88.31% to 91.53% at the preadmission stage. For HF/AMI patients, the accuracy ranged from 64.12% to 66.78% at the predischarge stage and 63.69% to 67.47% at the preadmission stage when a tolerance of two days was allowed.

3 A REVIEW ON SEMI-SUPERVISED LEARNING

Semi-supervised learning (SSL) constitutes of an extension of supervised and unsupervised learning, aiming to exploit the explicit classification information of labeled data with the information hidden in the unlabeled data for improving the classification performance. In contrast to traditional classification approaches, SSL utilizes large amount of unlabeled samples together with labeled samples to build an efficient and accurate classifier. More specifically, SSL methods utilize only a small proportion of the whole amount of data to be labeled for accomplishing their task, known as labeled ratio R which is defined by

$$R = \frac{\text{Number of labeled dinstances}}{\text{Number of all instances}}$$

and it is usually provided in percentage values (%). Subsequently, after the labeled ratio is defined, all the available data is split into two distinct subsets: the labeled (L) set and the unlabeled (U) set. Clearly, the set $L \cup U$ forms the training set. The generic representation of the examples included in each of these subsets is respectively defined as follows:

$$\begin{cases} x_L = \{\text{Feature set} \mid \text{Class}\} \\ x_U = \{\text{Feature set} \mid \text{Not known class}\} \end{cases}$$

In the literature, many SSL algorithms have been proposed with different philosophy on labeling of examples in U and their incorporation in L (Figure 1). Self-training, Co-training and Tri-training constitute the most representative and commonly utilized.

Self-training is generally considered as the simplest and one of the most efficient SSL algorithms. This algorithm is a wrapper based SSL approach which constitutes an iterative procedure of self-labeling unlabeled data. According to Ng and Cardie [20] "*self-training is a single-view weakly supervised algorithm*" which is based on its own predictions on unlabeled data to teach itself. Firstly, an arbitrary classifier is initially trained with a small amount of labeled data, constituting its training set which is iteratively augmented using its own most confident predictions of the unlabeled data. More analytically, each unlabeled instance which has achieved a probability over a specific threshold c is considered sufficiently reliable to be added to the labeled training set and subsequently the classifier is retrained.

It is worth noticing that the success of Self-training is heavily depended on the newly-labeled data based on its own predictions, hence its weakness is that erroneous initial predictions will probably lead the classifier to generate incorrectly labeled data [36].

Co-training [4] is a semi-supervised algorithm which is based on the strong assumption that feature space can be split into two different conditionally independent views and that each view is able to predict the classes perfectly [6, 27]. Under these assumptions, this algorithm assumes that it is more effective to predict the unlabeled instances by dividing the features of data into two separable categories. In this framework, two learning algorithms are trained separately for each view utilizing the initial labeled dataset and the most confident predictions of each algorithm on unlabeled data are used to augment the training set of the other algorithm through an iterative learning process. Essentially, co-training is a "*two-view weakly supervised algorithm*" since it uses the self-training approach on each view [20].

The classification efficacy and effectiveness of co-training is closely related with the appropriate selection of the two learning algorithms, as well as with the existence of two conditionally independent views. Nevertheless, the requirement of two sufficient and redundant views is a luxury hardly met in most scenarios and real-world tasks, therefore several extension of this algorithm have been developed, such as tri-training.

Tri-Training [35] consists of an improved version of Co-Training which overcomes the requirements for multiple sufficient an redundant feature sets. This algorithm is a bagging ensemble of three

Attribute	Values
Gender	male, female
Age	65 − 74, 75 − 84, > 85
Insurance type	IKA, OGA, TAYT, NAT, private, uninsuranced, indigent, other,
Residence altitude	0 − 100, 100 − 300, > 300 (m).
Residence urbanity	urban, semi-urban, rural.
Residence distance from hospital	0 − 15, 15 − 30, 30 − 45, > 45 (km)
Residence medical cover type	hospital, regional clinic, rural clinic.
Patient's day of admission	Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday.
Patient's month of admission	January, February, March, April, May, June, July, August, September, October, November, December.
ICD-10 diagnosis code	A00 − B99, C00 − D48, D50 − D89, E00 − E90, F00 − F99, G00 − G99, H00 − H59, H60 − H95, I00 − I99, J00 − J99, K00 − K93, L00 − L99, M00 − M99, N00 − N99, Q00 − Q99, R00 − R99, S00 − T98, V01 − Y98, Z00 − Z99, other.
Ward of nursing	cardiology, general surgery, orthopaedics, internal medicine.
Number of admissions in a ward	1, 2, . . . , 100.
Class	"1-2", "3-6", "over 6"

