IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Optimization of Warfarin Dosing Using Machine Learning: A Literature Review

Cheyeon Shin

Maria Carrillo High School, 6975 Montecito Blvd, Santa Rosa, CA, 95409, USA; cheyeons2024@gmail.com

ABSTRACT: Warfarin is an oral anticoagulant medication used to prevent conditions caused by blood clots. Due to the broad spectrum of variability among patients and the limited therapeutic margin, determining the appropriate dose for each individual is often difficult. To remediate the potentially fatal consequences that incorrect dosing can bring (such as thromboembolism or hemorrhaging), machine learning (ML) algorithms are used to enhance the accuracy of doses using clinical and pharmacogenetic variables. Researchers continue to build on previously published works to develop further methods for improving accuracy.

Despite the advantages that ML techniques can bring, there are certain limitations when evaluating their performance in real clinical settings. For instance, warfarin research lacks diversity in demographics because most current papers only deal with the Caucasian, African/African American, and Asian populations. Those who don't fall under these specific categories are labeled as having a "missing" or "mixed" ethnicity, which may not accurately represent their ethnic demographic. Incorrect identification of patient information leads to inaccurate dose predictions, which can have detrimental consequences. This results in a disparity in the accuracy of medical treatment, a prevalent problem in all aspects of medicine.

KEYWORDS: Computational Biology and Bioinformatics; Computational Pharmacology; Pharmacogenetics; Machine Learning; Warfarin Dose Prediction.

Introduction

Warfarin is an anticoagulant commonly used in patients experiencing conditions such as pulmonary embolism, stroke, atrial fibrillation, deep vein thrombosis, or heart attack.[1] To decrease the risk of blood clots, the patient's blood viscosity must be lowered precisely and consistently. Because the consequences of administering too little or too much of the medication are severe, determining the initial dose is imperative for proper treatment.[2] Typically, a 35mg/wk baseline dose is modified based on demographic variables such as gender, race, height, and weight. For example, a taller person would be administered a higher dose than someone of average height. To determine whether the 'correct' dose has been reached, the international normalized ratio (INR) is calculated using the prothrombin time (the time it takes for a clot to form in a blood sample).[3] For those using warfarin, an INR of 2.0 to 3.0 demonstrates that the patient is within the target range.[4] After receiving the initial dose, patients must revisit the hospital to test their INR and determine whether the dose needs to be adjusted to reach their target range. Not only is this process cumbersome and time-consuming, but patients who require an especially high or low dose may also experience severe repercussions. For example, providing anticoagulants can increase the risk of bleeding, but providing too little won't properly prevent blood clot formation.[5,6] This illustrates the importance of refining the accuracy of the initial warfarin dose, as the consequences of an incorrect dosage can be fatal. With the recent acceleration of technological advancements in the machine learning field, a wider variety of variables (such as pharmacogenetic data) can be considered, and a more accurate initial dose can be administered.

Machine learning is a branch of artificial intelligence that enables computer systems to automatically learn and improve from experience without being explicitly programmed. Its general approach involves algorithms that iteratively learn from data, identifying patterns and making predictions or decisions.[7] Specific types of algorithms include supervised learning, where models are trained on labeled data; unsupervised learning, which finds patterns in unlabeled data; and reinforcement learning, where agents learn to interact with an environment to achieve a goal.[8] In the context of warfarin dosing, machine learning algorithms have been increasingly applied to personalizing dosage decisions based on individual patient characteristics. Training datasets typically consist of patient data, while testing datasets evaluate the algorithm's performance on unseen data. Commonly used performance metrics include accuracy, precision, recall, and F1 score,[9] which assess the model's ability to correctly predict warfarin dosage requirements for optimal patient care.

The International Warfarin Pharmacogenetics Consortium (IWPC) released one of the first papers on optimizing warfarin dosing.[10] With a dataset containing more than 5700 patients collected from 21 research groups from 9 countries and 4 continents, this paper has provided a valuable resource for countless researchers around the world by contributing a large dataset readily available for testing as well as a reference for many articles to compare the performance of their model. However, because the data was collected from multiple organizations with different testing capacities and criteria, some information regarding certain variables is unaccounted for. This presents a challenge because researchers using this dataset

DOI: 10.36838/v6i5.19

must either predict these missing values or exclude the subject entirely.

