

## انتخاب ویژگی‌های مؤثر کلینیکی و ژنتیکی به منظور پیش‌بینی دوز وارفارین با استفاده از شبکه‌های عصبی مصنوعی

### چکیده

محمد کریم سهرابی\*  
علیرضا تاجیک

گروه کامپیوتر، واحد سمنان، دانشگاه آزاد اسلامی، سمنان، ایران.

دریافت: ۱۳۹۴/۰۶/۲۶ پذیرش: ۱۳۹۴/۰۹/۲۴ آنلاین: ۱۳۹۵/۰۱/۱۴

**زمینه و هدف:** وارفارین داروی ضدانعقاد می‌باشد که نقش آن پیشگیری از ایجاد لخته‌گی است. هدف از این مقاله، ارائه روشی مناسب برای انتخاب ویژگی‌های مهم کلینیکی و ژنتیکی و پیش‌بینی میزان دوز وارفارین بود. **روش بررسی:** این مطالعه تجربی، از اردیبهشت تا خرداد ۱۳۹۴ بر روی ۵۵۲ نفر از بیمارانی که در بیمارستان مرکز قلب تهران کاندید استفاده از وارفارین بودند، انجام گرفت. عوامل تأثیرگذار در میزان دوز استخراج شده، و روش‌های انتخاب ویژگی و شبکه‌های عصبی در نرم‌افزار MATLAB (MathWorks, MA, USA) پیاده‌سازی گردید. **یافته‌ها:** از بین دو الگوریتم به کار گرفته شده، الگوریتم بهینه‌سازی ازدحام ذرات دارای دقت مناسب‌تری بوده و برای میانگین مربعات خطا، جذر میانگین مربعات خطا و میانگین قدر مطلق خطا به ترتیب ۰/۰۲۶۲، ۰/۱۶۲۱ و ۰/۱۱۶۴ به دست آمد.

**نتیجه‌گیری:** بر اساس نتایج به دست آمده، با تعداد ویژگی‌های کمتر، می‌توان به دقت مناسبی از پیش‌بینی دوز وارفارین دست یافت. همچنین از این مدل می‌توان به‌عنوان یک سیستم پشتیبان تصمیم‌گیری بهره برد.

**کلمات کلیدی:** وارفارین، نرم‌افزارهای اطلاعات پزشکی، سیستم‌های کامپیوتری، شبکه‌های عصبی مصنوعی.

\* نویسنده مسئول: سمنان، کیلومتر ۵ جاده دامغان، دانشگاه آزاد اسلامی واحد سمنان.

تلفن: ۰۲۳-۳۳۶۵۴۰۴۰

E-mail: amir\_sohraby@yahoo.com

### مقدمه

زیادی در تشخیص دوز، تأثیرگذار هستند. با توجه به اینکه تنظیم دوز بر اساس روش آزمون و خطا بوده و بر اساس تغییر میزان دوز مصرفی در آزمایش‌های متوالی می‌باشد، روش‌های داده‌کاوی نقش به‌سزایی در پیش‌بینی دوز مصرفی می‌تواند داشته باشد. هدف از این پژوهش شناسایی ویژگی‌های تأثیرگذار در مصرف داروی وارفارین و سپس پیش‌بینی میزان دوز بر اساس شبکه‌های عصبی مصنوعی می‌باشد. از مدل ساخته شده، می‌توان به‌عنوان یک سیستم پشتیبان تصمیم‌گیری برای کمک به پزشکان استفاده نمود.

