Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2023; 9(4): 209-229
10.34172/jhbmi.2023.02

Development of a Pharmacogenomics Model based on Support Vector Regression with Optimal Features Selection Approach to Determine the Initial Therapeutic Dose of Warfarin Anticoagulant Drug

Maghsoudi Rouhollah[1], Mirzarezaee Mitra[2]*, Sadeghi Mehdi[3], Najar-Araabi Babak[4]

• **Received:** 8 Sep 2022 • **Accepted:** 21 Dec 2022

Introduction: Using artificial intelligence tools in pharmacogenomics is one of the latest bioinformatics research fields. One of the most important drugs that determining its initial therapeutic dose is difficult is the anticoagulant warfarin. Warfarin is an oral anticoagulant that, due to its narrow therapeutic window and complex interrelationships of individual factors, the selection of its optimal dose is challenging.

Method: Inaccuracy in determining the initial dose of warfarin will simply lead to thrombosis or severe bleeding and ultimately, patient death. Among the relatively successful methods of kernel-based estimation, comparison and identification of suitable kernels have not been researched. In the present research, while carefully examining this approach, different features of selection algorithms were analyzed based on expert opinions, and an appropriate subset of efficient predictor variables was identified for dose estimation.

Results: In the current study, a dataset collected by the International Warfarin Consortium was used. The results showed that the support vector machine with a suitable kernel and a subset of the proposed features can successfully predict the ideal dose of warfarin for a significant percentage of patients with an error of approximately 0.7 mg per week.

Conclusion: The estimation was conducted using the least squares version of the support vector regression based on a suitable kernel and feature selection strategy. In this way, a better approach for predicting the optimal therapeutic dose of warfarin was presented, which can significantly reduce the wrong dose error and its consequences.

Keywords: Pharmacogenomics, Initial Warfarin Dose Estimation, Feature Selection, Least Squares Support Vector Regression

• **Citation:** Maghsoudi R, Mirzarezaee M, Sadeghi M, Najar-Araabi B. Development of a Pharmacogenomics Model based on Support Vector Regression with Optimal Features Selection Approach to Determine the Initial Therapeutic Dose of Warfarin Anticoagulant Drug. Journal of Health and Biomedical Informatics 2023; 9(4): 209-29. [In Persian] doi: 10.34172/jhbmi.2023.02

1. PhD Student, Deportment of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
2. Assistant Professor, Department of Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
3. Associate Professor, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran
4. Full Professor, School of Electrical and Computer Engineering, University College of Engineering, University of Tehran, Iran
***Corresponding Author:** Mitra Mirzarezaee
Address: Science and Research Branch, Daneshgah Blvd., Simon Bolivar Blvd., Tehran, Iran
• **Tel:**021-44865179-82 • **Email:** mirzarezaee@srbiau.ac.ir

مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره نهم شماره چهارم. ص ۲۲۹-۲۰۹، ۱۴۰۱
10.34172/jhbmi.2023.02

توسعه مدل فارماکوژنومیکس مبتنی بر رگرسیون بردار پشتیبان با رویکرد انتخاب ویژگی‌های بهینه جهت تعیین دوز اولیه درمانی داروی ضد انعقادی وارفارین

روح‌اله مقصودی[۱]، میترا میرزارضایی[۲*]، مهدی صادقی [۳]، بابک نجار اعرابی[۴]

• دریافت مقاله: ۱۴۰۱/۶/۱۷ • پذیرش مقاله: ۱۴۰۱/۹/۳۰

مقدمه: فارماکوژنومیکس و استفاده از ابزارهای هوش مصنوعی در آن یکی از جدیدترین زمینه‌های تحقیقاتی بیوانفورماتیک است. یکی از داروهای بسیار مهم که تعیین دوز اولیه درمانی آن کار مشکلی است، داروی ضدانعقادی وارفارین (Warfarin) می‌باشد. وارفارین یک داروی ضد انعقاد خوراکی است که انتخاب دوز بهینه آن به دلیل پنجره درمانی باریک و روابط پیچیده فاکتورهای فردی، چالش برانگیز است. هدف این پژوهش تعیین دوز اولیه بهینه می‌باشد.

روش: تعیین دوز اولیه وارفارین با درجه صحت نامناسب، به سادگی منجر به ترومبوز و یا خونریزی شدید و در نهایت مرگ بیمار خواهد شد. در میان روش‌های مبتنی بر کرنل، مقایسه و شناسایی کرنل مناسب مورد بحث قرار نگرفته است. در این پژوهش ضمن بررسی دقیق این رویکرد، الگوریتم‌های مختلف انتخاب ویژگی را مورد آنالیز قرار داده و با تکیه به نظر خبرگان، زیرمجموعه مناسب از متغیرهای پیش‌بین مؤثر جهت تخمین دوز شناسایی خواهد شد.

نتایج: در این مطالعه از مجموعه داده‌ای جمع‌آوری شده توسط کنسرسیوم بین‌المللی وارفارین استفاده شده است. نتایج نشان می‌دهد که ماشین بردار پشتیبان با کرنل مناسب و زیرمجموعه ویژگی‌های پیشنهادی قادر است به طور موفقیت‌آمیزی دوز ایده‌آل وارفارین را برای درصد قابل توجهی از بیماران با خطایی حدود ۰/۷ میلی‌گرم در هفته پیش‌بینی کند.

نتیجه‌گیری: تخمین با نسخه حداقل مربعات رگرسیون بردار پشتیبان مبتنی بر کرنل مناسب و با یک استراتژی مناسب انتخاب ویژگی صورت گرفت. به این روش، رویکرد بهتری برای پیش‌بینی دوز بهینه درمانی وارفارین ارائه شده است که قادر است خطای دوزهای اشتباه و عواقب ناشی از آن را به طور قابل ملاحظه‌ای کاهش دهد.

کلیدواژه‌ها: فارماکوژنومیکس، تخمین دوز اولیه وارفارین، انتخاب ویژگی، رگرسیون بردار پشتیبان حداقل مربعات

• **ارجاع:** مقصودی روح اله، میرزارضایی میترا، صادقی مهدی، نجار اعرابی بابک. شناسایی برنامه‌های کامپیوتری بازتوانی شناختی مؤثر در ارتقاء توجه در کودکان و نوجوانان مبتلا به اختلال بیش فعالی- نقص توجه. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۱؛ ۹(۴): ۲۹-۲۰۹. doi: 10.34172/jhbmi.2023.02

۱. دانشجوی دکتری، گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۲. استادیار، گروه مهندسی کامپیوتر، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۳. دانشیار، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، تهران، ایران
۴. استاد تمام، دانشکده مهندسی برق و کامپیوتر، دانشکدگان فنی، دانشگاه تهران، تهران، ایران

* **نویسنده مسئول:** میترا میرزارضایی

آدرس: تهران، انتهای بزرگراه شهید ستاری، میدان دانشگاه، بلوار شهدای حصارک، دانشگاه آزاد اسلامی واحد علوم وتحقیقات

• **شماره تماس:** ۰۲۱-۴۴۸۶۵۱۷۹-۸۲ • **Email:** mirzarezaee@srbiau.ac.ir

مقدمه

امروزه یکی از حوزه‌های مهم علم پزشکی، پزشکی شخصی (Personalized Medicine) می‌باشد. پزشکی شخصی تجویز مقدار (دوز) مناسب یک دارو به بیمار متناسب با ویژگی‌های ژنتیکی و بالینی‌اش می‌باشد [۱]. اسنیپ‌ها SNP(single-nucleotide polymorphism) امروزه به عنوان عامل اصلی تنوع ژنتیکی انسان (Human Genetic Variability) به رسمیت شناخته می‌شوند و در حال حاضر یک منبع ارزشمند برای نگاشت صفات ژنتیکی پیچیده هستند [۲]. هزاران تغییر در DNA شناسایی شده است که با بیماری‌ها و صفات ژنتیکی در ارتباط هستند. با ترکیب دستاوردهای حوزه‌های ژنتیکی، فنوتایپی و پاسخ‌های دارویی است که پزشکی شخصی قادر خواهد بود تا درمان مؤثری به ازای ژنوتیپ خاص هر بیمار ارائه نماید. مرور تاریخ پزشکی متأسفانه تجویز داروهایی با پیامدهای ناخواسته را یادآور می‌شود؛ به عنوان نمونه، در دهه ۸۰، داروی درمان آنژین (Angine) تحت عنوان پرهگزیلین (Perhexiline)، سبب ایجاد مسمومیت کبدی و عصبی در تعدادی از بیماران شد [۴، ۳]. دانشمندان بعدها متوجه شدند که این مسمومیت در افرادی با یک چندریختی نادر و کمیاب تحت عنوان CYP2D6 اتفاق افتاده است. CYP2D6 آنزیمی است که در متابولیزم دارویی دخیل است. این ژن مسئول متابولیزم طیف وسیعی از داروهایی است که غالباً استفاده می شوند. ژنتیک نه تنها نقش اساسی در رخدادهای نامساعد و بد ایفا می‌کند، بلکه بر روی دوز دارویی بهینه یک فرد نیز تأثیرگذار است [۴، ۳]. به بیان دیگر می‌توان گفت به منظور انتخاب یک دارو و دوز مصرف آن، غالباً پزشکان از فاکتورهای بالینی نظیر سن، وزن یا بیماری ارگانی از بدن که درگیر است، استفاده می‌کنند و متأسفانه تفاوت‌های فردی مانند اطلاعات ژنتیکی که ممکن است در انتخاب یا دوز مصرف دارو تأثیر بگذارد، در بسیاری مواقع در نظر گرفته نمی‌شود. توانایی درک برخی از علت‌های نهفته پاسخ یا عدم پاسخ بیمار به یک داروی خاص یکی از زمینه‌های توسعه اخیر بوده است. فارماکوژنومیکس مسئول بررسی تأثیر اطلاعات ژنتیک انسانی بر روی پاسخ داروها است و هدف آن بهبود اثر بخشی دارو و کاهش عوارض جانبی آن می‌باشد [۵، ۳].

در میان انواع مختلف داروها، وارفارین گسترده‌ترین و مورداعتمادترین داروی ضدانعقادی است که در سرتاسر دنیا مورد استفاده قرار می‌گیرد. تخمین دوز اولیه صحیح وارفارین کار آسانی نیست و تعیین ناصحیح دوز منجر به عواقب جدی و شاید حتی مرگ شود. تحقیقات گسترده‌ای توسط کنسرسیوم بین‌المللی وارفارین (International Warfarin Pharmacogenetics Consortium)IWPC شامل جمع‌آوری یک گروه (cohort) غنی از جمعیت‌های مختلف و در ادامه اجرای الگوریتم‌های متعدد آماری و هوش مصنوعی برای تعیین دوز وارفارین صورت گرفته است[۶]؛ اساساً تخمین دوز صحیح اولیه وارفارین به صورت تجربی همراه با خطا و کار مشکلی است. علت مرگ درصد قابل توجه بیمارانی که جراحی هایی مرتبط با مشکلات قلبی انجام داده‌اند، به تعیین دوز ناصحیح وارفارین ارتباط دارد. مداخله فاکتورهای ژنتیکی شخص در کنار فاکتورهای بالینی و توسعه استراتژی که بتواند فارغ از خطاهای انسانی به تعیین دوز بهینه دست یابد، بسیار ضروری به نظر می‌رسد. مطالعات انجام شده در زمینه داروی وارفاین منجر به ارائه الگوریتم‌های بهبود یافته با امتیازاتی نسبت به الگوریتم بالینی محض مرسوم شده‌اند [۷، ۶].

یکی از برجسته‌ترین پژوهش‌ها در حوزه فارماکوژنومیکس و پیش‌بینی پاسخ دارویی، تحقیقی است که توسط IWPC برای تعیین دوز مناسب تحت یک جمعیت بسیار بزرگ از چندین کشور با ویژگی‌های ژنتیکی متنوع انجام شده است. بسیاری از پژوهش‌هایی که در حوزه تخمین دوز وارفارین کار کردند، این مطالعه را مرجع اصلی کارشان قرار داده‌اند و نتایجشان را با نتایج حاصله این تحقیق مقایسه نمودند [۶]. در این مطالعه، دوز مصرف وارفارین با استفاده از داده‌های بالینی و ژنتیکی بیماران تخمین زده شده است. استراتژی مشخصی برای انتخاب ویژگی‌ها معرفی نشده است. مدل نهایی مبتنی بر رگرسیون بردار پشتیبان با کمترین خطا انتخاب شد. عملکرد مدل فارماکوژنومیکس ارائه شده با یک مدل بالینی و یک مدل با دوز تجربی (دوز شروع ثابت ۵ mg/day) مورد مقایسه قرار گرفت که در هر دو معیار MAE و R^2 عملکرد بهتری داشته است.

