Improvement of Adequate Use of Warfarin for the Elderly Using Decision Tree-based Approaches*

K. E. Liu1; C.-L. Lo2; Y.-H. Hu2

1 Department of Economics, National Chung Cheng University, Chia-Yi, Taiwan;
2 Department of Information Management and Graduate Institute of Healthcare Information Management, National Chung Cheng University, Chia-Yi, Taiwan.

1. Introduction

Warfarin is one of the most commonly used anticoagulants in the prevention of diseases like thromboembolism [1, 2]. Among the cardiovascular prescription drugs used in the U.S., warfarin is ranked the fourth most frequently-used anticoagulant [3]. Oral absorption makes warfarin one of the most effective anticoagulants [4]; however, warfarin is also at the top of the list as having adverse drug events [5], mainly because of its narrow therapeutic range and high drug-to-drug interactions (DDIs) with nearly 250 medicines [6].

Hazardous risks from taking warfarin are more serious to the elderly for two reasons. First, elderly patients are more likely to take multiple-drugs to control chronic diseases; warfarin is recognized as a medicine with high DDIs. Second, due to physiological degradation, the physical condition of the elderly is usually worse than the young; dysfunction in organs like the liver and kidneys may cause difficulty in prescribing accurate warfarin dosage. It is worth noting that gerontology and geriatrics have become hot topics due to a striking change in the world’s older population: the aging population has increased dramatically reaching 15% worldwide and is predicted to surpass 20% in 2020 and climb to 31.86% in 2050; this rise in the aging population is worsening in the developed countries [7]. These factors make it increasingly important to improve the management of the clinical use of warfarin for the elderly.

Current studies have found that ethnicity is one of the most profound factors affecting warfarin dosage prediction. As summarized in Table 1, for example, the warfarin dosing algorithm proposed by Schelleman et al. [8] outperforms the ones proposed by other studies when Caucasians were investigated; however, their results from African Americans reveal that predicting performance using all proposed algorithms is quite similar. The International Warfarin Pharmacogenetics Consortium [9] collected over four thousand clinical cases from nine countries and the experimental results had almost identical conclusions, i.e., warfarin dose requirements vary across ethnic groups.

The majority of previous studies also considered both clinical and pharmacogenetic factors in the prediction of warfarin doses or treatment appropriateness. Two

Keywords
Warfarin, anticoagulant, decision support techniques, decision trees, health services for the elderly

Summary
Objectives: Due to the narrow therapeutic range and high drug-to-drug interactions (DDIs), improving the adequate use of warfarin for the elderly is crucial in clinical practice. This study examines whether the effectiveness of using warfarin among elderly patients can be improved when machine learning techniques and data from the laboratory information system are incorporated.

Methods: Having employed 288 validated clinical cases in the DDI group and 89 cases in the non-DDI group, we evaluate the prediction performance of seven classification techniques, with and without an Adaptive Boosting (AdaBoost) algorithm. Measures including accuracy, sensitivity, specificity and area under the curve are used to evaluate model performance.

Results: Decision tree-based classifiers outperform other investigated classifiers in all evaluation measures. The classifiers supplemented with AdaBoost can generally improve the performance. In addition, weight, congestive heart failure, and gender are among the top three critical variables affecting prediction accuracy for the non-DDI group, while age, ALT, and warfarin doses are the most influential factors for the DDI group.

Conclusion: Medical decision support systems incorporating decision tree-based approaches improve predicting performance and thus may serve as a supplementary tool in clinical practice. Information from laboratory tests and inpatients’ history should not be ignored because related variables are shown to be decisive in our prediction models, especially when the DDIs exist.
polymorphisms, CYP2C9 and VKORD1, have been validated by the U.S. Food and Drug Administration (FDA) as genetic factors affecting warfarin doses [23]. Other related researches [8, 9, 12, 13, 15–17, 20, 21, 22, 24] also show that approximately 50 to 60 percent of the variation can be explained by the above two genes. Although genetic testing may help to realize the relationships between the human gene and warfarin dosing risk, it requires a relatively longer time and a higher cost to obtain the report in clinical practice [25]. Recently, only a few studies [17, 26] have revealed that factors elicited from laboratory data can help determine warfarin usage. Because the collection of laboratory data is relatively simple and inexpensive, utilizing laboratory variables to develop prediction models is a possible future direction in clinical practice.