Table 1: Attributes description

classifiers, trained on the data subsets generated through bootstrap sampling from the original labeled training set. In case two of the classifiers agree on a prediction, then they label the unlabeled example with their prediction and augment the third classifier with the newly labeled example. The efficiency of the training process is based on the "*majority teach minority strategy*" which serves as an implicit confidence measurement, avoiding thereby the use of a complicated time consuming approach to explicit measure the predictive confidence.

In contrast to several SSL algorithms, Tri-training does not require different supervised algorithms as base learners which leads to greater applicability in many real world classification problems [9, 17, 36]. However, sometimes the performance of tri-training degradates thus three other issues must be taken into account [9]:

(1) Excessively restrictions, introduce further classification noise.
(2) Estimation of the classification error is unsuitable.
(3) Differentiation between initial labeled example and labeled of previously unlabeled example is deficient.

4 DATASET

The data set consisted of patients hospitalized in the General Hospital of Kalamata, Greece during the period between 2008 and 2012. We identified 4403 patients, aged over 65 years of both genders and diagnoses. Data cleansing and preprocessing operations involved the deletion of repeated records, irregularities, and irrelevancies and manipulation of records with missing and outlier data. Furthermore, records with the same admission and discharge date (i.e. resulting to 0 LoS) were excluded.

Table 1 presents the set of the thirteen (13) attributes utilized in our study concerning demographic, clinical, geographical and administrative factors. The first three (3) attributes are related with patient's personal information such as gender, age, and insurance type. Notice that each patient in Greece belongs to a specific insurance fund based on his occupation such as IKA, OGA, NAT, TAYT, or he/she has a private hearth insurance. The following four (4) geographical and demographic attributes concern the patient's residence altitude, urbanity and distance from the hospital as well as the medical cover of the residence. The last five (5) attributes are related with patients' pathological and clinical characteristics. These attributes concern the day and month of the patient's admission in the hospital and the number of patients which have been admitted that day. Additionally, the ward hospital in which the patient was admitted and the ICD-10 diagnosis code according to the World Health Organization [32] are usually the main reasons of patient's LoS. Finally, the patients were classified according to the number of days in the hospital utilizing a three-level classification scheme: "1-2" days, "3-6" days, "over 6" days.

5 EXPERIMENTAL RESULTS

In this section, we conduct a series of tests in order to evaluate the performance of the most popular and commonly used SSL algorithms: Self-training, Co-training and Tri-training. Each SSL algorithm was evaluated deploying as base learners: Naive Bayes (NB) [5], Multilayer Perceptron (MLP) [25], Sequential Minimum Optimization (SMO) [23], 3NN algorithm [1], C4.5 decision tree algorithm [24] and PART [8] as a typical rule-learning technique. These algorithms constitute the most effective and the most popular machine learning algorithms for classification problems [33]. Furthermore, the implementation code was written in JAVA, using WEKA Machine Learning Toolkit [12]. The configuration parameters for all SSL algorithms used in our experiments are presented in Table 2. Regarding the base learners, the default parameter settings included in the WEKA software were utilized in order to minimize the effect of any expert bias by not attempting to tune any of the algorithms to the specific datasets.

The classification accuracy of all learning algorithms was evaluated utilizing the standard procedure called stratified 10-fold cross-validation i.e. the data was separated into folds so that each fold had the same distribution of grades as the entire data set. Accuracy constitute as one of the most frequently and commonly utilized measures for assessing the overall effectiveness of a classification algorithm [26] is defined as the percentage of correctly classified instances. In order to study the influence of the amount of labeled data, four different ratios of the training data were used: 10%, 20%, 30% and 40%.

