Medical Decision Making for Warfarin Dosing Using Machine Learning Methods

BY

ASHKAN SHARABIANI

B.S., SHAHID BEHESHTI UNIVERSITY, 2009

M.S., SHARIF UNIVERSITY OF TECHNOLOGY, 2011

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Industrial Engineering and Operations Research in the Graduate College of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:

Houshang Darabi, Chair and Advisor
William Galanter, Medicine
David He, Mechanical and Industrial Engineering
Edith Nutescu, Pharmacy Practice
Adam Bress, Pharmacotherapy, University of Utah

This thesis is dedicated to my parents, Nasrin and Behrooz Sharabiani, and my wife Saba Rezaeian without whom this work would never have been accomplished.

ACKNOWLEDGMENTS

I would like to seize this opportunity and thank several people without whom this work would never be possible.

First and foremost, I would like to thank my mentor and adviser Professor Houshang Darabi. In the last four years, I had the honor and privilege to work under his supervision. I am sincerely grateful for his unconditional guidance and support throughout my work at the University of Illinois at Chicago. I also thank him for constantly pushing me towards being creative and working indefatigably towards utilizing my potentials. His work ethics, authenticity, and intelligence have always been a source of inspiration to me. I thank him for his insight and his friendship. He gave me the opportunity to succeed and believed in me through all the challenges in my work.

I would like to express my gratitude to my committee members. Especially, I would like to thank Dr. Bress, Dr. Galanter, Dr. Nutescu, and Dr. He whose support and knowledge have been invaluable in my research.

I also would like to thank my fellow lab mates at Process Mining and Intelligent System Analytics Team, Robert Andrzejewski, Tatiana Benavides Gallego, Anooshiravan Sharabiani, Fazle Karim, Maryam Teimoori, Ankush Bansal, Elnaz Douzali, Ashley Pimentel, Jillian Economy, and Samaneh Ebrahimi, who provided a convivial environment for me to work.

I am thankful to my brothers Arash and Anooshiravan Sharabiani for their constant support and love in my life. They have always believed in me and provided any means necessary for my success.

I am grateful to my wife, the love of my life, Saba Rezaeian. I am truly indebted for her love, her patience, and her support. She has always been a source of energy and hope for all my endeavors in these years at the University of Illinois at Chicago.

Last but not least, my deepest gratitude and love goes to my parents Nasrin and Behrooz Sharabiani for their never-ending love, support, and guidance in life. This work would never become possible without them. They have always encouraged me to pursue this degree and be patient and resilient especially at difficult moments.

TABLE OF CONTENTS

1 INTRODUCTION 1
1.1 What is Warfarin? 1
1.2 Importance of concentration on Warfarin Dosing 1
1.3 Chapter Synopsis 3
2 LITERATURE REVIEW 5
2.1 Dosing Methods for Warfarin 5
2.2 Dosing methods for specific cohort of patients 9
2.3 Hesitation Regarding Involving Genetic Data in Modeling 10
3 PRELIMINARIES 12
3.1 Machine Learning 12
3.2 Supervised Learning Methods 13
3.2.1 Linear Regression Modeling 13
3.2.2 Support Vector Machines 18
3.2.3 Relevance Vector Machines 23
3.2.4 Decision Trees 25
3.2.5 Shrinkage Regression 26
3.3 Model Evaluation 28
4 DOSE PREDICTION MODELS FOR PATIENTS OF SPECIFIC ETHNICITIES 30
4.1 Problem Definition 30
4.2 Data Set Description 30
4.3 Methods 31
4.4 Results 39
5 NEW METHODOLOGY TOWARDS MULTI-ETHNIC PREDICTION MODELING 40
5.1 Problem Definition 40
5.2 Data Set Description 40
5.3 Methods 42
5.4 Results 46
6 COMPANION CLASSIFICATION MODEL TO PREDICTION MODELS 47
6.1 Problem Definition 47
6.2 Data Set description 47

6.3 Methods 48
6.4 Results 51
7 A NEW APPROACH TOWARDS MINIMIZING THE RISK OF MIS-DOSING FOR POPULAR WARFARIN INITIATION DOSES PRESCRIBED BY THE PHYSICIANS 53
7.1 Problem Definition 53
7.2 Data set Description 53
7.3 Data Preprocessing & Visualization 56
7.4 Methods 61
7.5 Conclusion 64
8 CONCLUSION AND FUTURE WORKS 65
8.1 FUTURE WORKS 67
CITED LITERATURE 68

LIST OF TABLES

Table 1. Variables and Coefficients for Gage CL 6
Table 2. Variables and Coefficients for Gage PKG 6
Table 3. Variables and Coefficients for IWPC CL 8
Table 4. Variables and Coefficients for IWPC PKG 8
Table 5. List of features in the dataset. 31
Table 6. Prediction performance of Neural Network 35
Table 7. Prediction performance of Support Vector Machine 36
Table 8. Model Coefficients for the regression model 36
Table 9. Model Coefficients for after implementing the stepwise model selection 37
Table 10. Results for multivariable Regression 37
Table 11. IWPC data set description 41
Table 12. Classification results for RVM 44
Table 13. Comparing the prediction accuracy of the proposed methodology with IWPC Cl and Gage Cl models 45
Table 14. Comparing the performance of different classification models on the Test set 49
Table 15. Comparing the prediction accuracy of the IWPC CL model on original and shrunken test sets 50
Table 16. Categorical variables in the data set 54
Table 17. Continuous variables in the data set 55
Table 18. Test of Hypothesis Results 56
Table 19. Model Coefficients 62
Table 20. Estimates Coefficients of the linear model without involving IDP in modeling 63
Table 21. Comparing the performance of the revised values of IDP with the original values of IDP and Gage CL model 63

LIST OF FIGURES

Figure 1. Factors Affecting Warfarin's Activity 2
Figure 2. Geometry of the least squares fitting 15
Figure 3.The separating hyper plane 19
Figure 4. Soft Margin Classifiers 21
Figure 5. Evaluation Process 29
Figure 6. Single hidden layer feed-forward neural network 32
Figure 7. Functionality of the neural network 32
Figure 8. Density graph of the stable dose after selecting the desired dose range 34
Figure 9. Density graph of the stable dose 34
Figure 10. Artificial Neural Network 35
Figure 11. Checking the model assumptions 38
Figure 12. Comparing the performance of different models. 38
Figure 13. The proposed two stage modeling 43
Figure 14. The proposed methodology for using the IWPC clinical model 48
Figure 15. IDP Vs. Therapeutic Dose 57
Figure 16. Pareto chart for popular IDPs 57
Figure 17. Popular IDPs Vs. Therapeutic Dose 58
Figure 18. Comparing the distribution of Therapeutic Dose for Popular IDPs using Boxplots 58
Figure 19. Distribution of the percentage error 59
Figure 20. Distribution of percentage error at each level of popular IDPs 60
Figure 21. Correlation Matrix 61

LIST OF ABBEREVAIATIONS

ANN	Artificial Neural Network
AA	African American
IWPC	International Warfarin Pharmacogenetics Consortium
PKG	Pharmacogenetics
CL	Clinical
DT	Decision Tree
SVM	Support Vector Machines
RVM	Relevance Vector Machines
DS	Decision System
BSA	Body Surface Area
Mg	Milligram
Wk	Week
DVT	Deep Vein Thrombosis
PE	Pulmonary Embolism
LMWH	Low Molecular Weight Heparin
IDP	Initial Dose Prescribed by the Physician

SUMMARY

This thesis has four main contributions. A brief introduction to the four contributions is presented as follows.

The first contribution of this thesis is that it provides a new warfarin prediction model for patients of specific ethnicity (African-American (AA) patients). After examining three powerful machine learning–based methods (Artificial Neural Networks, Support Vector Regression, and Multivariate Linear Regression), a regression model is developed for AA patients which outperforms four popular dose prediction models in the literature known as IWPC Clinical model, IWPC Pharamacogenetic model, Gage Clinical model, and Gage Pharamacogenetic model.

The second contribution is that it presents a new methodology for developing prediction models for Warfarin dosing. The proposed methodology estimates the initial dose for Warfarin in two stages. In the first stage, using relevance vector machines, the patients are classified into two classes; patients requiring high doses (>30mg/wk) and patients who require low doses ($\leq$30mg/wk). In the second stage, for each class, using two different regression models, the dose is predicted. The proposed model was examined against Gage, IWPC Clinical models, the regression model for AA patients that was mentioned above, and the fixed-dose approach. It outperformed all of them in terms of prediction accuracy.

The third contribution is developing a companion model for IWPC Clinical model. IWPC Clinical model is one of the most widely used prediction models in application. The companion model functions as a decision support system which helps clinicians to identify the patients for whom using the IWPC Clinical is most beneficial. It is expected that using the proposed

companion model decreases the risk of misdosing (Overdosing/ Underdosing) by IWPC Clinical model significantly.

The fourth contribution of this paper is the development of an approach to estimate the amount of percentage error for initial doses prescribed by the physicians using shrinkage methods. By applying this estimation, the prescribed doses were revised accordingly. It was shown that by revising physicians' doses, the resulting doses are much more accurate than the original values of doses and the values predicted by Gage Clinical model. This approach is promising and warrants further study that may produce a functional clinical decision support system to assist with initial dosing of Warfarin.

1 INTRODUCTION

In this Chapter the introduction of Warfarin and the significance of concentration on this drug are presented.

1.1 What is Warfarin?

Warfarin is one of the most commonly prescribed drugs in the United States (Kirley et al. 2012). This drug was initially invented in 1954 as a pesticide for mice and rats. Warfarin has been found to be quite effective to avoid blood thrombosis (formation of blood clots inside blood vessels). Since 1954, by approving the effectiveness of this drug, it has been prescribed and used commonly. This drug is the most popular and most widely prescribed oral anticoagulant in America. Although this drug has been proven to have significant impact for preventing thrombosis, its treatment has been quite challenging. With existence of several competitors in the market (Connolly et al. 2009)(Patel et al. 2011)(Granger et al. 2011)(Mega 2011), in 2011, more than 33 million prescriptions were dispensed in United States (Informatics 2011).

1.2 Importance of concentration on Warfarin Dosing

Determination of the optimal dose for this drug is quite challenging considering its narrow therapeutic index and the substantial inter-patient variability in dose requirements to attain ideal anticoagulation (Elaine M Hylek et al. 2007). This means that mis-dosing (overdosing/under dosing) puts patients at risk of thrombosis, such as deep vein thrombosis or pulmonary embolism for under dosing, and bleeding for overdosing. For the time being, this drug ranks as the major drug-related cause of adverse effects resulting in hospitalization among the elderly (Palareti et al. 1996). Warfarin dose is determined based on a blood test called as International Normalized

Ratio (INR), which measures anticoagulation activity (Hutten et al. 2000). An INR of 2 to 3 is targeted for most indications. If the INR surpasses 3, the patient is at higher risk for bleeding. If the INR falls below 2, the patient is at increased risk for thrombosis (E. M. Hylek et al. 2006)(Wittkowsky 2004). The risk of bleeding or thrombosis with Warfarin is highest during the initial months of treatment. There are several factors affecting the activity of Warfarin, including age, body size, co-morbidities, genetic variants in the drug metabolizing enzyme, CYP2C9, and the drug target, VKORC1. In 2007, UFDA (US Food and Drug Administration), has suggested to modify Warfarin labels by providing information regarding VKORC1 and CYP2C9 variants (Brian F. Gage and Lesko 2008). One of the most important factors which affect Warfarin's activity is patients' diets. The level of consumption of vitamin K, which is mainly stored in green vegetables such as Broccoli, Cabbage, Parsley, and Apiaceae, have a significant impact on this drug's activity. See Figure 1.

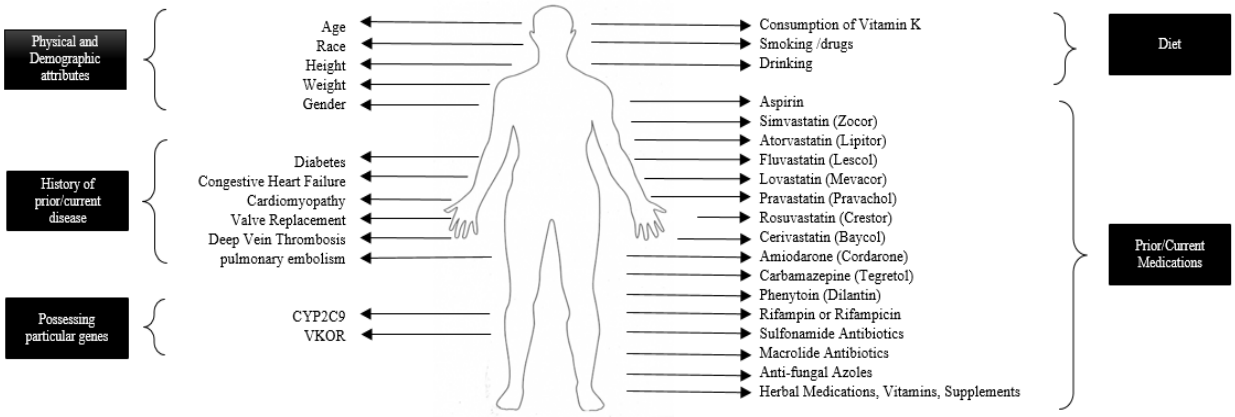


Figure 1. Factors Affecting Warfarin's Activity

The process of Warfarin treatment initiates by determination of the initial dose by the clinicians (Physicians, Nurses, etc.). The initial dose will then be refined according to the result of its corresponding INR test, which indicates the level of coagulation. This phase is known as

dose refinement. The process of dose refinement continues once the maintenance dose (Therapeutic Dose) is reached. An appropriate choice of initial dose will shorten the length of dose refinement process and also decrease the risk of unfavorable outcomes for patients.

Considering the variety of different factors affecting Warfarin's activity, estimating the initial dose is very critical. Therefore, different clinicians approach the dosing problem from different perspectives. One of the popular methods for Warfarin treatment is known as Loading Dose procedure. In this procedure, a dose higher than the desired maintenance dose will be prescribed and then it will be decreased gradually to reach the maintenance dose.

The time to reach the maintenance dose is dependent on how fast this drug is removed from the system. Therefore, if initial dose is close to the maintenance dose, it will take almost five-times the half-life of Warfarin for reaching the maintenance dose (Eriksson and Wadelius 2012).

Another approach is to use mathematical models for prediction of the initial dose for each patient. There are different mathematical models in the literature which are trained by the data of different cohorts of patients. The mathematical models range from traditional statistical models to more advance machine learning models. The major focus of this thesis is to develop new mathematical models or improve the performance of existing popular models in the literature from different perspectives. In section 1.3 the structure of materials in this thesis is presented.

1.3 Chapter Synopsis

In Chapter 2, a comprehensive review of the used mathematical models for predicting the initial dose for Warfarin is presented. In Chapter 3, the required mathematical background for the methodologies which were applied in the thesis are presented; starting from introducing machine

learning methods, supervised learning, and five powerful methods in the family of supervised learners (Regression Modeling, Decision Trees, Support Vector Machines, Relevance Vector Machines, and Shrinkage Methods) to evaluating the modeling results are discussed. In Chapter 4, development of a new prediction model for African-American patients is explained. In Chapter 5, a novel methodology for developing prediction model is presented. This methodology functions in two stages which uses a classification method in the first stage and prediction models in the second stage. In Chapter 6, developing a companion classification model for IWPC Clinical model is described. The developed model functions as the identifier of the appropriate cohort for using IWPC Clinical model. Finally, in Chapter 7, a new approach towards choosing an appropriate initial dose is presented. In the proposed approach, using the shrinkage methods, the amount of percentage error for doses prescribed by the physicians are estimated and the doses are revised accordingly. It is shown that the modified doses are much more accurate than the original values and doses predicted by the Gage Clinical model.

2 LITERATURE REVIEW

In this Chapter, a comprehensive review of mathematical models for Warfarin dosing that have been proposed in the literature is presented.

2.1 Dosing Methods for Warfarin

In 2005, Sconce et al, proposed a PKG model containing the variables Age, Height, CYP2C9, and VKOR1. They used the data of 297 patients for their derivation cohort and the data of 38 patients for the validation cohort. The resulting model provided a satisfactory level of fitness (R^2 = 55%) (Sconce et al. 2005). In 2008, a research led by Dr. Brian Gage (from Department of Internal Medicine, Washington University School of Medicine, St. Louis, Missouri, USA) developed two prediction models for the Warfarin initiation dose. They used the data of 1,015 as their derivation cohort and 292 patients in validation cohort. 83% of the data that they used for modeling constitutes the data of White patients. The first model that was proposed by this team was a Clinical model (CL) and the other model was the Pharmacogenetic model (PKG). These models are known as "Gage Models". The variables that were applied in the CL model were BSA (Body Surface Area), target INR, Smoking status, Age, Amiodarone, and DVT/PE (Deep Vein Thrombosis/Pulmonary Embolism). However, in PKG model several variables for genomic data were utilized. In Table 1-2 the coefficients for both models are presented. It must be noted that in the data preprocessing phase, the response variable (Maintenance Dose) was transformed using logarithmic transformation. Therefore, after applying both models, the results have to be exponentiated to get transformed back to the original format. They evaluated their models' performance with respect to the level of fitness (R^2) and Median absolute prediction error, mg/day after applying the model on the validation set. The R^2 for the

CL model was 17% and for PKG mode 54%. Also, the median absolute error for the CL model was 1.5 mg/day and for the PKG model 1.0 mg/day (B F Gage et al. 2008).

In 2009, International Warfarin Pharmacogenetics Consortium (IWPC) research team also collected the data of 5052 patients. The data was collected from 21 research teams in 9 countries over 4 continents. They used 80% of the data set (4043 patients) as their derivation cohort and the remaining 20% (1009 patients) as their validation set.