Anzabi Zadeh و همکاران یک مدل مبتنی بر یادگیری تقویتی عمیق برای تخمین دوز به خصوص برای بیماران حساس به وارفارین ارائه کردند. برای غلبه بر مسئله حجم نمونه نسبتاً کوچک در کارآزمایی‌های دوز، از مدل فارماکوکینتیک/فارماکودینامیک(pharmacokinetic/phar macodynamic)PK/PD وارفارین برای شبیه‌سازی دوز پاسخ بیماران مجازی استفاده کردند. به‌کارگیری الگوریتم پیشنهادی بر روی بیماران آزمایش مجازی نشان می‌دهد

[DOI: doi: 10.34172/jhbmi.2023.02]

که این مدل از مجموعه‌ای از پروتکل‌های دوز پذیرفته شده بالینی با یک حاشیه بسیار بهتر عمل می‌کند [۸]. Bontempi به پایش INR(International Normalized Ratio) پرداختند که یک جزء ضروری برای مدیریت درمان بیماری ترومبوتیک است. این مطالعه یک مدل نیمه تجربی INR را به عنوان تابعی از زمان و درمان‌های اختصاص داده شده (وارفارین و ویتامین k) ارائه می‌کند. با توجه به سایر روش‌ها، این مدل قادر به توصیف INR با استفاده از تعداد محدودی از پارامترها است و قادر به توصیف تغییرات زمانی INR توصیف‌شده هم می‌باشد. روش ارائه شده دقت زیادی در کالیبراسیون مدل نشان داد [(صحت (دقت)]: ۰/۲٪ (۰/۱٪) تا ۱/۲٪ (۰/۳٪) برای فاکتورهای انعقادی، از ۵٪ (۹٪) تا ۹/۷٪ (۱۲٪) برای وارفارین. پارامترهای مرتبط و ۳۸٪ (۴۰٪) برای پارامترهای مرتبط با ویتامین K. مقادیر اخیر دو نتیجه مهم دربردارد: اول این که توانست INR را به درستی تخمین بزند. دوم این که یک پارامتر نیمه‌تجربی عددی را معرفی می‌کند که قادر است پاسخ INR/دوز را با شرایط فیزیولوژیکی و محیطی بیماران مرتبط کند [۹].

Liu و همکاران یک چارچوب مبتنی بر یادگیری گروهی برای تخمین دوز نگهدارنده وارفارین با رویکرد ایجاد متغیرهای متقاطع (با تکنیک ضرب کارتزین) در یک مجموعه داده‌ای ناقص ارائه نمودند. مجموعه داده‌ای مورد استفاده از یک گروه بیماران تحت نظر بیمارستان شین‌هوا وابسته به دانشکده پزشکی دانشگاه شانگهای جمع‌آوری شده بود. ابتدا متغیرهای منفرد با یک تحلیل تک متغیره بررسی شدند و تنها متغیرهای معنی‌دار آماری را (p-value<۰/۰۵) شامل می‌شوند. سپس یک روش مهندسی ویژگی جدید را برای تولید خودکار متغیرهای متقاطع پیشنهاد دادند. تأثیر هر کدام را با روش رگرسیون گام به گام ارزیابی نموده و تنها ویژگی‌های مؤثر و مهم انتخاب می‌شوند. مدل ساخته شده برای زیرگروه‌هایی با دوز متوسط و بالا خوب عمل کرده است و این که در زیرجمعیت چینی مشابه از مدل IWPC عملکرد بهتری داشته است. از مجموعه داده‌ای محلی با ۳۷۷ بیمار توانستند به R^2 با مقدار متوسط حدوداً ۷۵٪ برسند [۱۰].

Truda و Marais یک چارچوب جدید نرم‌افزاری مبتنی بر پایتون برای ارزیابی مدل‌های تخمین دوز وارفارین روی مجموعه داده‌های متعدد توسعه دادند. این مطالعه دقت امیدوارکننده‌ترین الگوریتم‌ها را در مجموعه داده‌های IWPC و مجموعه داده‌های بالینی جدید بیماران آفریقای جنوبی ارزیابی کرد. بردارهای پشتیبان و رگرسیون خطی بهترین عملکردها را در هر دو مجموعه داده داشتند، در حالی که شبکه‌های عصبی یکی از بدترین عملکردها در هر دو مجموعه داده را داشتند [۱۱]. Xie و همکاران، ۱۸ روش مختلف تعیین دوز وارفارین را برای یک گروه از بیماران چینی مورد مقایسه قرار دادند و به الگوریتم ایده‌آلی که در همه معیارهای ارزیابی برتری داشته باشد، دست نیافتند. با این وجود نشان دادند الگوریتم‌های مبتنی بر جمعیت جغرافیایی ممکن است برای پیش‌بینی دوزهای پایدار وارفارین در بیماران محلی مناسب‌تر باشند [۱۲].

Altay و همکاران مسئله تعیین دوز وارفارین را از دید گسسته و در قالب یک مسئله دسته‌بندی حل کردند. مجموعه داده‌ای از بیماران ترک هم شامل داده‌های ژنتیکی و هم غیرژنتیکی می‌باشد و از دو الگوریتم بیزین و (K-Nearest Neighbors) KNN استفاده کردند و به ترتیب به صحت دسته‌بندی ۵۰/۰۱٪ و ۵۰/۵۲٪ دست یافتند [۱۳]. Sharabiani و همکاران یک سیستم کامپیوتری برای تعیین محدوده کاربرد دوز بالینی وارفارین توسعه دادند. این الگوریتم هوش مصنوعی از یک ماشین بردار پشتیبان با کرنل چندجمله‌ای استفاده نموده است. مجموعه داده‌ای مورد استفاده همان IWPC بوده و چندین روش دسته‌بندی شامل شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان با کرنل‌های خطی، چندجمله‌ای و گاوسی مورد بررسی قرار گرفتند و بهترین نتیجه برای ماشین بردار پشتیبان با کرنل چندجمله‌ای با میزان صحت ۶۹/۷٪ به‌دست آمده است. فرآیند خاصی برای رویه انتخاب ویژگی اتخاذ نکرده بودند و مدل توسعه داده شده صرفاً در برگیرنده ویژگی‌های بالینی بوده است و ویژگی‌های ژنوتایپ را در نظر نگرفته بودند [۱۴].

Rad و همکاران [۱۵] تأثیر فاکتورهای دموگرافیک و چندریختی ژن VKORC1 1639 G>A را بر روی یک جمعیت از بیماران ایرانی که تحت درمان وارفارین بودند بررسی کردند. جمعیت مورد مطالعه شامل ۹۵ بیمار با میانگین سنی تقریباً ۶۲ سال و واریانس حدوداً ۱۳ سال می‌باشد. نتیجه این پژوهش حاکی از آن است که بیماران مسن‌تر به دوز درمانی کمتری از وارفارین به نسبت افراد کم سن‌تر نیاز دارند. مشابه این پژوهش، در خاورمیانه هم چند کار علمی در خصوص تأثیر چندریختی‌های CYP2C9 (آنزیم CYP2C9 آنزیم اصلی دخیل در آشکارسازی وارفارین می‌باشد، در حالی‌که VKORC1 برای باز تولید ویتامین k کاهش یافته حیاتی است.) و VKORC1

[DOI: doi: 10.34172/jhbmi.2023.02]

بر روی بیماران پاکستانی و فلسطینی صورت گرفته است [۱۷، ۱۶]. Ayesh و همکاران [۱۶] یک الگوریتم برای پیش‌بینی دوز پایدار وارفارین مبتنی بر ژنوتایپ‌های CYP2C9*2 و VKORC1 -1639 G>A ارائه نمودند. سه الگوریتم فارماکوژنتیک، بالینی و دوز ثابت بر روی ۱۰۱ بیمار فلسطینی به کار گرفته شد. الگوریتم IWPC از دو الگوریتم دیگر عملکرد بهتری داشته است. عملکرد الگوریتم فارماکوژنومیکس برای دو معیار MAE و R^2 به ترتیب ۸/۹ و ۰/۳۵۰ بوده است.

Cho و همکاران [۱۸] از یک الگوریتم رگرسیون برای پیش‌بینی دوز وارفارین در گروهی از بیماران کره‌ای استفاده نمودند. این گروه مشتمل بر نمونه‌هایی از ۱۲۹ بیمار می‌باشد که در بازه INR ۱/۵–۳ قرار دارند. صحت پیش‌بینی الگوریتم با ۹ الگوریتم این حوزه از جمله IWPC، مقایسه شد. بهترین نتیجه توسط رگرسیون خطی چندمتغیره (multiple linear regression)MLR حاصل شد. برای تحلیل MLR، چهار متغیر شامل Age، Body Weight، ژنوتایپ‌های CYP2C9*3 و VKORC1 1173 انتخاب شدند. نتایج کار این پژوهش در مقایسه با بهترین نتیجه، ۰/۰۵ در معیار R^2 بهبود داشته است. Pavani و همکاران [۱۹] به دنبال نشان دادن تأثیر انواع Vitamin K (K1, K2, K3) در تعیین دوز وارفارین بودند. از یک مدل نوروفازی (Neuro-fuzzy) برای پیش‌بینی دوز وارفارین استفاده کردند. ۱۱۵ بیمار (۶۳ مرد و ۵۲ زن) در گروه سنی ۷۵–۱۲ سال برای این مطالعه به خدمت گرفته شدند. مدل نوروفازی ایجاد شده به (mean squared error) MSE حدود ۰/۰۰۰۰۲۹رسیده است.

Hamberg و همکاران [۲۰] به جای تخمین دوز اولیه وارفارین از یک رویه بیزین برای پیش‌بینی INR استفاده نمودند. در حقیقت یک نرم‌افزار تحت محیط جاوا توسعه دادند که با دریافت ورودی‌هایی از هر دو دسته بالینی و ژنوتایپ نظیر سن، وزن و چندریختی‌های CYP2C9 و VKORC1 اقدام به پیش‌بینی INR به عنوان خروجی می‌کند. تمامی پژوهش‌هایی که در بالا اشاره شد سعی کردند از هر دو دسته ویژگی‌های بالینی و ژنتیکی برای توسعه مدل پیش‌بین استفاده کنند؛ اما Sharabiani و همکاران [۲۱] یک روش جدید در تعیین دوز وارفارین برای افراد بالغ پیشنهاد دادند که فقط متمرکز به استفاده از متغیرهای بالینی می‌باشد. ابتدا بیماران با استفاده از (Relevance Vector RVM(Machine به دو دسته تفکیک شدند؛ دسته اول بیمارانی است که به دوز بالا یعنی mg/week ۳۰< و دسته دوم هم به دوز پایین یعنی mg/week ۳۰≥ نیازمندند. در فاز بعدی، دوز بهینه هر بیمار توسط دو مدل بالینی رگرسیون پیش‌بینی شد. مجموعه داده‌ای مورد استفاده IWPC بود. مدل پیشنهادی بالینی در دو معیار MAE و RMSE به ترتیب ۸٪ و ۱۶٪ نسبت به IWPC بالینی بهبود داشته است. در پژوهش Karaca و همکاران [۲۲] یک مجموعه داده‌ای جدید از جمعیت ترک شامل فاکتورهای ژنتیکی و غیرژنتیکی جمع‌آوری شد. پارامترهای تأثیرگذار با یک تحلیل رگرسیون تک متغیره انتخاب شدند. ژنوتایپ‌های VKORC1، سن، BSA و Aortic Valve Replacement پیش‌بین‌های مهم تعیین دوز بودند البته در افرادی که سابقه Hemorrhagic یا Thromboembolic نداشتند. سپس از رویکرد MLR برای تعریف بهترین مدل تخمین دوز وارفارین استفاده گردید. روش حاضر با تحقیقاتی که غالباً مبتنی بر مجموعه داده‌ای شامل بیش از ۱۰۰۰ نمونه بودند، مقایسه شد. این پژوهش توانست دوز ناصحیح بیمارانی با سابقه اتفاقات Hemorrhagic عمده را تشخیص دهد. Santos و همکاران [۲۳] از MLR برای پیش‌بینی دوز وارفارین در دو گروه از بیماران برزیلی استفاده کردند. گروه اول شامل ۸۳۲ بیمار برای ساخت مدل و گروه دوم شامل ۱۳۳ بیمار برای اعتبارسنجی می‌باشد. از ۹ ویژگی جنسیت، سن، BMI، نژاد/رنگ، درصد افراد سیگاری، اندیکاسیون دارویی و ژنتیکی استفاده شده است. الگوریتم توسعه داده شده وقتی بر روی یک جمعیت برزیلی متمرکز می‌شود، از صحت بالاتری نسبت به IWPC برخوردار است. این الگوریتم در معیار R^2 حدود ۱۰٪ از IWPC دقیق‌تر بوده است.