### روش بررسی

این مطالعه، از اردیبهشت تا خرداد ۹۴ بر روی ۵۵۲ نفر از

وارفارین (نام تجاری کمادین)، یکی از شناخته شده‌ترین داروهای خوراکی می‌باشد.<sup>۱</sup> از این دارو، برای جلوگیری از ترومبوز و انسداد جریان خون در اختلالات فیبریلاسیون دهلیزی، آمبولی ریه و تعویض دریچه قلب استفاده می‌شود.<sup>۲</sup> اندازه‌گیری دوز وارفارین توسط تست خونی International normalized ratio (INR) انجام می‌شود که مدت زمان تشکیل لخته‌گی خون را نشان می‌دهد.<sup>۳</sup> مقدار INR بیمارانی باید در محدوده درمانی نگه داشته شود زیرا افزایش یا کاهش آن باعث افزایش خطر مرگ و میر می‌گردد.<sup>۴</sup> مشکلی که در تجویز وارفارین وجود دارد، چگونگی نیل به اهداف درمانی برای هر بیمار می‌باشد. فاکتورهای کلینیکی و ژنتیکی

وارفارین قرار می‌گیرند، دارای یک دفترچه PT می‌باشند. در این دفترچه، دلیل مصرف وارفارین، میزان INR، دوز تجویز شده و تاریخ انجام آزمایش ثبت می‌شود.

بیمارانی که فاقد اطلاعات ژنتیکی بودند، از مطالعه حذف شدند. همچنین بیمارانی که INR آنها در محدوده هدف قرار گرفته و سه ماه به‌صورت پیاپی در محدوده درمان قرار گرفته باشند، حفظ شده و بقیه از پژوهش کنار گذاشته شدند. بر اساس تحقیقات انجام گرفته، اگر سه INR پشت سر هم در محدوده هدف باشد، به این معنی می‌باشد که بیمار به این دوز پاسخ داده است.<sup>۵</sup>

انتخاب ویژگی، یکی از مهمترین عوامل تأثیرگذار در دقت دسته‌بندی می‌باشد.<sup>۶</sup> یک روش مؤثر و کارآمد در انتخاب بهترین ویژگی‌ها، می‌تواند منجر به درک بهتر داده‌ها، حذف ویژگی‌های نامرتبط، کاهش هزینه محاسبات و بهبود پیش‌بینی شود.<sup>۷</sup>

در این مقاله از رویکرد Wrapper برای انتخاب ویژگی‌ها و معیار ارزیابی دقت طبقه‌بندی استفاده شده است. دو روش جستجو به‌صورت تصادفی (مبتنی بر الگوریتم ژنتیک) و به‌صورت مکاشفه‌ای (مبتنی بر الگوریتم PSO) به‌کار گرفته می‌شود و قانون توقف برای الگوریتم انتخاب ویژگی‌ها، حداکثر تعداد تکرار الگوریتم و انتخاب تعداد مشخصی از ویژگی‌ها می‌باشد.

در جدول ۱، ویژگی‌های مرتبط با دوز وارفارین نشان داده شده است. سعی داریم با تبدیل مسئله انتخاب ویژگی، به یک مسئله بهینه‌سازی، حل آن را با کمک الگوریتم‌های هوشمند انجام دهیم. در این مسئله از شبکه‌های عصبی مصنوعی و کاربرد آن در رگرسیون استفاده شد. الگوریتم ژنتیک ماهیت تصادفی داشته و با هر بار اجرا، زیرمجموعه‌ای از ویژگی‌ها را تولید کرده و در اختیار شبکه عصبی قرار می‌دهد. تابع هدف کمینه‌سازی خطای شبکه می‌باشد. از الگوریتم Particle Swarm Optimization (PSO) به روش Stochastic Focusing Search (SFS) استفاده شده است. روش SFS، با یک مجموعه خالی آغاز می‌شود. در گام اول یک ویژگی اضافه شده و بر اساس تابع هدف مورد ارزیابی قرار می‌گیرد. در گام دوم، بقیه ویژگی‌ها به‌صورت تکی به مجموعه اضافه شده و مجموعه جدید مورد ارزیابی قرار می‌گیرد. توقف الگوریتم زمانی انجام می‌گیرد که تمام ویژگی‌ها به مجموعه اضافه شده باشند.<sup>۸</sup> برای پیاده‌سازی الگوریتم‌ها، از نرم‌افزار MATLAB (MathWorks, Inc., Natick

جدول ۱: ویژگی‌های مؤثر در وارفارین (تعداد کل بیماران: ۵۵۲ نفر)