Table 1. Variables and Coefficients for Gage CL

Variable Name	*Corresponding Coefficient in the model*
Intercept	0.613
BSA	0.425
Age	-0.0075
African-American Race	0.156
Target INR	0.216
Amiodarone	-0.257
Smoking Status	0.108
DVT/PE	DVT/PE

Table 2. Variables and Coefficients for Gage PKG

Variable Name	*Corresponding Coefficient in the model*
Intercept	0.9751
BSA	0.4317
Age	-0.00745
African-American Race	− 0.0901
Target INR	0.2029
Amiodarone	− 0.2538
VKOR3673G>A	− 0.3238
CYP2C9*3	− 0.4008
CYP2C9*2	− 0.2066
Smoking Status	0.0922
DVT/PE	0.0664

After performing the data preprocessing, several modeling techniques were implemented on the data for reaching the best model. The prediction models were ordinary linear regression,

multivariate adaptive regression splines, support vector regression, regression trees, model trees, least-angle regression, and Lasso Regression. Among those modeling techniques, the linear regression model appeared to be the most effective. They developed two linear regression models (IWPC CL and IWPC PKG). The response variable was transformed by logarithmic transformation and square-root transformation. However, the square-root transformation was selected for modeling. Instead of using the actual values for Age, the Age-Decade was applied (1 represented 10-19 years old, 2 represented 20-29, etc.). Also, actual values for Height and Weight were utilized in the model instead of BSA. In addition, a new variable entered the model as Enzyme Inducer Status which takes the value of 1, if the patients consumed any of the following drugs: carbamazepine, phenytoin, rifampin, or rifampicin, otherwise it takes the value of 0. In addition, three binary variables were involved in the model indicating whether Race, VKORC1, or CYP2C9 are missing or not. In Table 3-4 the variables in each model and their corresponding coefficients are presented. They assessed the performance of each algorithm in three categories; patients requiring less than or equal to 21 mg per week, between 21 to 49 mg per week, and more than or equal to 49 mg per week. The models were compared against the fixed-dose approach (35 mg per week) in each category. The proposed modeling approaches were significantly more accurate than the fixed dose approach for patients requiring less than or equal to 21 mg per week, or more than or equal to 49 mg per week. These categories constitute 46.2% of the population. In both categories, the PKG model appeared to be more accurate than the CL model (Klein et al. 2009).

Table 3. Variables and Coefficients for IWPC CL

Variable Name	*Corresponding Coefficient in the model*
Intercept	4.0376
Age (Decades)	-0.2546
Height (Cm)	0.0118
Weight (Kg)	0.0134
Asian	-0.6752
Black	0.406
Missing or Mixed Race	0.0443
Enzyme Inducer Status	1.2799
Amiodarone	-0.5695

Table 4. Variables and Coefficients for IWPC PKG

Variable Name	*Corresponding Coefficient in the model*
Intercept	5.6044
Age (Decades)	-0.2614
Height (Cm)	0.0087
Weight (Kg)	0.0128
VKORC1^A/G	-0.8677
VKORC1A/A	-1.6974
VKORC1 genotype unknown	-0.4854
CYP2C9*1/*2	-0.5211
CYP2C9*1/*3	-0.9357
CYP2C9*2/*2	-1.0616
CYP2C9*2/*3	-1.9206
CYP2C9*3/*3	-2.3312
CYP2C9 genotype unknown	-0.2188
Asian	-0.1092
Black	-0.276
Missing or Mixed Race	-0.1032
Enzyme Inducer Status	1.1816
Amiodarone	-0.5503

According to “Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 Genotypes and Warfarin Dosing” which was published in 2011 by

Johnson *et al.*, the models proposed by Gage and IWPC are the most recommended models for predicting warfarin initiation dose (Johnson et al. 2011).

The majority of patients in the data sets that were both used by IWPC and Gage *et al.*, were Caucasian. Therefore, the performance of models were significantly less accurate for patients of different ethnicities; namely African-American and Asian patients. This biased modeling procedure is also evident in works of Wadelius et al.(Wadelius et al. 2007)(Wadelius et al. 2009), Limdi et al. (Limdi et al. 2008)(Limdi et al. 2010), and Shellman et al. (Schelleman et al. 2008)(Schelleman, Limdi, and Kimmel 2008). This limitation called for developing prediction models which produce accurate results for patients of specific ethnicities.

In the next section, the models that were developed for these specific cohorts of patients are presented.

2.2 Dosing methods for specific cohort of patients

A PKG model was developed by Hernandez et al using the cohort of 349 AA patients. The developed model was compared to IWPC models (CL and PKG) and its outperformance was proven (Hernandez et al. 2014).

Grossi et al. developed a PKG prediction model using Artificial Neural Networks (ANN) using the data of 377 patients. The patients were all Caucasian and over the age of 18. Their model outperformed the models developed by IWPC and Gage (Grossi et al. 2013).

Cosgun et al. examined three powerful machine learning based models in developing PKG models for AA patients. The methods were Random Forest Regression, Boosted Regression

Tree, and Support Vector Regression. They compared their models with popular prediction models in terms of level of fitness (R^2)(Cosgun, Limdi, and Duarte 2011).

Oztaner et al, developed a Bayesian estimation framework for developing PKG models. They examined their procedure both on IWPC data set and a local data set of Turkish patients (N=107). The proposed methodology was examined against famous prediction models and it was proven that the model provides a better level of fitness (Serdar Oztaner et al. 2014).

Xu *et al.*, also developed a refined PKG model for Chinese patients. By incorporating additional genes in the modeling, the proposed models outperformed the conventional PKG model (with CYP2C9 and VKOR1) and the fixed dose approach (3 mg/day) in terms of level of fitness (Xu *et al.* 2012).

2.3 Hesitation Regarding Involving Genetic Data in Modeling

Involving the genetic factors in dose prediction has been a challenging procedure. Applying PKG in practice requires the availability of genetic data. Acquiring such data is not feasible for most institutions in the world. Therefore, a major hesitation towards applying the genetic factors in modeling exists.

In 2013, two randomized and controlled trials for evaluating the performance of PKG models were published. The study known as EU-PACT (European Pharmacogenetics of Anticoagulation Therapy) found that the modified version of the IWPC model (PKG) outperformed the conventional one (Pirmohamed *et al.* 2013). In a different study known as COAG (Clarification of Optimal Anticoagulation through Genetics), it was found that by involving the genetic factors in the models, no more benefit can be achieved than CL models (Kimmel *et al.* 2013). The major limitation found in both studies was that the population of

patients involved were predominantly European (They only included the principle genetic determinants of Warfarin dosing for European patients; vitamin K epoxide reductase complex 1 (VKORC1) − 1639 G>A (rs9923231), cytochrome P450 2C9 (CYP2C9) *2, and CYP2C9*3 polymorphisms). However, for patients of other ethnicities the hesitation for performance of genetic factors remains. Drozda *et al*. investigated the involvement of important genetic factors for AA patients such as (CYP2C9*5, CYP2C9*6, CYP2C9*8, CYP2C9*11 alleles and rs12777823 G>A genotype) in the modeling. Using the cohort of 274 AA patients, they found out that removing the genetic variables from modeling results in a massive increase in prediction error (Drozda *et al.* 2015). In a study known as 'Marshfield Clinic Research Foundation (MCRF)', Burmester et al. investigated the time to reach the therapeutic dose on two patient cohorts. They proved that Pharmacogenetic factors did not accelerate the process of reaching the therapeutic dose (Burmester et al. 2011).

No robust conclusions were achieved from these studies regarding the involvement of Pharmacogenetic factors on Warfarin dosing. Detailed investigation of the above-mentioned studies are presented in some reviews (Scott and Lubitz 2014)(Cavallari and Nutescu 2014).

In 2013, Yang et al., investigated the influence of VKOR1 and CYP2C9 genotypes on the risk of hemorrhagic[1] complications for patients who are under Warfarin treatment (Yang et al. 2013). They performed a meta-analysis using 22 publications and concluded that "both CYP2C9 and VKORC1 genotypes are associated with an increased risk for warfarin over-anticoagulation, with VKORC1 c. −1639 G >A more sensitive early in the course of anticoagulation. CYP2C9*3 is the main genetic risk factor for Warfarin hemorrhagic complications" (Yang et al. 2013).

[1] Pertaining to bleeding or the abnormal flow of blood.

3 PRELIMINARIES

In this Chapter, the required mathematical background that has been applied in Chapters 4-7 is presented.

3.1 Machine Learning

The science of learning plays a crucially important role in different fields such as Artificial Intelligence (AI), Statistics, and Data Mining. Machine Learning (ML) which is one of the branches of AI, is about developing algorithms to assist computers to learn similar to human beings (Hastie et al. 2009). ML is also known as Statistical Learning to statisticians and mathematicians' community. This field aims to develop and study algorithms for learning from the data. The data set that is used in the learning process is known as the training set. According to the nature of the data in the training set and the scope of the study, different types of learning might be of interest. If the data set contains one or more target variables (which function as outputs) that we are interested to describe their current behavior and estimate its future behavior using other variables (which function as inputs) in the data set, the type of learning will be a Supervised Learning. The target variable is also known as Label, Response Variable, and Dependent Variable. Subsequently, if the target variables are missing in the data set the type of learning will be an Unsupervised Learning (Jiawei and Kamber 2001). There is also a third class of learning which is known as Semi-supervised learning in which the target variable is partially available. However, this class of learning is out of the scope of this thesis.

In the next sections, the Supervised Learning methods will be explored, specifically the methods that were applied in different Chapters of this document, in detail.

3.2 Supervised Learning Methods

As mentioned in 3.1, when the data set contains the target variable(s), the nature of the learning will be a Supervised Learning. There are numerous methods in the family of supervised learners which differ from each other in different perspectives. When the target variable takes continuous values (Quantitative variable), the type of prediction will be Regression (or Prediction) and when it takes discrete values (Qualitative variable) the type of prediction will be known as classification. In sections 3.2.1 - 3.2.5, there are powerful prediction and classification techniques which are applied in the later Chapters are presented.

3.2.1 Linear Regression Modeling

Assuming that Y is the response variable in our data set and X= (X_1, X_2,..., X_p) are the set of explanatory variables, in linear regression modeling it is presumed that the *E(Y|X)* is linear or linear model is an appropriate approximation for it. This assumption, although might seem too simple, enables the analysts to create interpretable and efficient models. In terms of prediction, these models sometimes create more accurate results than famous nonlinear and complex models. In this section, the application of linear regression for prediction is only discussed; however, they can also be applied for classification.

The linear regression model has the form

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \qquad (1)$$

The β_j s in (1) are known as the model parameters or the coefficients and the X_j s might come from different sources such as quantitative inputs, different transformations of the quantitative inputs such as square-root and log transformation, basis expressions like ($X_2 = X_4^{\,3}$)

, dummy coding a qualitative inputs, or the interactions between variables like ($X_2 = X_4 \cdot X_5$). It is extremely important to note that when entering a categorical variable into the modeling, they have to be converted to dummy variables. For example, when entering a variable Race in the modeling, which takes discrete values (White, Hispanic, African-American, and Asian), three binary variables are created for three values and leave the fourth one as the reference.

Therefore in regression, the goal is to estimate the parameters β using the data points in the training set

$$\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\} \tag{2}$$

For each case *i*, $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ represents the vector of measurements for each feature. One of the most popular methods for estimating the model parameters $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$ is the least squares in which the set of coefficients that minimize the residuals sum of squares will be selected.

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{3}$$

The criterion for least squares method to succeed is that x_i 's should be drawn randomly from the population or the y_i 's are conditionally independent given the x_i 's .The geometry of the least squares fitting is displayed in Figure 2.

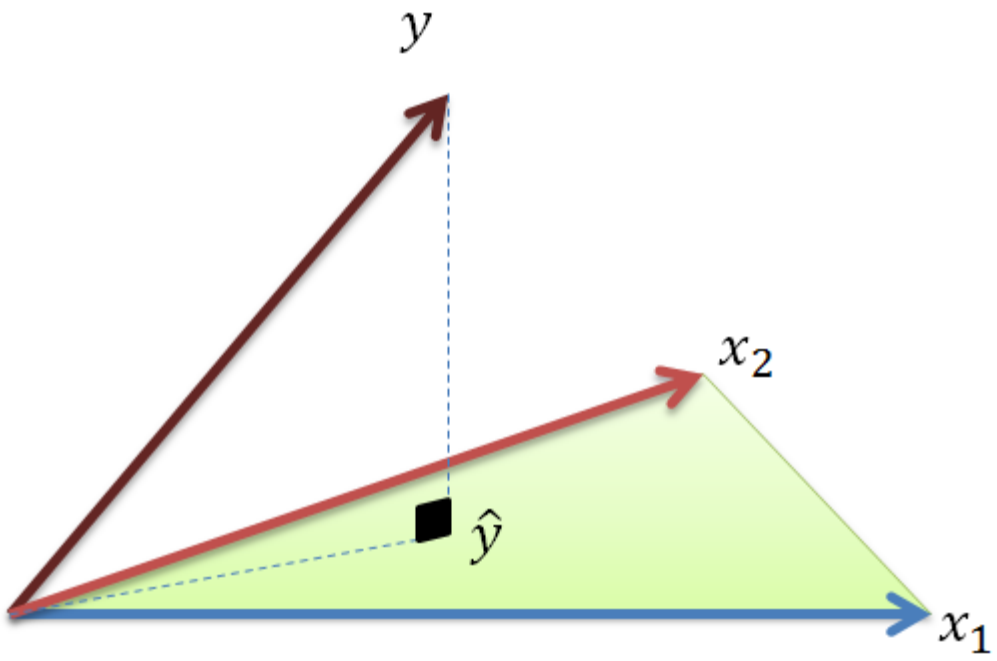


Figure 2. Geometry of the least squares fitting

In order to minimize (3), it must be noted that X is matrix with N × $(P + 1)$ dimentions. Therefore, the $RSS(\beta)$ can be written as

$$RSS(\beta) = (y - X\beta)^T (y - X\beta) \tag{4}$$

By differentiating (4) with respect to β, we have

$$\begin{aligned} \frac{\partial RSS}{\partial \beta} &= -2X^T (y - X\beta) \\ \frac{\partial^2 RSS}{\partial \beta \partial \beta^T} &= 2X^T X \end{aligned} \tag{5}$$

$X^T X$ will be a positive definite matrix if X is full rank column matrix and therefore, by setting the derivatives to zero we have

$$X^T(y - X\beta) = 0 \tag{6}$$

And

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{7}$$

Using the estimated vector of coefficients, the prediction values will be

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y \tag{8}$$

In which $H = X(X^T X)^{-1} X^T$ is known as the Hat Matrix. The Hat Matrix computes the orthogonal projection of *y* and therefore it is also called the projection matrix.

There are major assumptions in regression modeling which must be validated unless the reliability of the built model will be under question.

The important assumption about y_i s are that they are uncorrelated and have a constant variance σ^2. The variance-covariance matrix of β can be easily driven from (8), given by

$$Var(\hat{\beta}) = (X^T X)^{-1}\sigma^2 \tag{9}$$

and estimated by

$$\hat{\sigma}^2 = \frac{1}{N - p - 1}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{10}$$

The reason for choosing *N-p-1* instead of *N* in (10) is to make σ^2 and unbiased estimator. Another assumption is that the deviation of *y* around its expected values are Gaussian and additive

$$\begin{aligned} Y &= E(Y \mid X_1, X_2, ..., X_p) \\ &= \beta_0 + \sum_{j=1}^{p} X_j \beta_j + \varepsilon \end{aligned} \tag{11}$$

The ε has the Gaussian distribution with mean of zero and a constant variance: $\varepsilon \sim N(0, \sigma^2)$.

Based on (11) it is easy to show that

$$\hat{\beta} \sim N(\beta, (X^T X)^{-1}\sigma^2) \tag{12}$$

The reason for investigating the distributional properties of β is to perform different tests of hypothesis and develop confidence intervals for each β_j. For example, to test that if $\beta_j = 0$, we use the Z-score

$$z_j = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} \tag{13}$$

where v_j is the *j*th element of the diagonal of $(X^T X)^{-1}$ matrix.

After developing a regression model, the regression assumptions should be examined using the diagnostic tests.

3.2.2 Support Vector Machines

Among numerous classifiers that are proposed in machine learning literature, Support Vector Machine (SVM) is one of the most popular classification techniques. This model was first introduced by Vapnik in 1992 (Vapnik and Vapnik 1998). SVMs use a simple linear method applied to the data but in a high-dimensional feature space which is non-linearly associated to the input space (Steinwart and Christmann 2008).

In a typical classification problem, the data set consists of several features X_1, X_2,..., X_L and one or several variables for labels C_1,C_2,...,Cp. The goal is to develop a model to assign the objects (data points) to their classes. In a two class classification problem (C_1 and C_2), the objective is to develop a classifier using the N data points in the training set. Therefore for each point in the training set $\{x_n\}_{n=1}^{N}$ a label $z_n \in \{-1,1\}$, n = 1,..., N should be estimated. The classifier is defined as

$$y(x;w) = w^T \phi(x) + b \tag{14}$$

or

$$y(x;w) = \sum_{i=1}^{M} w_i \phi_i(x) + b \tag{15}$$

where $w \in \mathrm{R}^M$ is the weight vector, and $b \in \mathrm{R}$ is the constant and $\phi(.)$ is the transformation function. The predicted labels are computed using the sgn*(.)* function; *sgn(y(x))*. Assuming the data is linearly separable, there exists a vector $w(w^*)$ and $b(b^*)$ which yield a *hyperplane* that completely separates the data to two disjoint areas. This *hyperplane* is called the decision boundary (D) and the predicted labels for the data points and the value of $y(x_n)$ have the same sign; ($z_n y(x_n) > 0$; $\forall x_n \in \mathrm{R}^D$ and $z_n \in \{-1,1\}$). The minimum distance of the points in the training

set to *D* is called the *margin* (See Figure 3) which is computed using $\min_{n\in\{1,\ldots,N\}} \frac{z_n y(x_n)}{||w||}$; $\|\cdot\|$ is the L^2- norm.