Isma'eel و همکاران [۲۴] از دو روش رگرسیون خطی مبتنی بر مدل کمترین مربعات (Least–squares Model) و شبکه‌های عصبی مصنوعی برای پیش‌بینی دوز مصرفی وارفارین استفاده کردند. این مدل‌ها بر روی داده‌های جمع‌آوری شده از ۱۷۴ بیمار داوطلب اعمال شدند. داده‌ها مبتنی بر فاکتورهای بالینی و ژنوتایپ هستند. نتایج مدل فارماکوژنتیک مبتنی بر ANN از دقت بهتر و مقدار R بزرگ‌تری از دیگر مدل‌های مبتنی بر LSM برخوردار می‌باشد. الگوریتم ارائه شده درصد نمونه‌های تخمینی اشتباه را ۷/۰۴٪ کاهش داد. Grossi و همکاران [۲۵] هم پیش‌بینی دوز بهینه وارفارین را با استفاده از ویژگی‌های دموگرافیک، بالینی و ژنتیکی (شامل چند ریختی‌های CYP2C9 و VKORC1) از ۳۷۷ بیمار به کمک شبکه عصبی انجام دادند. دوز وارفارینی که توسط شبکه عصبی

[DOI: doi: 10.34172/jhbmi.2023.02]

پیش‌بینی شد با دوزهای واقعی توسط رگرسیون خطی تک متغیره مقایسه شد. بهترین نتیجه برای حالت mg/week ≥۲۱ با MAE= ۳/۸۶ و R^2=۰/۶۷ به دست آمده است. روش مطرح شده در این مطالعه با سه پژوهش دیگر از جمله IWPC مقایسه شده است. این مطالعه ۷۰ درصد بیماران را به درستی دسته‌بندی کرد که حدود ۲۰ درصد از بهترین نتایج مقالات مقایسه شده بهتر بوده است.

Pavani و همکاران [۲۶] از هر دو دسته متغیرهای پیش‌بین بالینی و ژنتیکی شامل سن، جنس، BMI، سطح ویتامین k پلاسما، وضعیت تیروئید و ۱۰ متغیر ژنتیکی به عنوان ورودی یک سیستم مبتنی بر شبکه عصبی استفاده کردند. مدل توسعه داده شده دوز وارفارین را در ۷۴/۵٪ از بیمارانی که INR کمتر از ۲ و همچنین برای ۸۳/۳٪ از بیمارانی که INR بیشتر از ۳/۵ داشتند به درستی پیش‌بینی کرده بود. Öztaner و همکاران [۲۷] از یک چارچوب تخمینی بیزین برای پیش‌بینی دوز استفاده کردند. رویکرد بیزین پیشنهادی را با چندین روش مبتنی بر رگرسیون خطی مقایسه کردند. در بهترین حالت به R^2 حدوداً ۵۷٪ رسید که بهبود ۱۲٪ نسبت به IWPC داشته است.

Sharabiani و همکاران [۲۸] سه ابزار پیش‌بینی شامل رگرسیون چند متغیره، رگرسیون بردار پشتیبان و شبکه‌های عصبی مصنوعی را برای پیش‌بینی دوز بهینه وارفارین برای بیماران آفریقایی–آمریکایی استفاده نمودند. در نهایت رگرسیون چند متغیره با مقادیر ۱۴/۵۱ و ۱۲/۲ به ترتیب برای معیارهای RMSE و MAE به عنوان مدل نهایی تخمین ارائه شد. Hu و همکاران [۲۹] هم مشابه پژوهش شربیانی و همکاران [۲۱] صرفاً از ویژگی‌های بالینی برای توسعه یک مدل تخمین دوز وارفارین مبتنی بر یک رویکرد یادگیری نظارت شده استفاده کردند. مجموعه داده‌ای شامل ۵۸۷ نمونه از یک گروه تایوانی بوده است. علاوه بر استفاده از تعدادی تکنیک نظارت شده، برای دستیابی به صحت پیش‌بینی بالاتر از تکنیک‌های Bagging و Voting هم استفاده کردند. بین همه مدل‌های پیش‌بین، Bagged Voting (MAE=0.210, $\sigma(E)$ = 0.357) با چهار دسته‌بند و Bagged SVR (MAE=0.210, $\sigma(E)$ = 0.366) به خاطر MAE و $\sigma(E)$ به خاطر MAE و $\sigma(E)$ کمینه به عنوان دو مدل تخمینی مؤثر پیشنهاد شدند.

اغلب روش‌های یادگیری ماشینی یا آماری که در مدل‌سازی‌ها به کار گرفته شدند شامل RF(Random Forest)، MDR (Multifactor-Dimensionality Reduction)، ANN(Artificial Neural Network) و Support Vector Machine(SVM) هستند [۳۰]. در برخی از مطالعات هم الگوریتم مشخصی توسعه داده نشد بلکه به صورت مشخص به توسعه یک بسته نرم‌افزاری پرداخته شد. روش‌هایی که برای تخمین دوز اولیه وارفارین به کار گرفته می‌شوند روش‌های آماری یا الگوریتم‌های هوشمندی هستند که آنالیز خاصی در حوزه انتخاب ویژگی صورت نگرفته است و متناسب با ماهیت مجموعه داده‌ای شخصی‌سازی نشده است. هدف این تحقیق توسعه یک رویکرد یادگیری مبتنی بر کرنل می‌باشد تا بتواند با دقت بالاتری میزان دوز وارفارین مصرفی را تخمین بزند. این امر عوارض جانبی مصرف دارو را کاهش داده و احتمال مرگ بیماران را پایین می‌آورد. در این پژوهش تلاش شده مهم‌ترین ویژگی‌های بالینی، دموگرافیک و ژنتیکی بیمار شناسایی و طرح درمان مناسب اتخاذ شود.

روش

مجموعه داده

مجموعه داده‌ای مورد استفاده متعلق به IWPC می‌باشد. این مجموعه به همراه ابرداده (Metadata) آن از پایگاه دانش فارماکوژنومیکس http://www.pharmgkb.org/page/iwpc قابل دریافت است. مجموعه داده‌ای IWPC شامل ۵۷۰۰ نمونه و ۶۸ ویژگی می‌باشد. این مجموعه داده‌ای از ۲۱ گروه تحقیقاتی از نه کشور در چهار قاره جمع‌آوری شده است. از این‌رو مدل‌های توسعه داده شده مبتنی بر این مجموعه داده‌ای بسیار کلی‌تر (More Universal) از مدل‌های ایجاد شده از گروه‌های محلی خواهد بود و پتانسیل استفاده برای گروه وسیع‌تر و متنوع‌تری از بیماران را دارد. گروهی که در مطالعه کنسرسیوم بین‌المللی وارفارین [۶] استفاده شد، شامل ۵۰۵۲ نمونه می‌باشد که INR هدفی بین ۲ تا ۳ داشتند. این داده‌ها شامل اطلاعاتی از ویژگی‌های دموگرافیک، اندیکاسیون اولیه برای درمان وارفارین (Primary Indication for Warfarin Treatment)، دوز درمانی پایدار وارفارین (Stable Therapeutic Dose of Warfarin)، INR درمانی (INRایی که با یک دوز وارفارین پایدار به‌دست آمد)، INR هدف (Desired INR)، استفاده همزمان از چندین دارو (بر اساس این که کدام‌ها INR را افزایش می‌دهند یا کدام‌ها

[DOI: doi: 10.34172/jhbmi.2023.02]

کاهش می‌دهند، گروه‌بندی شدند)، سابقه ابتلا به بیماری‌های زمینه‌ای نظیر دیابت، وضعیت سیگاری بودن، حضور متغیرهای ژنتیکی CYP2C9(*1, *2 , *3) و VKORC1 و نژاد هستند. متغیرهای پیش‌بین مذکور به عنوان ورودی‌های سیستم تعریف خواهند شد و خروجی تخمینی هم دوز اولیه مصرفی وارفارین می‌باشد.

رگرسیون بردار پشتیبان

در سال‌های اخیر استفاده از (Support Vector Machine) SVM به دلیل قابلیت تعمیم‌پذیری فوق‌العاده‌اش بسیار مورد توجه واقع شده است [۳۳-۳۱]. یکی از مؤثرترین روش‌های یادگیری ماشین نسخه‌ای از SVM تحت عنوان (Support SVR(Vector Regression بوده است [۳۵، ۳۴]. این تکنیک در سال‌های اخیر به عنوان یک تکنیک قدرتمند برای حل مسائل رگرسیون غیرخطی و تقریب توابع ظهور و بروز پیدا کرده است. محققین فراوانی با بهینه‌سازی پارامترهای SVR به‌وسیله الگوریتم‌های فراابتکاری، ابزارهای پیش‌بین خوبی جهت پیش‌بینی در زمینه‌های مختلف ایجاد نمودند که البته مزایا و معایب خاص آن الگوریتم‌ها را نیز داشته‌اند [۴۵-۳۶]؛ اما در کنار تمامی این کاربردها و قابلیت‌ها، نسخه‌ای از SVR توسط Suykens و Vandewalle [۴۶] تحت عنوان LSSVR معرفی شده که همان نسخه حداقل مربعات SVR می‌باشد. الگوریتم LSSVR نه تنها مزایای SVR نظیر قابلیت تعمیم‌پذیری بالا را دارا است بلکه قابلیت‌های اضافه‌تری نظیر این که در حالت استفاده از کرنل پارامترهای کمتری باید بهینه شوند هم دارد؛ بنابراین LSSVR، سریع تر و مقاوم‌تر از SVR در یک کار یکسان خواهد بود. مدل نهایی LSSVR برای تخمین تابع به صورت نوشته می‌شود:

$$y = f(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x) + b \qquad (۱)$$

که متغیرهای α و b از یک سیستم خطی به دست می‌آیند. اندیس i تعداد نمونه‌های فضای مسئله می‌باشد. یکی از نقاط قوت پارادایم‌های یادگیری مبتنی بر کرنل قابلیت آن‌ها در پشتیبانی بیشتر از بازنمایی‌هایی از داده‌ها است [۴۷]. مدل غیرخطی در فضای ورودی به یک مدل خطی در فضای ویژگی مرتبط می‌گردد. انتخاب تابع هسته مناسب برای موفقیت همه الگوریتم‌های مبتنی بر هسته حیاتی است، چرا که هسته پایه‌گذار دانش اولیه و پیشین خوبی در مورد ماهیت یک مسئله است. در حقیقت اثرگذاری SVM به انتخاب تابع هسته مناسب و تخمین مناسب مقادیر پارامترهای هسته بستگی دارد. در حالت کلی یک کرنل K به صورت ضرب داخلی بین تصاویر دو بردار ورودی است که پس از یک نگاشت به یک فضای هیلبرت با ابعاد بالا (گذر از تابع تبدیل $\boldsymbol{\varphi}$) به صورت زیر تعریف می‌شود:

$$k(x_1,x_2) =< \varphi(x_1).\varphi(x_2) > \qquad (۲)$$

در جدول ۱ برخی توابع هسته معروف‌تر و پرکاربردتر اشاره شده در مطالب قبل لیست شده‌اند [۴۸]. در بین این توابع هسته، تابع گاوسین یا RBF(Radial Basis Function) شناخته شده‌ترین (البته نه همواره بهترین) انتخاب در اکثر کارهای پژوهشی است.

جدول ۱: برخی توابع هسته معروف [۴۴]، a و b ثابت هستند؛ d عدد توان چندجمله‌ای است، σ هم width تابع RBF هست.

نام تابع	فرمول‌بندی ریاضی
خطی	$k(x, x_i) = x_i^T x$
چندجمله‌ای	$k(x, x_i) = (ax_i^T x + b)^d$
سیگموئید	$k(x, x_i) = tanh(ax_i^T x + b)$
تابع پایه شعاعی گاوسی (RBF)	$k(x, x_i) = \exp(-\|x_i - x\|/2\sigma^d)$

[DOI: doi: 10.34172/jhbmi.2023.02]

البته هر یک از کرنل‌ها با توجه به ماهیتشـــان برای مســـئله خاصـــی مناســـب خواهند بود. کرنل‌های چندجمله‌ای برای داده‌های گسسته و اسمی (Discrete and Nominal) بهتر هستند. علاوه بر این، چون این مجموعه‌های داده‌ای به صورت نرمال بر اساس آزمون‌های هیستوگرام توزیع نمی‌شوند، بنابراین به نظر می‌رسد کرنل چندجمله‌ای برای داده‌هایی که به صورت نرمال توزیع نشده‌اند و شامل مقادیر گسسته و اسمی هستند، مناسب‌تر است. از طرفی کرنل گاوسی برای مجموعه داده‌ای پیوسته مناسب است [۴۹].

روش پیشنهادی

رویکرد پیشنهادی این مطالعه از پنج مرحله اساسی مشتمل بر فاز پیش‌پردازش، اســتخراج ویژگی‌های مناســب، انتخاب کرنل مناســـب، تعبیه این مدل یادگیری جدید به اســـتراتژی‌های پیش‌بین و در نهایت تخمین دوز دارو با مدل ساخته شده است. گام‌های الگوریتم پیشنهادی با جزئیات بیشتر از دیاگرام بلوکی شکل ۱ تبعیت می‌کند.