This literature review will explore five articles that develop and analyze machine learning algorithms to improve warfarin dosing accuracy. The articles were chosen from the Google Scholar search engine, and all five meet the following criteria: use data from the IWPC cohort, encompass multiple ethnic populations, and consider pharmacogenetic data. They are placed in chronological order (from 2009 to 2023) to demonstrate the shift in trends over time as both the field of machine learning and our understanding of medicine have progressed.

Discussion

There have been countless studies performed (using a variety of methods and datasets) on the issue of optimizing the initial warfarin dose. As discussed previously, creating and publishing the IWPC dataset can be seen as one of the 'starting points' of this branch of research because it allowed researchers easy access to an expansive data set. Although Gage *et al*. released a study in 2008, one year before the IWPC article, their dataset only included 1307 patients compared to the 5052 patients in the IWPC dataset.[11] Additionally, not as many articles have used the data from Gage *et al*. to build their algorithm, demonstrating once again the significant impact and contribution of the IWPC cohort on research regarding machine learning-based warfarin dosing. Because a very large sample size is needed to train most machine learning algorithms, having a cohort of 5000+ participants played a crucial role in furthering this field of study.

IWPC (2009):

Klein *et al*.[10] develop three models: comparing a fixed-dose approach, a clinical data-driven approach, and a pharmacogenetic data-driven approach. The fixed-dose approach administers 35 mg of warfarin per week to each patient, regardless of the patient's demographic variables. In contrast, the latter two models use machine learning algorithms explicitly formulated to adjust the dose based on patient information. The term 'clinical data' encompasses variables such as age, height, weight, race (Asian, Black/African American, or Mixed/Missing), enzyme inducer status, and amiodarone status. The pharmacogenetic model uses the variables in the clinical model but additionally considers the single nucleotide polymorphisms (SNPs) of cytochrome P450, family 2, subfamily C, polypeptide 9 (*CYP2C9*), and vitamin K epoxide reductase complex, subunit 1 (*VKORC1*). These SNPs, like race or age, influence warfarin dose requirements.

Klein decided to omit patients who had any missing data, so a total of 4043 patients were divided into groups depending on whether they needed a "low dose" (≤21 mg/week), "intermediate dose" (>21 and <49 mg/week), or "high dose" (≥49 mg/week) of warfarin for stable therapeutic anticoagulation. 75% of these patients were used to develop the two algorithms, and the remaining 25% were used to validate the algorithms based on whether the prediction was within 20% of the actual stable therapeutic dose. Klein observed that the most accurate doses were predicted through the pharmacogenetic algorithm, specifically in the low-dose or high-dose groups. The second-most optimal algorithm was the clinical model, putting the fixed-dose approach in last place. Based on his findings, Klein concluded that the inclusion of genetic information is beneficial when estimating warfarin dose.

Sharabiani et al. (2014):

In contrast to Klein, Sharabiani *et al*.[12] conclude that there isn't enough evidence to prove that pharmacogenetic data yields higher accuracy in warfarin dosing and conducts an experiment using only clinical data. Sharabiani explains that even if pharmacogenetic data provides more accurate dose predictions, it's often impractical to gather genetic data of individuals and perform tests in real clinical settings before giving patients a warfarin dose.

To deal with the missing values from the IWPC dataset, Sharabiani used the K-nearest neighbor method to predict the values, classifying data based on the proximity between data points. The patients were assigned into groups based on whether they needed a "low dose" (≤30 mg/wk) or a "high dose" (>30 mg/wk). 50% of the patients were used to develop the algorithm, and the remaining 50% were used to validate it. Sharabiani's clinical model yielded greater performance when compared to the fixed-dose approach.