Most importantly, the populations in previous studies cover a wide range of ages and only a few studies [27–30] emphasized the design of warfarin prediction models designated for elderly inpatients. Since poor metabolism and other medical complications are commonly seen in the elderly, development of an accurate warfarin prediction model especially for the elderly is imperative. On the other hand, most of the related studies [8, 11–18, 20, 22] on the prediction of adequate warfarin doses utilized statistical modeling such as linear regression (LR) or logistic regression (LGR); we intend to introduce machine learning-based techniques, including C4.5 [35], k-nearest neighbors (kNN), classification and regression tree (CART), random forest (RF), multi-layer perceptron neural network with back-propagation (MLP), and support vector machine (SVM), to enhance the decision-making process in clinical practice. Therefore, the purpose of this paper is to conduct a comparative study to show the superiority of incorporating machine learning-based techniques in predicting the adequacy of warfarin usage for elderly inpatients. Separating the elderly inpatient data into two groups, with DDIs and without DDIs, our experimental results reveal that prediction accuracy can be significantly improved when an effective classification model is employed, showing a promising tool in computer-assisted warfarin management for clinical practice.

2. Materials and Methods

2.1 Data

In this study, we collected complete records of inpatients 65 years old or older who received warfarin therapy in a medical center in southern Taiwan from January 2005 to December 2009. A washout period of three months is considered to eliminate the influence of warfarin treatment from the previous period, resulting in an exclusion of inpatients’ records if they had any warfarin treatment before April 1, 2005. Each clinical record contains demographics, such as gender, age, and weight. In addition, the adequate dosage of warfarin for inpatients varies if they have symptoms such as congestive heart failure (CHF) and thyrotoxicosis [4]. Both of these conditions can be identified by the inpatient’s historical diagnosis codes in physician orders. As to the inpatients’ liver and kidneys functions, metabolism and other medical complications are commonly seen in the elderly, development of an accurate warfarin prediction model especially for the elderly is imperative. On the other hand, most of the related studies [8, 11–18, 20, 22] on the prediction of adequate warfarin doses utilized statistical modeling such as linear regression (LR) or logistic regression (LGR); we intend to introduce machine learning-based techniques, including C4.5 [35], k-nearest neighbors (kNN), classification and regression tree (CART), random forest (RF), multi-layer perceptron neural network with back-propagation (MLP), and support vector machine (SVM), to enhance the decision-making process in clinical practice. Therefore, the purpose of this paper is to conduct a comparative study to show the superiority of incorporating machine learning-based techniques in predicting the adequacy of warfarin usage for elderly inpatients. Separating the elderly inpatient data into two groups, with DDIs and without DDIs, our experimental results reveal that prediction accuracy can be significantly improved when an effective classification model is employed, showing a promising tool in computer-assisted warfarin management for clinical practice.
such information can be effectively collected from a laboratory information system (LIS); the two most well-known indicators, alanine amino-transferase (ALT) [31] and serum creatinine (SCr) [32], are utilized in this study.

To build a classification model, each inpatient in the dataset is associated with a particular class label, i.e. adequate or inadequate warfarin therapy. In clinical practice, physicians usually examine the international normalized ratio (INR) value to adjust the dosage of warfarin. However, there is no unified consensus on the optimal INR treatment range [34]. Based on the British Society for Haematology and the American College of Chest Physicians, the recommended range of the INR value is between 2 and 3; but You et al. [33] suggested that the INR value for Asians should be controlled at 1.8–2.4 to decrease the possibility of thromboembolism. Combining previous studies and the suggestions by the clinicians at the case medical center, the target range of warfarin was set at 1–3 and the warfarin therapy was defined to be adequate if there existed at least three records of the INR measurement in the duration of hospital stay and all the recorded INRs fell within 1–3 between the time of initial warfarin dose and the discharge; otherwise, the inpatient was classified as inadequate.