۱- پیش پردازش

گروهی که برای این مطالعه آماده شـــد، شـــامل یک زیرگروه ۵۰۵۲ بیمار از کل ۵۷۰۰ رکورد مجموعه داده‌ای IWPC بود. یعنی در ابتدا رکوردهایی انتخاب شدند که مقدار فیلد Target INR در بازه ۳-۲ قرار داشـته باشـد، یعنی نهایتاً ۵۰۵۲ رکورد در گام اول بود. قبل از ورود به فاز کدینگ فیلدهای IWPC باید چند کار انجام شود؛ رکوردهای دربرگیرنده ویژگی‌هایی که تعداد مقادیر NA کم یا بســـیار کمی دارند و ویژگی‌هایی که تعداد مقادیر NA بیش از ۵۰٪ دارند، حذف شـدند. در گام بعد، ستون‌ها و فیلدهای دموگرافیک و ژنوتایپ گسسته به صورت عددی کد شدند. ستون‌های پیوسته هم که وضعیت مشخصی دارند؛ در نهایت گام آخر پیش پردازش داده‌ها، مدیریت داده‌های گمشده و پر کردن آن‌ها می‌باشد. البته قابل ذکر است که برای پر کردن فیلدهای ژنتیکی به صورت جداگانه‌ای عمل شد. برای جایگذاری مقادیر گم شده ویژگی‌های پیوسته از میانگین همان ستون و برای داده‌های گسسته هم از مد استفاده شد و این جایگزینی تحت نرم‌افزار SPSS نسخه ۱۶ انجام گرفت.

۲- انتخاب ویژگی

مسـئله بعدی انتخاب زیرمجموعه‌ای از ویژگی‌های مناسـب‌تر اسـت. ویژگی‌های مجموعه داده‌ای در حالت کلی شـامل سـه دسـته ویژگی‌های بالینی، ژنوتایپ و دموگرافیک هسـتند که در جدول‌های ۲، ۳ و ۴ جزئیات تمام ویژگی‌ها مشـاهده می‌شـود. ویژگی‌های دموگرافیک در جدول ۲ آمده است.

جدول ۲: ویژگی‌های دموگرافیک

ویژگی	توضیح
جنسیت	مرد، زن یا ناشناخته
نژاد	بر اساس اطلاعات خوداظهاری بیماران و دسته‌های نژادی
قومیت	همانند نژاد
سن	به عنوان یک فاکتور مهم در میزان دوز مؤثر است.

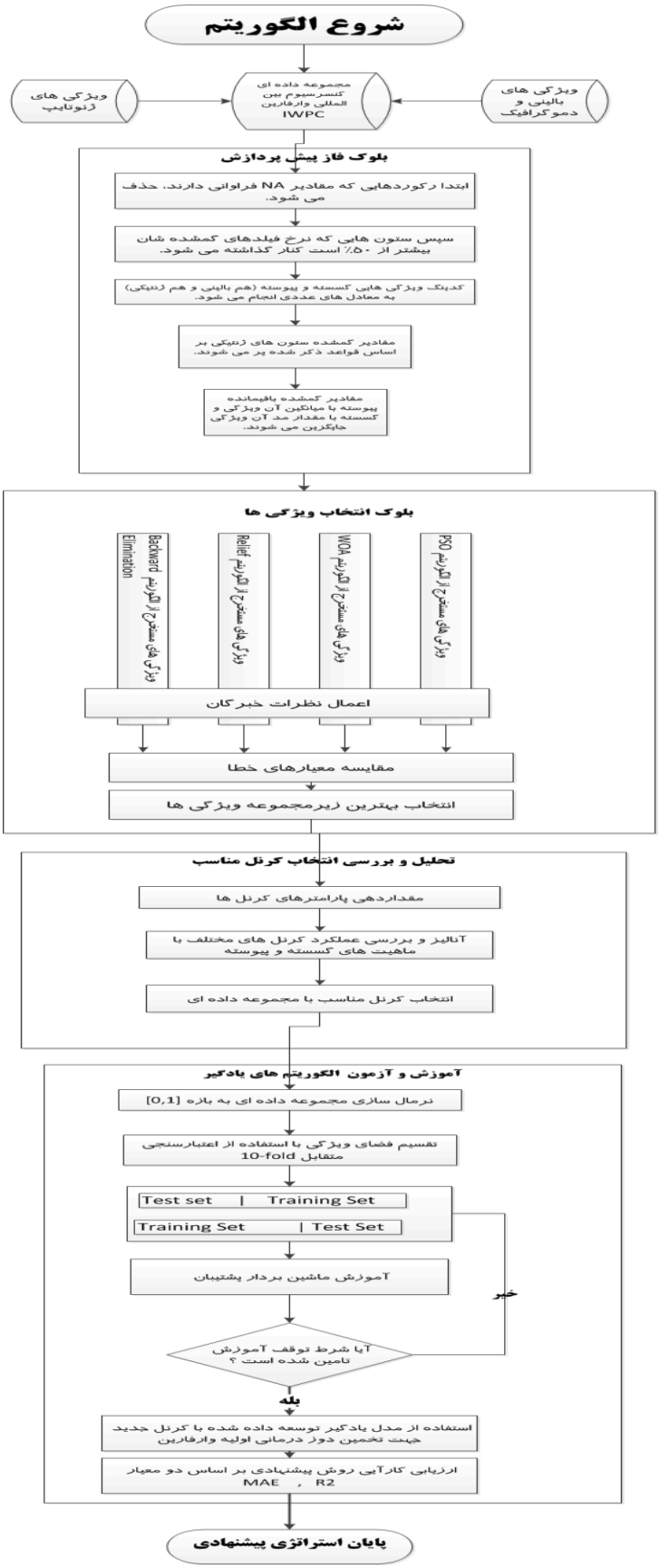
دسـته دوم ویژگی‌های فنوتیپی هسـتند که در بیشـتر تحقیقات ذیل هـمان ویـژگی‌های بـالینی ذکر می‌شـوند و مـقادیر آن در جدول ۳ نمایش داده شده است.

جدول ۳: ویژگی‌های فنوتایپ

ویژگی	توضیح (یا مقدار)
دوز درمانی وارفارین	دوزی که به بر حسب میلی‌گرم در هفته به بیمار داده می‌شود.
وضعیت سیگاری بودن	معادل بله، نه، و یا معلوم نیست.

دسته سوم ویژگی‌های بالینی هستند که در این پژوهش برای بحث دقیق‌تر در فاز انتخاب ویژگی به چهار دسته بیماری‌های همزمان، مصرف و تداخلات دارویی، مشکلات و اعمال جراحی مرتبط با قلب و سایر فاکتورهای بالینی تفکیک شده است.

[DOI: doi: 10.34172/jhbmi.2023.02]



شکل ۱: دیاگرام بلوکی رویکرد پیشنهادی

[DOI: doi: 10.34172/jhbmi.2023.02]

جدول ۴: ویژگی‌های بالینی

ویژگی		توضیح
بیماری‌های همراه (لیست بیماری‌هایی که ممکن است بیمار به آن‌ها مبتلا باشد، یک یا چند بیماری)		بیماری مزمن انسداد ریوی، هیپرلیپیدمی، فشار خون، بیماری مزمن انسداد ریوی، بیماری کبد الکلی، بیماری سل، سنگ کیسه صفرا، آرتروز، هپاتیت B، هپاتیت C، کبد چرب، افسردگی، سکته مغزی، نقرس، کم‌کاری تیروئید، پرکاری تیروئید، پروستات، سرطان
اعمال جراحی و مشکلات قلبی		نارسایی احتقانی قلب، تعویض دریچه آئورت؛ فیبریلاسیون دهلیزی، بیماری عروق کرونر، سکته، تنگی آئورت[۱]، نارسایی آئورت سه دریچه‌ای[۲]، بطن چپ گشاد شده[۳]، نارسایی روماتیسمی میترال[۴]، نارسایی قلبی[۵]، انسداد شریان[۶] وضعیت پس از تعویض دریچه میترال[۷]، بیماری قلبی عروقی فشار خون بالا
دیابت		
مصرف داروها		Aspirin, Acetaminophen, Was Dose of Acetaminophen >1300mg/day, Simvastatin, Atorvastatin, Fluvastatin, Pravastatin, osuvastatin, Cerivastatin, Amiodarone, Carbamazepine, Phenytoin, Rifampin, Sulfonamide Antibiotics, Macrolide Antibiotics including erythromycin, azithromycin, and clarithromycin, Herbal Medications, Vitamins and Supplements including garlic, ginseng, danshen, donquai, vitamins, zinc, iron, magnesium.
سایر	قد، وزن، اندیکاسیون درمان وارفارین[۸]	
	INR هدف	معادل مقدار INR هدف می‌باشد و یا مقدار ثبت نشده‌ای دارد.
	بیمار به دوز پایدار وارفارین رسید.[۹]	معادل بله، نه، و یا معلوم نیست.
	INR دوز درمانی گزارش شده[۱۰]	مقدار INR ایی که بر دوز درمانی وارفارین گزارش شد

1 - aortic stenosis
2 - trivalve aortic regurgitation
3 - dilated left ventricle
4 - rheumatic mitral insufficiency
5 - heart failure
6 - artery occlusion
7- status post mitral valve replacement
8 - Indication for Warfarin Treatment
9 - Subject Reached Stable Dose of Warfarin
10- INR on Reported Therapeutic Dose of Warfarin

دسته چهارم که بسیار مهم است، ویژگی‌های ژنوتایپ می‌باشد. همچنان‌که از جدول ۵ مشخص هست تکلیف کار با CYP2C9 روشن است؛ اما VKORC1 چندین اسنیپ مختلف دارد. غالب ارجاعات به ژنوتیپ VKORC1 برای rs9923231 اشاره دارد. در این پژوهش هم این اسنیپ به عنوان دومین ویژگی ژنوتایپ انتخاب شد و مقادیر گمشده آن بر اساس قواعد زیر پر خواهند شد. برای ویژگی‌های ژنوتایپ، از مقاله کنسرسیوم بین‌المللی وارفارین [۶] و سایر تحقیقات این حوزه آمده است که آزمایشگاه‌های مختلف، اسنیپ‌های VKORC1 را ژنوتایپ کردند و از اسنیپ rs9923231 از بین اسنیپ‌های موجود استفاده کردند. در این مطالعه هم همین رویه استفاده شد و از بین ستون‌های آلل VKORC1، اسنیپ rs9923231 انتخاب و مقادیر فیلدهای NA با استفاده از اسنیپ‌های دیگر و ویژگی Race پر می‌شود. جایگزینی مقادیر گمشده بر اساس حدوداً ده قاعده اگر–آنگاه می‌باشد. قاعده زیر یک نمونه از آن‌ها می‌باشد.

If Race is not “Black or African American” or “Missing or Mixed Race” and rs2359612=’C/C’ then impute rs9923231=’G/G’

[DOI: doi: 10.34172/jhbmi.2023.02]

جدول ۵: ویژگی‌های ژنوتایپ

ویژگی	مقادیر
ژنوتیپ‌های CYP2C9	*1, *2, *3, *4, *5, *6, *7, *8, *9, *10, *11, *12, *13 or NA
VKORC1 -1639 G>A (3673), rs9923231, G/A	A/A, A/G, G/G or NA
VKORC1 497T>G (5808), rs2884737, A/C	G/G, G/T, T/T or NA
VKORC1 1173 C>T (6484), rs9934438, A/G	C/C, C/T, T/T or NA
VKORC1 1542G>C (6853), rs8050894, C/G	C/C, C/G, G/G or NA
VKORC1 3730 G>A (9041), rs7294, A/G	A/A, A/G, G/G or NA
VKORC1 2255C>T (7566), rs2359612, A/G	C/C, C/T, T/T or NA
VKORC1 -4451 C>A (861), rs17880887, A/C	A/A, A/C, C/C or NA

تکنیک‌های انتخاب ویژگی در روشی که جستجو را در فضای اضافه شده زیرمجموعه ویژگی در انتخاب مدل ترکیب می‌کنند، متفاوت‌اند [۳۰]. در فاز انتخاب ویژگی از سه رویکرد متفاوت پوشنده (الگوریتم Backward Elimination)، الگوریتم‌های فراابتکاری (الگوریتم وال و بهینه سازی گروهی ذرات) و از دسته الگوریتم‌های مبتنی بر فیلتر از الگوریتم Relief استفاده شد که در جداول ۶، ۷ و ۸ هر کدام از این مجموعه ویژگی‌ها قابل مشاهده است. در روش Backward Elimination برای کار انتخاب ویژگی، معیارهای خطا به ترتیب با حذف هر یک از زیرمجموعه‌های فوق و تک تکِ ویژگی‌های آن‌ها توسط رگرسیون بردار پشتیبان مورد سنجش قرار گرفته است. مزایای روش پوشنده شامل تعامل بین جستجوی زیرمجموعه ویژگی و توامان انتخاب مدل و همچنین قابلیت محاسبه وابستگی‌های یک ویژگی است. ابتدا در جدول ۶ و ۷ ویژگی‌هایی که پس از اجرای الگوریتم‌های فراابتکاری وال و بهینه‌سازی گروهی ذرات به‌دست آمده است، مشاهده می‌شود.