ویژگی	مشخصات آماری
سن (سال)	۱۶-۸۸ (۵۸/۸۳±۱۲/۷۲)
جنسیت	زن ۲۹۱ (%۵۳) مرد ۲۶۱ (%۴۳)
قد (m)	۱/۳۸-۱/۹۷ (۱۶۳±۸/۴)
وزن (kg)	۴۱-۱۱۹ (۷۱/۵۰±۱۱/۰۴)
خونریزی	خیر ۴۶۵ (%۸۴) بله ۸۷ (%۱۶)
دیابت	خیر ۴۶۲ (%۸۳/۵) بله ۹۰ (%۱۶/۵)
چربی خون	خیر ۳۱۴ (%۵۷) بله ۲۳۸ (%۴۳)
قند خون	خیر ۲۹۰ (%۵۲/۵) بله ۲۶۲ (%۴۷/۵)
مصرف سیگار	خیر ۴۴۸ (%۸۱) بله ۳۵ (%۱۹)
تیروید	خیر ۵۰۷ (%۹۲) بله ۴۵ (%۸)
نوع بیماری	Valvular ۳۶۵ (%۶۶) Non-Valvular ۱۸۷ (%۳۴)
مصرف دارو متداخل	خیر ۳۶ (%۶/۵) بله ۵۱۷ (%۹۳/۵)
VK9934438Hinf1	Wild ۱۹۷ (%۳۵/۵), Hetro ۲۵۸ (%۴۶/۵) Mutant ۹۷ (%۱۸)
VK1639Msp1	Wild ۱۹۲ (%۳۵), Hetro ۲۵۴ (%۴۶) Mutant ۱۰۶ (%۱۹)
Factor7Msp1	Wild ۳۰۸ (%۵۶), Hetro ۲۱۰ (%۳۸) Mutant ۳۴ (%۸)
NQO1	Wild ۴۱۴ (%۷۵), Hetro ۱۳۶ (%۲۵)
exon3EcorV	Wild ۳۱۶ (%۵۷), Hetro ۱۶۹ (%۳۱) Mutant ۶۷ (%۱۲)
EPHX4Rsa1	Wild ۳۴۶ (%۶۲/۵), Hetro ۱۹۰ (%۳۴/۵) Mutant ۱۶ (%۳)
haplotype2c9	*1/*1 (317), *1/*3 (103), *2/*3 (17), *1/*2(68), *3/*3(15), *2/*2(32) ۵، ۹۷/۵۰

بیمارانی که در بیمارستان مرکز قلب تهران کاندیدای مصرف وارفارین بودند، انجام گرفت. هر کدام از بیمارانی که تحت درمان توسط

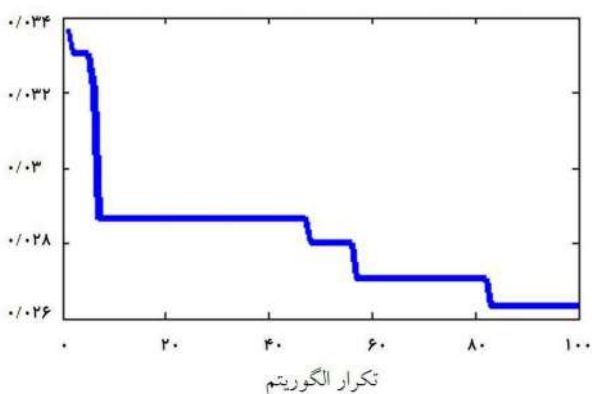
ترکیب ۰/۸ و نرخ جهش ۰/۳ در نظر گرفته شده است. در نمودار ۲ نیز، نمودار خروجی SFS، به ازای ویژگی‌های مسئله نشان داده شده است. برای الگوریتم ژنتیک شاخص‌های میانگین مربعات خطا، جذر میانگین مربعات خطا و میانگین قدر مطلق خطا به ترتیب ۰/۰۲۶۹، ۰/۱۷۷۱ و ۰/۱۳۴۹ به دست آمد همچنین برای الگوریتم PSO نیز به ترتیب مقادیر ۰/۰۲۶۲، ۰/۱۶۲۱ و ۰/۱۱۶۴ حاصل گردید. همانطور که مشاهده می‌شود، از ترکیب شبکه‌های عصبی و PSO به دقت مناسب‌تری جهت پیش‌بینی دوز، دست یافته شده است.