Algorithm	Parameters
Self-training	MaxIter = 40. c = 95%.
Co-training	MaxIter = 40. Initial unlabeled pool = 75.
Tri-training	No parameters specified.

Table 2: Parameter specification for all the SSL methods employed in our experiments

Tables 3 present the classification performance of each SSL algorithm utilizing 10%, 20%, 30% and 40%, respectively as labeled data ratio and the best accuracy for each base learner is highlighted in bold style. Additionally, a more representative visualization of the classification performance of the compared classifiers is presented in Figure 2.

SSL algorithm	Ratio			
	10%	20%	30%	40%
Self-training (NB)	63.09%	62.96%	**63.57%**	62.55%
Co-training (NB)	63.03%	63.09%	62.71%	63.03%
Tri-training (NB)	**63.18%**	**63.32%**	63.12%	**63.23%**
Self-training (MLP)	63.21%	62.12%	62.82%	62.93%
Co-training (MLP)	**63.93%**	64.45%	**64.00%**	**64.48%**
Tri-training (MLP)	63.82%	**64.46%**	62.64%	63.50%
Self-training (SMO)	63.77%	62.93%	62.71%	61.59%
Co-training (SMO)	**64.07%**	63.07%	62.44%	62.86%
Tri-training (SMO)	63.82%	**64.27%**	**63.93%**	**64.23%**
Self-training (3NN)	**62.84%**	62.53%	61.93%	62.48%
Co-training (3NN)	62.07%	60.84%	61.77%	62.41%
Tri-training (3NN)	62.77%	**62.82%**	**62.71%**	**62.55%**
Self-training (C4.5)	64.39%	**64.98%**	**64.83%**	**65.30%**
Co-training (C4.5)	63.66%	62.18%	62.82%	62.62%
Tri-training (C4.5)	**64.59%**	64.21%	64.18%	64.86%
Self-training (PART)	62.34%	62.46%	**62.75%**	62.68%
Co-training (PART)	**62.82%**	62.37%	62.00%	62.68%
Tri-training (PART)	62.55%	**62.82%**	62.57%	**63.34%**

Table 3: Classification accuracy of all SSL algorithms

The number of wins of each one of the tested methods according to the ratio of labeled data in the training set is presented in Table 4, while the best scores are highlighted in bold. It should me mentioned that draw cases between algorithms have not been encountered. The aggregated results illustrate that Tri-training is by far the most effective method since it exhibits the most wins with a labeled ratio of 20% and 40%. Moreover, Co-training and Self-training report the most wins with a labeled ratio 10% and 30%, respectively followed by Tri-training.

SSL algorithm	Ratio			
	10%	20%	30%	40%
Self-training	1	1	**3**	1
Co-training	**3**	0	1	1
Tri-training	2	**5**	2	**4**

Table 4: Total wins for each SSL algorithm

In machine learning, the statistical comparison of multiple algorithms over multiple data sets is fundamental and it is usually carried out by means of a statistical test [17]. Therefore, in order to evaluate the rejection of the hypothesis that all the algorithms perform equally well for a given level and highlight the existence of significant differences between our proposed algorithm and the classical SSL algorithms, we utilized the non-parametric Friedman Aligned Ranking (FAR) [13] test. Notice that, since the test is non-parametric, it does not require commensurability of the measures across different data sets, it does not assume normality of the sample means and it is robust to outliers. Moreover, the Finner post hoc test [7] with a significance level $\alpha = 0.05$ was applied a post hoc procedure to detect the specific differences among the algorithms.

Tables 5, 6, 7 and 8 present the information of the statistical analysis performed by nonparametric multiple comparison procedures over 10%, 20%, 30% and 40% of labeled data, respectively. The best(lowest) ranking obtained in each FAR test determines the control algorithm for the post hoc test. Clearly, Tri-training reports the best performance due to better probability-based ranking and higher classification accuracy.