جدول ۶: ویژگی‌های استخراج شده توسط الگوریتم بهینه‌سازی وال

دموگرافیک		سن، نژاد (Race)، قومیت (Ethnicity)، جنسیت (Gender)
بالینی	مصرف داروها	Amiodarone, Acetaminophen or Paracetamol (Tylenol), Cerivastatin (Baycol), Aspirin, Atrovastatin, Fluvastatin, Lovavastatin, Pravastatin, Rosavastatin, Carbamazepine, Sulfonamide, Antifungal, Herbal medications, INR on therapeutic dose, Indication for drug treatment
	مرتبط با مشکلات قلبی	Congestive heart failure, Valve replacement
	سایر ویژگی‌ها	قد، دیابت، Subject reached to stable dose
ژنوتایپ		CYP2C9

در ادامه ویژگی‌هایی که توسط الگوریتم بهینه‌سازی گروهی ذرات PSO (Particle Swarm Optimization) استخراج شده‌اند لیست می‌گردد. در زیرمجموعه مصرف داروهای همزمان مشابه الگوریتم وال هست، اما در سایر زیرگروه ویژگی‌ها تفاوت‌هایی دارند.

[DOI: doi: 10.34172/jhbmi.2023.02]

جدول ۷: ویژگی‌های استخراج شده توسط الگوریتم PSO

بالینی	دموگرافیک	سن (Age)، قومیت (Ethnicity)، جنسیت (Gender)
	مصرف داورها	Amiodarone, Acetaminophen or Paracetamol (Tylenol), Cerivastatin (Baycol), Aspirin, Atrovastatin, Fluvastatin, Lovavastatin, Pravastatin, Rosavastatin, Carbamazepine, Sulfonamide, Antifungal, Herbal medications, INR on therapeutic dose
	مرتبط با مشکلات قلبی	Congestive heart failure, Valve replacement
	سایر ویژگی‌ها	قد، وزن، دیابت، سیگاری بودن
ژنوتایپ		CYP2C9, VKORC1

اگر به دو جدول ۶ و ۷ توجه شـــود فقدان برخی ویژگی‌های ا سا سی م شهود ا ست. به عنوان مثال در ویژگی‌های ا ستخراج شـــده توســـط الگوریتم وال وزن، ژن VKORC، Target INR حذف شـده‌اند و توسـط الگوریتم PSO هم ویژگی‌هایی نظیر نژاد، Target INR، Subject reached to stable dose کنار گذاشته شده‌اند که اتفاقاً، بر اساس نظر خبرگان این ویژگی‌ها حائز اهمیت ه ستند. در ادامه ویژگی‌هایی که تو سط الگوریتم Relief انتخاب شده، در جدول ۸ آمده ا ست. در این روش ویژگی‌های مرتبط با اعمال جراحی قلب اســاســاً دیده و انتخاب ن شده‌اند که با توجه به این که دامنه و سیعی از م صرف وارفارین به بیماران قلبی و عروقی تعلق دارد، این حذف نقطه مثبتی به نظر نمی‌رسد.

جدول ۸: ویژگی‌های استخراج شده توسط الگوریتم Relief

بالینی	دموگرافیک	سن (Age)، قومیت (Ethnicity)، نژاد (Race)، جنسیت (Gender)
	مصرف داورها	Indication for drug treatment, Acetaminophen or Paracetamol (Tylenol), Aspirin, Atrovastatin, Carbamazepine, Phenytin, Antifungal, Herbal medications
	مرتبط با مشکلات قلبی	-
	سایر ویژگی‌ها	قد، وزن، سیگاری بودن
ژنوتایپ		CYP2C9, VKORC1

در آزمایش آخر به جای بررسی ویژگی‌ها به صورت جداگانه، آن‌ها در قالب یک رویکرد پوشنده در تعامل با مدل تخمین‌گر سنجیده شد؛ اگرچه که ممکن است این رویکرد، بار محاسباتی بیشتری به همراه داشته باشد، اما با توجه به اهمیت فاکتور دقت و صحت پیش‌بینی نسبت به فاکتورهای سرعت یا مصرف حافظه، این مسامحه منطقی به نظر می‌رسد. در فرآیند انتخاب ویژگی، هم از رویکرد کامپیوتری و هم از نظر متخصصان خبره در زمینه تجزیه و تحلیل پزشکی و هم تجربه مقالات مشابه در این زمینه استفاده شد. در رویکرد کامپیوتری، مبنا یک پارادایم دو سطحی برای انتخاب ویژگی‌های مناسب‌تر بود. در سطح اول، مجموعه ویژگی‌های کلی واکنش‌های دارویی، عوارض و جراحی‌های قلب، ژنوتیپ، وضعیت سیگاری بودن، بیماری‌های زمینه‌ای مانند دیابت و ... با هم استفاده شد. به این صورت که تحت استراتژی Backward Elimination، ابتدا در سطح بالاتر، شروع به حذف هر یک از این زیر مجموعه ویژگی‌های اصلی و سنجش نتیجه تخمین شد. یعنی اگر وجود این مجموعه

[DOI: doi: 10.34172/jhbmi.2023.02]

از ویژگی‌ها خطای پیش بینی را بهبود بخشید، آن را نگه داشته و وارد سطح دوم شد. در سطح دوم، هر یک از ویژگی‌های زیر مجموعه این مجموعه کلی را یکی پس از دیگری با همان استراتژی Backward Elimination برداشته تا تأثیر مثبت یا منفی آن‌ها در برآورد دوز وارفارین ارزیابی گردد. این کار با استفاده از ابزار پیش‌بینی LSSVR انجام می‌گردد. متغیرهایی که حذفشان منجر به افزایش MAE می‌شوند، به عنوان پارامترهای ورودی برای مدل نهایی انتخاب شد. اگر نتایج را بهبود ندهند یا احتمالاً برآورد را بدتر کنند، طبیعتاً حذف می‌شوند.

جدول ۹: مقایسه و ارزیابی زیرمجموعه‌های مختلف ویژگی‌ها

زیرمجموعه ویژگی‌ها	R^2	RMSE	MAE
حضور تمام ویژگی‌ها	۰/۶۲۷۶۱	۰/۰۳۹۵۵۲	۰/۰۲۸۵۷۴
حضور ویژگی‌های بالینی و ژنوتایپ و سیگاری بودن، عدم حضور تداخلات دارویی و عوارض قلبی	۰/۶۲۱۳۷	۰/۰۳۹۸۲۱	۰/۰۲۸۵۰۴
حضور ویژگی‌های بالینی و ژنوتایپ و تداخلات دارویی و دیابت، عدم حضور عوارض قلبی	۰/۶۳۸۴۶	۰/۰۳۸۶۷۸	۰/۰۲۸۰۱
حضور ویژگی‌های بالینی و ژنوتایپ و عوارض قلبی، عدم حضور تداخلات دارویی	۰/۶۳۲۵۲	۰/۴۰۶۷۲	۰/۲۸۶۵۶

بر اساس مقایسه‌ای که در جدول ۹ مشاهده می‌شود استفاده از همه ویژگی‌ها نتیجه خوبی به همراه ندارد، بنابراین باید برخی از ویژگی‌ها حذف شود. در عین حال حذف کلی زیرمجموعه تداخلات دارویی هم منجر به نتیجه بدتری می‌شود؛ بنابراین این زیرمجموعه را حفظ نموده و باید از بین ویژگی‌های دارویی مناسب‌ترین‌ها را انتخاب کرد. حضور تمام ویژگی‌های عوارض قلبی هم مناسب نیستند، اما در عین حال حذف همه آن‌ها هم نتیجه را بهبود نمی‌بخشد، در نتیجه باید این جا هم در عین حال که کلیت این زیرمجموعه حفظ می‌ماند، تعدادی از ویژگی‌ها حذف شوند. با توجه به نتایج ردیف‌های سه و چهار جدول ۹ به نظر می‌رسد به نوعی زیرمجموعه تداخلات دارویی مهم‌تر از عوارض قلبی باشند. ویژگی سیگاری بودن در حالت‌های مختلف تأثیر مثبتی ندارد و کنار گذاشته می‌شود. دیابت در نتیجه تأثیرگذار است. در نهایت ترکیب زیر به عنوان بهترین زیرمجموعه از ویژگی‌ها گزینش شد.

جدول ۱۰: ترکیب نهایی ویژگی‌های دموگرافیک، بالینی و ژنوتایپ

دموگرافیک		سن، نژاد، قومیت، جنسیت
بالینی	مصرف داورها	Amiodarone, Acetaminophen or Paracetamol (Tylenol), Cerivastatin (Baycol), Aspirin, Indication for drug treatment
	مرتبط با مشکلات قلبی	Congestive heart failure, Valve replacement
	سایر ویژگی‌ها	قد، وزن، دیابت، Target INR، Subject reached to stable dose، INR on reported therapeutic dose
ژنوتایپ		CYP2C9, VKORC1

ترکیب ویژگی‌های جدول ۱۰ برای ایجاد مدل نهایی به کار گرفته شدند. فلسفه اولیه انتخاب متغیرها رجوع به ادبیات قبلی و مشابه این حوزه تحقیقاتی بود. اما سپس بر اساس مدل پیشنهادی در مرحله انتخاب ویژگی، متغیرهای مؤثر شناسایی شدند. پس از انتخاب ویژگی‌های مناسب توسط هر استراتژی، نتایج دوز اولیه وارفارین توسط مدل رگرسیون پیشنهادی به صورت جداگانه برای هر یک از مجموعه ویژگی‌های استخراج شده تخمین زده خواهد شد. در جدول ۱۱ نتایج مذکور مشاهده می‌شود.

[DOI: doi: 10.34172/jhbmi.2023.02]

جدول ۱۱: مقایسه نتایج برای مجموعه ویژگی‌های استخراج شده متفاوت در تخمین دوز اولیه توسط LSSVR مبتنی بر کرنل گاوسی

مجموعه ویژگی‌ها	R^2	MAE	RMSE	MSE
WOA	۰/۶۰۵۷	۰/۱۵۸۵۱	۰/۲۰۱۲۷	۰/۰۴۰۵۱
PSO	۰/۵۲۶۸	۰/۲۶۷۲۴	۰/۳۲۹۹۸	۰/۱۰۸۸۹
Relief	۰/۵۸۲۱	۰/۲۴۱۷۲	۰/۳۰۳۱۹	۰/۰۹۱۹۲۷
رویکرد پیشنهادی مبتنی بر Backward Elimination	۰/۶۴۵۲	۰/۳۰۹۱۷	۰/۳۳۶	۰/۰۰۱۱۲۸۲

۳-استخراج تابع کرنل متناسب با ماهیت دادگان مسئله

یکی از اصلی‌ترین مسائل در استفاده از کرنل‌ها، تشخیص و تعیین کرنل مناسب برای مسئله مورد نظر می‌باشد. در این مطالعه سعی شد تا کرنل مناسب مجموعه داده‌ای انتخاب شود. فرضیه اولیه برای این سناریو این است که با توجه به این که بخش زیادی از ویژگی‌ها (چه بالینی و چه ژنوتایپ) ماهیت گسسته دارند، کرنل‌هایی با ماهیت گسسته نظیر چندجمله‌ای برای این کار مناسب‌تر باشد.

تغذیه کرنل جدید به یک روش یادگیر

یکی از استراتژی‌های تخمین این پژوهش مبتنی بر نسخه حداقل مربعات ماشین بردار پشتیبان بوده است، که بر اساس مطالعه IWPC عملکرد بهتری نسبت به سایر روش‌ها داشته است. کرنل انتخاب شده منطبق بر دادگان موجود IWPC

$$p' = \frac{p - p_{min}}{p_{max} - p_{min}} \quad (۳)$$

۴- معیارهای ارزیابی

$$R - square = 1 - \sum_i \frac{(y_i - f(\bar{x}_i))^2}{(y_i - \bar{y})^2} \quad (۴)$$

که y_i پاسخ iام، $f(\bar{x}_i)$ حاصل پیش بینی مدل برای رکورد iام و $\bar{y}$ هم میانگین پاسخ‌ها است.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \quad (۵)$$

می‌باشد. با این تفاسیر، می‌توان از هر تکنیک یادگیری که مبتنی بر کرنل باشد در ترکیب با این مدل شخصی‌سازی‌شده استفاده نمود. با توجه به عملکرد مناسب‌تر رویکردهای ماشین بردار پشتیبان در تخمین دوز اولیه وارفارین، استراتژی پیشنهادی ایجاد یک مدل هیبریدی از کرنل مناسب و روش‌های ماشین بردار پشتیبان خواهد بود.