Furthermore, DDIs could create potential risks to the elderly due to polypharmacy. Warfarin is recognized to have DDIs with more than 250 medicines and thus it is imperative to consider DDIs in warfarin dosage determination [6]. In the selected hospital, we found a list of forty medicines causing severe DDIs with warfarin in accordance with the Drug Interaction Facts and the DDI information released by the Department of Health in Taiwan (Appendix). Based on the aforementioned definitions of variables, clinical data of all inpatients were collected, filtered, and preprocessed; a total of 288 validated clinical cases with DDIs as well as 89 cases without DDIs were obtained, resulting in 377 valid clinical cases in total. It is worth noting that laboratory factors were utilized in this study for descriptive statistics for both DDI and non-DDI groups are similar, i.e., from 0.25 to 6 mg, oratory factors were utilized in this study.

### 2.2 Experimental Design for the Classification Systems

To build an adequate dosage evaluation system for the elderly, this study adopts Weka 3.7.3 open-source data mining software (www.cs.waikato.ac.nz/ml/weka) to investigate the performance of the classification techniques, including J48 (C4.5 in Weka), IBk (kNN in Weka), SimpleCART (CART in Weka), RandomForest (RF in Weka), MultilayerPerceptron (MLP in Weka), SMO (SVM in Weka), SimpleLogistic (LGR in Weka). Moreover, this study further employs Adaptive Boosting (AdaBoost in short), one of the most popular classifier ensembles, to enhance the predictive power of the classical classification techniques [41]. Previous studies showed that several classification algorithms in conjunction with AdaBoost achieve higher classification accuracy than individual base classifiers [42–45]. In Weka, the AdaBoost can be performed by utilizing the AdaBoostM1 module. Table 2 lists the specific parameter values selected for each classification technique in Weka.

A previous study showed that the class imbalance problem deteriorates the performance of classification techniques [46]; thus a resample module in Weka is adopted to modify the proportions of two classes to be almost identical. In addition, some useful instances in the adequate class may not be chosen by the resample method, resulting in the loss of valuable information for classifications. Therefore, the random resample technique is applied thirty times to construct datasets; for each generated dataset, ten-fold cross-validation is then applied in all the experimental evaluations.

### 3. Results

#### 3.1 Descriptions of the Related Variables

Variables used in this study and their descriptive statistics for both DDI and non-DDI groups are shown in Table 3 for comparative purpose. As mentioned earlier, there were 288 inpatient cases drawn from the group with DDIs and 89 cases without DDIs, resulting in 377 valid clinical cases in total. It is worth noting that laboratory factors were utilized in this study including the ALT and SCr. The thyrotoxicosis variable was removed because no subject has such symptoms. Even though ranges of warfarin dosages taken by both groups are similar, i.e., from 0.25 to 6 mg, the inadequacy rate of the DDI group of inpatients is 58.68% whereas that of the non-DDI group is 40.45%. The higher inadequacy rate for the DDI group could be a
result of the interference of taking multiple medicines and thus adequate warfarin usage decisions should be made with discretion.

As to personal characteristics in both groups, ranges of age distribution are similar; however, the mean age for the DDI group is close to 78 years old, which is about two years older than the average of the non-DDI group. Meanwhile, weight distribution is similar to the age. The mean weight for the group of DDIs is higher than that of the non-DDI group. As to gender, the female proportions in both groups are almost the same, accounting for about 57%. According to the CHF, the proportions in both groups are very close to each other; SCr scores also show the similarity in both groups. However, summary statistics of the ALT indicator show that the mean scores of the ALT are 42.2 and 29.2, respectively, for both DDI and non-DDI groups.

### 3.2 Evaluation Results

The evaluation results of seven different classification techniques for both DDI and non-DDI groups are presented in Table 4; comparisons of adding or not adding AdaBoost are also included. Specificity, sensitivity, overall accuracy and AUC are utilized to compare the evaluation performance for the adequacy prediction. In Table 4, we only report mean and standard deviation of the thirty generated data-sets for ease of explanation; other summary statistics are available upon request from the authors.

Several sets of intriguing comparative results are revealed. First, decision tree-based classifiers including C4.5, CART and RF outperform other classifiers in prediction accuracy. For the non-DDI group (upper panel of Table 4), the average values of sensitivity, specificity, accuracy and AUC using the decision tree-based approaches as well as the MLP are all higher than those using the kNN, LGR and SVM, whether AdaBoost is incorporated or not. For example, the accuracy measures are all above 72% for the former classifiers but below 67% for the latter ones. As for the DDI group, the decision tree-based approaches also perform well according to the four selected evaluation indicators whereas performance of MLP cannot match that of the decision tree-based classifiers regardless of AdaBoost. Therefore, we can conclude that predicting ability of the decision tree-based approaches is better than other classifiers, especially when DDI is considered.