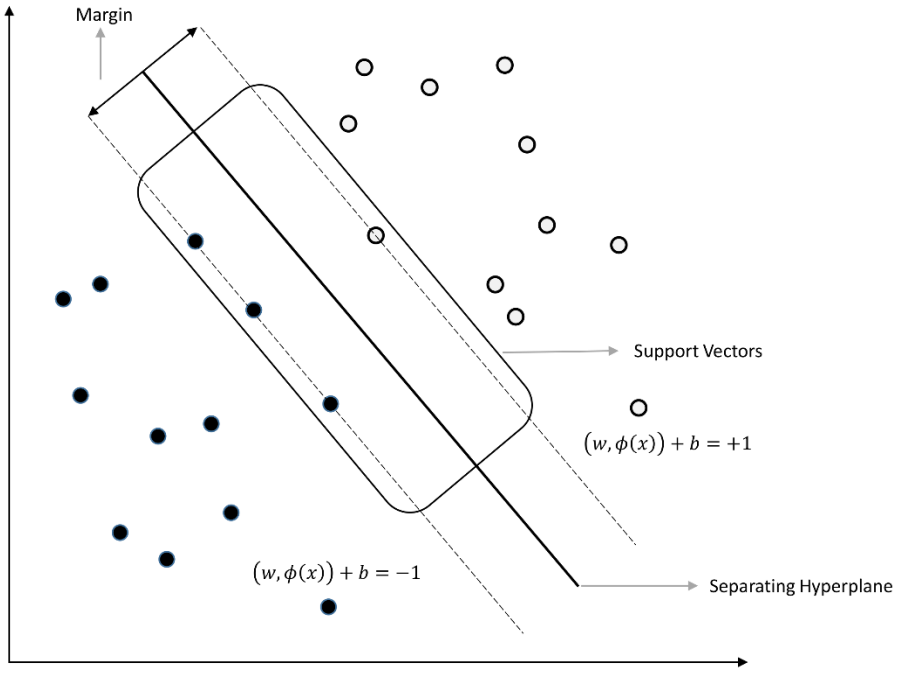


Figure 3.The separating hyper plane

The objective in *SVM* is choosing the values for *W* and b which maximizes the *margin* and also minimizes the classification error. The values for w^* and b^* are yielded by solving the following optimization problem

$$\max_{w \in \mathbb{R}^M,\ b \in \mathbb{R}} \left\{ \frac{1}{\|w\|} \min_{n \in \{1,\ldots,N\}} [z_n (w^T \phi(x_n) + b)] \right\} \tag{16}$$

The w^* and b^* which are resulted from (16) are also the solutions to the following minimization problem (17).

$$\min_{w \in \mathbb{R}^M,\ b \in \mathbb{R}} \frac{1}{2} \|w\|^2 \quad \textit{subject to} \tag{17}$$

$$z_n(w^T\phi(x_n) + b) \geq 1$$

where $x_n \in \mathbb{R}^D, z_n \in \{-1,1\}$, and $n = 1, \ldots, N$

The optimization problem in (17) can also be solved by applying Lagrange multipliers ($\lambda_n \in$ R, n = 1,...,N). The Lagrangian formation of (17) is

$$\mathcal{L}(w, b, \lambda) = \frac{1}{2}\left|\left|w\right|\right|^2 - \sum_{n=1}^{N} \lambda_n[z_n(w^T\phi(x_n) + b) - 1] \tag{18}$$

The first-order conditions for optimality in (18) are $\sum_{n=1}^{N} \lambda_n z_n \phi(x_n) = w$ and $\sum_{n=1}^{N} \lambda_n z_n = 0$.

After applying the conditions, the dual form of (17) will be resulted as follows(19).

$$\begin{aligned} &\max_{\lambda \in \mathbb{R}^N} \mathcal{L}(\lambda) \\ &\textit{subject to} \\ &\lambda_n \geq 0, n = 1, \ldots, N \\ &\sum_{n=1}^{N} \lambda_n z_n = 0 \end{aligned} \tag{19}$$

Where $\mathcal{L}(\lambda) \triangleq \sum_{n=1}^{N} \lambda_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n \lambda_m z_n z_m k(x_n, x_m)$ and $k(x, x') = \phi^T(x)\phi(x')$ is called the kernel function. The KKT (Karush-Kuhn–Tucker) conditions for optimality of optimization problems in (17, 19)

are $\lambda_n \geq 0$, $z_n y(x_n) - 1 \geq 0$, and $\lambda_n(z_n y(x_n) - 1) = 0$ where n = 1,...,N.

Those data points for which the corresponding λ_n is non-zero are called *support vectors.* These points play a crucial role in classifying new points.

If the points in the data set are not linearly separable, by using slack variables ($\xi_n \geq 0$) the concept of soft-margin classifiers (See Figure 4) will be defined. In this family of classifiers, by assigning

a penalty for the points that lay on the wrong side of the boundary, the optimization problem in (17) will be rewritten as follow in (20)

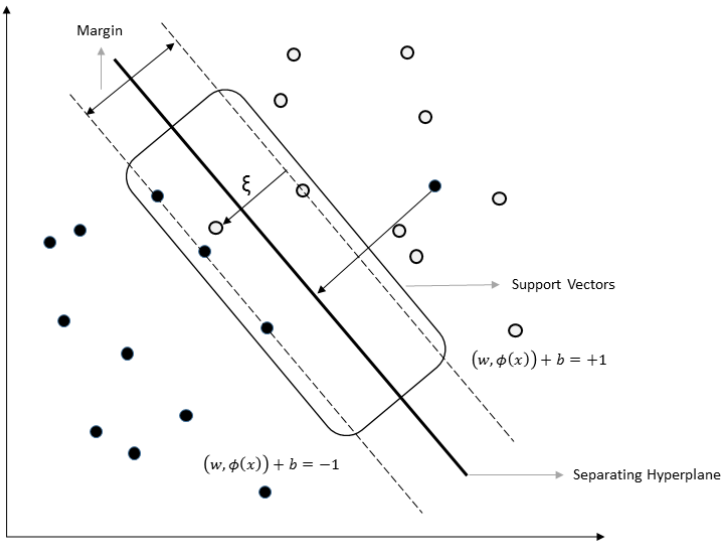


Figure 4. Soft Margin Classifiers

$$\min_{w \in \mathbb{R}^M, b \in \mathbb{R}, \xi \in \mathbb{R}^N} C \sum_{n=1}^{N} \xi_n + \frac{1}{2}||w||^2$$
$$\textit{subject to}$$
$$z_n y(x_n) \geq 1 - \xi_n, \quad n = 1, \dots, N \tag{20}$$
$$\xi_n \geq 0, n = 1, \dots, N$$

$C > 0$ is called the *complexity parameter*. The Lagrangian method can again be applied for solving (20) which has the form (21)

$$\mathcal{L}(w, b, \lambda, \xi) = \frac{1}{2}||w||^2 + C \sum_{n=1}^{N} \xi_n - \sum_{n=1}^{N} \lambda_n (z_n y(x_n) - 1 + \xi_n) - \sum_{n=1}^{N} \mu_n \xi_n \tag{21}$$

where $w = \sum_{n=1}^{N} \lambda_n z_n \phi(x_n)$, $0 = \sum_{n=1}^{N} \lambda_n z_n$, $\lambda_n = C - \mu_n$, $n = 1, \dots, N$, and $\lambda_n \geq 0$.

The dual form of this optimization problem is presented in (22)

$$\begin{aligned} &\min_{\lambda \in \mathbb{R}^N} \mathcal{L}(\lambda) \\ &\textit{subject to} \\ &0 \le \lambda_n \le C, n = 1,\ldots,N \\ &\sum_{n=1}^{N} \lambda_n z_n = 0 \end{aligned} \tag{22}$$

When the data space is not linearly separable, SVMs use a suitable mapping $\langle \Phi \rangle$ of the input data values to a higher dimensional feature space which will be regulated by the kernel function. The data set will be linearly separable in the transformed space. The kernel function returns the inner product of two images of x and x′, i.e., k (x, x') $=\langle \Phi(x), \Phi(x') \rangle$. Based on the nature of the data set, different kernel functions can be most effective: i.e. the polynomial kernel $K(x, x') = (\langle x, x' \rangle + 1)^2$, Multi-Layer Perceptron $K(x, x') = \tanh(\langle x, x' \rangle + \vartheta)$, Gaussian RBF Kernel $K(x, x') = \exp\left(-\frac{||x-x'||^2}{2\delta^2}\right)$, ANOVA kernel $K(x, x') = \sum_{k=1}^{n} \mathrm{Exp}(-\sigma(x^k - x'^k)^2)^d$, etc.

The major drawbacks of *SVM* are:

- The linear growth of the number of support vectors with the number of data points in the training set.
- Providing a hard binary decision. In most applications it would be much more useful when the level of certainty is addressed when classifying new objects.
- It is necessary to estimate the *C* (complexity parameter) which requires the cross-validation.

To overcome the above-mentioned shortcomings, in the next section the *Relevance Vector Machines (RVM)* will be introduced.

3.2.3 Relevance Vector Machines

Relevance Vector Machines (RVM) belong to the family of sparse Bayesian learners. This method, which can be used for both classification and regression, was introduced by Tipping (Tipping 2001). One of the most important advantages of *RVM* is its ability for handling classification problems when the cost of misclassification is different for different classes. In a classification problem, *RVM* assigns a class membership probability for a given point (x); $p(C_k|x,X,Z)$ where X is the feature set and Z is the set of labels in the training set. Assuming that the posterior probability of a target variable in C_1 is calculated by

$$p(z_n = 1|x_n,w) = \frac{1}{1+e^{-(x_n{}^T\phi(x)+b)}},\ n = 1,...,N \tag{23}$$

we will configure the likelihood function (LF). Using $\sigma(.)$ for the logit function, the right side of (23) can be denoted as $\sigma(y(x_n))$. Therefore, in our binary classification problem, the LF is

$$p(Z|X,w) = \prod_{n=1}^{N} p(z|x_n,w) = \prod_{n=1}^{N} \sigma\big(y(x_n)\big)^{z_n} \left(1-\sigma\big(y(x_n)\big)\right)^{1-z_n} \tag{24}$$

The weight parameters (w) in (24) have a Gaussian distribution with a mean of zero. However the variance of each w_i $i = 1,...,M$ could be different. So, the prior distribution of the weight vector will be

$$p(w|\alpha) = \prod_{n=1}^{M} \mathcal{N}(w_n; 0, \alpha_n^{-1}) \tag{25}$$

where α_i, $i = 1,\ldots,M$ is known as hyperparameters and are the inverse of the Gaussian distribution variance. For any new point (x) the posterior probability can be calculated as $p(z|x,X,Z)$. This probability is computed by marginalizing the $p(z,x,X,Z,w,\alpha)$;

$$p(z|x,X,Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(z|x,X,Z,w,\alpha) \times p(w|x,X,Z,\alpha)p(\alpha|x,X,Z)dwd\,\alpha \qquad (26)$$

Solving (26) can be done by using approximation, in which the vector of α will be used as a constant (α^*). α^* is the value which maximizes the $p(Z|X,\alpha)$. Therefore, (26) will be equal to

$$\int_{-\infty}^{\infty} p(z|x,X,Z,w,\alpha^*)p(w|x,X,Z,\alpha^*)\,dw \qquad (27)$$

Furthermore, $p(w|(x,X,Z,\alpha) = \frac{p(Z|x,X,w,\alpha)p(w|x,X,\alpha)}{p(Z|x,X,\alpha)} = \frac{p(Z|X,w)p(w|\alpha)}{p(Z|X,\alpha)}$. This probability should also be approximated. The approximation process aims to detect the vector of w which maximizes $p(w|x,X,Z,\alpha)$. The maximization problem (w^*) is

$$\max_{w\,\in\,\mathbb{R}^M}\{\ln\big(p(Z|X,w)p(w|\alpha)\big) - \ln p(Z|X,\alpha)\} \qquad (28)$$

and the marginal LF $p(Z|X,\alpha)$ will be

$$\int_{-\infty}^{\infty} p(Z|X,w,\alpha)p(w|X,\alpha)dw = \int_{-\infty}^{\infty} p(Z|X,w)p(w|\alpha)dw \qquad (29)$$

which, using the Laplace approximation method, is equivalent to

$$p(Z|X,w^*)p(w^*|\alpha)(2\pi)^{N/2}(det\Sigma)^{1/2} \qquad (30)$$

The Σ in (30) is the covariance matrix of the Gaussian approximation. Using the approximation method, the vector of α and w will be estimated. Surprisingly enough, the value of α for most weights go to infinity which will result in minimizing w to zero. Therefore, this process will yield

a much sparser model. The points in the training set for which the corresponding w is non-zero are called the relevance vectors.

3.2.4 Decision Trees

Decision trees (DT) are also a powerful family of classifiers. A Decision tree is a collection of rules which are configured as a tree. The process of creating the Decision tree starts with picking the variables one by one and determining the criteria for splitting them. Each node in the tree represents a feature in the data set which can take either a continuous or a categorical value. In order to clarify this method, the following definitions are needed;

Definition 1: Tree Root: The first feature that is chosen and is placed on top of the tree is known as root.

Definition 2: Tree Leaves: The class labels which will be placed at the bottom of the tree are known as leaves.

Definition 3: Tree Branches: The conjunction of attributes which will lead to the leaves (Classes).

Definition 4: Recursive Partitioning: The process of splitting the data set into subsets based on the value of one attribute and repeating this process on each resulted subset.

DT aims to classify the points in the data set by sorting them down from the root node to the leaf node. The process of choosing the attributes is to get the nodes with highest purity. There are several indexes to measure the purity in a node such as: Gain-Ratio, Information-Gain, Gini-Index, and Accuracy. One of the most popular indexes to quantify the level of purity in each node is the node's Entropy. In a multi-class classification situation the Entropy is defined as:

$$H(S) = \sum_{c \in C} -P_c Log_2 P_c. \quad (31)$$

C represents the set of classes in the data set and *Pc* represents the proportion of points of class c in subset S. Information gain is the reduction in Entropy:

$$Gain(S,A) = H(S) - \sum_{v \in values(A)} \frac{|Sv|}{|S|} H(Sv) \quad (32)$$

Where *values* (A) indicates the set of all possible values that the attribute *A* can take. In addition, *Sv* in (32) represents the subset of *S* for which attribute *A* contains value *v*. (Prabhu et al. 2007)

3.2.5 Shrinkage Regression

An alternative approach to least square method (and ridge regression) towards estimating a linear model's coefficients is lasso (Least Absolute Shrinkage and Selection Operator). The objective in lasso is to minimize the residual sum of square subject to the summation of the absolute values of coefficients to be less than a constant.

$$Argmin\left\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_j \beta_j x_{ij} \quad)^2\right\}$$
$$\text{Subject to} \quad (33)$$
$$\sum_j |\beta_j| \leq \lambda$$

One of the most important characteristics associated to lasso is that it enforces some coefficients to be exactly equal to zero and hence it results in a simpler model. However,

by choosing a significantly large value for , this property will be nullified (and lasso regression will be the regular least square model). Therefore, an appropriate choice of λ is quite critical. Because of this important attribute, the variable selection and modeling phases take place simultaneously. This idea can be considered as a major improvement over ridge regression where some coefficients will tend to zero but not exactly zero (See 34).

$$Argmin\left\{\sum_{i=1}^{N}(y_i-\beta_0-\sum_j \beta_j x_{ij}\ \ )^2\right\} \quad \text{Subject to} \quad \sum_j {\beta_j}^2 \le \lambda \tag{34}$$

Another major advantage of lasso is its interpretability. As opposed some more complex nonlinear models such as neural networks, lasso will result in an interpretable model which is very important especially in clinical studies. For a detailed study over lasso see Tibshirani's original paper [(Tibshirani,1996)])

3.3 Model Evaluation

There are several methods to evaluate a classification method. A confusion matrix is a tabulated presentation of correctly or incorrectly classified points in the data set. The definition of the cell values in the confusion matrix is presented below:

- True positives (TP): number of positive examples that were predicted correctly
- False positives (FP): number of positive examples that were predicted incorrectly
- True negatives (TN): number of negative examples that were predicted correctly
- False negatives (FN): number of negative examples that were predicted incorrectly.

The measures that were considered to pick the best model are:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{33}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{34}$$

$$Specificity = \frac{TN}{TN+FP} \tag{35}$$

$$Precision+ = \frac{TP}{TP+FN} \tag{36}$$

$$Precision- = \frac{TN}{TN+FP} \tag{37}$$

In Figure 5, the model evaluation process for classification problems is displayed.

In prediction models, the method's prediction accuracy is evaluated based on RMSE (Root Mean Squared Error); $\sqrt{mean[(ActualValue - PredictedValue)^2]}$ and MAE (Mean Absolute Error); mean ($|ActualValue - PredictedValue|$).

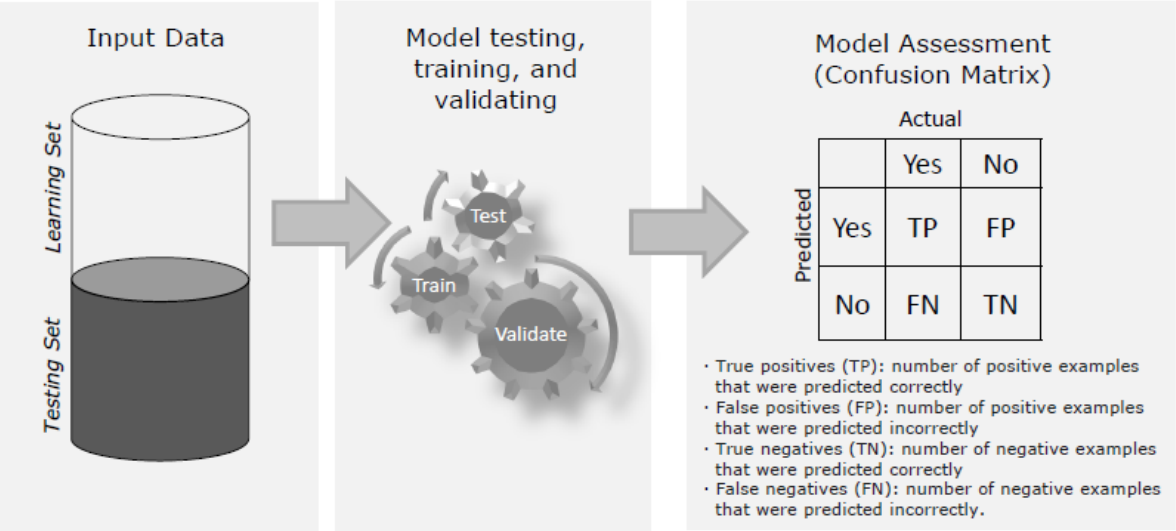


Figure 5. Evaluation Process

4 DOSE PREDICTION MODELS FOR PATIENTS OF SPECIFIC ETHNICITIES

4.1 Problem Definition

As mentioned in Chapter 2, several dose prediction models for estimating the initial dose of Warfarin are proposed in the literature. The key problem at hand is that one does not know how well these commonly used dose prediction equations perform in patients of African descent. This is because the data sets used to build the models contained only 10% of patients of African descent.