ویژگی‌های انتخابی برای الگوریتم ژنتیک شامل سن، جنسیت، قد، وزن، پلی مورفیسم ژن EcorV، exon3 و haplotype2c9 می‌باشد. همچنین این ویژگی‌ها برای الگوریتم PSO عبارتند از: سن، جنسیت، قد، وزن، مصرف دارو متداخل، پلی مورفیسم ژن VK9934438HinfI، VK1639Msp1 و exon3EcorV و haplotype2c9.

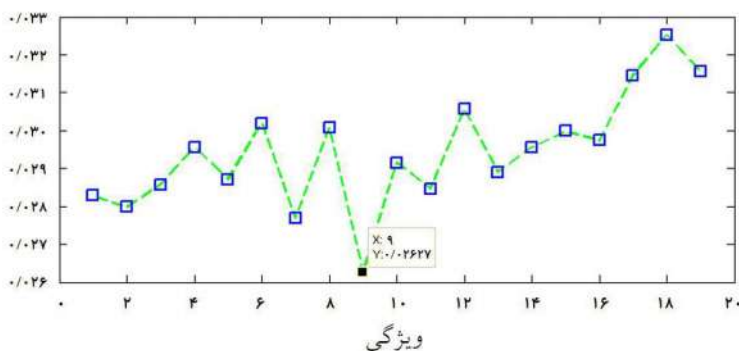
(MA, USA) استفاده شد. از آنجایی که پاسخ‌های شبکه عصبی در هر بار تکرار به صورت تصادفی می‌باشد، تعداد تکرار را به پنج بار افزایش داده و میانگین خطا را برای هر زیرمجموعه تولید شده محاسبه شد. تعداد نوروهای لایه میانی را ۱۲ در نظر گرفته و از تابع یادگیری Trainlm در جعبه ابزار شبکه‌های عصبی، استفاده شد. همچنین به منظور ارزیابی شبکه، از شاخص‌های Mean Squared Error (MSE)، Root Mean Squared Error (RMSE) و Mean Absolute Error (MAE) استفاده شده است.

### یافته‌ها

در نمودار ۱، نمودار خروجی الگوریتم ژنتیک برای ۱۰۰ تکرار، بر حسب معیار MSE نشان داده شده است. جمعیت اولیه ۲۰، نرخ



نمودار ۱: خروجی الگوریتم ژنتیک



نمودار ۲: خروجی الگوریتم SFS

آفریقایی می‌باشد، این پژوهش بر پایه Multivariable Regression Model (MRM) و شبکه‌های عصبی مصنوعی (ANN) و (SVR) Support Vector Regression، برای پیش‌بینی دوز، ارائه می‌دهد. نشان داده شده است که MRM از نظر درستی نسبت به روش‌های دیگر بهینه‌تر است.<sup>۱۲</sup>

در پژوهشی دیگر، با اینکه نشان داده می‌شود روش Multi-Layer Perceptron (MLP) نسبت به روش رگرسیون خطی عملکرد بهتری دارد، اما به این نتیجه دست می‌یابد که بدون استفاده از اطلاعات ژنتیکی نمی‌توان به دقت مناسبی دست یافت.<sup>۱۳</sup>

از آنجایی که برای تشخیص دوز مصرفی بیمار، به تجربه بالای پزشک نیاز می‌باشد، مدل ارائه شده می‌تواند در پیش‌بینی آن نقش مهمی داشته باشد. هدف از این تحقیق، ابتدا انتخاب ویژگی‌های مهم بر اساس الگوریتم‌های ژنتیک و بهینه‌سازی ازدحام ذرات و سپس پیش‌بینی توسط شبکه‌های عصبی مصنوعی بوده است. براساس نتایج به‌دست آمده، با به‌کارگیری الگوریتم بهینه‌سازی ازدحام ذرات، به خروجی بهتری دست یافتیم. همچنین الگوریتم ژنتیک به‌کار گرفته شده نیز تا حدود زیادی به تعداد تکرار بستگی دارد. در آینده نیز می‌توان از جایگزینی شبکه‌های عصبی با الگوریتم‌های دیگر، به‌منظور ارزیابی مدل بهره برد.