#### Table 3 Summary statistics of variables for the non-DDI and DDI groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-DDI group</th>
<th>DDI group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequacy</td>
<td>Yes/No</td>
<td>Yes/No</td>
</tr>
<tr>
<td>warfarin dose (mg)</td>
<td>0.25 to 5</td>
<td>μ = 1.08, σ = 0.61</td>
</tr>
<tr>
<td>Gender</td>
<td>Male/Female</td>
<td>Male: 38/ Female: 51</td>
</tr>
<tr>
<td>Age (years)</td>
<td>65 to 97</td>
<td>μ = 76.21, σ = 6.82</td>
</tr>
<tr>
<td>Weight (kgs)</td>
<td>39 to 84</td>
<td>μ = 60.83, σ = 10.62</td>
</tr>
<tr>
<td>CHF</td>
<td>Yes/No</td>
<td>Yes/5/ No: 84</td>
</tr>
<tr>
<td>ALT</td>
<td>8 to 195</td>
<td>μ = 29.15, σ = 25.58</td>
</tr>
<tr>
<td>SCr</td>
<td>0.42 to 9.7</td>
<td>μ = 1.46, σ = 1.43</td>
</tr>
</tbody>
</table>

Note: CHF: Congestive heart failure; ALT: Alanine amino transferase; SCr: Serum creatinine
of clinical practice. For each generated dataset, we calculated the score of each input variable based on the gain ratio of each input variable associated with the dependent variable in Weka 3.7.3 and then ranked all input variables. Table 5 lists the rankings of input variables for both DDI and non-DDI groups. As shown, weight, CHF, and gender are identified to be the top three critical variables affecting prediction accuracy for the elderly group without DDI in our experiment, while age, ALT, and warfarin doses are the most influential factors for the DDI group. Therefore, our results show that utilizing the laboratory information may assist clinical practitioners to make adequate decisions for the elderly, especially for the DDI group.

Extraction of crucial factors for the non-DDI and DDI groups of elderly inpatients can provide supplementary knowledge for making a proper decision for clinical practice. From discussions with professional clinicians and pharmacists from the sampled hospital, explanations for the ranking of the important factors are summarized here. Considering potential DDIs with warfarin (one of the most incurable high-alert medications), the metabolic burden of the kidneys and liver and/or CHF of elderly inpatients should be the primary concern for practitioners. On the other hand, as for the elderly inpatients without DDIs, drug use is less complicated and therefore burdens on renal function and/or of liver disease are relatively less serious; as a consequence, personal characteristics such as weight can be used as determinants for the prescription of high-alert medication. To sum up, the evaluation results using decision tree-based approaches can provide supplementary knowledge to make effective clinical decisions on warfarin dosages for the elderly.

4. Discussion

Prediction accuracy of warfarin has been a heated topic among the high-alert drugs [3]. Statistical models such as the LR and LGR are commonly utilized in the literature [8, 12, 16, 17, 26]; until recently, prediction ability has been improved using other approaches like data mining and machine learning [11, 19, 48]. Using the combined techniques of each classifier and AdaBoost, our study reinforces the conclusion that prediction models incorporating decision tree-based techniques can be more efficient and accurate than other classifiers. Therefore, decision tree-based approaches could be widely adopted to assist clinicians in making better medication decisions.