There are key differences between African American and non-African American patients, such as differential distribution and effects of influential genetic traits that effect warfarin dose. The ultimate goal of this line of work is to generate race/ethnicity specific dose prediction equations that out perform traditional equations which ultimately result in better dose predictions for these patients. Therefore, the objective of this Chapter is to explore different modeling approaches in a cohort of warfarin treated African Americans and compare their performance of predicting a stable Warfarin dose. The ultimate goal is to generate warfarin dose prediction models that account for ethnic differences and ultimately outperform (more accurate) existing equations.

4.2 Data Set Description

The Cavallari group has developed a database of over 326 warfarin-treated African Americans containing each patient's observed stable dose and a rich set of covariates. The data were collected over a ten-year period at the University of Illinois Hospital and Health Science System, Chicago, IL. For each patient, several features were measured that are presented in Table 5.

Table 5. List of features in the dataset.

Variable Number	*Variable Name/Type*
1	Age (continuous value)
2	Sex (Male=1, Female=0)
3	Height (continuous value)
4	Weight (continuous value)
5	Amiodarone (AMIO)(Yes=1,No=0)
6	Smoking (Yes=1,No=0)
7	DVT/PE (Yes=1,No=0)
8	Diabetes (Yes=1,No=0)
9	Cancer (Yes=1,No=0)
10	Hypertension[HTN] (Yes=1,No=0)
11	Stable INR (continues values)
12	Stable Warfarin Dosage

4.3 Methods

Three machine learning techniques were used for the prediction of warfarin dosing of African American patients. The techniques include Artificial Neural Networks, Support Vector Regression models, and Linear Multivariate Regression. The modeling details and comparisons are provided below.

One of the prediction methods that was used in this Chapter is Artificial Neural Networks. The concept of brain-style computation was originally rooted over 60 years ago in the research of McCulloch and Pitts (1943)(McCulloch and Pitts 1943) and furthermore in (1949)(Hebb 1949).. The basic structure of the neural network is presented in Figure 6.

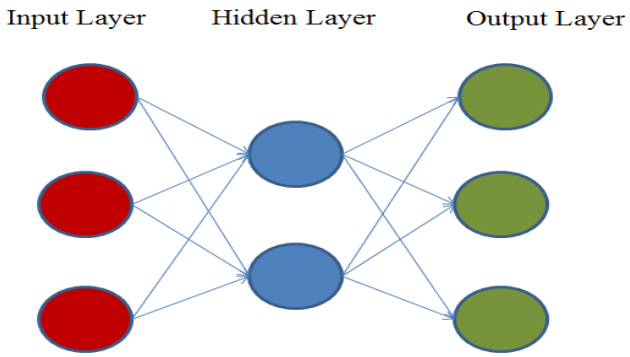


Figure 6. Single hidden layer feed-forward neural network

The fundamental components of neural networks are known as neurons. Each processing unit is considered by an activity level which represents a node's level of polarization; an output which represents a node's firing rate; the input and output connections and a bias value (Fullér 2000). There are weights associated to each connection which determine the strength of the effect of the input on the unit's activation level (See Figure 7).

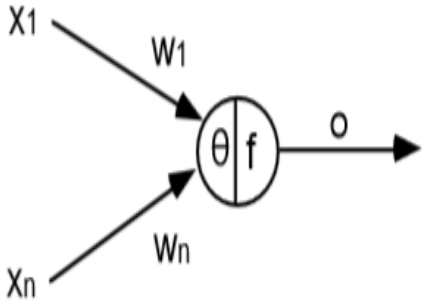


Figure 7. Functionality of the neural network

In Figure 7, W_i denotes the weight from one input node to its associated output node. Therefore we can describe our system as $f(\sum_{i=1}^{n} w_j x_j)$. The challenge is to find the appropriate W_i's.

The function f is called the activation function. The weights are adjusted in the process of perceptron learning method (for details see (Yegnanarayana 1994)).

The second method that was applied in this Chapter was Support Vector Regression. The theory behind Support Vector Machines was discussed in Chapter 3.

By changing the loss function in (20), which is called the ξ -insensitive loss function, SVMs can also be used for regression (Karatzoglou, Meyer, and Hornik 2006). This loss function will only accept the error terms which are greater than a predefined threshold (ξ).

$$\begin{aligned} &\text{minimize } t(\mathbf{w}, \xi) = \frac{1}{2}\ \|w\|^2 + \frac{C}{m}\sum_{i=1}^{m}(\xi_i + \ \xi_i^*) \\ &\text{subject to } (\langle \Phi(x_i), \mathbf{w}\rangle + b) - y_i \leq\ \epsilon - \ \xi_i \\ &y_i - (\langle \Phi(x_i), \mathbf{w}\rangle + b) \leq\ \epsilon - \ \xi_i^* \\ &\xi_i^*\ \geq 0 \qquad (i = 1, \dots, m) \end{aligned} \tag{38}$$

The third model that was used for modeling was Linear Multivariate Regression. When performing the modeling, 80% of the data set was selected randomly for our training set and the remaining 20% was used for validating our models. Before starting the modeling, the data from patients with a stable INR between 2 and 3 were selected, and then the data cleansing was performed. All the missing values were imputed using K-nearest neighbor method. To avoid any collinearity between Height and Weight, BSA (Body Surface Area) was used. In the process of modeling, patients whose stable dose was between 25 mg/wk and 65 mg/wk were selected since 90% of our patients had the stable dose within this range (Figure 8-9).

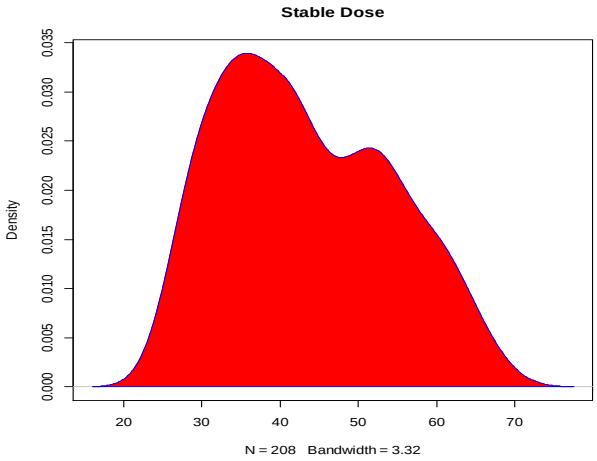


Figure 8. Density graph of the stable dose after selecting the desired dose range

By selecting the aforementioned range, our data became significantly more amenable (Figure 9).

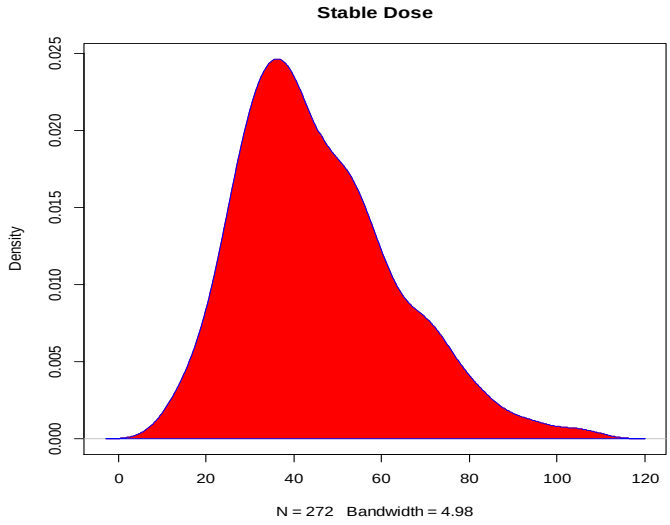


Figure 9. Density graph of the stable dose

The first model was built upon designing an artificial neural network with 2 hidden layers. The back propagation for calibrating the weights was used. (See Figure 10). The minimum error at 1% and the epoch at 5,000,000 were set and the activation function was tanh(s).

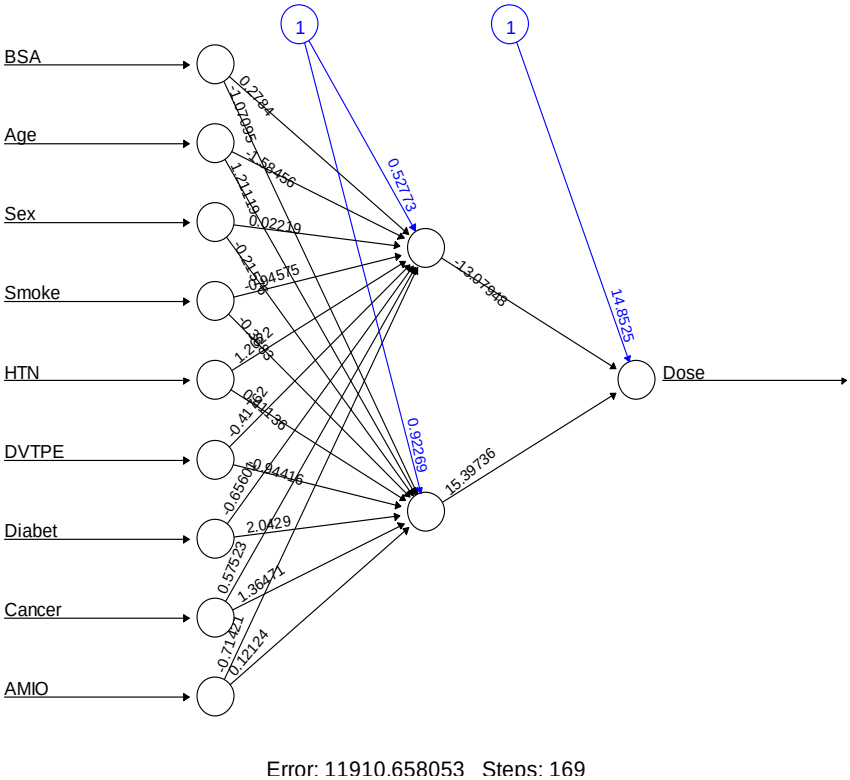


Figure 10. Artificial Neural Network

After testing numerous different numbers for hidden layers therefore, the model with 2 hidden layers turned out to have the minimum validation error.

In Table 6, RMSE (root mean squared error) and MAE (mean absolute error) are shown.

Table 6. Prediction performance of Neural Network

Method	*RMSE*	*MAE*
Neural Network	21.6	20.2

The next model that was tested is Support Vector Regression. As mentioned in Chapter 3, this method works by identifying the support vectors in the training set and then using those vectors,

the model's coefficients will be calibrated. The Gaussian RBF kernel $k(x,x') = exp(-\sigma |x - x'|^2)$ was used. The result of this model is presented in Table 7.

Table 7. Prediction performance of Support Vector Machine

Method	*RMSE*	*MAE*
Support Vector Regression	17.3	16.1

The last model that was tested was Multiple Linear Regression. The model coefficients were calibrated using the Least Square technique. To support the assumption of normality, the logarithmic transformation was implemented on the stable doses. In Table 8, the estimations for model coefficients along with their P-values are presented.

Table 8. Model Coefficients for the regression model

Variable Name	*Corresponding Coefficient in the model*	*P-Values*
(Intercept)	3.755	~ 0
BSA	0.124	0.038
Age	-0.004	0
Sex	0.013	0.736
Smoke	0.066	0.138
HTN	-0.068	0.096
DVT/PE	-0.001	0.98
Diabetes	0.051	0.164
Cancer	-0.032	0.79
Amiodarone	-0.187	0.043

Since some of the variables are not significant at the 0.05 level of confidence, the step-wise method was used to select the best subset for modeling. The resulting model coefficients are presented in Table 9.

Table 9. Model Coefficients for after implementing the stepwise model selection

Variable Name	*Corresponding Coefficient in the model*	*P-Values*
(Intercept)	3.778	~ 0
BSA	0.117	0.035
Age	-0.004	0
Smoke	0.066	0.031
HTN	-0.068	0.089
Diabetes	0.051	0.061
Amiodarone	-0.192	0.034

The model assumptions were examined through investigating the residuals, the assumption of independency and normality of residuals were confirmed (See Figure 11). In parts a and b in Figure 11. both the raw values of residuals and the squared root of the standardized values of residuals versus the fitted values are graphed. The stability of the variance of random error is quite evident. In part c, using the Q-Q plot, the validity of the assumption of normality of random errors is illustrated. In part d, the robustness of the model in presence of leverage point using Cook's Distance is examined.

The model performance was quite satisfactory. In Table 10. the model performance is presented.

Table 10. Results for multivariable Regression

Method	*RMSE*	*MAE*
Multivariable Regression	14.51	12.2

The three new models were examined against the IWPC and Gage models. In Figure 12., all six models are compared. While Artificial Neural Network and Support Vector Machine methods perform poorly compared to the available techniques (IWPC and GAGE), the multi regression model outperforms the IWPC and Gage models.

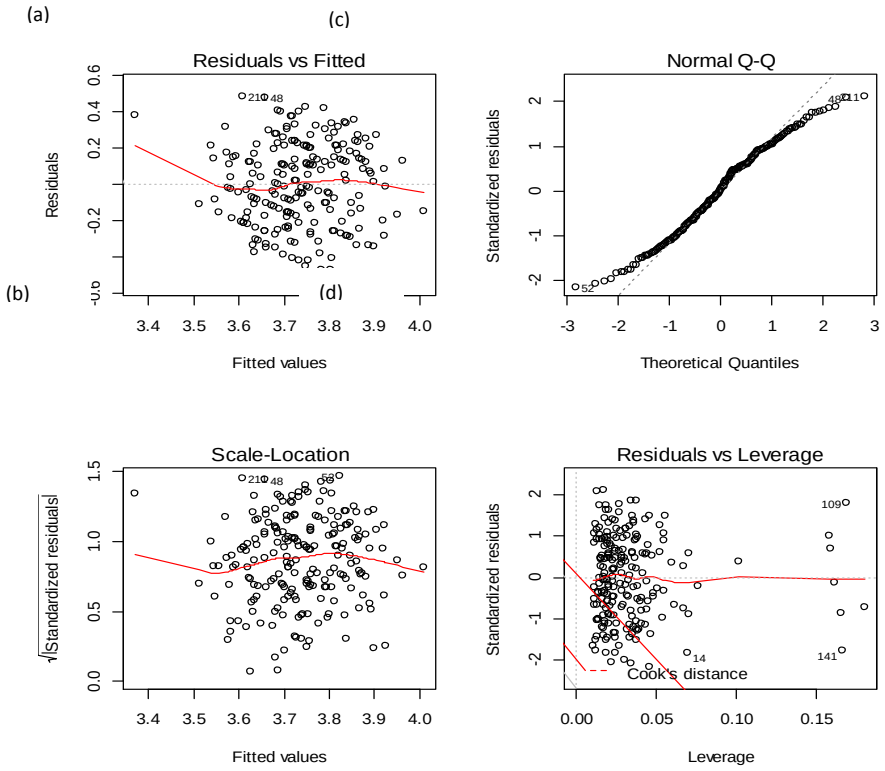


Figure 11. Checking the model assumptions

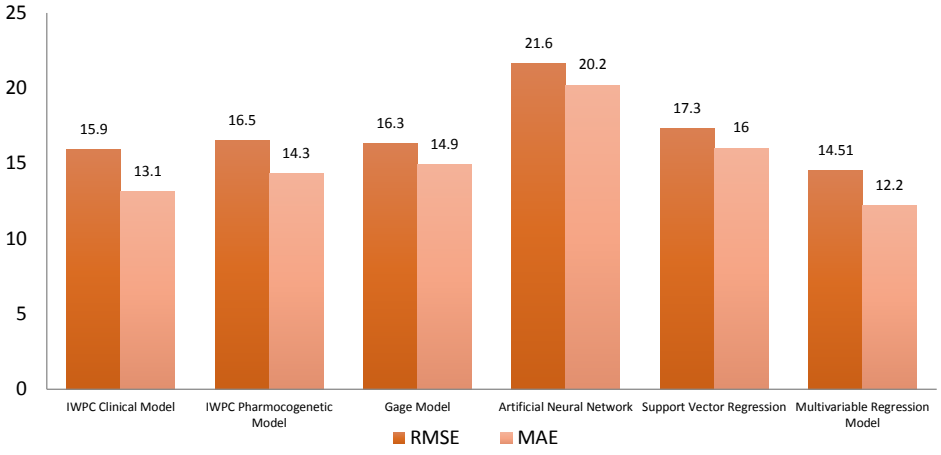


Figure 12. Comparing the performance of different models.

4.4 Results

In this Chapter, machine learning techniques were used to develop a new model for the prediction of the optimal warfarin dosing for African American patients. To develop the new model, three different machine learning techniques were specifically examined, namely support vector machines, neural networks, and multivariable regression. It was shown that the new model has a better prediction accuracy than the existing popular dosing algorithms. Therefore, the new model would be safer in determining the optimal warfarin dosing for African American patients.

5 NEW METHODOLOGY TOWARDS MULTI-ETHNIC PREDICTION MODELING

5.1 Problem Definition

In this Chapter, a novel method for warfarin dosing is developed. In this proposed methodology, the patients are primarily categorized into two classes. Class 1 contains patients who need doses of > 30 mg/wk. Class 2 contains those patients who need doses of ≤ 30 mg/wk. In the next stage, dose prediction takes place for each class individually. This method was compared with the most popular dose prediction models in the literature along with the method proposed in Chapter 4 and its outperformance in terms of prediction accuracy was proved.