سپاسگزاری: این مقاله حاصل طرح تحقیقاتی تحت عنوان "به‌کارگیری روش‌های داده‌کاوی به‌منظور تشخیص دوز وارفارین در بیماران تحت درمان بیمارستان مرکز قلب تهران" مصوب دانشگاه علوم پزشکی و خدمات درمانی تهران در سال ۱۳۹۳ به کد ۷۴۷ می‌باشد که با حمایت دانشگاه علوم پزشکی و خدمات درمانی تهران اجرا شده است. نویسندگان بدین وسیله مراتب قدردانی و تشکر را از جناب آقای شایان ضیائی و سرکار خانم لیلا پورقلی که در جمع‌آوری داده‌ها همکاری داشته‌اند، اعلام می‌دارند.

در این مطالعه، از بین ۱۹ ویژگی تأثیرگذار در تعیین دوز وارفارین، ویژگی‌های مهم را شناسایی کرده و به‌دقت مناسبی در میزان دوز مصرفی دست یافتیم. پژوهش حاضر نشان می‌دهد که با تعداد ویژگی‌های کمتری می‌توان به پیش‌بینی دوز وارفارین دست یافت. تحقیقات گذشته، تأثیرگذاری اکثر ویژگی‌های شناسایی شده در این مقاله را اثبات می‌کند. به‌طور کلی با کاهش ویژگی‌ها و شناسایی عوامل مؤثر و با صرف زمان و هزینه کمتر، به تشخیص سریعتری دست پیدا می‌کنیم. این تحقیق کمک زیادی به پزشک در تشخیص دوز مصرفی بیمار می‌کند. از این مدل می‌توان به‌عنوان یک سیستم پشتیبان تصمیم برای پزشک معالج استفاده نمود. در ادامه به شرح کارهای انجام گرفته در زمینه انتخاب ویژگی و پیش‌بینی دوز وارفارین می‌پردازیم. در روشی که بر پایه الگوریتم ژنتیک و در رابطه با تابع ارزیابی فازی ارائه شده است، روش فازی ارائه شده، به‌منظور ارزیابی کیفیت ویژگی‌های انتخاب شده مورد استفاده قرار می‌گیرد. در مقایسه با الگوریتم‌های دسته‌بندی، روش مطرح شده دارای پیچیدگی کمتری می‌باشد.<sup>۹</sup> در پژوهشی دیگر از روش انتخاب ویژگی بر روی داده‌های ماموگرافی به‌منظور تشخیص سرطان سینه استفاده شده است. الگوریتم ژنتیک به‌عنوان روش جستجو و شبکه‌های عصبی و روش‌های آماری برای ارزیابی استفاده شده است. در نتیجه آن شبکه‌های عصبی به‌دقت بالاتری دست یافته است.<sup>۱۰</sup> در پژوهشی که بر پایه شبکه‌های بی‌زین انجام شد، یک سیستم پشتیبان تصمیم‌گیری به‌منظور تنظیم نگهداری دوز وارفارین و فواصل پیگیری بیمار، طراحی گردید. در نتیجه این تحقیق، به دقت ۸۴ درصدی برای تنظیم دوز و ۷۹٪ برای فواصل پیگیری دست یافت.<sup>۱۱</sup> پژوهشی دیگر بر پایه روش‌های یادگیری ماشین برای بیماران با نژاد آمریکایی و

## References

1. US Food and Drug Administration. FDA approves updated warfarin (Coumadin) prescribing information; new genetic information may help providers improve initial dosing estimates of the anticoagulant for individual patients [Internet]. 2007 Aug 16 [cited 2016 Mar 15]; Available from: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2007/ucm108967.htm>
2. Hirsh J, Fuster V, Ansell J, Halperin JL; American Heart Association/American College of Cardiology Foundation. American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. *J Am Coll Cardiol* 2003;41(9):1633-52.
3. Bussey HI. How to monitor the dosage of warfarin. *Heart Dis Stroke* 1993;2(5):388-92.