Statistical models of root mean squared error (RMSE) and mean absolute error (MAE) were used to evaluate the prediction accuracy. Sharabiani's model yielded the values 11.6 mg/week and 8.4 mg/week, respectively. For both of these mathematical models, a lower value signifies higher accuracy in prediction. Therefore, Sharabiani's clinical model was found to outperform that of the IWPC clinical model, as the IWPC model produced an MAE value of 10.1 mg/week (Klein used MAE and r^2 instead of RMSE, so these values can't be directly compared). However, it's important to consider that this compares to the IWPC clinical model, which performed poorly compared to the IWPC pharmacogenetic model. While the IWPC clinical model had an MAE of 10.1 mg/week, the IWPC pharmacogenetic model had an MAE of 8.8 mg/week. Sharabiani's work is significant because it highlights the limits of applying pharmacogenetic models in real life and proposes a solution using clinical models to resolve this problem.

Ma et al. (2018):

The most notable aspect of Ma *et al*.'s[13] work is that they use a stacked generalization framework, which utilizes multiple machine learning algorithms and improves dose prediction accuracy. Ma also incorporates more variables when calculating doses. While Klein considers the variables genotype, age, height, weight, race, enzyme inducers, and amiodarone, Ma considers diabetes mellitus, heart failure, valve replacement, use of statins, and smoking status in addition to all the variables that Klein included.

A total of 5743 subjects from the IWPC cohort were included in Ma's study. 80% of these subjects were used as the training set, while the remaining 20% were used as a validation set. Another unique aspect of Ma's work is the division of subgroups prior to testing—he uses the categories of low, intermediate, and high dose, much like other articles; however, he defines these terms differently depending on race. For example, while the "low" category represents those who take <15.0 mg/wk of warfarin in Asians, the same category repre-

DOI: 10.36838/v6i5.19

sents those who take <30.0 mg/wk of warfarin in Africans/African Americans.

Ma notes that multivariate linear regression (MLR) models, popular among researchers seeking to optimize warfarin dosing, can be flawed because they attempt to provide a linear solution for a non-linear dataset. Predictive accuracy was measured with MAE, and both of Ma's stacked generalization models outperformed his MLR models with an MAE of 8.31 mg/week as opposed to an MAE of 8.53 mg/week. This is impressive because not only were Ma's own stacked generalization models more accurate than his MLR models, but they even outperformed previous models created by Klein or Sharabiani.

Truda et al. (2021):

Truda *et al.*[14] explain the difficulty of replicating and comparing results with previously published research for a couple of reasons: the lack of open-source code for many articles, variation in testing methods, and differences in data. Firstly, the need for open-source code suggests that the machine learning algorithms used in each article are described textually rather than as a sharable and replicable code. This leads to problems when comparing with other authors, making it challenging to analyze the proficiency of accuracy, and prevents researchers from comfortably building off each other's work. Furthermore, each of the testing methods has fundamental differences between them that may result in inconsistencies. For instance, some use pharmacogenetic data, while others use clinical; some incorporate various variables, while others only consider a few; and some validate their findings through mathematical models like the percent of the cohort with a mean percentage within 20% of the actual dose, while others use root mean squared error (RMSE) and mean absolute error (MAE). Additionally, all researchers in this literature review who used the IWPC cohort for their dataset exclude different patients based on different standards of determining outliers.

To remedy the aforementioned problems, Truda creates a software framework called "Warfit-learn," which allows smoother accessibility of the source code for those interested. When comparing his model to the IWPC model, he found that his MAE of 8.59 mg/week was higher than Klein's MAE of 8.5 mg/week. Although Truda's predictions weren't as accurate as the previously published models, he fulfilled his initial intentions by establishing the Warfit-learn framework, which allowed warfarin research to be conducted more efficiently by opening up the availability of prediction models to the general public. In addition, Truda also demonstrated extensive research on the comparison of different ML techniques, including the following: linear regression (LR), support vector regression (SVR), ridge regression (RR), boosted regression trees (BRT), gradient-boosted trees (GBT), and neural network (NN). Out of these, he found that LR and SVR models produced the most accurate predictions, while NN models produced the worst.