5.2 Data Set Description

The data set that was used in this Chapter is the IWPC data set which is a well-known multi-ethnic warfarin data set. This data set is one of the most widely used and publically available warfarin data sets, as evident by its citations in the literature (SM Oztaner et al. 2014). The missing values in the data set were imputed using the K-nearest Neighbor (KNN) method with k=1 (Hastie et al. 2009). The variables whose percentage of missing values were more than 50% were not involved in the model. The variables used in the modeling were only the clinical and demographic variables which are presented in Table 11. In order to develop a robust prediction model, the *CRISP-DM* methodology was followed in order to build our models (Wirth and Hipp 2000). 50% of the data points were randomly selected to comprise the training set (*derivation cohort*) and the remaining 50% were assigned to the testing set (*validation cohort*). The data in the test set were used for the models' performance in dealing with unseen data points.

Table 11. IWPC data set description[1]

Continuous Variables		
Target International Normalized Ratio	Mean	2.5
	Std. Deviation	0.1
	Minimum	1.8
	Maximum	3.5
Body Surface Area	Mean	1.94
	Std. Deviation	0.3
	Minimum	1.2
	Maximum	3.4

Categorical Variables							
Gender	Values	Frequency	Percent	Amiodarone	Values	Frequency	Percent
	0	1822	43.00%		0	3984	94.03%
	1	2415	57.00%		1	253	5.97%
Race	Values	Frequency	Percent	Carbamazepine	Values	Frequency	Percent
	1	2663	62.85%		0	4195	99.01%
	2	656	15.48%		1	42	0.99%
	3	918	21.67%	Phenytoin	Values	Frequency	Percent
Deep Vein Thrombosis and Pulmonary Embolism (DVT/PE)	Values	Frequency	Percent		0	4197	99.06%
	0	3846	90.77%		1	40	0.94%
	1	391	9.23%	Rifampin	Values	Frequency	Percent
Diabetes	Values	Frequency	Percent		0	4231	99.86%
	0	3500	82.61%		1	6	0.14%
	1	737	17.39%	Sulfonamide Antibiotics	Values	Frequency	Percent
Congestive Heart Failure	Values	Frequency	Percent		0	4214	99.46%
	0	3492	82.42%		1	23	0.54%
	1	745	17.58%	Macrolide Antibiotics	Values	Frequency	Percent
Valve Replacement	Values	Frequency	Percent		0	4225	99.72%
	0	3243	76.54%		1	12	0.28%
	1	994	23.46%	Anti-fungal Azoles	Values	Frequency	Percent
Aspirin	Values	Frequency	Percent		0	4210	99.36%
	0	3199	75.50%		1	27	0.64%
	1	1038	24.50%	Smoker	Values	Frequency	Percent
Simvastatin	Values	Frequency	Percent		0	3733	88.10%
	0	3608	85.15%		1	504	11.90%
	1	629	14.85%	Enzyme	Values	Frequency	Percent
Atorvastatin	Values	Frequency	Percent		0	4150	97.95%
	0	3810	89.92%		1	87	2.05%
	1	427	10.08%		Values	Frequency	Percent

Fluvastatin	Values	Frequency	Percent	Age	1	9	0.21%
	0	4220	99.60%		2	94	2.22%
	1	17	0.40%		3	189	4.46%
Lovastatin	Values	Frequency	Percent		4	444	10.48%
	0	4153	98.02%		5	806	19.02%
	1	84	1.98%		6	1023	24.14%
Pravastatin	Values	Frequency	Percent		7	1133	26.74%
	0	4121	97.26%		8	511	12.06%
	1	116	2.74%		9	28	0.66%
Rosuvastatin	Values	Frequency	Percent				
	0	4208	99.32%				
	1	29	0.68%				

[1] The variable Gender takes 0 for Female patients and 1 for Male patients. The variable Race takes 1 for White, 2 for African-American, and 3 for Asian patients. Consumption of any drug or possession of any disease is indicated with 1 and 0 otherwise. The variable Age is coded in Age-decade format (1 represents 10-19 years old, 2 represents 20-29 etc.).

5.3 Methods

The dose prediction method that is proposed in this Chapter contains two phases. In the first phase, the data points in the test will be assigned to two classes. The first class contains patients who require doses of > 30 *mg/wk (High-Required-Dose (HRD))* and the second class contains the patients who need doses of ≤ 30 *mg/wk (Low-Required-Dose (LRD))*.

The selected cut-off point (30 mg/wk) was derived from the validation process in which the data in the Learning set was divided randomly into Training and Validation sets. Different values (15, 20, 30, 35, 40, 45, and 50 mg/wk) were selected and examined to identify the threshold that maximized the classification accuracy. The optimal threshold, 30 mg/wk, from the validation process, was applied in the modeling procedure.

This phase is performed using a classification technique which incorporates Relevance Vector Machines *(RVM)*. In the second phase, the optimal dose for each patient will be predicted by two regression clinical models which are customized for each class of patients (See Figure 13).

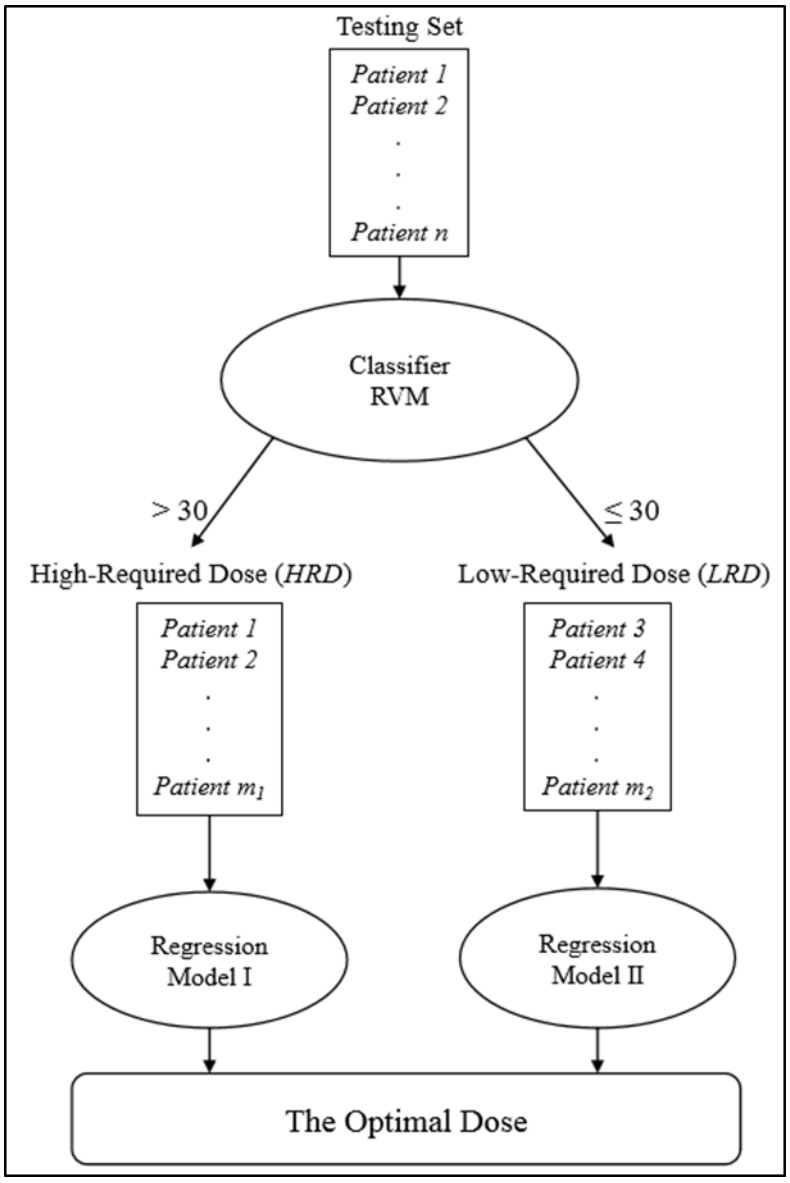


Figure 13. The proposed two stage modeling

The classification and the regression models are created using the data points in the learning set. Each data point in the learning set got labeled as 0 (LRD patients) or 1 (HRD patients)

depending on the value of the therapeutic dose. Now by considering the generated labels as the new response variable, the nature of the problem transforms to classification. A classification model (RVM) is trained using the data in the learning set. Additionally, the points in the learning set are assigned to two groups according to their label and a regression model for each group is generated.

As it is shown in Figure 13., when the points are labeled as 1 or 0 by the classification model, they will be entered into the second phase which is the prediction phase. Using the RVM model, the data points in the testing set were classified to HRD and LRD classes and two regression models were developed for each class separately. The models are presented below.

Model for HRD Class (Model I):

Predicted Dose =Exp(2.85332 -0.07370 X Race -0.06513 X Age +0.10246 X DVT/PE + 0.05766 X Diabetes + 0.03742 X VR - 0.08763 X Lovastatin-0.12542X Amiodarone + 0.13207 X TargetINR + 0.12403X Enzyme + 0.34487 X BSA)

Model for HRD Class (Model II):

Predicted Dose = Exp(3.44056 - 0.03649XRace - 0.04820 X Age+ 0.05059 X DVT/PE - 0.03060 X Aspirin - 0.06150 X Amiodarone -0.20356 X AfungalAzoles +0.05744 X Smoker + 0.10923 X Enzyme + 0.24601 X BSA)

In the cross-validation phase, the trained models were applied on the data points in the testing set. The classification results for the two models are presented in Table 12.

Table 12. Classification results for RVM

Method	*Accuracy*	*Sensitivity*	*Specificity*	*Precision +*	*Precision -*
RVM	0.66	0.63	0.73	0.81	0.5

After classifying the points in the test set, 49% of the points were assigned to HRD class and 51% to LRD class. The proposed method's prediction accuracy got evaluated based on RMSE and MAE. The prediction results are presented in Table 13.

Table 13. Comparing the prediction accuracy of the proposed methodology with IWPC Cl and Gage Cl models

Methods	*RMSE*	*MAE*
The Proposed Methodology	11.6	8.4
IWPC Cl	13.8	9.1
Gage Cl	12.2	9.9
Sharabiani	18.1	12.7
Fixed-dose approach	18.7	12.3

As it is evident in Table 13., the proposed methodology for predicting the warfarin dose outperforms the IWPC cl model for 16% in terms of RMSE and 8% in terms of MAE. It also outperforms the Gage Cl model for 5% in terms of RMSE and 16% in terms of MAE. The proposed method was also compared with fixed-dose approach (35 mg/wk) and the prediction model proposed in (Sharabiani et al. 2013). The method resulted in significantly lower RMSE and MAE than both models (37%, 31% less than the fixed dose approach and 35%, 33% less than Sharabiani's method in terms of RMSE and MAE respectively).

The major limitations for comparing developed models with one another in warfarin dosing literature can be viewed from three perspectives. First, the variables that were involved in the reference model should be available in the data set that is used for developing new models. For example, not all variables that were applied in (Grossi et al. 2013) are available in IWPC data set. Therefore, it is not possible to measure the performance of the proposed model developed by (Grossi et al. 2013) on IWPC data set. The second limiting factor is the use of genetic variables in the model such as (Grossi et al. 2013) and (SM Oztaner et al. 2014). As discussed in the

Introduction, there is a serious hesitation towards applying such models in practice. Specially, applying these models require the data of quite costly variables, which are not available to most institutes around the world. Therefore, when developing clinical models, their performance must be compared to the existing clinical models. Thirdly, some models are developed targeting specific cohorts of patients (patients of different ethnicities, age groups, etc). Therefore, comparing models which target general public with these special models will result in a biased conclusion. For, example, in IWPC data set, African-American patients constitute about 16% of the whole population. Therefore, applying the models which are developed for African-American patients (such as (Sharabiani et al. 2013), (Cosgun, Limdi, and Duarte 2011)) will result in an expected underperformance than general models.

5.4 Results

The significance of prescribing an accurate initial dose for warfarin is undeniably important. Therefore several mathematical models have been proposed in order to predict the optimal dose for each patient. In this Chapter, a novel methodology for predicting the initial dose is proposed, which only relies on patients' clinical and demographic data. In this method, the patients are assigned to either one of two classes in the first phase. The patients who require doses of > 30 mg/wk belong to the first class and the the patients who need doses of ≤ 30 mg/wk belong to the second class. This phase is implemented using (RVM). Then, each patient's dose will be determined using one of the two regression clinical models which are customized for each class. The proposed methodology outperformed two popular existing clinical prediction models (IWPC Cl, Gage Cl , and Sharabiani models) in addition to fixed-dose approach) in terms of prediction

accuracy. The methodology which is proposed in this work can be extended by investigating the best classifiers for patients of specific ethnicities.

6 COMPANION CLASSIFICATION MODEL TO PREDICTION MODELS

6.1 Problem Definition

Considering the predominant uncertainty in using the Pharmacogenetic models in practice, in this Chapter, the concentration is aligned towards one of the most popular and generally used clinical models; the IWPC Cl model. Although, it has been reported that this model performs the best for patients with therapeutic range of less than or equal to 21 to more than or equal to 49 mg/wk, since the therapeutic dose is not evident in early stages of the treatment, a companion classification model is proposed to help the clinicians to identify the patients whom are compatible with this dosing model.

Using a sample of 4,237 patients, a companion classification model to one of the most popular dosing algorithms (IWPC clinical model) is proposed, which identifies the appropriate cohort of patients for applying this model. The proposed model will function as a clinical decision support system which assists clinicians in dosing. A classification model using Support Vector Machines, with a polynomial kernel function is developed to determine if applying the dose prediction model is appropriate for a given patient. The IWPC clinical model will only be used if the patient is classified as “Safe for the model” by the classification model.

6.2 Data Set description

The data set that was used in this Chapter is the multi-ethnicity (IWPC) data set which is comprehensively described is section 5.2.

6.3 Methods

As mentioned in the previous section, the prediction model which we applied in the system development process is the IWPC clinical model. The variables, their corresponding coefficients, and their units are presented in Table 1. For all patients in the data set, the dose prediction value using the IWPC model was generated. If the difference between the prediction value and the therapeutic dose is more than 15 mg/week (|Therapeutic Dose – IWPC Clinical | > 15 mg week), the patient will be labeled as 'High-risk' patient otherwise he/she will be labeled as 'Safe for the model'. The objective is to develop a classification model to detect the High-risk patients, See Figure 14.

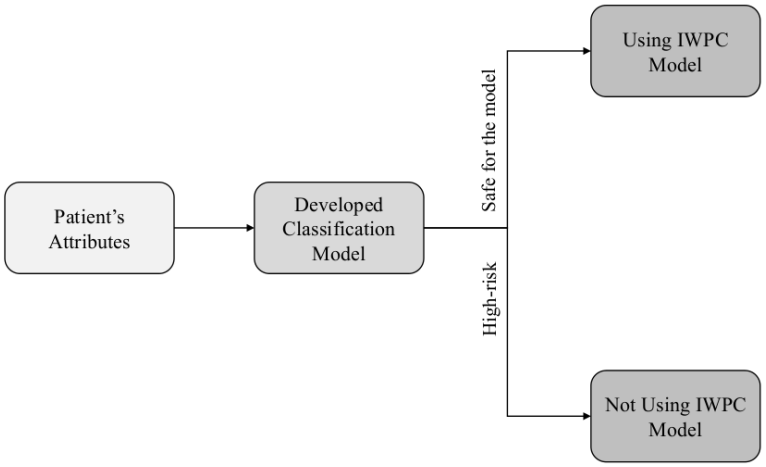


Figure 14. The proposed methodology for using the IWPC clinical model

For establishing a reliable model and testing its performance against the out-of-sample data points, the data set was assigned to Learning (50%) and Testing (50%) sets. The choice of 15 mg/wk as a threshold was yielded through the validation phase where after trying different

thresholds, on the data points in the training set, the threshold 15 mg/wk resulted in the maximum classification accuracy. Several classification models were examined using K-fold cross validation with k=10 on the learning set. The sensitivity, specificity, and accuracy were used in comparing the classification models. The sensitivity and specificity were characterized by the balanced accuracy (Hastie et al. 2009); Balanced Accuracy = (sensitivity + specificity)/2. After developing the model, it can be applied to determine if the patient is compatible with IWPC Cl model or not, and use the dosing model only if he/she is classified as 'Safe for the model'. After labeling the patients using the classification model, if the patient is classified as "Safe for the model", the clinician has a choice to apply IWPC Cl.

Several classification methods were examined using the test data set and were compared based on their Accuracy and Balanced accuracy. The classification methods are Decision Trees (DT) with several parameter settings for minimum size for leaves, depth of the tree and minimum branch size, logistic regression, Naïve Bayes, Artificial Neural networks, SVM with linear kernel, SVM with Gaussian kernel, and SVM with a polynomial kernel. The classification results are presented in Table 14.

Table 14. Comparing the performance of different classification models on the Test set

Model Name	**Accuracy**	**Sensitivity**	**Specificity**	**Balanced Accuracy**
DT(2,20,4)	*51.9%*	*48.6%*	*55.2%*	*51.9%*
DT(4,10,5)	*50.7%*	*45.3%*	*56.0%*	*50.7%*
DT(4,20,7)	*51.7%*	*45.0%*	*58.4%*	*51.7%*
DT(10,20,20)	*52.4%*	*45.7%*	*59.1%*	*52.4%*
Naïve Bayes	*56.9%*	*42.9%*	*70.9%*	*56.9%*
Neural Nets	*50.0%*	*100.0%*	*0.0%*	*50.0%*
SVM(Linear)	*50.0%*	*100.0%*	*0.0%*	*50.0%*
SVM(Sigmoid)	*50.0%*	*0.0%*	*100.0%*	*50.0%*
SVM(Polynomial)	*59.0%*	*61.2%*	*55.1%*	*58.2%*

The SVM with polynomial kernel was selected as the best model as it had the highest accuracy, 59.0% and performed acceptably in both Specificity and Sensitivity, thus having the highest balanced accuracy, 58.2%.

The SVM with a polynomial Kernel was applied to the patients in the test set to classify patients as either ‘Safe for model’ or ’High-risk’. Once the patients were classified as ’High-risk’, they were eliminated from the test set. For the remaining patients (Shrunken test set), the IWPC clinical model was used to predict the initial dose.

In Table 15., the prediction accuracy of the IWPC clinical model was compared between the original test set and the shrunken test set based on RMSE and MAE.