Algorithm	Friedman	Finner post-hoc test	
	Ranking	p-value	Null Hypothesis
Tri-training	7.6667	-	-
Co-training	10.1667	0.417304	accepted
Self-training	10.6667	0.025321	rejected

Table 5: FAR test and Finner post hoc test (labeled ratio 10%)

Algorithm	Friedman	Finner post-hoc test	
	Ranking	p-value	Null Hypothesis
Tri-training	5.6667	-	-
Self-training	10.3333	0.130009	accepted
Co-training	12.500	0.026621	rejected

Table 6: FAR test and Finner post hoc test (labeled ratio 20%)

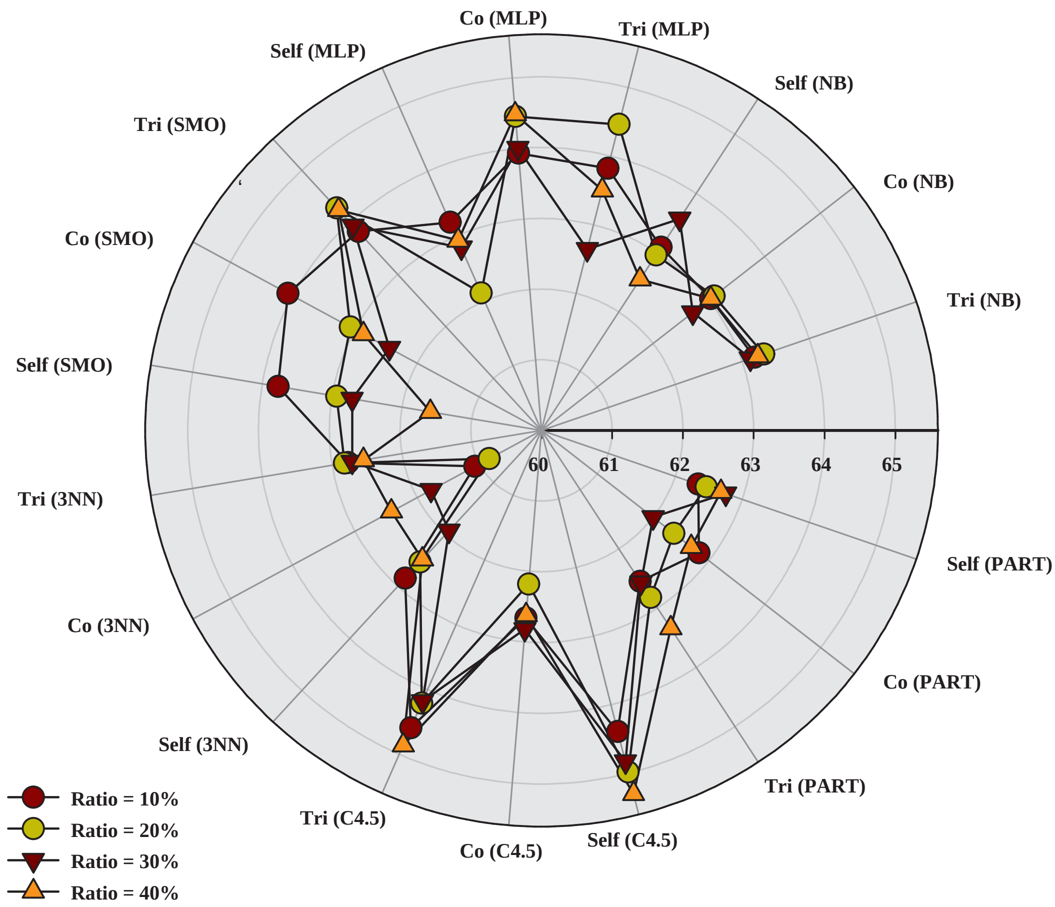

Figure 2: Labeled ratio comparison

Algorithm	Friedman Ranking	Finner post-hoc test	
		p-value	Null Hypothesis
Tri-training	6.0000	-	-
Self-training	9.3333	0.130009	accepted
Co-training	13.1667	0.026621	rejected

Table 7: FAR test and Finner post hoc test (labeled ratio 30%)

Algorithm	Friedman Ranking	Finner post-hoc test	
		p-value	Null Hypothesis
Tri-training	6.0000	-	-
Co-training	10.4167	0.151870	accepted
Self-training	12.0833	0.048417	rejected

Table 8: FAR test and Finner post hoc test (labeled ratio 40%)

Finally, in order to illustrate the classification performance of the SSL algorithms we evaluate their best reported performance for each base learner with the corresponding supervised algorithms trained with 100% of the training set. The results presented in Table 9 illustrate that SSL algorithms are comparatively better than the respective supervised algorithms, relative to all base learners.