Jahmunah et al. (2023):

Jamunah *et al.*[15] differentiate their study by using a machine learning method called "deep learning," which mimics how the human brain thinks. The multiple layers of neural networks function similarly to the neurons in a human brain, which allows deep learning models to recognize complex patterns within a dataset.[16] Additionally, Jahmunah has a unique method of testing and validating their algorithm; unlike Klein and Ma, Jahmunah uses the combined dataset (all demographics) to train the data, then tests the algorithm within members of a specific race. This allows them to input more patient data while training the model, increasing the sample size. To mathematically validate its findings, MAE and percentage of the cohort within 20% of the actual dose are used.

Like Ma, Jahmunah continues the trend of incorporating more variables when training the models. In addition to the general clinical variables (age, gender, ethnicity, height, and weight), Jahmunah considered factors including cardiomyopathy, stroke, Type II diabetes, smoking status, and ten concomitant medications. Similar to Sharabiani, Jahmunah noticed that their model performed significantly better in Asian populations than in other ethnic groups. This could be explained by the fact that 90% of the Asian population within the IWPC dataset carried the VKORC1-1639 A polymorphism, making them a more homogenous population. In contrast, only 38% of Whites and 11% of Africans/African Americans carried the polymorphism. Unsurprisingly, the lowest MAE achieved was from the Asian population, producing a value of 7.6 mg/week. As Truda pointed out, many published articles regarding the optimization of warfarin lack practicality due to the large variability in datasets and methodology that each author chooses. Jahmunah reflects this observation and points out that in a real-world setting, the data available on a patient may be very limited and will most likely vary depending on the capacity of medical facilities.

■ Conclusion

Machine learning has been previously used to optimize the dosage of medications other than warfarin,[17] such as vancomycin (an antibiotic that treats bacterial infections)[18] and lamotrigine (an anticonvulsant that treats seizures).[19] Similar limitations were observed in studies using ML to improve the dose accuracy of these drugs: overfitting of models led to the need for external validation, small sample sizes presented questions on the generalizability of models, and specific demographics performed better than others.

Klein's[10] (IWPC) study from 2009 explored the possibility of genetic data in improving warfarin dosing accuracy and concluded that pharmacogenetic algorithms yield closer predictions than clinical ones. In contrast, Sharabiani[12] introduced a prediction model in 2014 dealing only with clinical data due to the practical limits of collecting genetic data from patients. Using this clinical model, Sharabiani produced a lower MAE value than Klein's model, demonstrating improved dose prediction accuracy. Despite this, research from Ma[13] in 2018, Truda[14] in 2021, and Jahmunah[15] in 2023 all start with the assumption that pharmacogenetic data should be included in their study. All three of these models outperform Klein's model, highlighting that dose prediction accuracy increases as our understanding of machine learning and medicine expands. There is a continuous debate about whether or not the inclusion of pharmacogenetic data significantly improves dose prediction compared to a clinical model. Still, the underlying agreement is that using genetic information doesn't decrease the efficiency

and accuracy of models. From the 2010s to the early 2020s, there has been a slight shift in research direction: rather than comparing the performance of pharmacogenetic models to that of clinical models, more and more researchers incorporate genetic data into their models as a fundamental aspect and compare their performance to that of models published in previous articles.

There are many limitations to the research methods explored throughout this literature review despite all five articles using data from the IWPC cohort. Although the same dataset is used, the sample size varied between studies because of the inconsistent exclusion criteria of patients with missing information on certain variables. Such differences resulted from the fact that data was collected from multiple different organizations. These missing values were either extrapolated based on other patient information or the patient was omitted entirely. In the former case, an inaccurate estimation of a value could jeopardize the results of the entire experiment. The latter presented the risk of using too small of a sample size. Thus, it was critical that researchers find a balance between ensuring a large sample size and generating correct predictions about patient data. Specifically for crafting machine learning algorithms, a large sample size is imperative to prevent overfitting, which occurs when a model yields a prediction that adheres too specifically to the given dataset and thus doesn't apply to larger, more general datasets when tested.[20] Increasing the sample size is often the easiest and most effective solution to this phenomenon.