Table 15. Comparing the prediction accuracy of the IWPC CL model on original and shrunken test sets

Original Test Set	
Number of data points	2119
Error (RMSE)	23.0
Error (MAE)	16.6
Shrunken Test Set	
Number of data points	1271
Error (RMSE)	17.8
Error (MAE)	14.03

After applying the proposed classification of "High-risk" or "Safe for the model", the model’s prediction error improved from 23.0 to 17.8 (5.2 absolute, 23% relative) for RMSE and similarly for the MAE method, improved from 16.6 to 14.0 (2.6 absolute, 15% relative). In the shrunken test set, 40% of the patients were labeled as “High-risk”. The proportion of patients that would be considered “High-risk” in any new set of patients cannot be determined prospectively and this is something that would need to be watched if this system were to be used on a new cohort of patients.

Clinically the knowledge of whether the patient was classified as "safe for model" or "high-risk" can be used to help decide on the use of clinical pharmacists, which are often a limited resource in healthcare settings. The "high-risk" patients may be the ones that a limited number of pharmacists are assigned to help with anticoagulation. Most patients being started on warfarin do not require continued admission until the INR is stable due to the use of low molecular weight heparin (LMWH). Knowledge of stratification of patients as “high-risk” for a poor dose could potentially be used to help decide the delay from discharge to the first visit for ambulatory monitoring of INR.

6.4 Results

In this Chapter, a novel methodology for identifying patients appropriate for the IWPC clinical model is proposed, functioning as a companion to IWPC clinical model. The multi-ethnicity (IWPC) data set was used to develop, examine, and ultimately select the best classification model to identify the ‘Safe for model’ patients; the patients for whom the difference between the prediction by IWPC clinical model and their therapeutic dose is less than 15 mg/wk, and ‘High-risk’ patients; the patients for whom the difference between the prediction by IWPC clinical model and their therapeutic dose is more than 15 mg/week. A support vector machine with a polynomial kernel function was found to be the best performing classification model. The patients classified as ‘High-risk’ were eliminated from the test set. For the remaining patients, the IWPC clinical model is used for predicting the initial dose. The performance of the approach was tested using RMSE and MAE comparisons on the original test set and the shrunken test set. The RMSE value improved by 23% and the MAE value by 15%. The application of the proposed methodology can be extended to the prediction models which are developed for specific ethnic groups and children.

The ability of this system to predict which patients may be appropriate or inappropriate for the IWPC model may have many clinical applications. This system could be used to help decide on the use of clinical pharmacists in assistance with warfarin dosing. The "high-risk" patients may be chosen as requiring pharmacy assistance in a situation with limited clinical pharmacists. In addition, stratification of patients as “High-risk” for a poor dose could potentially be used to help decide the delay from discharge to the first visit for ambulatory monitoring of INR.

7 A NEW APPROACH TOWARDS MINIMIZING THE RISK OF MIS-DOSING FOR POPULAR WARFARIN INITIATION DOSES PRESCRIBED BY THE PHYSICIANS

7.1 Problem Definition

In clinical practices, in order to determine the initiation dose of Warfarin, the clinicians face several alternatives. They can use the loading method, the dose prediction models that are proposed in the literature, or rely solely on their knowledge and experience. In this Chapter two objectives were pursued. The first objective is to minimize the risk of mis-dosing when the clinicians prescribe the initial dose based on their known judgment. The risk of mis-dosing is defined as the significant percentage difference between the initial dose and the therapeutic dose. Since the definition of a "significant percentage difference" is subject to individual interpretation, the proposed procedure is examined based on different scenarios. The proposed model estimates the amount of percentage error which can be either positive (in case of overdose) or negative (in case of under dose). Once the amount of percentage error is estimated, the initial dose can be modified accordingly. It is shown that by using the proposed method, the risk of mis-dosing decreases significantly.

7.2 Data set Description

The dataset which was used for this project contain the data of 150 warfarin-treated patients in the University of Illinois at Chicago Hospital who had reached the therapeutic dose in

their course of treatment. Numerous variables about these patients were measured. The variables in the data set and their frequencies are presented in Tables 16-17.

Table 16. Categorical variables in the data set

Variable Name	*Values*	*Code*	*Frequency*	*Percentage*	*Variable Name*	*Values*	*Code*	*Frequency*	*Percentage*
Race	African American	1	79	53%	Smoking	Current Smoker	1	13	9%
	Hispanic	2	34	23%		Never Smoker	2	107	71%
	White	3	18	12%		Ex-Smoker	3	30	20%
	Asian	4	4	3%	EtOH	Yes	1	24	16%
	Other	5	15	10%		No	2	119	79%
Gender	Male	1	67	45%		Missing	NA	7	5%
	Female	2	83	55%	Illicit	Yes	1	6	4%
Liver Disease	Yes	1	3	2%		No	2	144	96%
	No	2	125	83%	Hypertension	Yes	1	86	57%
	Missing	NA	22	15%		No	2	64	43%
warfarin indication (WI)	A.fib	1	25	17%	Angina	Yes	1	1	1%
	DVT	2	53	35%		No	2	149	99%
	PE	3	34	23%	Myocardial Infarction	Yes	1	3	2%
	TKA/THA	4	13	9%		No	2	147	98%
	MVR	5	1	1%	Percutaneous Coronary Intervention (PCI)	Yes	1	6	4%
	CVA	6	4	3%		No	2	144	96%
	Other	7	20	13%	coronary artery bypass graft(CABG)	Yes	1	5	3%
Goal INR	2-3	1	136	91%		No	2	145	97%
	2.5-3.5	2	3	2%	Atrial fibrillation or flutter	Yes	1	11	7%
	1.8-2.5	3	11	7%		No	2	139	93%
Amioadarone	Yes	1	5	3%	Diabetes mellitus(DM)	Yes	1	48	32%
	No	2	144	96%		No	2	102	68%
	Missing	NA	1	1%	Stroke	Yes	1	11	7%

Bactrim	Yes	1	1	1%		No	2	139	93%
	No	2	148	99%	Chronic Renal Insufficiency	Yes	1	15	10%
	Missing	NA	1	1%		No	2	135	90%
Azole	Yes	1	1	1%	Chronic Obstructive Pulmonary Disease (COPD)	Yes	1	7	5%
	No	2	148	99%		No	2	143	95%
	Missing	NA	1	1%	Asthma	Yes	1	18	12%
Which Statin?(ST)	None	0	93	62%		No	2	132	88%
	Simva	1	14	9%	Valvular Heart Disease	Yes	1	1	1%
	Atrova	2	23	15%		No	2	149	99%
	Prava	3	7	5%	Sickle Cell	Yes	1	3	2%
	Lova	4	8	5%		No	2	147	98%
	Rosuva	5	4	3%	Cancer History	Yes	1	12	8%
	Missing	NA	1	1%		No	2	138	92%
Dialysis	Yes	1	8	5%	pulmonary Embolism (PE)	Yes	1	5	3%
	No	2	142	95%		No	2	144	96%
Rheumatoid Arthritis	Yes	1	1	1%		Missing	NA	1	1%
	No	2	149	99%	Dyslipidemia	Yes	1	53	35%
Collagen vascular disease	Yes	1	2	1%		No	2	97	64%
	No	2	148	99%	heart failure (HF)	Yes	1	15	10%
Deep vein thrombosis(DVT)	Yes	1	10	7%		No	2	135	90%
	No	2	140	93%	Peripheral vascular disease (PVD)	Yes	1	7	4%
						No	2	143	95%

Table 17. Continuous variables in the data set

Continuous Variables	***Unit***	***Number of Missing***	***Mean***	***Median***	***Sd***	***Min***	***Max***
Therapeutic Dose (Label)	mg/day	0	5.68	2.87	5.1	0.9	16.8
Initial Dose Prescribed By the Physician(IDP)	mg/day	2	6.12	2.59	5	1	16
Percentage Error		2	0.26	0.7	0.12	-0.84	4.83
Age		0	54.29	17.82	57	18	91
Height(Ht)	cm	0	168.28	10.35	169	142.2	195
Weight(Wt)	kg	0	89.9	31.12	83	40	220
Creatinine Clearance (CrCl)	ml/min	2	64.79	36.32	63.65	3.6	146.5
Albumin	g/dl	17	3.12	0.65	3.2	1.4	4.3
Aspartate Aminotransferase(AST)	u/L	22	33.56	41.04	22	9	379
Alanine Aminotransferase(ALT)	u/L	22	25.88	24.85	19	5	199
Baseline INR		1	1.18	0.14	1.2	1	1.8

In the next section, the set of steps for data preprocessing and data visualization are presented.

7.3 Data Preprocessing & Visualization

In order to measure the impact of initial dose on the trend of prescribed doses the following algorithm was used:

- N = Number of patients in the data set
- ni = Number of prescribed doses for patient i to reach the therapeutic dose
- $D_{pi} = [d_1, d_2, \dots, d_{ni}]$; profile of prescribed doses to patient i
- CID_i =Complexity Index for D_{pi}. ; CID (D_{pi}) [2] $= \frac{\sqrt{\sum_{j=1}^{ni-1}(d_j - d_{j+1})^2}}{ni}$

1. *For patients 1:N compute CID_i store in Profile Complexity Vector(PCV); PCV=[CID_1, CID_2,…, CID_N]*
2. *Perform the following test of Hypothesis:*

$$H_0: \mu_{PCV} = 0$$
$$H_1: \mu_{PCV} > 0$$

3. *Rejecting the null hypothesis indicates that the prescribed doses in patients profile fluctuate significantly.*

In order to perform the hypothesis test which is mentioned in step 3. of the algorithm above, the t-student test was performed. The test results are presented in Table 18.

Table 18. Test of Hypothesis Results

Test Results
t.test(res,mu=0, alternative = c("One. Sided"))
One Sample t-test
t = 17.3833, p-value < 2.2e-16

Based on the p-value in Table 18., it is safe to reject the null hypothesis with 95% level of confidence.

[2] CID stands for complexity-invariant distance which is designed to estimate the fluctuation level in a time series.

In an ideal dosing setting, the initial doses prescribed by the physicians have to be reasonably close to the therapeutic dose. In Figure 15., the correlation between these two variables is presented. The red line on Figure 15. indicates the ideal dosing scenario for each patient.

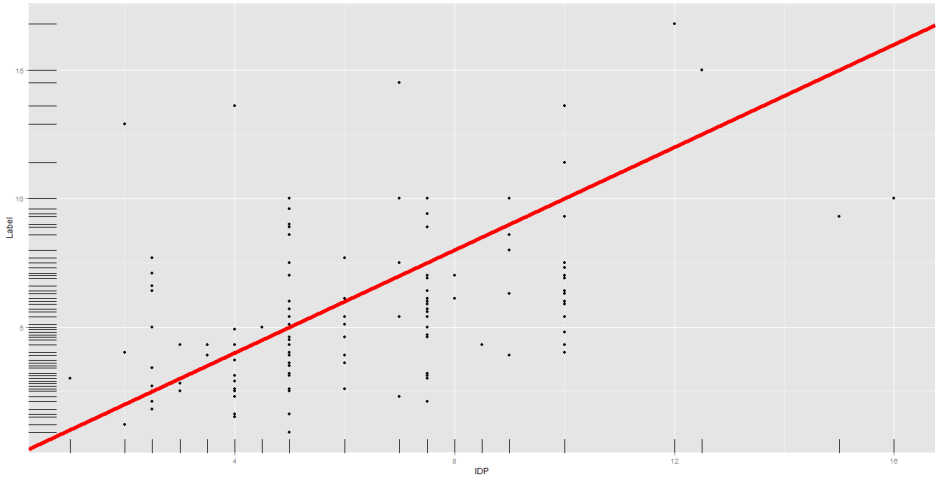
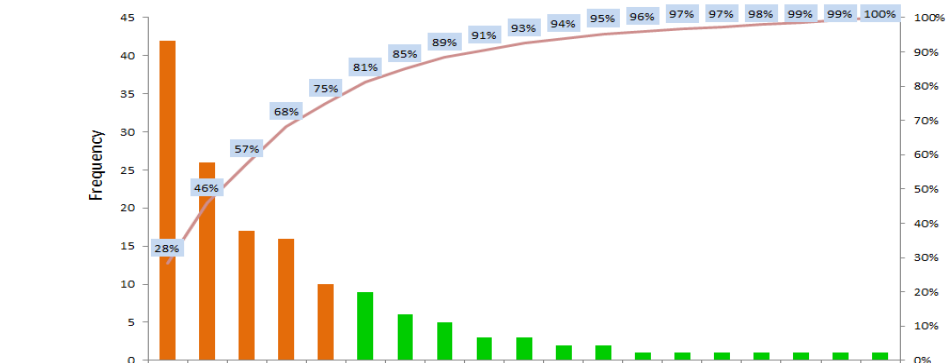


Figure 15. IDP Vs. Therapeutic Dose

It is evident that most physicians tend to prescribe doses at popular discrete dose values. Hence a pareto chart for measuring this tendency is created in Figure 16.



IDP	5	7.5	4	10	2.5	6	7	9	2	3	3.5	8	1	4.5	8.5	12	12.5	15	16
Frequency	42	26	17	16	10	9	6	5	3	3	2	2	1	1	1	1	1	1	1
Cumulitive Percentage	28%	46%	57%	68%	75%	81%	85%	89%	91%	93%	94%	95%	96%	97%	97%	98%	99%	99%	100%

Figure 16. Pareto chart for popular IDPs

As it is presented in Figure 16., 75% of patients in the dataset received dose values of 2.5, 4, 5, 7.5, 10 mg/day. Therefore, by focusing on the patients who have received those doses, the objective is to estimate the percentage error at each dose value. In Figure 17., the distribution of the therapeutic dose at each level of the IDP is presented. Additionally, in Figure 18., a box plot for each level is created.

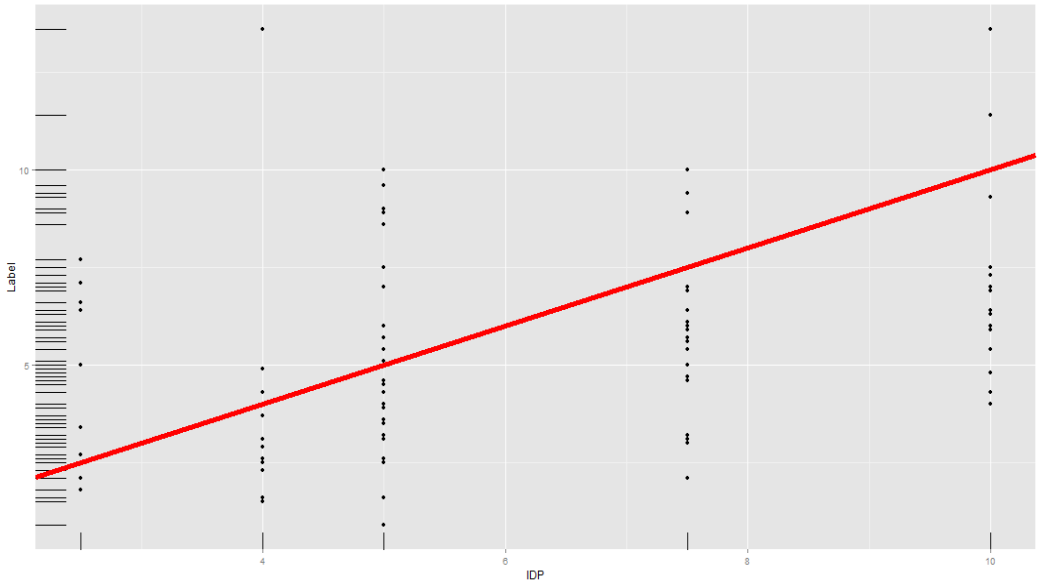


Figure 17. Popular IDPs Vs. Therapeutic Dose

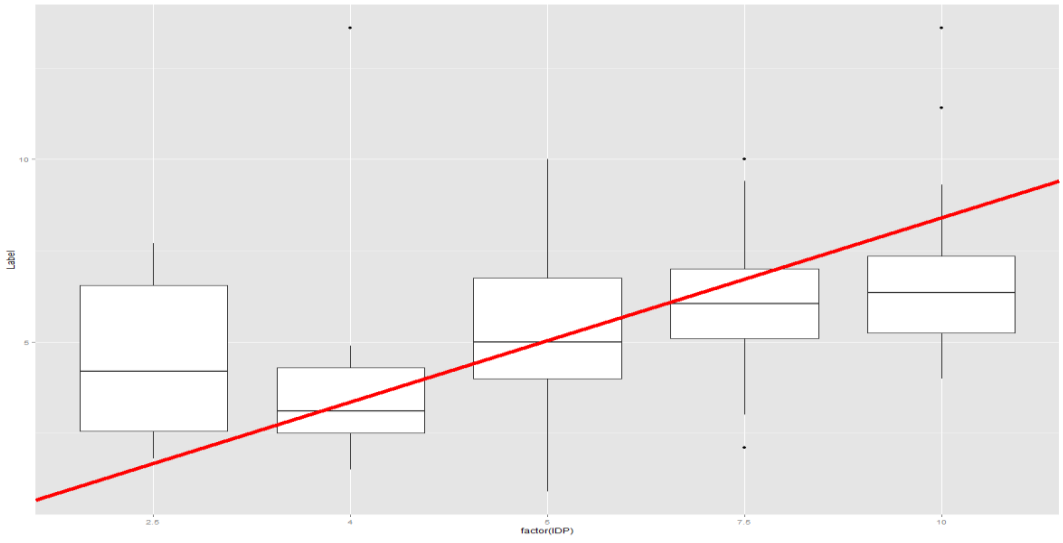


Figure 18. Comparing the distribution of Therapeutic Dose for Popular IDPs using Boxplots

Using the initial dose which was prescribed by the clinicians and the value of the therapeutic dose, the amount of percentage error is calculated. The frequency of patients with different amounts of associated percentage error is presented in Figure 19. By a subjective definition of a significant percentage difference, the patients whom are at high risk/ low risk of mis-dosing can be identified. For instance, in Figure 19. it is assumed that 20% percentage difference is a significant difference and it is shown by dark vertical lines.