نتایج تجربی

نرمال‌سازی داده‌ها

با توجه به محتوای مجموعه داده‌ای، ویژگی‌ها دارای بازه‌های متفاوتی هستند. همین مسئله تأثیر هر ویژگی در خروجی ابزار پیش‌بین را با توجه به در نظر گرفتن پارامترهای وزن یکسان به خوبی منعکس نخواهد کرد. برای اعمال تأثیر یکسان، داده‌ها در بازه $[0,1]$ نرمال می‌شود.

در این مطالعه از معادله ۳ برای نرمال سازی p استفاده شد.

برای این که مقایسه روش پیشنهادی و مقالات مشابه در این حوزه در یک بستر یکسانی صورت پذیرد، از دو معیار ارزیابی MAE و R^2 که در غالب پژوهش‌ها آمده است برای ارزیابی نتایج استفاده شد. البته چون در این مطالعه به تخمین دوز به دید پیوسته و رگرسیونی توجه شد، معیارهای ارزیابی در این حالت هم غالباً دو معیار مذکور هستند.

[DOI: doi: 10.34172/jhbmi.2023.02]

که y_i خروجی واقعی، m تعداد نمونه‌های آموزشی و $\hat{y}_i$ خروجی تخمین زده شده می‌باشد.

۵- نتایج شبیه‌سازی

نتایج به دست آمده از روی متوسط ۱۰ بار اجرای متفاوت (10 fold cross validation) الگوریتم‌ها محاسبه شده است. برای مقایسه نتایج به‌دست آمده با نتایج مطالعات مشابه گذشته، به دو صورت عمل شد. جهت صحت‌سنجی و آزمون اولیه مدل پیشنهادی مبتنی بر رگرسیون بردار پشتیبان حداقل مربعات، ابتدا نتایج تخمین LSSVR مبتنی بر کرنل مرسوم گوسی، با دو روش مرسومی که در غالب تحقیقات این حوزه به کار گرفته شدند، یعنی رگرسیون خطی و رگرسیون بردار پشتیبان با همان کرنل گوسی مقایسه شد. در ادامه به ترتیب در شکل (a)۱-۵، (b)۱-۵ و (c)۱-۵ نتایج تخمین دوز با استفاده از رگرسیون خطی، رگرسیون بردار پشتیبان و رگرسیون بردار پشتیبان حداقل مربعات مشاهده شد. رگرسیون بردار پشتیبان بعد از رگرسیون خطی، جزو روش‌های تقریباً پرکاربرد در این حوزه بوده است. ضمن این‌که در شکل‌های ۲ هیستوگرام و روند خطا و نتایج تخمین این دو روش هم قابل رؤیت است. اگر به نمودار R^2 در سه شکل ۲ توجه شود، مشخص است که رگرسیون خطی و SVR (به خصوص SVR) دچار مشکل overprediction شده‌اند و از خط برازش انحراف جدی دارند، در صورتی‌که این انحراف از میانگین در نمودار LSSVR بسیار کمتر مشاهده می‌شود. در واقع شاخص R^2 به عنوان یکی از شاخص‌های برازش مدل، قدرت پیش‌بینی متغیر وابسته را براساس متغیرهای مستقل نشان می‌دهد. اگر مقدار این شاخص از ۰/۶ بیشتر باشد نشان می‌دهد متغیرهای مستقل تا حد زیادی توانسته‌اند تغییرات متغیر وابسته را تبیین کنند. علاوه بر این اگر به نمودار هیستوگرام خطای سه شکل مذکور هم توجه شود انحراف دو روش رگرسیونی اول مشخص است. در جدول ۱۲ سه روش مذکور بر اساس سه معیار MSE، MAE و R^2 با هم مقایسه شدند. همچنان که مشخص است دو روش اول عملکردی مشابه هم دارند، با این تفاوت که رگرسیون خطی در دو معیار بهتر عمل کرده است، اما نتایج شبیه‌سازی LSSVR مؤید مطالب عنوان شده در بخش مواد و روش‌ها در خصوص برتری نسخه حداقل مربعات رگرسیون بردار پشتیبان می‌باشد. در ادامه فازهای شبیه‌سازی، LSSVR با استفاده از کرنل‌های مختلف و کرنل مناسب مورد مقایسه قرار گرفت.

جدول ۱۲: مقایسه دقت پیش‌بینی سه الگوریتم رگرسیون خطی، رگرسیون بردار پشتیبان و رگرسیون بردار پشتیبان حداقل مربعات

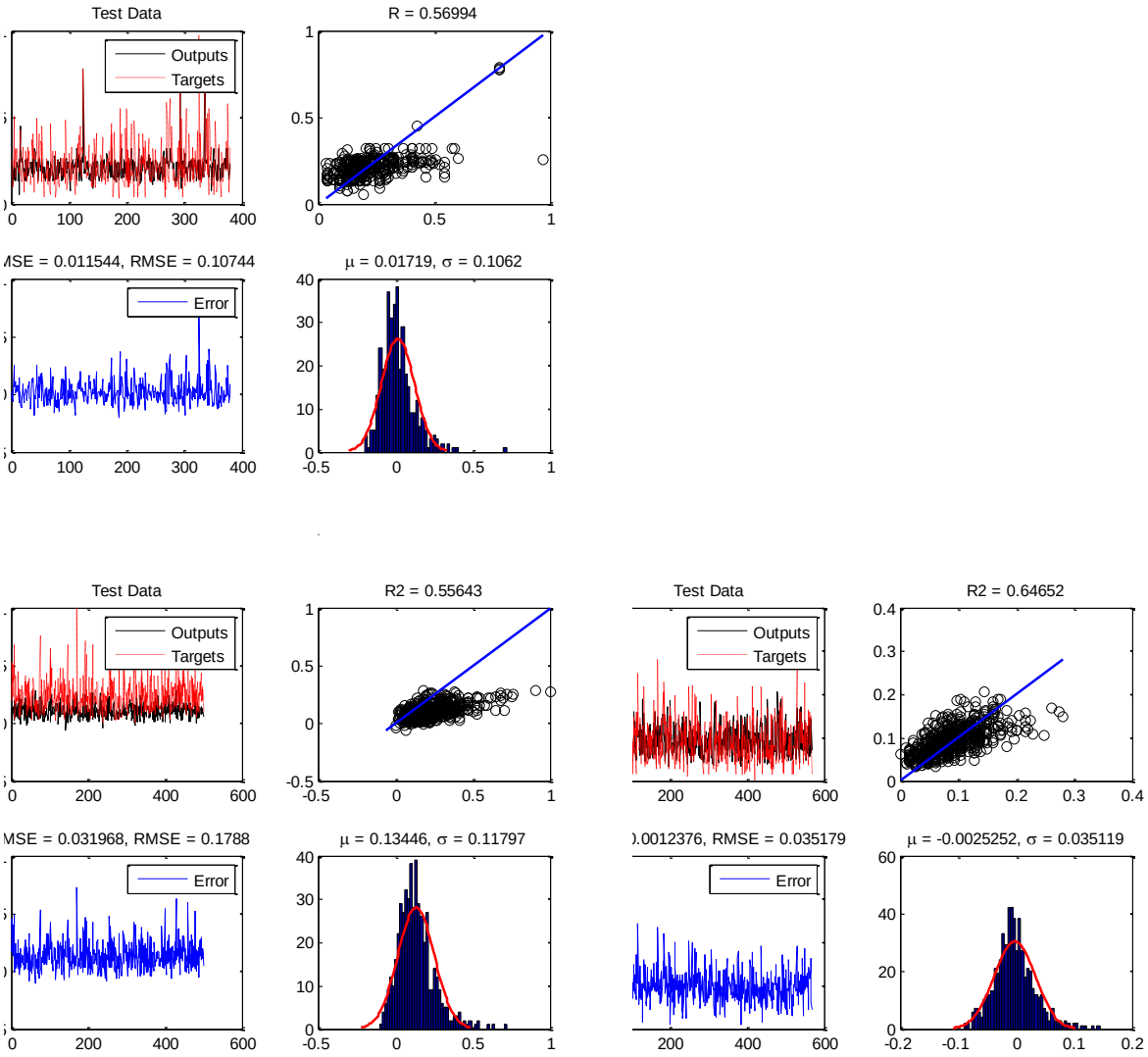
روش	MSE	MAE	R^2
رگرسیون خطی	۰/۰۱	۰/۰۸۶۸۷۷	۰/۵۶۹۹۴
SVR با هسته گوسی	۰/۰۳	۰/۱۲۱۲۷	۰/۵۵۶۴۳
LSSVR با هسته گوسی	۰/۰۰۱	۰/۰۳۰۹۱۷	۰/۶۴۶۵۲

در کنار مقایسه این روش‌ها با یکدیگر، نتایج روش پیشنهادی مبتنی بر کرنل‌های مختلف هم با یکدیگر مقایسه شد. مقایسه نتایج شبیه‌سازی در دو جدول جداگانه ۱۳ و ۱۴ بررسی شد. ابتدا نتایج تخمین دوز توسط استراتژی اصلی LSSVR با کرنل‌های مختلف در جدول ۱۳ مقایسه و تحلیل شد و سپس نتایج تخمین مبتنی بر بهترین کرنل پیشنهادی با دو روش یادگیری دیگر شامل شبکه‌های عصبی مصنوعی و استراتژی یادگیری عمیق، که ماهیت یادگیری کرنل ندارند؛ اما در پژوهش‌های این حوزه هم سابقاً توسط محققین به کار گرفته شدند، هم مقایسه شد. ضمن این توضیح که MAELD، MAEMD و MAEHD به ترتیب میانگین قدرمطلق خطا برای سه حالت دوز کم، متوسط و زیاد می‌باشد. علاوه بر مقایسه مقادیر تخمینی بر اساس معیارهای فوق، یک فاصله اطمینان (95% Confidence Interval) ۹۵٪ هم برای هر روش تخمین محاسبه شد تا یک بررسی سختگیرانه‌تر و دقیق‌تر اعمال شود. این فاصله اطمینان بر اساس برآورد پارامتر خروجی و واریانس تخمین زده شده محاسبه می‌شود. پارامتر میانگین جامعه (یا میانگین خروجی بهینه) ۰/۰۸۹۵۹۵ است که در فاصله اطمینان به‌دست آمده برای روش پیشنهادی (بر اساس ستون CI در جدول ۱۳) قرار دارد. پارامترهای شبکه عصبی عمیق (DNN) هم به صورت تعداد ورودی‌ها ۱۸، تعداد سلول‌ها ۱۰۰، تعداد تکرار الگوریتم ۵۰۰، تابع فعال‌ساز سیگموئید، نرخ آموزش ۰/۰۰۳ و الگوریتم آموزش هم گرادیان نزولی در نظر گرفته شدند. در شبکه عصبی MLP استفاده شده هم پارامترها به این صورت مقداردهی شدند که تعداد ورودی‌ها ۱۸، دو لایه مخفی به ترتیب دارای نودهای ۲۰ و ۱۰ تعریف شد، توابع فعال‌ساز logsig و pureline در نظر

[DOI: doi: 10.34172/jhbmi.2023.02]

گرفته شدند، الگوریتم آموزش Levenberg-Marquardt، نرخ آموزش ۰/۰۰۹ و تعداد تکرارها هم ۲۰۰۰ تنظیم شدند.

همان‌طور که در جدول ۱۳مشاهده می‌شود، کرنل چند جمله‌ای و شبکه عصبی یک فاصله اطمینان ۹۵٪ را ارائه می‌دهد. البته درست است که روش پیشنهادی از نظر پوشش فاصله اطمینان با یک روش دیگر همپوشانی دارد، اما فاصله اطمینان کرنل پیشنهادی کمتر از روش دیگر است که نشان‌دهنده دقت بهتر الگوریتم پیشنهادی است. نکته دیگری که در خصوص جدول نتایج ۱۳ وجود دارد این است که با توجه به مباحث مطرح شده در بخش دوم، اگر در یک مجموعه داده‌ای تعداد ویژگی‌های گسسته بیشتر از پیوسته باشد، کرنل‌هایی از جنس گسسته نظیر خطی و چندجمله‌ای عملکرد بهتری خواهند داشت. مجموعه ویژگی‌های نهایی که مطابق جدول ۱۰ به دست آمده‌اند، ۱۳ مورد گسسته و ۵ مورد پیوسته هستند. این‌جا هم مشاهده می‌شود که نتایج این نوع کرنل از کرنل‌های پیوسته‌ای نظیر RBF بهتر شده است. اگرچه ERBF نتایجی نزدیک به چندجمله‌ای داشته است، اما نتوانست فاصله اطمینان را پوشش دهد.