	Supervised	Self-training	Co-training	Tri-training
NB	62.54%	63.57%	63.09%	63.32%
MLP	61.78%	63.21%	64.45%	64.46%
SMO	62.57%	63.77%	64.07%	64.27%
3NN	61.66%	62.84%	61.77%	62.82%
C4.5	64.77%	64.98%	62.82%	64.59%
PART	63.43%	62.75%	62.82%	62.82%

Table 9: Classification accuracy comparison

6 CONCLUSIONS

In this work, we evaluated the classification performance of semi-supervised algorithms for predicting the LoS in hospitalized patients. Our experimental results illustrated that semi-supervised algorithms can improve the classification accuracy utilizing a few labeled and many unlabeled data for developing reliable prediction models. Furthermore, it is worth mentioning that the patients' attributes utilized in our work do not constitute a conclusive list. An extension can introduce new attributes and other criteria which were not in the current database, but are collectable by medical staff and may potentially influence the performance and the quality of the prediction.

Since the experimental results are quite encouraging, a next step could be the development of a decision-support tool concerning the prediction of the expected LoS in order to assist and support healthcare policy planning. Moreover, another direction for a future research would be to enlarge our database with data from more hospitals and more years and apply machine learning methods to predict LoS and extract the factors affecting it among various types of patients.

REFERENCES

[1] D. Aha. 1997. *Lazy Learning*. Dordrecht: Kluwer Academic Publishers.

[2] J. Bai, A. Fügener, J. Schoenfelder, and J.O. Brunner. 2018. Operations research in intensive care unit management: a literature review. *Health care management science* 21, 1 (2018), 1–24.

[3] A. Belderrar and A. Hazzab. 2017. Hierarchical Genetic Algorithm and Fuzzy Radial Basis Function Networks for Factors Influencing Hospital Length of Stay Outliers. *Healthcare Informatics Research* 23, 3 (2017), 226–232.

[4] A. Blum and T. Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *11th annual conference on Computational learning theory*. ACM, 92–100.

[5] P. Domingos and M. Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29 (1997), 103–130.

[6] J. Du, C.X. Ling, and Z.H. Zhou. 2011. When does co-training work in real data? *IEEE Transactions on Knowledge and Data Engineering* 23, 5 (2011), 788–799.

[7] H. Finner. 1993. On a monotonicity problem in step-down multiple test procedures. *J. Amer. Statist. Assoc.* 88, 423 (1993), 920–923.

[8] E. Frank and I.H. Witten. 1998. Generating Accurate Rule Sets Without Global Optimization. In *15th International Conference on Machine Learning*. 144–151.

[9] T. Guo and G. Li. 2012. Improved tri-training with unlabeled data. *Software Engineering and Knowledge Engineering: Theory and Practice* (2012), 139–147.

[10] D.H. Gustafson. 1968. Length of Stay: Prediction and Explanation. *Health Services Research* 3, 1 (1968), 12–34.

[11] P.R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi. 2013. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Healthcare informatics research* 19, 2 (2013), 121–129.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. 2009. The WEKA data mining software: An update. *SIGKDD Explorations Newsletters* 11 (2009), 10–18. Issue 1.

[13] J.L. Hodges and E.L. Lehmann. 1962. Rank methods for combination of independent experiments in analysis of variance. *The Annals of Mathematical Statistics* 33, 2 (1962), 482–497.

[14] D.J. Huang, L.Z. Xie, and Y. Qiu. 2016. Analysis of factors affecting the length of hospital stay for patients with diabetes. *Experimental and Clinical Endocrinology & Diabetes* 124, 1 (2016), 5–10.

[15] X. Jiang, X. Qu, and L.B. Davis. 2010. Using Data Mining to Analyze Patient Discharge Data for an Urban Hospital.. In *Proceedings of International Conference on Data Mining*. 139–144.

[16] N. Khajehali and S. Alizadeh. 2017. Extract critical factors affecting the length of hospital stay of pneumonia patient by data mining (case study: an Iranian hospital). *Artificial Intelligence in Medicine* 83 (2017), 2–13.