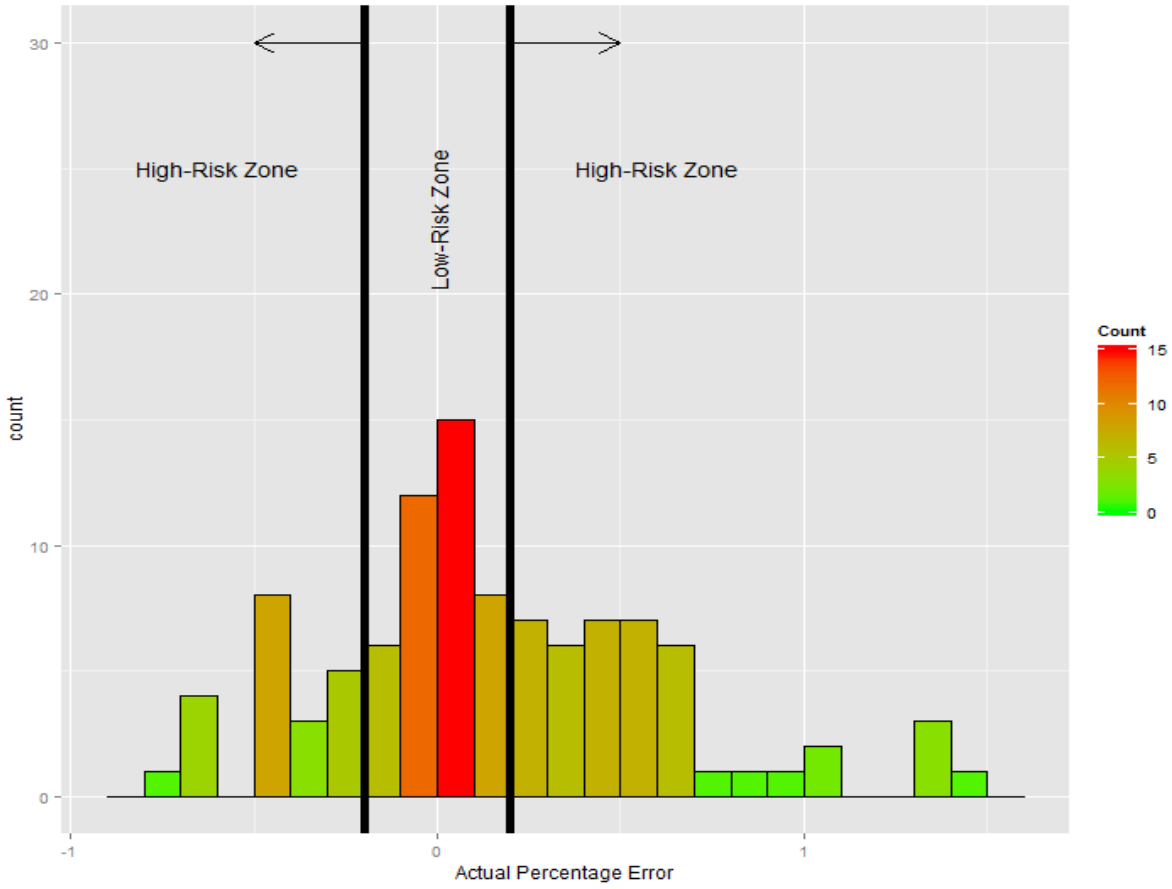


Figure 19. Distribution of the percentage error

Another point of interest is to identify the ranges of prescribed initial dose where higher values of percentage error occur. In Figure 3. the relationship between the initial dose and the

percentage error is presented. Additionally, using a polynomial local regression, the fitted line describing their relationship along with its prediction confidence interval is presented in Figure 20. The size of each point in Figure 20. is proportional to the amount of percentage error. It is evident from Figure 20. that the frequency of higher values of percentage error tends to increase at higher values of initial dose.

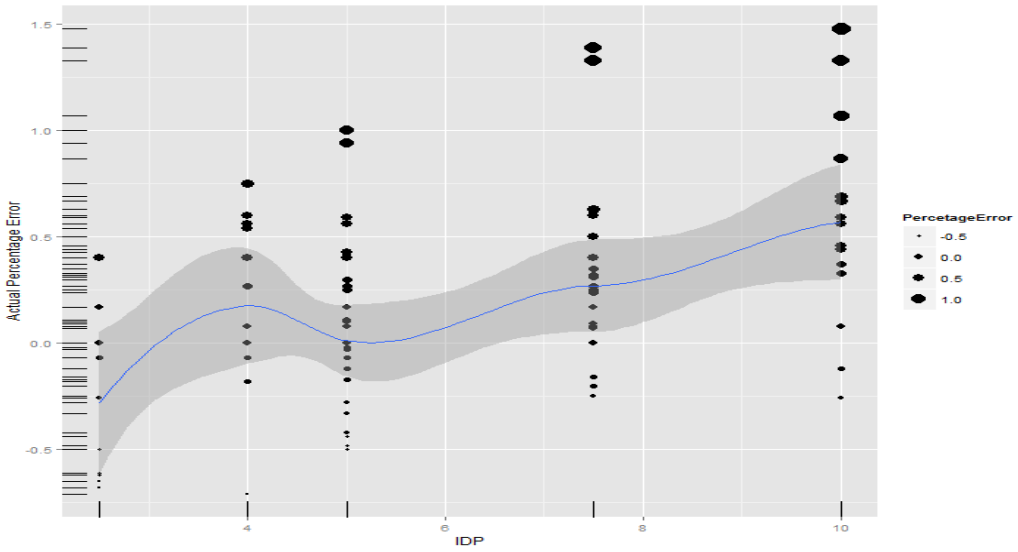


Figure 20. Distribution of percentage error at each level of popular IDPs

Our goal is to develop a prediction model which assigns potential risk of mis-dosing to any prescribed initial dose. Therefore, in order to identify the linear dependency among the variables, the correlation matrix was created and is presented in Figure 21.

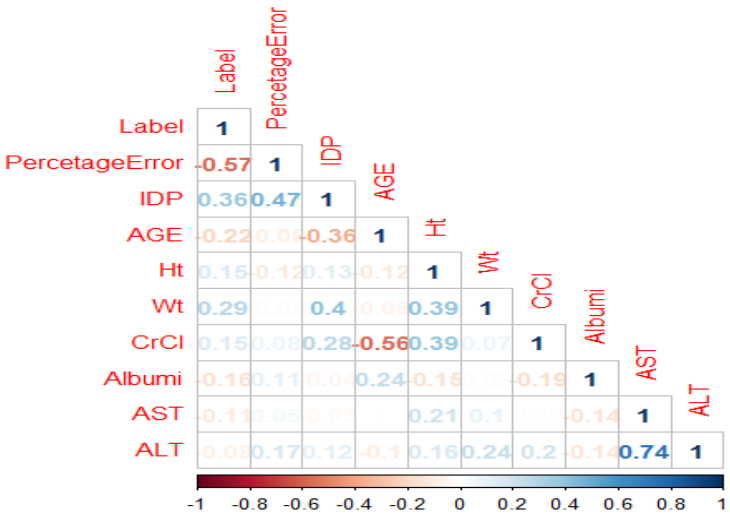


Figure 21. Correlation Matrix

In order to avoid collinearity in modeling, the variables that had the correlation more than or equal to 85%, were defined as highly correlated and only one of them was entered in the modeling phase.

The data points which had missing values for their therapeutic dose were eliminated from the dataset and the missing values for other variables were imputed using KNN (K=1) method. The outliers in the data set were defined as those who had extremely high or low values for therapeutic dose (more than 90 or less than 10 mg/wk). The outliers constitute about 6% of the data set and were eliminated. In the next section the modeling process along with the results are presented.

7.4 Methods

Considering that there exists significant number of variables in the data set compared to the number of data points in the data set, it is needed to select the best subset of variables. Therefore, using shrinkage methods the process of variable selection and developing a prediction model took place simultaneously. Accordingly, the categorical variables in the data

set were transformed into multiple binary dummy variables with one level kept out as the reference. After dividing the data randomly to derivation and validation cohorts (60% / 40%) the optimal prediction model was developed using LASSO. The optimal value of λ was selected by performing the k-fold cross validation (k=10). The resulting prediction model is presented in Table 18.

Table 19. Model Coefficients

Model Coefficients

IDP	AGE	Ht	Wt	CrCl	Albumi	AST	BaselieIR
0.105	-0.001	0.000	-0.003	0.001	0.069	0.001	0.000
Race2	Race3	Race5	Geder2	WI2	WI3	WI4	WI5
0.268	-0.153	0.293	0.159	-0.052	-0.172	-0.073	0.000
WI6	WI7	GoalIR2	GoalIR3	ST1	ST2	ST3	ST4
0.309	-0.036	0.000	0.000	0.210	0.120	0.000	0.396
Smokig2	Smokig3	EtOH2	HT2	DM2	Asthma2	Dyslipidemia2	
0.000	0.160	-0.031	-0.089	0.022	-0.064	0.009	

After developing the prediction model using the training set, its performance was evaluated on the testing set. Therefore, for every data point in the testing set the amount of percentage error was estimated. By defining a given threshold for determination of the significant percentage error, it can be decided whether it is need to revise IDP or use it as is. According to the estimated percentage error, the prescribed initial dose can be revised.

$$Revised\ Dose = (1 - Estimated\ Percentage\ Error) \times IDP$$

Therefore, the resulting revised initial dose values were compared against the original initial dose along with the Gage model in terms of RMSE. Additionally, in order to examine the impact of involving IDP in the modeling process, a new prediction model was developed

with IDP being eliminated from the feature set. The developed model coefficients are presented in Table 19.

Table 20. Estimates Coefficients of the linear model without involving IDP in modeling

Model Coefficients							
(Intercept)	AGE	Ht	Wt	CrCl	Albumi	AST	Baseline INR
-0.026	-0.001	-0.003	0.005	0.003	0.037	0.003	-0.437
Race2	Race3	Race5	Geder2	WI2	WI3	WI4	WI5
0.315	-0.248	0.507	0.109	0.094	0.019	-0.166	0
WI6	WI7	GoalIR2	GoalIR3	ST1	ST2	ST3	ST4
0.376	-0.026	-0.698	0.145	0.331	0.080	-0.172	0.608
Smokig2	Smokig3	EtOH2	HT2	DM2	Asthma2	Dyslipidemia2	
0.214	0.441	-0.039	-0.293	0.039	-0.184	0.191	

Based on the results presented in Table 20., revising the initial doses prescribed by the clinicians will result in much more accurate estimations than the original dose values (RMSE = 2.38), the prediction values made by Gage model (RMSE = 2.05), and the developed linear model without involving IDP in modeling (RMSE = 2.68).

Table 21. Comparing the performance of the revised values of IDP with the original values of IDP and Gage CL model

Threshold	***RMSE***	***Outperformance than the Original IDP***	***Outperformance than the Gage model***	***Outperformance than the linear model without involving IDP***
0.1	1.65	31%	20%	38%
0.15	1.76	26%	14%	34%
0.2	1.77	26%	13.7%	34.0%
0.25	1.9	20%	7.3%	29.1%
0.3	1.96	18%	4.4%	26.9%
0.35	1.96	18%	4.4%	26.9%
0.4	2.06	13%	-0.5%	23.1%

7.5 Conclusion

In this Chapter, an intelligent clinical decision support system for prescribing the initial dose of warfarin is presented. In the proposed procedure the amount of percentage error for initial doses prescribed by the physicians are estimated using shrinkage methods. By applying this estimation, the prescribed doses were revised accordingly. It was shown that by revising physicians' doses, the resulting doses are much more accurate than the original values of doses and the values predicted by Gage Clinical model. This approach is promising and warrants further study that may produce a functional clinical decision support system to assist with initial dosing of warfarin. The major limitation of this analysis is the small sample size that was used in its derivation. This limits the clinical implementability of our specific findings, however the method is novel and should be tested in larger data sets.

8 CONCLUSION AND FUTURE WORKS

In this thesis, four major contributions towards increasing the efficiency of determination of the Warfarin initial dose are presented. After introducing Warfarin and discussing the significance of concentration of this drug in Chapter 1, a comprehensive review of the contributions in Warfarin dosing in the literature is mentioned in Chapter 2.

The necessary mathematical background for exploring the machine learning methods that were utilized in Chapters 4-7 were discussed in Chapter 3. After presenting a holistic view towards machine learning methods, the particular prediction and classification methods of interest (Multivariate regression, Decision Tree, Support Vector Machines, Relevance Vector Machines, Shrinkage methods) were discussed.

In Chapter 4, the process of developing customized prediction models for patients of specific ethnicities is discussed. Using the data of African-American patients at the University of Illinois at Chicago hospital, a prediction model for estimating the initial dose of Warfarin was developed. It is proven that the developed model has a better performance in terms of prediction accuracy than popular methods in the literature known as IWPC and Gage models.

In Chapter 5, a novel procedure for determining the initial dose was introduced. In the proposed procedure, the patients are initially labeled as High-Required Dose and Low-Required Dose. This phase is done using Relevance Vector Machines. After labeling the patients, a separate regression model for each class was developed. Using the proposed methodology it was proven that a more accurate estimation for warfarin initial dose can be achieved than IWPC CL, Gage Cl and the loading method.

In Chapter 6, a companion classification model for IWPC Clinical model was developed using Support Vector Machines with a polynomial kernel function. The classification model labels patients as Safe for the model or High-risk patients. Once the patients are classified, the IWPC Clinical model will only be used for patients who are labeled as Safe-for the model. The remaining patients will be eliminated from the validation set. It was shown that by applying this procedure, the model's performance increases significantly. The choice of Support Vector Machines with a polynomial kernel function occurred after examining several classification models and choosing the method that yielded the best performance on the derivation set in terms of prediction accuracy.

In Chapter 7, a new idea towards determination of the initial Warfarin dose was introduced. By involving physicians' opinion on the initial dose in the modeling phase it was shown that much more accurate results can be achieved. The idea was to estimate the percentage error of doses prescribed by the physicians in practice for each individual patient. Based on this estimation, the prescribed dose might get revised accordingly (increases, decreases, kept unaltered). It was shown that the modified doses are significantly more accurate than the original dose values prescribed by the physicians and the predictions made by the Gage CL model. Additionally, the performance of the proposed procedure was compared against a linear prediction model developed without containing the physicians' doses and its outperformance was proven.

8.1 FUTURE WORKS

The future developments of the works discussed in this thesis can be categorized into three categories.

First, as it was presented in Chapter 4 and 7, the idea of developing customized dosing protocols for each institution works more efficiently than applying generic techniques. Therefore, by utilizing several significant factors such as the local patients' dominant race and the approaches taken by the clinicians at each institution, more efficient prediction models can be derived than popular dosing algorithms in the literature.

Second, the idea of concentration on the initial dose can be generalized to the dose refinement phase. By determination of appropriate later doses (doses prescribed after the initial dose) the negative impact of an inappropriate initial dose can be weakened. Additionally, by studying the trend of prescribed doses until reaching the therapeutic dose, the process of dosing can be done more efficiently in order to increase the likelihood of keeping the patients INR in the therapeutic range.

Lastly, by developing a dynamic modeling framework, the choice of the prediction/classification models or the models' parameters can be modified by increasing the training set. In the current modeling setting, using a static data set, models are created and applied in practice. However, by linking the data analytics engine to the hospital's data warehouse and updating the models by collecting more data points, more robust models can be achieved by evolving the derivation sets.

CITED LITERATURE

Burmester, James K, Richard L Berg, Steven H Yale, Carla M Rottscheit, Ingrid E Glurich, John R Schmelzer, and Michael D Caldwell. 2011. "A Randomized Controlled Trial of Genotype-Based Coumadin Initiation." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 13 (6): 509–18. doi:10.1097/GIM.0b013e31820ad77d.

Cavallari, L H, and E a Nutescu. 2014. "Warfarin Pharmacogenetics: To Genotype or Not to Genotype, That Is the Question." *Clinical Pharmacology and Therapeutics* 96 (1): 22–24. doi:10.1038/clpt.2014.78.

Connolly, Stuart J, Michael D Ezekowitz, Salim Yusuf, John Eikelboom, Jonas Oldgren, Amit Parekh, Janice Pogue, et al. 2009. "Dabigatran versus Warfarin in Patients with Atrial Fibrillation." *The New England Journal of Medicine* 361 (12): 1139–51.

Cosgun, Erdal, Nita a Limdi, and Christine W Duarte. 2011. "High-Dimensional Pharmacogenetic Prediction of a Continuous Trait Using Machine Learning Techniques with Application to Warfarin Dose Prediction in African Americans." *Bioinformatics (Oxford, England)* 27 (10): 1384–89. doi:10.1093/bioinformatics/btr159.

Drozda, Katarzyna, Shan Wong, Shitalben R. Patel, Adam P. Bress, Edith a. Nutescu, Rick a. Kittles, and Larisa H. Cavallari. 2015. "Poor Warfarin Dose Prediction with Pharmacogenetic Algorithms That Exclude Genotypes Important for African Americans." *Pharmacogenetics and Genomics* 25 (2): 73–81. doi:10.1097/FPC.0000000000000108.

Eriksson, Niclas, and Mia Wadelius. 2012. "Prediction of Warfarin Dose: Why, When and How?" *Pharmacogenomics* 13 (4): 429–40. doi:10.2217/pgs.11.184.

Fullér, Robert. 2000. *Introduction to Neuro-Fuzzy Systems*. Vol. 2. Springer.

Gage, B F, C Eby, J a Johnson, E Deych, M J Rieder, P M Ridker, P E Milligan, et al. 2008. "Use of Pharmacogenetic and Clinical Factors to Predict the Therapeutic Dose of Warfarin." *Clinical Pharmacology and Therapeutics* 84 (3): 326–31. doi:10.1038/clpt.2008.10.

Gage, Brian F., and Lawrence J. Lesko. 2008. "Pharmacogenetics of Warfarin: Regulatory, Scientific, and Clinical Issues." In *Journal of Thrombosis and Thrombolysis*, 25:45–51.

