

## ارائه روش هیبریدی نوین برای پیش‌بینی شاخص کل قیمت بورس اوراق بهادار

دیاکو درودی<sup>۱</sup>، سید بابک ابراهیمی<sup>۲</sup>

**چکیده:** روند تغییرات شاخص کل قیمت سهام، همواره به‌عنوان یکی از ملاک‌های سرمایه‌گذاری مدنظر قرار می‌گیرد. به دلیل وجود دو مؤلفه‌ای غیرخطی و متلاطم سری زمانی شاخص قیمت، در این پژوهش سعی شده مدل هیبریدی نوینی ارائه شود که بتواند روند حرکتی و تغییرات شاخص را با بیشترین دقت پیش‌بینی کند. در این مدل ابتدا با استفاده از تبدیل موجک، سری زمانی شاخص به شش سری زمانی مجزایی که ویژگی‌های غیرخطی و متلاطم شاخص مدنظر را نمایندگی می‌کند، تفکیک می‌شود. در ادامه، سری‌های زمانی استخراج‌شده با رفتار غیرخطی، با استفاده از ترکیب مدل ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات و سری‌های زمانی مبتنی بر رفتار متلاطم شاخص کل با بهره‌گیری از مدل GJR پیش‌بینی می‌شوند؛ سپس با جمع نتایج به‌دست‌آمده از پیش‌بینی دو مؤلفه‌ای غیرخطی و متلاطم شاخص قیمت، سری زمانی شاخص کل قیمت برآورد می‌شود. نتایج به‌دست‌آمده نشان می‌دهد مدل هیبریدی ارائه‌شده این پژوهش در مقایسه با سایر روش‌های پیش‌بینی، خطای کمتری داشته و از دقت بیشتری برخوردار است.

**واژه‌های کلیدی:** بهینه‌سازی ازدحام ذرات، تبدیل موجک، ماشین بردار پشتیبان، مدل GJR

۱. دانشجوی کارشناسی ارشد مهندسی مالی، دانشگاه صنعتی خواجه‌نصیرالدین طوسی، تهران، ایران

۲. استادیار گروه مهندسی مالی، دانشگاه صنعتی خواجه‌نصیرالدین طوسی، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۵/۰۴/۳۰

تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۰۹/۱۷

نویسنده مسئول مقاله: دیاکو درودی

E-mail: ddorodi@mail.kntu.ac.ir

### مقدمه

امروزه بازارهای مالی تأثیر ویژه‌ای بر اقتصاد کشورها دارند. بازارهای مالی از دیدگاه اقتصادی به دو بخش بازار پول و بازار سرمایه دسته‌بندی می‌شوند. بازار سرمایه که مختص منابع بلندمدت است، امکان مبادله منابع را میان دارندگان مازاد منابع با متقاضیان کسری منابع فراهم می‌آورد (نصیر طوسی، آقابرگی، وای و لینگ‌نگو، ۲۰۱۵). بازار بورس به‌عنوان یکی از ارکان بازار سرمایه، هم برای سرمایه‌گذاران کلان و هم برای عموم مردم در دسترس است. انتخاب رویکرد مناسب در سرمایه‌گذاری برای بسیاری از فعالان بازار بورس عاملی مهم و حیاتی است. در سرمایه‌گذاری، بررسی، تجزیه و تحلیل اطلاعات بازار و آگاهی از روند تغییرات اطلاعات در آینده، می‌تواند به کسب سود برای سرمایه‌گذار منجر شود.

در این میان شاخص به‌عنوان یک معیار آماری، قابلیت مقایسه وضعیت کنونی را نسبت به گذشته فراهم آورده و بررسی و تحلیل آن اطلاعات مفیدی را به کارشناسان و افراد ذی‌ربط در آن حوزه ارائه می‌دهد. شاخص‌های قیمت سهام در تمام بازارهای مالی دنیا، به‌مثابه یکی از مهم‌ترین معیارهای سنجش عملکرد بورس اوراق بهادار، اهمیت زیادی دارند و شاید مهم‌ترین دلیل توجه روزافزون به آنها، این نکته باشد که شاخص‌های یادشده از تجمیع حرکت‌های قیمتی سهام تمام شرکت‌ها یا طبقه خاصی از شرکت‌های موجود در بورس، به‌دست می‌آیند (ژیان ژو وانگ، ژنگ، گو و ژو ژی وانگ، ۲۰۱۱).

داده‌های تاریخی نشان می‌دهد ویژگی‌های پیچیده شاخص کل قیمت، مانند غیرخطی بودن، عدم قطعیت، نوسان و پویایی، پیش‌بینی آن را دشوار می‌کند و نتایج پیش‌بینی را با عدم قطعیت زیادی مواجه می‌سازد که خود تأثیر شایان توجهی در بازده سرمایه‌گذاران، صندوق‌های سرمایه‌گذاری، نهادهای سرمایه‌گذاری و سایر فعالان این حوزه به همراه دارد (خسروی‌نژاد و شعبانی، ۱۳۹۳). به همین دلیل، در پژوهش حاضر، مدل هیبریدی نوینی برای پیش‌بینی شاخص کل قیمت، با استفاده از مزایای و خصوصیات مدل‌های اقتصادسنجی و روش‌های هوش مصنوعی، در به تصویر کشیدن ویژگی‌های غیرخطی و متلاطم شاخص کل قیمت ارائه شده است. از مهم‌ترین مدل‌های غیرخطی که در سال‌های اخیر در بازارهای مالی بسیار استفاده شده و به نتایج مطلوبی نیز دست یافته، الگوریتم ماشین بردار پشتیبان<sup>۱</sup> است. کاربرد ماشین بردار پشتیبان در علم اقتصاد، اغلب در زمینه پیش‌بینی و دسته‌بندی متغیرها در بازارهای پولی و مالی از قبیل قیمت‌های سهام و نرخ ارز بوده است. ماشین بردار پشتیبان برای پیش‌بینی سری‌های

زمانی در شرایط غیر ایستایی متغیرها، بی‌توجیه بودن روش‌های کلاسیک یا پیچیدگی سری‌های زمانی، بسیار استفاده شده است (تیان، ۲۰۱۵).