[17] I.E. Livieris, K. Drakopoulou, V. Tampakas, T. Mikropoulos, and P. Pintelas. 2018. Predicting secondary school students' performance utilizing a semi-supervised learning approach. *Journal of Educational Computing Research* (2018).

[18] L. Luigi, A. di Giorgio, and A.F. Dragoni. 2015. Length of Stay Prediction and Analysis through a Growing Neural Gas Model. In *Proceedings of the 4th International Workshop on Artificial Intelligence and Assistive Medicine*. 11–21.

[19] A. Morton, E. Marzban, G. Giannoulis, A. Patel, R. Aparasu, and I.A. Kakadiaris. 2014. A Comparison of Supervised Machine Learning Techniques for Predicting Short-Term In-Hospital Length of Stay among Diabetic Patients. In *13th International Conference on Machine Learning and Applications*. 428–431.

[20] V. Ng and C. Cardie. 2003. Weakly supervised natural language learning without redundant views. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 94–101.

[21] I. Nouaouri, A. Samet, and H. Allaoui. 2015. Evidential data mining for length of stay (LOS) prediction problem. In *IEEE International Conference on Automation Science and Engineering 2015*. IEEE, 1415–1420.

[22] V.U. Panchami and N. Radhika. 2014. A novel approach for predicting the length of hospital stay with DBSCAN and supervised classification algorithms. In *5th International Conference in Applications of Digital Information and Web Technologies*. 207–212.

[23] J. Platt. 1999. Using sparseness and analytic QP to speed training of support vector machines. In *Advances in neural information processing systems*, M.S. Kearns, S.A. Solla, and D.A. Cohn (Eds.). MA: MIT Press, 557–563.

[24] J.R. Quinlan. 1993. *C4.5: Programs for machine learning*. Morgan Kaufmann, San Francisco.

[25] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. 1986. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. Rumelhart and J. McClelland (Eds.). Cambridge, Massachusetts, 318–362.

[26] M. Sokolova, N. Japkowicz, and S. Szpakowicz. 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *Australian conference on artificial intelligence*, Vol. 4304. 1015–1021.

[27] S. Sun and F. Jin. 2011. Robust co-training. *International Journal of Pattern Recognition and Artificial Intelligence* 25, 07 (2011), 1113–1126.

[28] P.F. Tsai, P.C. Chen, Y.Y. Chen, H.Y. Song, H.M. Lin, F.M. Lin, and Q.P. Huang. 2016. Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. *Journal of Healthcare Engineering* (2016), 1–11.

[29] I.W.M. Verburg, A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek, and N.F. de Keizer. 2017. Which models can I use to predict adult ICU length of stay? A systematic review. *Critical care medicine* 45, 2 (2017), 222–231.

[30] I.W.M. Verburg, N.F. de Keizer, E. de Jonge, and N. Peek. 2014. Comparison of Regression Methods for Modeling Intensive Care Length of Stay. *PLoS One* 9, 10 (2014).

[31] S. Walczak, R.J. Scorpio, and W.E. Pofahl. 1998. Predicting Hospital Length of Stay with Neural Networks. In *Proceedings of the 11th International FLAIRS Conference*. 333–337.

[32] World Health Organization. 2004. *International statistical classification of diseases and related health problems*. Vol. 1. World Health Organization.

[33] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A.F.M. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand, and D. Steinberg. 2008. Top 10 algorithms in data mining. *Knowledge and information systems* 14, 1 (2008), 1–37.

[34] C.S. Yang, C.P. Wei, C.C. Yuan, and J.Y. Schoung. 2010. Predicting the length of hospital stay of burn patients: Comparisons of prediction accuracy among different clinical stages. *Decision Support Systems* 50, 1 (2010), 325–335.

[35] Z.H. Zhou and M. Li. 2005. Tri-training: Exploiting unlabeled data using three classifiers. *IEEE Transactions on knowledge and Data Engineering* 17, 11 (2005), 1529–1541.

[36] X. Zhu and A.B. Goldberg. 2009. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning* 3, 1 (2009), 1–130.