Granger, Christopher B, John H Alexander, John J V McMurray, Renato D Lopes, Elaine M Hylek, Michael Hanna, Hussein R Al-Khalidi, et al. 2011. "Apixaban versus Warfarin in Patients with Atrial Fibrillation." *The New England Journal of Medicine* 365 (11): 981–92. doi:10.1056/NEJMoa1107039.

Grossi, Enzo, Gian Marco Podda, Mariateresa Pugliano, Silvia Gabba, Annalisa Verri, Giovanni Carpani, Massimo Buscema, Giovanni Casazza, and Marco Cattaneo. 2013. "Prediction of Optimal Warfarin Maintenance Dose Using Advanced Artificial Neural Networks." *Pharmacogenomics* 15 (1). Future Medicine: 29–37. doi:10.2217/pgs.13.212.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. 2009. *The Elements of Statistical Learning*. Vol. 2. Springer.

Hebb, D. O. 1949. *The Organization of Behaviour*. *Organization*. Wiley & Sons.

Hernandez, W, E R Gamazon, K Aquino-Michaels, S Patel, T J O'Brien, a F Harralson, R a Kittles, et al. 2014. "Ethnicity-Specific Pharmacogenetics: The Case of Warfarin in African Americans." *The Pharmacogenomics Journal* 14 (3): 223–28. doi:10.1038/tpj.2013.34.

Hutten, Barbara A., Martin H. Prins, Michael Gent, Jeff Ginsberg, Jan G. P. Tijssen, and Harry R. Buller. 2000. "Incidence of Recurrent Thromboembolic and Bleeding Complications Among Patients With Venous Thromboembolism in Relation to Both Malignancy and Achieved International Normalized Ratio: A Retrospective Analysis." *J. Clin. Oncol.* 18 (17): 3078–83. http://jco.ascopubs.org/content/18/17/3078.short.

Hylek, E. M., J. D'Antonio, C. Evans-Molina, C. Shea, L. E. Henault, and S. Regan. 2006. "Translating the Results of Randomized Trials into Clinical Practice: The Challenge of Warfarin Candidacy Among Hospitalized Elderly Patients With Atrial Fibrillation." *Stroke*. doi:10.1161/01.STR.0000209239.71702.ce.

Hylek, Elaine M, Carmella Evans-Molina, Carol Shea, Lori E Henault, and Susan Regan. 2007. "Major Hemorrhage and Tolerability of Warfarin in the First Year of Therapy among Elderly Patients with Atrial Fibrillation." *Circulation* 115 (21): 2689–96. doi:10.1161/CIRCULATIONAHA.106.653048.

Informatics, I M S Institute for Healthcare. 2011. "The Use of Medicines in the United States: Review of 2011." IMS Health and the IMS Institute for Healthcare Informatics Parsippany, NJ.

Jiawei, Han, and Micheline Kamber. 2001. "Data Mining: Concepts and Techniques." *San Francisco, CA, Itd: Morgan Kaufmann* 5.

Johnson, J A, L Gong, M Whirl-Carrillo, B F Gage, S A Scott, C M Stein, J L Anderson, et al. 2011. "Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 Genotypes and Warfarin Dosing." *Clin Pharmacol Ther* 90 (4). American Society of Clinical Pharmacology and Therapeutics: 625–29. http://dx.doi.org/10.1038/clpt.2011.185.

Karatzoglou, A, D Meyer, and K Hornik. 2006. "Support Vector Machines in R." *Journal of Statistical Software* 15 (9): 1–28. papers3://publication/uuid/833C3464-37B5-436F-BF1B-9CE8BAC83424.

Kimmel, Stephen E, Benjamin French, Scott E Kasner, Julie a Johnson, Jeffrey L Anderson, Brian F Gage, Yves D Rosenberg, et al. 2013. "A Pharmacogenetic versus a Clinical Algorithm for Warfarin Dosing." *The New England Journal of Medicine* 369 (24): 2283–93. doi:10.1056/NEJMoa1310669.

Kirley, Kate, Dima M. Qato, Rachel Kornfield, Randall S. Stafford, and G. Caleb Alexander. 2012. "National Trends in Oral Anticoagulant Use in the United States, 2007 to 2011." *Circulation: Cardiovascular Quality and Outcomes* 5 (5): 615–21.

Klein, T E, R B Altman, N Eriksson, B F Gage, S E Kimmel, M-T M Lee, N A Limdi, et al. 2009. "Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data." *The New England Journal of Medicine* 360 (8): 753–64. doi:10.1056/NEJMoa0809329.

Limdi, Nita A, T Mark Beasley, Michael R Crowley, Joyce A Goldstein, Mark J Rieder, David A Flockhart, Donna K Arnett, Ronald T Acton, and Nianjun Liu. 2008. "VKORC1 Polymorphisms, Haplotypes and Haplotype Groups on Warfarin Dose among African-Americans and European-Americans." Future Medicine.

Limdi, Nita A., Mia Wadelius, Larisa Cavallari, Niclas Eriksson, Dana C. Crawford, Ming Ta M Lee, Chien Hsiun Chen, et al. 2010. "Warfarin Pharmacogenetics: A Single VKORC1 Polymorphism Is Predictive of Dose across 3 Racial Groups." *Blood* 115 (18): 3827–34.

McCulloch, Warren S, and Walter Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics* 5 (4). Springer: 115–33.

Mega, Jessica L. 2011. "A New Era for Anticoagulation in Atrial Fibrillation." *The New England Journal of Medicine*.

Oztaner, Serdar, Tugba Taskaya Temizel, S Erdem, and Mahmut Ozer. 2014. "A Bayesian Estimation Framework for Pharmacogenomics Driven Warfarin Dosing: A Comparative Study." *IEEE Journal of Biomedical and Health Informatics* 2194 (c). doi:10.1109/JBHI.2014.2336974.

Oztaner, SM, T Taskaya Temizel, SR Erdem, and M Ozer. 2014. "A Bayesian Estimation Framework for Pharmacogenomics Driven Warfarin Dosing: A Comparative Study."

Palareti, G, N Leali, S Coccheri, M Poggi, C Manotti, a D'Angelo, V Pengo, et al. 1996. "Bleeding Complications of Oral Anticoagulant Treatment: An Inception-Cohort, Prospective Collaborative Study (ISCOAT). Italian Study on Complications of Oral Anticoagulant Therapy." *Lancet* 348 (9025): 423–28. doi:10.1016/S0140-6736(96)01109-9.

Patel, Manesh R, Kenneth W Mahaffey, Jyotsna Garg, Guohua Pan, Daniel E Singer, Werner Hacke, Günter Breithardt, et al. 2011. "Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation." *The New England Journal of Medicine* 365 (10): 883–91.

Pirmohamed, Munir, Girvan Burnside, Niclas Eriksson, Andrea L Jorgensen, Cheng Hock Toh, Toby Nicholson, Patrick Kesteven, et al. 2013. "A Randomized Trial of Genotype-Guided Dosing of Warfarin." *The New England Journal of Medicine* 369 (24): 2294–2303. doi:10.1056/NEJMoa1311386.

Prabhu, Sharmila. Data mining and warehousing. New Age International, 2007.

Schelleman, Hedi, Jinbo Chen, Zhen Chen, Jason Christie, Craig W Newcomb, Colleen M Brensinger, Maureen Price, Alexander S Whitehead, Carmel Kealey, and Caroline F Thorn. 2008. "Dosing Algorithms to Predict Warfarin Maintenance Dose in Caucasians and African Americans." *Clinical Pharmacology & Therapeutics* 84 (3). Nature Publishing Group: 332–39.

Schelleman, Hedi, Nita A Limdi, and Stephen E Kimmel. 2008. "Ethnic Differences in Warfarin Maintenance Dose Requirement and Its Relationship with Genetics." *Pharmacogenomics* 9 (9). Future Medicine: 1331–46. doi:10.2217/14622416.9.9.1331.

Sconce, Elizabeth A, Tayyaba I Khan, Hilary A Wynne, Peter Avery, Louise Monkhouse, Barry P King, Peter Wood, Patrick Kesteven, Ann K Daly, and Farhad Kamali. 2005. "The Impact of CYP2C9 and VKORC1 Genetic Polymorphism and Patient Characteristics upon Warfarin Dose Requirements: Proposal for a New Dosing Regimen." *Blood* 106 (7): 2329–33. doi:10.1182/blood-2005-03-1108.

Scott, Stuart A, and Steven A Lubitz. 2014. "Warfarin Pharmacogenetic Trials: Is There a Future for Pharmacogenetic-Guided Dosing?" *Pharmacogenomics* 15 (6). Future Medicine: 719–22.

Sharabiani, Ashkan, Houshang Darabi, Adam Bress, Larisa Cavallari, Edith Nutescu, and Katarzyna Drozda. 2013. "Machine Learning Based Prediction of Warfarin Optimal Dosing for African American Patients." *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, August. Ieee, 623–28. doi:10.1109/CoASE.2013.6653999.

Steinwart, Ingo, and Andreas Christmann. 2008. *Support Vector Machines*. Springer.

Tibshirani, R, and Royal Statistical Society. 1996. "Regression and Shrinkage via the Lasso." *J R Stat Soc, Ser B* 58 (1). Chapman and Hall/CRC: 267–88.

Tipping, ME. 2001. "Sparse Bayesian Learning and the Relevance Vector Machine." *The Journal of Machine Learning Research*.

Vapnik, VN, and V Vapnik. 1998. *Statistical Learning Theory*.

Wadelius, Mia, Leslie Y Chen, Jonatan D Lindh, Niclas Eriksson, Mohammed J R Ghori, Suzannah Bumpstead, Lennart Holm, Ralph McGinnis, Anders Rane, and Panos Deloukas.

2009. “The Largest Prospective Warfarin-Treated Cohort Supports Genetic Forecasting.” *Blood* 113 (4). American Society of Hematology: 784–92.

Wadelius, Mia, Leslie Y. Chen, Niclas Eriksson, Suzannah Bumpstead, Jilur Ghori, Claes Wadelius, David Bentley, Ralph McGinnis, and Panos Deloukas. 2007. “Association of Warfarin Dose with Genes Involved in Its Action and Metabolism.” *Human Genetics* 121 (1): 23–34.

Wirth, R, and J Hipp. 2000. “CRISP-DM: Towards a Standard Process Model for Data Mining.” *Proceedings of the 4th International Conference on the …*.

Wittkowsky, Ann K. 2004. “Effective Anticoagulation Therapy: Defining the Gap between Clinical Studies and Clinical Practice.” *The American Journal of Managed Care* 10 (10 Suppl): S297–306; discussion S312–17.

Xu, Qiang, Bin Xu, Yuxiao Zhang, Jie Yang, Lei Gao, Yan Zhang, Hongjuan Wang, Caiyi Lu, Yusheng Zhao, and Tong Yin. 2012. “Estimation of the Warfarin Dose with a Pharmacogenetic Refinement Algorithm in Chinese Patients Mainly under Low-Intensity Warfarin Anticoagulation.” *Thrombosis and Haemostasis* 108 (6): 1132–40. doi:10.1160/TH12-05-0362.

Yang, Jie, Yanming Chen, Xiaoqi Li, Xiaowen Wei, Xi Chen, Lanning Zhang, Yuxiao Zhang, et al. 2013. “Influence of CYP2C9 and VKORC1 Genotypes on the Risk of Hemorrhagic Complications in Warfarin-Treated Patients: A Systematic Review and Meta-Analysis.” *International Journal of Cardiology* 168 (4). Elsevier Ireland Ltd: 4234–43. doi:10.1016/j.ijcard.2013.07.151.

Yegnanarayana, B. 1994. “Artificial Neural Networks for Pattern Recognition.” *Sadhana* 19 (2). Springer India: 189–238.

VITA

Ashkan Sharabiani

SEL 4209, 842 W Taylor, Chicago, IL 60607. Phone: (312) 547 0360
Email: ashara2@uic.edu

Education

University of Illinois at Chicago(UIC) (GPA: 3.83/4) — **Chicago, IL**, Aug. 2011- Present

PhD Candidate in Industrial Engineering and Operations Research

Thesis: "Medical Decision Making for Warfarin Dosing Using Machine Learning Methods"

Sharif University of Technology — **Tehran, Iran**, Sept. 2009- June. 2011

MSc. Industrial Engineering

Thesis: "Designing an Intelligent Heart Disease Prediction System via Data Mining Techniques"

Shahid Beheshti University — **Tehran, Iran**, Sept. 2004- June. 2009

BSc. Statistics

Technical Skills

Machine Learning, Big Data Analysis, High Dimensional Statistical Analysis, Probabilistic Graphical Models, Predictive Modeling, Parallel Computing, Time Series Analysis, System Performance Analysis, Project Management, Process Mining

Computer Skills

Programming, Systems, and Databases:	Hadoop, Pig, Mahout, Spark, MongoDB, SQL Server, Java, Python, HTML5, CSS, JavaScript, SAP
Mathematical Analysis and Simulation:	R, MATLAB , Python Scikit, Python Numpy
Statistical and Data Mining Packages:	Rapid miner, XLminer ,SPSS Clementine, SPSS, Minitab, ARENA
Data Visualization:	R ggplot2, Python Matplotlib library, Tableau

Professional Experience

Research/Teaching Assistant at University of Illinois at Chicago — **Chicago, IL** — Aug. 2012- Present

- Developed medical decision support systems using machine learning techniques for Warfarin dosing which decreased the misdosing risk by 34%.
- Developed and implemented efficient methods to read big data sets and trained machine learning methods using Hadoop, Python, and R which resulted in 75% accuracy in predicting seizure for epileptic patients.
- Created a physician assistant software to detect high-risk patients for Warfarin dosing

  which identified patients with 85% prediction accuracy.
- Developed a trace-based system to predict the visit pattern and delivery location of Medicaid obstetric patients at University of Illinois Hospital & Health Sciences System which resulted in 76% prediction accuracy.
- Retrieved and analyzed the massive data sets of patients of University of Illinois Outpatient care center for 2000-2012 to evaluate the performance of each clinic and proposed efficient solutions for increasing productivity.
- Established a probabilistic prediction framework for estimating the engineering students' grades using students' prior performance which resulted in 85% prediction accuracy.
- Created a dashboard for analyzing massive data of students in University of Illinois at Chicago in 2000-2014 and designed a report

generating software using interactive spreadsheets and VBA.

- Held 15 workshops on Programing and Data visualization in R
- Held 5 workshops on SQL server systems
- Developed an intelligent scoring software to grade residents in Illinois Hospital & Health Sciences System based on patients' reported pain score profiles
- Assisted more than 10 graduate and undergraduate students in applying statistical and data mining methods in their degree thesis
- Developed a website for the Department of Mechanical Industrial Engineering's accreditation file management system using JavaScript, CSS, and HTML5
- Led 25 students in collecting workflow data in Mile Square Health Center and analyzing the video and numerical data.
- Led 30 students in collecting workflow data in Women's Health Center clinic and analyzing the numerical data.

Publications

- Ashkan Sharabiani, Edith A. Nutescu, William L. Galanter, Houshang Darabi, "A New Approach towards Minimizing the Risk of Mis-Dosing for Popular Warfarin Initiation Doses Prescribed by the Physicians ",under review by Journal of Thrombosis and Thrombolysis, 2015.
- Ashkan Sharabiani, Adam Bress, Elnaz Douzali, Houshang Darabi, "Revisiting Warfarin Dosing Using Machine Learning Techniques",

Published in Computational and Mathematical Methods in Medicine, 2015.

- Ashkan Sharabiani, Adam Bress, Houshang Darabi, “A Computer-Aided System for Determining the Application Range of a Clinical Warfarin Dosing Algorithm Using Support Vector Machines with Polynomial Kernel Function“, under review by International Journal of Medical Informatics, 2015.
- Anooshiravan Sharabiani, Ashkan Sharabiani, Houshang Darabi, “A Novel Bayesian and Chain Rule Model on Symbolic Representation for Time Series Classification”, under review by journal of Knowledge and information system, 2015.
- Ashkan Sharabiani, Fazle Karim, Anooshiravan Sharabiani, Mariya Atanasov, and Houshang Darabi. "An enhanced Bayesian network model for prediction of students' academic performance in engineering programs." In Global Engineering Education Conference (EDUCON), 2014 IEEE, pp. 832-837. IEEE, 2014.
- Ashkan Sharabiani, Houshang Darabi, Adam Bress, Larisa Cavallari, Edith Nutescu, and Katarzyna Drozda, "Machine learning based prediction of warfarin optimal dosing for African American patients" In Automation Science and Engineering (CASE), 2013 IEEE International Conference on, pp. 623-628. IEEE, 2013.
- Ashkan Sharabiani, Houshang Darabi. "Observation policies for patient and resource tracking in outpatient clinics." In Automation Science and

Engineering (CASE), 2012 IEEE International Conference on, pp. 532-537. IEEE, 2012.

Presentations

- Ashkan Sharabiani, Houshang Darabi," Educational performance analysis of African American engineering students: a process mining approach", ISERC Conference, Nashville, TN, May, 2015.
- Ashkan Sharabiani, Maryam Teimoori, Anooshiravan Sharabiani, Fazle Karim, Houshang Darabi," Comparing trace-based and time series prediction modelling for estimating the enrollment in engineering courses" , ISERC Conference ,Nashville, TN, May,2015.
- Ashkan Sharabiani, "An enhanced Bayesian network model for prediction of students' academic performance in engineering programs." In Global Engineering Education Conference (EDUCON), Istanbul, Turkey, 2014.
- Ashkan Sharabiani,"R programing in analyzing big data", Institute of Industrial Engineers, UIC, Dec 2014.
- Ashkan Sharabiani,"Medical decision making using machine learning methods", Institute of Industrial Engineers, UIC, Nov 2014.
- Ashkan Sharabiani,"Machine learning based prediction of warfarin optimal dosing for African American patients" In Automation Science and Engineering (CASE), 2013 IEEE International Conference, WI, Madison, 2013.
- Ashkan Sharabiani, Houshang Darabi," Analysis of Low Risk

Obstetrics Patients Workflow through Process Mining", ISERC Conference ,Puerto Rico, May, 2013.

- Ashkan Sharabiani, Houshang Darabi, "Causal Effect Models for the Prediction of Location Dependent Demand for Primary Care Service", ISERC Conference, Puerto Rico, May, 2013.