Even though a decision support system developed by using decision tree-based approaches can be used as a supplementary tool for clinical practitioners, incorporation of the most profound factors related to warfarin dose requirements is actually the key to quality prediction. Several factors are found to be crucial in warfarin dosages including genetic, clinical, environmental, and medication factors [15, 26, 30]. Our study shows that information from using laboratory tests as well as inpatients’ history should not be ignored because related variables (e.g. CHF and ALT) are shown to be decisive in our prediction models. As pointed out by Winkelmayer et al. and Schmucker [49, 50], the functions of the kidneys and liver of inpatients should be seriously considered while using high-alert drugs like warfarin. Other related studies [6, 19, 26] showed that DDI effects need to be taken into account especially when medicines with narrow therapeutic range are used. To our knowledge, only a
few studies utilized the LIS data in making warfarin dosage prediction; this study serves as one of a few studies [17, 26] that make use of the LIS data to demonstrate its advantages in prediction modeling. As presented in the previous section, ALT contributes the second most in determining the adequacy of warfarin prediction for elderly inpatients with DDIs following age. Compared with other clinical features like weight and gender, this study provides significant evidence for utilizing the relatively inexpensive information collected by the LIS.

If DDIs were not considered a serious problem to the elderly, on the other hand, our results show that weight and gender of the elderly inpatient are among the top three most influential factors affecting the prediction accuracy of warfarin usage; CHF is also on the list. This finding indicates that whether or not warfarin usage is adequate depends mainly on the inpatient's weight as long as this inpatient does not take other medications which cause drug interaction with warfarin. However, non-DDI inpatient cases from the selected medical center in southern Taiwan accounted for less than one fourth of the elderly inpatients; the majority of the elderly inpatients (over 75% in our study) were classified into the DDI group. To sum up, our study provides evidence showing the importance of the LIS data, whose benefits were underestimated in constructing prediction models in the literature.

Based on our evaluation findings for each DDI and non-DDI group, decision rules induced by the decision tree-based algorithm can be used for developing a drug-dosage decision support system. Because the knowledge extracted by the decision tree-based algorithm is in the form of IF-THEN rules, the expert systems can be easily built by incorporating the rules into an expert system shell [51]. Clinicians can refer to the rules summarized by historical prescription records before prescribing warfarin to the elderly inpatient. As a consequence, decision rules developed from our prediction models can be of assistance to clinical practitioners for making better medication decisions for elderly inpatients.

5. Conclusions

Inappropriateness of drug use is commonly seen in elderly patients not only because of medical problems caused by aging but because of the high probable DDIs resulting from various chronic diseases. Prediction on adequate usage of medicines, especially high-alert drugs like warfarin, is extremely important for the elderly because improper treatment of these high-alert medicines would increase adverse drug events and may cause severe morbidity and mortality.

This study responds to the challenge of predicting appropriate warfarin prescriptions for the elderly by developing dosage decision support systems. Specifically, we applied seven classification techniques, as well as their extensions using AdaBoost techniques for improving predictive performance. According to our analysis of 377 inpatient cases in Taiwan, all classifiers predict the adequacy of warfarin more accurately than does the clinical physicians' subjective decision. The overall evaluation results verify that DDI is a critical factor in warfarin dosage decision-making. In addition, the decision tree-based techniques with AdaBoost are suggested as the most effective prediction models in this study.

Considering the complicated characteristics of warfarin, this study shows that decision tree-based techniques can serve as a supplementary tool due to their superior performance in predicting adequacy. Even though DDIs are so common for the elderly that they complicate the effectiveness of using warfarin, our study provides sufficient evidence to support the assumption that risks from inadequate use of warfarin can be dramatically reduced, and thus, the improvement in the safe use of high-alert drugs will be of benefit to both clinicians and elderly patients.

Some constraints are addressed because they may restrict implications from our study; these limitations may provide clues for future studies. First, we collected cases from one hospital only in southern Taiwan; one should be cautious making inferences from samples of this hospital to all elderly inpatients in Taiwan. A random sampling scheme from multiple hospitals may lead to a collection of better representatives and may strengthen the validity of generalization from the evaluation results. Second, a laboratory information system is illustrated to be of importance in determining the adequacy of warfarin dosage for the elderly. Even though we only utilized two variables from the LIS, due to the difficulty of access to more detailed data of inpatients in the selected hospital, we provided evidence to show the importance of making the best use of the LIS in clinical practice. Predictability may be improved if models can include more variables in the LIS, which is most often the least expensive way to retrieve inpatient's information. Finally, an interventional study in which patients are initiated based on the rules in expert systems may be conducted as future work.

Acknowledgments

This research was supported in part by National Science Council of the Republic of China under the Grant NSC 99-2410-H-194-065 and NSC 101-2914-I-194-005-A1.

References