In selecting the most accurate or efficient warfarin prediction model, there is a problem resulting from the limited availability of patient genetic information in real-world settings. Due to factors such as disparities between medical facilities and equipment availability, researchers may need access to genetic testing equipment. On the other hand, while pharmacogenetic models may improve warfarin dosing accuracy, they lack practicality when applied to clinical situations. This disparity is highlighted explicitly in developing countries where medical care is not yet advanced enough to provide accurate and timely genetic testing. This is especially important because administration of warfarin can be such a time-sensitive procedure.

Another real-world limitation is the minimal number of demographics included within the IWPC dataset. Because the categories for race were limited to Caucasian, African/African American, Asian, and mixed/unknown, specific populations (such as the Hispanic population) are much less likely to have an accurate dose prediction. Additionally, because the primary demographic that uses warfarin tends to be the elderly (generally 65 years or older), research on warfarin dosing for children or adolescents is limited. As a result, further exploration of this aspect is needed to ensure the safety and comfort of all warfarin users.

Acknowledgment

I would like to acknowledge Ms. Seo Kim for her patience, guidance, and support throughout my writing of this literature review.

References

1. Wysowski, D.; Nourjah, P.; Swartz, L. Bleeding Complications with Warfarin Use: A Prevalent Adverse Effect Resulting in Regulatory Action. https://pubmed.ncbi.nlm.nih.gov/17620536/(accessed 2023-07-01).
2. Budnitz, D.; Shehab, N.; Kegler, S.; Richards, C. Medication Use Leading to Emergency Department Visits for Adverse Drug Events in Older Adults. https://pubmed.ncbi.nlm.nih.gov/18056659/ (accessed 2023-07-29).
3. Besselaar, A.; Barrowcliffe, T.; HouBouyan-Reveillard, L.; Jespersen, J.; Johnston, M.; Poller, L.; Tripodi, A. Guidelines on Preparation, Certification, and Use of Certified Plasmas for ISI Calibration and INR Determination. https://www.sciencedirect.com/science/article/pii/S1538783622169424 (accessed 2023-07-13).
4. Shikdar, S.; Vashisht, R.; Bhattacharya, P. International Normalized Ratio (INR). https://www.ncbi.nlm.nih.gov/books/NBK507707/ (accessed 2023-07-24).
5. Garcia, D.; Regan, S.; Henault, L. Risk of Thromboembolism with Short-Term Interruption of Warfarin Therapy. https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/413677 (accessed 2023-07-17).
6. Rosand, J.; Eckman, M.; Knudsen, K. The Effect of Warfarin and Intensity of Anticoagulation on Outcome of Intracerebral Hemorrhage. https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/216977 (accessed 2023-07-17).
7. Sarker, I. Machine Learning: Algorithms, Real-World Applications and Research Directions https://link.springer.com/article/10.1007/s42979-021-00592-x (accessed 2024-04-13).
8. Morales, E.; Escalante, H. A brief introduction to supervised, unsupervised, and reinforcement learning https://www.sciencedirect.com/science/article/abs/pii/B9780128201251000178 (accessed 2024-04-13).
9. Japkowicz, N.; Shah, M. Performance Evaluation in Machine Learning | SpringerLink https://link.springer.com/chapter/10.1007/978-3-319-18305-3_4 (accessed 2024-04-13).
10. Klein, T. E. Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data. *New England Journal of Medicine* **2009**, *360* (8), 753–764. DOI:10.1056/nejmoa0809329.
11. Gage, B.; Eby, C.; Johnson, J.; Deych, E.; Rieder, M.; Ridker, P.; Milligan, P.; Grice, G.; Lenzini, P.; Rettie, A.; Aquilante, C.; Grosso, L.; Marsh, S.; Langaee, T.; Farnett, L.; Voora, D.; Veenstra, D.; Glynn, R.; Barrett, A.; McLeod, H. Use of Pharmacogenetic and Clinical Factors to Predict the Therapeutic Dose of Warfarin. Clinical Pharmacology & *Therapeutics* **2008**, *84* (3), 326–331. DOI:10.1038/clpt.2008.10.
12. Sharabiani, A.; Bress, A.; Douzali, E.; Darabi, H. Revisiting Warfarin Dosing Using Machine Learning Techniques. https://www.hindawi.com/journals/cmmm/2015/560108/ (accessed 2023-07-01).
13. Ma, Z.; Wang, P.; Gao, Z.; Wang, R.; Khalighi, K. Ensemble of Machine Learning Algorithms Using the Stacked Generalization Approach to Estimate the Warfarin Dose. *PLOS ONE* **2018**, *13* (10). DOI:10.1371/journal.pone.0205872.
14. Truda, G.; Marais, P. Evaluating Warfarin Dosing Models on Multiple Datasets with a Novel Software Framework and Evolutionary Optimisation. https://www.sciencedirect.com/science/article/pii/S1532046420302628 (accessed 2023-07-05).
15. Jahmunah, V.; Chen, S.; Oh, S.; Acharya, U.; Chowbay, B. Automated Warfarin Dose Prediction for Asian, American, and Caucasian Populations Using a Deep Neural Network. https://www.sciencedirect.com/science/article/abs/pii/S0010482523000136 (accessed 2023-07-03).
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. https://www.nature.com/articles/nature14539 (accessed 2023-08-03).
17. Poweleit, E. A.; Vinks, A. A.; Mizuno, T. Artificial Intelligence and Machine Learning Approaches to Facilitate Therapeutic Dru