شکل ۲: آنالیز همبستگی دوزهای تخمین زده شده و واقعی و نمودار خروجی‌های تخمین داده‌های آزمایش و هیستوگرام و روند خطا برای رگرسیون خطی(a)، رگرسیون بردار پشتیبان (b) و رگرسیون بردار پشتیبان حداقل مربعات (c) با کرنل گوسی

[DOI: doi: 10.34172/jhbmi.2023.02]

نکته دیگر جدول ۱۳ این است که مدل پیشنهادی در تخمین دوز برای بیماران دوزهای کم و زیاد عملکرد مناسبی داشته، یعنی MAELD و MAEHD و همچنین Total MAE را از سایر روش‌ها مناسب‌تر تخمین زده است و این بهبود عملکرد مشهود است، اما خروجی مدل برای دسته دوز متوسط بهتر از تمامی دیگر روش‌ها نشده است، اگرچه این تفاوت عملکرد معنی‌دار و قابل توجه نیست. ذکر این نکته ضروری است که اساساً مسئله تخمین دوز توسط یک مدل رگرسیونی و با ماهیت پیوسته صورت گرفته است، اما در حقیقت یکی از دلایل سه کلاسه کردن دوز به کم (≥ ۲۱ mg)، متوسط (۲۱> و ۴۹>) و زیاد (۴۹ ≤) جهت ارزیابی و صحت‌سنجی مدل پیشنهادی در تخمین دوز برای سه دسته مختلف از بیماران نیازمند (یعنی بیماران نیازمند به دوز کم یا متوسط یا زیاد) بوده است و این تقسیم‌بندی را تقریباً تمامی تحقیقات حوزه وارفارین از جمله مطالعه کنسرسیوم هم انجام داده‌اند. مرور مطالعات نشان داد برخی مدل‌ها مثلاً در تخمین دوزهای کم و متوسط خوب عمل کرده‌اند؛ اما در تخمین مقادیر بالای وارفارین عملکرد خوبی نداشته‌اند و یا در دو دسته دیگر خوب عمل کرده و در یک دسته نامناسب عمل کرده‌اند. در حقیقت با این دسته‌بندی یک ارزیابی از مدل تخمین‌گر برای دوزهای سه‌گانه به دست می‌آید. DNN به طور کلی عملکرد خوبی نداشت. اگرچه DNN در برخی معیارها خوب عمل کرد، اما کارکرد تصادفی این روش مشهود است چرا که فاصله اطمینان را با توجه ستون CI پوشش نمی‌دهد. در مقابل، MLP عملکرد نسبتاً بهتری داشت. با این وجود، در اکثر معیارهای خروجی، نسبت به روش پیشنهادی ضعیف‌تر بود. روش‌هایی که پیاده‌سازی شدند عبارت‌اند از رگرسیون خطی، SVR، MLP، DNN و LSSVR، که بهترین نتیجه با LSSVR حاصل شد و طبیعتاً نسخه‌ای که از کرنل چندجمله‌ای استفاده شده است پاسخ بهتری ارائه کرده است. در تحلیل بعدی، مقایسه نتایج مدل پیشنهادی با نتایج مطالعات IWPC و [۶،۱۱،۵۰،۵۱] حالت دوز تجربی ثابت صورت گرفت. علت انتخاب این مقالات به دلیل اشتراک در استفاده از مجموعه داده‌ای IWPC و انجام مقایسه‌ها در یک بستر یکسان بود.

جدول ۱۳: مقایسه دقت پیش‌بینی LSSVR مبتنی بر کرنل‌های خطی، چند جمله‌ای، RBF، ERBF و شبکه عصبی مصنوعی و شبکه عصبی مصنوعی عمیق در تخمین دوز اولیه

روش تخمین	CI	R^2	RMSE	MAEHD	MAEMD	MAELD	Total MAE
LSSVR(ERBF)	۰/۸۹۹۱۸ – ۰/۹۵۳۴۸	۰/۶۶۸۷۲	۱۰/۹۶۶۶	۱۷/۵۸۶۶	۶/۱۹۳۶	۸/۴۰۴۷	۸/۰۷۷۷
LSSVR(RBF)	۰/۹۰۲۶۸ – ۰/۹۵۶۸۴	۰/۶۵۵۲۲	۱۱/۱۵۲۲	۱۷/۷۳۱۹	۶/۵۱۸۷	۸/۴۲۳۴	۸/۲۸۷۱
LSSVR(Linear)	۰/۰۹۰۰۰۱ – ۰/۰۹۵۲۱۷	۰/۶۵۲۲۸	۱۱/۱۹۳۰	۱۷/۷۵۳۱	۶/۳۱۱۳	۸/۹۳۴۵	۸/۳۳۱۱
LSSVR(Poly)	۰/۰۸۶۴۸۳ – ۰/۰۹۲۱۶۳	۰/۶۶۸۳۲	۱۰/۴۸۱۷	۱۷/۱۷۲۹	۶/۳۶۶۲	۸/۳۴۴۱	۸/۰۴۷۲
شبکه عصبی مصنوعی (MLP)	۰/۸۶۷ – ۰/۰۹۲۴	۰/۴۰۴۰	۱۲/۳۵۲۵	۱۷/۷۰۹۳	۷/۱۹۱۶	۷/۷۹۱۲	۸/۷۴۷۱
شبکه عصبی عمیق (DNN)	۰/۰۸۹۷ – ۰/۰۹۹۳	۰/۴۵۷۳	۱۷/۹۲۲۹	۵/۵۳۶۷	۴/۷۶۰۳	۷/۰۱۶۷	۵/۴۳۱۶

همچنان‌که از جدول ۱۴ مشهود است، روش پیشنهادی برای تخمین دوز اولیه درمانی وارفارین عملکرد بهتری از مدل تعریف شده توسط IWPC داشته است؛ این بهبود عملکرد حدود ۰/۷ میلی‌گرم در هفته از دوز اشتباه دارو کاسته است. ضمن اینکه در مقایسه با جدیدترین پژوهشی هم که از IWPC استفاده کرده بود [۱۱]، ۰/۵۸ میلی‌گرم در هفته بهبود داشته است. این بهبود در خطای درمان، برای دو پژوهش بعدی حدوداً ۱ میلی‌گرم در هفته می‌باشد. به عبارتی در مقایسه با دو پژوهش مشابه در این حوزه هم بهبود معیارهای پیش‌بینی مشهود است. رویکرد این کار هم‌چنین با رویه دوز ثابت هم مقایسه شده است که در هر

[DOI: doi: 10.34172/jhbmi.2023.02]

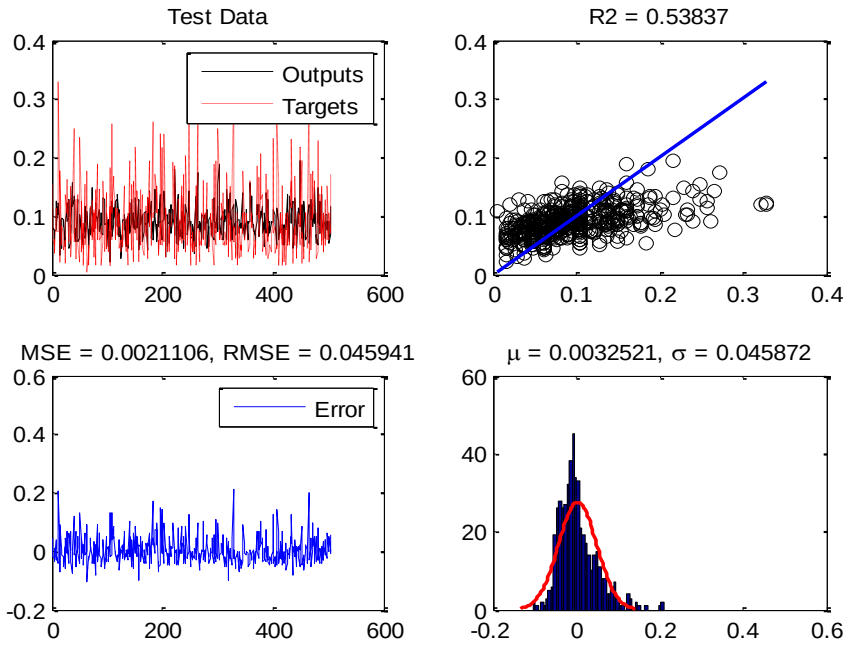
دو معیار موجود پرواضح است که استفاده از یک استراتژی که از هر دو جنس داده‌های بالینی و ژنتیکی استفاده نماید، به مراتب بهتر از حالت سنتی دوز ثابت می‌باشد. در همین راستا در انتهای بخش تحلیل نتایج، کارآیی مدل فارماکوژنومیکس پیشنهادی با مدلی که شامل صرفاً ویژگی‌های بالینی باشد، مقایسه شد. به عبارتی این مقایسه تأثیر حضور و یا عدم حضور متغیرهای ژنوتایپ را به خوبی نشان می‌دهد. شکل ۳ نمودارهای تخمین، ضریب تعیین، هیستوگرام و روند خطای مدل بالینی محض را نشان می‌دهد. در جدول ۱۵ هم دو رویکرد فارماکوژنومیکس و بالینی بر اساس چهار معیار خطا با هم مقایسه شدند، نتایج بهتر مدل فارماکوژنومیکس مشهود است.

جدول ۱۴: مقایسه صحت پیش‌بینی روش پیشنهادی در مقایسه با مدل‌های IWPC[۶]، منبع [۵۲،۵۰،۱۱] و رویکرد دوز ثابت برای گروه اعتبارسنجی

روش‌ها	MAE	$RMSE$	R^2
روش پیشنهادی	۸/۱	۱۰/۴۴	۰/۶۵
IWPC	۸/۸	–	۰,۴۳
مدل منبع [۱۱]	۸,۵۸	–	–
مدل منبع[۵۱]	۰/۹		۰/۵۴
مدل منبع [۵۰]	۰/۹	–	۰/۴۸
رویکرد دوز ثابت (۳۵ میلی گرم در هفته)	۱۳/۲		۰

جدول ۱۵: مقایسه صحت پیش بینی دو مدل بالینی و فارماکوژنومیک

روش	MSE	RMSE	MAE	R^2
فارماکوژنومیک	۰/۰۰۱	۰/۰۳۵	۰/۰۲۵۵۶۴	۰/۶۵
بالینی	۰/۰۰۲	۰/۰۴۵	۰/۰۳۵۹۲	۰/۵۳



شکل ۳: آنالیز همبستگی دوزهای تخمین زده شده و واقعی، نمودار خروجی‌های تخمین داده‌های آزمایش و روند خطا با استفاده استراتژی پیشنهادی بالینی

[DOI: doi: 10.34172/jhbmi.2023.02]

بحث و نتیجه‌گیری

در پژوهش حاضر یک استراتژی تخمین دوز وارفارین با رویکرد فارماکوژنومیکس توسعه داده شد. این روش توانسته است تخمین‌های مناسبی با نرخ خطای مناسب ارائه کند. مدل حاضر پیش‌بینی‌های موفق دوز را برای ۶۵٪ از بیماران انجام داده است. پژوهش‌های محدودی برای تخمین دوز اولیه وارفارین از رویه‌های مبتنی بر کرنل استفاده کردند، این پژوهش اطلاعات ژنتیکی، دموگرافیک و بالینی را به عنوان متغیرهای پیش‌بین به LSSVR با کرنل مناسب به خدمت گرفته است و با توجه کثرت ویژگی‌های گسسته و اهمیت ویژگی‌های ژنوتایپ، کرنل چندجمله‌ای انتخاب شد که مطابق با استدلال‌های مطرح شده به لحاظ نظری هم کرنل مناسبی تلقی می‌شد. از طرفی با مجموعه داده‌ای وجود داشت که ویژگی‌های متعدد و متنوعی داشت و با بررسی چندین روش انتخاب ویژگی از رویکردهای مختلف، مناسب‌ترین ویژگی‌ها استخراج شدند. برای استخراج ویژگی صرفاً به روش‌های هوش مصنوعی اکتفا نکرده و نظرات خبرگان حوزه داروسازی و پزشکی نظیر قلب و عروق هم مورد توجه قرار گرفت. عملکرد پیش‌بینی و اعتبارسنجی مدل فارماکوژنومیکس پیشنهادی با مدل فارماکوژنومیکی که توسعه کنسرسیوم وارفارین معرفی شده بود و نیز چند پژوهش معتبر حوزه تخمین دوز وارفارین مقایسه شد که نتایج مناسب‌تری ارائه گردید. یک مقایسه و تحلیل دیگر بین مدل فارماکوژنتیک پیشنهادی و حالت بالینی محض هم صورت گرفت که نتایج حاکی از برتری مدل نهایی فارماکوژنومیکس بودند. یکی از نقاط ضعفی که می‌تواند توسط پژوهش‌های آتی مورد بررسی قرار گیرد و به خاطر محدودیت‌های موجود، در این پژوهش انجام نشده است، ساخت و ایجاد یک کرنل نگاشت شخصی‌سازی شده از مجموعه داده‌ای بوده است. تعریف این کرنل نیازمند محاسبات ریاضیاتی پیچیده آماری خواهد بود. در راستای انجام این پژوهش جلسات علمی متعددی با اساتید حوزه پزشکی به خصوص متخصصین قلب و عروق مرکز قلب مازندران برگزار شد، از آنجایی که بخش وسیعی از کاربرد داروی وارفارین به بیماران قلبی مرتبط می‌باشد، نتایج این پژوهش می‌تواند به عنوان یک ابزار کاربردی تخمینی در اختیار این مراکز قرار بگیرد.