4. Odén A, Fahlén M. Oral anticoagulation and risk of death: a medical record linkage study. *BMJ* 2002;325(7372):1073-5
5. Carlquist JF, Horne BD, Muhlestein JB, Lappé DL, Whiting BM, Kolek MJ, et al. Genotypes of the cytochrome p450 isoform, CYP2C9, and the vitamin K epoxide reductase complex subunit 1 conjointly determine stable warfarin dose: a prospective study. *J Thromb Thrombolysis* 2006;22(3):191-7.
6. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157-82.
7. Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng* 2014;40(1):16-28.
8. Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. *Pattern Recogn Lett* 1994;15:1119-25.
9. Chakraborty B. Genetic algorithm with fuzzy fitness function for feature selection. In: Proceedings of the 2002 IEEE International Symposium on Industrial Electronics, ISIE, 2002.
10. Zhang P, Verma B, Kumar K. Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection. *Pattern Recogn Lett* 2005;26(7):909-19.
11. Yet B, Bastani K, Raharjo H, Lifvergrenb S, Marshc W, Bergman B. Decision support system for Warfarin therapy management using Bayesian networks. *Decis Support Sys* 2013;55(2):488-98.
12. Sharabiani A, Darabi H, Bress A, Cavallari L, et al. Machine Learning Based Prediction of Warfarin Optimal Dosing for African American Patients. IEEE International Conference on Automation Science and Engineering (CASE), 2013.
13. Solomon I, Maharshak N, Chechik G, Leibovici L, Lubetsky A, Halkin H, et al. Applying an artificial neural network to warfarin maintenance dose prediction. *Isr Med Assoc J* 2004;6(12):732-5.

## Effective feature selection of clinical and genetic to predict warfarin dose using artificial neural network

Mohammad Karim Sohrabi  
Ph.D.\*  
Alireza Tajik M.Sc. Student.

Department of Computer Engineering,  
Semnan Branch, Islamic Azad  
University, Semnan, Iran.

### Abstract

Received: 17 Sep. 2015 Accepted: 15 Dec. 2015 Available online: 02 Apr. 2016

**Background:** Warfarin is one of the most common oral anticoagulant, which role is to prevent the clots. The dose of this medicine is very important because changes can be dangerous for patients. Diagnosis is difficult for physicians because increase and decrease in use of warfarin is so dangerous for patients. Identifying the clinical and genetic features involved in determining dose could be useful to predict using data mining techniques. The aim of this paper is to provide a convenient way to select the clinical and genetic features to determine the dose of warfarin using artificial neural networks (ANN) and evaluate it in order to predict the dose patients.

**Methods:** This experimental study, was investigate from April to May 2014 on 552 patients in Tehran Heart Center Hospital (THC) candidates for warfarin anticoagulant therapy within the international normalized ratio (INR) therapeutic target. Factors affecting the dose include clinical characteristics and genetic extracted, and different methods of feature selection based on genetic algorithm and particle swarm optimization (PSO) and evaluation function neural networks in MATLAB (MathWorks, MA, USA), were performed.

**Results:** Between algorithms used, particle swarm optimization algorithm accuracy was more appropriate, for the mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE) were 0.0262, 0.1621 and 0.1164, respectively.

**Conclusion:** In this article, the most important characteristics were identified using methods of feature selection and the stable dose had been predicted based on artificial neural networks. The output is acceptable and with less features, it is possible to achieve the prediction warfarin dose accurately. Since the prescribed dose for the patients is important, the output of the obtained model can be used as a decision support system.

**Keywords:** computer systems, neural networks, medical informatics application, warfarin.

\* Corresponding author: Semnan, Damghan 5 km Road, Islamic Azad University of Semnan, Semnan, Iran.  
Tel: +98- 23- 33654040  
E-mail: amir\_sohraby@yahoo.com