DOI: 10.36838/v6i5.19

g Management and Model-Informed Precision Dosing. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10378651/ (accessed 2023-09-22).
18. Imai, S.; Takekuma, Y.; Miyai, T.; Sugawara, M. A New Algorithm Optimized for Initial Dose Settings of Vancomycin Using Machine Learning. https://www.jstage.jst.go.jp/article/bpb/43/1/43_b19-00729/_article/-char/ja/ (accessed 2023-07-29).
19. Zhu, X.; Huang, W.; Lu, H.; Wang, Z.; Ni, X.; Hu, J.; Deng, S.; Tan, Y.; Li, L.; Zhang, M.; Qiu, C.; Luo, Y.; Chen, H.; Huang, S.; Xiao, T.; Shang, D.; Wen, Y. A Machine Learning Approach to Personalized Dose Adjustment of Lamotrigine Using Noninvasive Clinical Parameters. https://www.nature.com/articles/s41598-021-85157-x (accessed 2023-08-15).
20. Ying, X. An Overview of Overfitting and Its Solutions. https://iopscience.iop.org/article/10.1088/1742-6596/1168/2/022022meta (accessed 2023-08-20).
21. Cilluffo, G.; Fasola, S.; Ferrante, G.; Malizia, V.; Montalbano, L.; La Grutta, S. Machine Learning: An Overview and Applications in Pharmacogenetics. https://www.mdpi.com/2073-4425/12/10/1511 (accessed 2023-07-16).
22. Liu, R.; Li, X.; Zhang, W.; Zhou, H.-H. Comparison of Nine Statistical Model Based Warfarin Pharmacogenetic Dosing Algorithms Using the Racially Diverse International Warfarin Pharmacogenetic Consortium Cohort Database. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0135784 (accessed 2023-07-10).
23. Zadeh, S.; Street, W.; Thomas, B. Optimizing Warfarin Dosing Using Deep Reinforcement Learning. https://www.sciencedirect.com/science/article/abs/pii/S1532046422002726 (accessed 2023-07-20).
24. Zhang, F.; Liu, Y.; Ma, W.; Zhao, S.; Chen, J.; Gu, Z. Nonlinear Machine Learning in Warfarin Dose Prediction: Insights from Contemporary Modelling Studies. https://www.mdpi.com/2075-4426/12/5/717 (accessed 2023-08-19).

Author

Cheyeon Shin is a senior at Maria Carrillo High School in Santa Rosa, California. She is interested in pursuing a higher degree in molecular, cellular, and developmental biology.