تعارض منافع

در مطالعه حاضر هیچ گونه تضاد منافعی وجود نداشته است.

References

1. Xu R, Wang Q. A semi-supervised approach to extract pharmacogenomics-specific drug–gene pairs from biomedical literature for personalized medicine. Journal of Biomedical Informatics 2013;46(4):585-93. https://doi.org/10.1016/j.jbi.2013.04.001

2. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. Bioinformatics 2011;27(13):1741-8. https://doi.org/10.1093/bioinformatics/btr295

3. Karczewski KJ, Daneshjou R, Altman RB. Chapter 7: pharmacogenomics. PLoS Computational Biology 2012;8(12):e1002817. https://doi.org/10.1371/journal.pcbi.1002817

4. Wagner MJ. Pharmacogenetics and personal genomes. Personalized Medicine 2009;6(6):643-52.

5. Godman B, Finlayson AE, Cheema PK, Zebedin-Brandl E, Gutiérrez-Ibarluzea I, Jones J, et al. Personalizing health care: feasibility and future implications. BMC Medicine 2013;11:1-23.

6. International Warfarin Pharmacogenetics Consortium; Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 2009; 360:753-64. doi: 10.1056/NEJMoa0809329

7. Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, Anderson JL, Kimmel SE, Lee MT, Pirmohamed M, Wadelius M. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. Clinical Pharmacology & Therapeutics 2011;90(4):625-9. https://doi.org/10.1038/clpt.2011.185

8. Anzabi Zadeh S, Street WN, Thomas BW. Optimizing warfarin dosing using deep reinforcement learning. Journal of Biomedical Informatics 2023;137:104267. https://doi.org/10.1016/j.jbi.2022.104267

9. Bontempi M. Semi-empirical anticoagulation model (SAM): INR monitoring during Warfarin therapy. Journal of Pharmacokinetics and Pharmacodynamics 2022:1-2.

10. Liu Y, Chen J, You Y, Xu A, Li P, Wang Y, et al. An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set. Computers in Biology and Medicine 2021;131:104242. https://doi.org/10.1016/j.compbiomed.2021.104242

11. Truda G, Marais P. Evaluating warfarin dosing models on multiple datasets with a novel software framework and evolutionary optimisation. Journal of Biomedical Informatics 2021;113:103634. https://doi.org/10.1016/j.jbi.2020.103634

12. Xie C, Xue L, Zhang Y, Zhu J, Zhou L, Hang Y, et al. Comparison of the prediction performance of different warfarin dosing algorithms based on Chinese

[DOI: doi: 10.34172/jhbmi.2023.02]

patients. Pharmacogenomics 2020;21(1):23-32. doi: 10.2217/pgs-2019-0124
13. Altay O, Ulas M, Mahmut OZ, Ece GE. An expert system to predict warfarin dosage in turkish patients depending on genetic and non-genetic factors. 7th International Symposium on Digital Forensics and Security (ISDFS); 2019 Jun 10-12; Barcelos, Portugal: IEEE; 2019. p. 1-6. doi: 10.1109/ISDFS.2019.8757526
14. Sharabiani A, Bress A, Galanter W, Nazempour R, Darabi H. A computer-aided system for determining the application range of a warfarin clinical dosing algorithm using support vector machines with a polynomial kernel function. 15th International Conference on Automation Science and Engineering (CASE); 2019 Aug 22-26; Vancouver, BC, Canada: IEEE; 2019. p. 418-23. doi: 10.1109/COASE.2019.8842932
15. Rad F, Hamidpour M, Dorgalaleh A, Poopak B. The effect of demographic factors and VKORC1 1639 G> A genotypes on estimated warfarin maintenance dose in Iranian patients under warfarin therapy. Indian Journal of Hematology and Blood Transfusion 2019;35:167-71.
16. Ayesh BM, Abu Shaaban AS, Abed AA. Evaluation of CYP2C9-and VKORC1-based pharmacogenetic algorithm for warfarin dose in Gaza-Palestine. Future science OA 2018;4(3):FSO276. doi: 10.4155/fsoa-2017-0112
17. Qayyum A, Najmi MH, Mansoor Q, Irfan M, Naveed AK, Hanif A, Kazmi AR, Ismail M. Frequency of common VKORC1 polymorphisms and their impact on warfarin dose requirement in Pakistani population. Clin Appl Thromb Hemost 2018; 24(2): 323–9. doi: 10.1177/1076029616680478
18. Cho SM, Lee KY, Choi JR, Lee KA. Development and comparison of warfarin dosing algorithms in stroke patients. Yonsei Medical Journal 2016;57(3):635-40. doi: https://doi.org/10.3349/ymj.2016.57.3.635
19. Pavani A, Naushad SM, Lakshmitha G, Nivetha S, Stanley BA, Malempati AR, Kutala VK. Development of neuro-fuzzy model to explore gene–nutrient interactions modulating warfarin dose requirement. Pharmacogenomics 2016;17(12):1315-25.
20. Hamberg AK, Hellman J, Dahlberg J, Jonsson EN, Wadelius M. A Bayesian decision support tool for efficient dose individualization of warfarin in adults and children. BMC Medical Informatics and Decision Making 2015;15(1):1-9.
21. Sharabiani A, Bress A, Douzali E, Darabi H. Revisiting warfarin dosing using machine learning techniques. Computational and Mathematical Methods in Medicine 2015;2015. https://doi.org/10.1155/2015/560108
22. Karaca S, Bozkurt NC, Cesuroglu T, Karaca M, Bozkurt M, Eskioglu E, Polimanti R. International warfarin genotype-guided dosing algorithms in the Turkish population and their preventive effects on major and life-threatening hemorrhagic events. Pharmacogenomics 2015;16(10):1109-18.
23. Santos PC, Marcatto LR, Duarte NE, Gadi Soares RA, Cassaro Strunz CM, Scanavacca M, Krieger JE, Pereira AC. Development of a pharmacogenetic-based warfarin dosing algorithm and its performance in Brazilian patients: highlighting the importance of population-specific calibration. Pharmacogenomics 2015;16(8):865-76.
24. Isma'eel HA, Sakr GE, Habib RH, Almedawar MM, Zgheib NK, Elhajj IH. Improved accuracy of anticoagulant dose prediction using a pharmacogenetic and artificial neural network-based method. European Journal of Clinical Pharmacology 2014;70:265-73.
25. Grossi E, Podda GM, Pugliano M, Gabba S, Verri A, Carpani G, Buscema M, Casazza G, Cattaneo M. Prediction of optimal warfarin maintenance dose using advanced artificial neural networks. Pharmacogenomics 2014;15(1):29-37.
26. Pavani A, Naushad SM, Kumar RM, Srinath M, Malempati AR, Kutala VK. Artificial neural network-based pharmacogenomic algorithm for warfarin dose optimization. Pharmacogenomics 2016;17(2):121-31.
27. Öztaner SM, Temizel TT, Erdem SR, Özer M. A Bayesian estimation framework for pharmacogenomics driven warfarin dosing: a comparative study. IEEE Journal of Biomedical and Health Informatics 2014;19(5):1724-33. doi: 10.1109/JBHI.2014.2336974
28. Sharabiani A, Darabi H, Bress A, Cavallari L, Nutescu E, Drozda K. Machine learning based prediction of warfarin optimal dosing for African American patients. In2013 IEEE International Conference on Automation Science And Engineering (CASE); 2013 Aug 17-20; Madison, WI, USA: IEEE; 2013. p. 623-8. doi: 10.1109/CoASE.2013.6653999
29. Hu YH, Wu F, Lo CL, Tai CT. Predicting warfarin dosage from clinical data: A supervised learning approach. Artificial Intelligence in Medicine 2012;56(1):27-34.
https://doi.org/10.1016/j.artmed.2012.04.001
30. Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. Bioinformatics 2010;26(4):445-55. https://doi.org/10.1093/bioinformatics/btp713
31. Ji W, Liu D, Meng Y, Xue Y. A review of genetic-based evolutionary algorithms in SVM parameters optimization. Evolutionary Intelligence 2021;14:1389-414.
32. Álvarez-Alvarado JM, Ríos-Moreno JG, Obregón-Biosca SA, Ronquillo-Lomelí G, Ventura-Ramos Jr E, Trejo-Perea M. Hybrid techniques to predict solar radiation using support vector machine and search optimization algorithms: a review. Applied Sciences 2021;11(3):1044.
33. Lengua MA, Quiroz EA. A systematic literature review on support vector machines applied to classification. In2020 IEEE Engineering International Research Conference (EIRCON); 2020 Oct 21-23; Lima, Peru: IEEE; 2020. p. 1-4. **doi:** 10.1109/EIRCON51178.2020.9254028
34. Vapnik VN. Statistical Learning Theory Hardcover. 1st ed. NewYork: Wiley-Interscience; 1998.
35. Vapnik V. The Nature of Statistical Learning Theory. 2th ed. USA: Springer; 1999.

[DOI: doi: 10.34172/jhbmi.2023.02]

36. Hong WC. Application of chaotic ant swarm optimization in electric load forecasting. Energy Policy 2010;38(10):5830-9. https://doi.org/10.1016/j.enpol.2010.05.033
37. Hong WC, Dong Y, Zheng F, Lai CY. Forecasting urban traffic flow by SVR with continuous ACO. Applied Mathematical Modelling 2011;35(3):1282-91. https://doi.org/10.1016/j.apm.2010.09.005
38. Niu D, Wang Y, Wu DD. Power load forecasting using support vector machine and ant colony optimization. Expert Systems with Applications 2010;37(3):2531-9. https://doi.org/10.1016/j.eswa.2009.08.019
39. Mustaffa Z, Yusof Y, Kamaruddin SS. Enhanced artificial bee colony for training least squares support vector machines in commodity price forecasting. Journal of Computational Science 2014;5(2):196-205. https://doi.org/10.1016/j.jocs.2013.11.004
40. Pourbasheer E, Riahi S, Ganjali MR, Norouzi P. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity. European Journal of Medicinal Chemistry 2009;44(12):5023-8. https://doi.org/10.1016/j.ejmech.2009.09.006
41. Zhang WY, Hong WC, Dong Y, Tsai G, Sung JT, Fan GF. Application of SVR with chaotic GASA algorithm in cyclic electric load forecasting. Energy 2012;45(1):850-8. https://doi.org/10.1016/j.energy.2012.07.006
42. Ju FY, Hong WC. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. Applied Mathematical Modelling 2013;37(23):9643-51. https://doi.org/10.1016/j.apm.2013.05.016
43. Selakov A, Cvijetinović D, Milović L, Mellon S, Bekut D. Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. Applied Soft Computing 2014;16:80-8. https://doi.org/10.1016/j.asoc.2013.12.001
44. Fei SW, Wang MJ, Miao YB, Tu J, Liu CL. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. Energy Conversion and Management 2009;50(6):1604-9. https://doi.org/10.1016/j.enconman.2009.02.004
45. Li-Xia L, Yi-Qi Z, Liu XY. Tax forecasting theory and model based on SVM optimized by PSO. Expert Systems with Applications 2011;38(1):116-20. https://doi.org/10.1016/j.eswa.2010.06.022
46. Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters 1999;9:293-300.
47. Kondor R, Jebara T. A kernel between sets of vectors. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003); Washington DC: 2003. p. 361-8.
48. Jiang BT, Zhao FY. Particle swarm optimization-based least squares support vector regression for critical heat flux prediction. Annals of Nuclear Energy 2013;53:69-81. https://doi.org/10.1016/j.anucene.2012.09.020
49. Ali S, Smith-Miles KA. A meta-learning approach to automatic kernel selection for support vector machines. Neurocomputing 2006;70(1-3):173-86. https://doi.org/10.1016/j.neucom.2006.03.004
50. Saleh MI, Alzubiedi S. Dosage individualization of warfarin using artificial neural networks. Molecular Diagnosis & Therapy 2014;18:371-9.
51. Gage BF, Eby C, Johnson JA, Deych E, Rieder MJ, Ridker PM, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. Clinical Pharmacology & Therapeutics 2008;84(3):326-31. https://doi.org/10.1038/clpt.2008.10

[DOI: doi: 10.34172/jhbmi.2023.02]