تاکنون مدل‌ها و روش‌های مختلفی برای مدل‌سازی تلاطم ارائه شده که از جمله آنها می‌توان به مدل‌های نوسان خودرگرسیو شرطی اشاره کرد که ابتدا توسط انگل در سال ۱۹۸۲ معرفی شد و پس از وی، بولرسلو آن را به مدل‌های گارچ تعمیم داد. مدل‌های واریانس شرطی تاکنون مهم‌ترین و پرکاربردترین روش برای تجزیه و تحلیل و پیش‌بینی سری‌های زمانی مالی شناخته شده‌اند. در بازار مالی، ممکن است تأثیر شوک‌های منفی و مثبت بر تلاطم، متفاوت باشد. بررسی‌ها نشان می‌دهد شوک‌های منفی قیمت در مقایسه با شوک‌های مشابه ولی مثبت، تأثیر بیشتری بر تلاطم دارند (پاتون و انگل، ۲۰۰۱). به همین دلیل در این پژوهش از مدل GJR برای مدل‌سازی عدم تقارن در تأثیر شوک‌ها بر واریانس سری زمانی شاخص کل قیمت استفاده شده است.

نظریه تبدیل موجک<sup>۱</sup>، توانایی تجزیه سری‌های زمانی با مقیاس‌های مختلف را دارد و برای نوین‌سازی داده‌ها، افزایش دقت پیش‌بینی و تشخیص روند در سری‌های زمانی با تغییرات ناگهانی در مقایسه با روش‌های دیگر بهتر عمل می‌کند (محمدی و سعیدی، ۱۳۹۰). همچنین تبدیل موجک برای تجزیه ویژگی غیرخطی و متلاطم شاخص به کار می‌رود. از ترکیب این مدل‌ها در پیش‌بینی شاخص کل قیمت استفاده شده است.

## پیشینه نظری پژوهش

### تبدیل موجک

موجک‌ها توابع ریاضی هستند که داده‌ها را به اجزای فراوانی (فرکانس) تفکیک کرده و هر جزء را با نمایش متناسب با مقیاس آن جزء مطالعه می‌کنند. یکی از مزایای موجک‌ها، توان بالای تحلیل آنها در شرایطی است که سیگنال‌ها با گسستگی و جهش‌های سریع همراه‌اند.

در این تبدیل، پنجره‌ای که در حوزه زمان - فرکانس تعریف می‌شود، بر خلاف تبدیل فوریه دارای ابعاد ثابت نیست؛ بلکه بر حسب تغییرات زمانی سیگنال، این پنجره با وفق دادن خود با سیگنال قادر خواهد بود که به دقت در تفکیک زمان - فرکانس بالاتری از سیگنال دست یابد. در تبدیل موجک، تابع پنجره را موجک مادر نامیده و با  $\psi(t)$  نشان می‌دهیم. تبدیل موجک تابع

$f(t)$  بر حسب تابع موجک مادر  $\psi(t)$  در واقع تصویر تابع  $f(t)$  بر حسب تابع پایه  $\frac{1}{\sqrt{a}}\Psi^*\left(\frac{t-b}{a}\right)$  است و به صورت رابطه ۱ تعریف می‌شود.

$$\frac{1}{\sqrt{a}}\Psi^*\left(\frac{t-b}{a}\right) = \int_{-\infty}^{+\infty} f(t)\Psi^*\left(\frac{t-b}{a}\right) dt \quad \text{رابطه ۱}$$

$\Psi^*$  مزدوج مختلط موجک  $\psi$  است و متغیر  $a$  موجب گسترش یا تراکم تابع  $\psi_{a,b}(t)$  در حوزه زمان و متغیر  $b$  موجب تغییر موقعیت زمانی  $\psi_{a,b}(t)$  می‌شود. در تحلیل موجک ضریب  $C(a, b)$  به عنوان ضریبی که میزان شباهت سیگنال  $f(t)$  را با موجک مادر نشان می‌دهد، معرفی می‌شود. مهم‌ترین مرحله در آنالیز تبدیل موجک، انتخاب موجک مادر است. هرچه مرتبه مشتق‌پذیری تابع موجک مادر بالاتر باشد یا به بیان دیگر، هر چه تعداد ممان‌های صفر شونده موجک مادر بیشتر باشد، توانایی تبدیل موجک در تشخیص نقاط ویژه بیشتر است (دبنت و شا، ۲۰۰۲: ۱۹-۱۰).

### ماشین بردار پشتیبان

این الگوریتم را نخستین بار محقق به نام فیشر در سال ۱۹۶۳ با هدف کاهش خطای طبقه‌بندی داده‌های آموزشی، ارائه کرد. تکامل یافته آن نیز در سال ۱۹۹۵ توسط ولادیمیر وپنیک بر پایه نظریه آماری یادگیری مسئله ماشین بردار پشتیبان ارائه شد. این الگوریتم در مباحثی مانند رگرسیون، طبقه‌بندی، خوشه‌بندی و به طور کلی تقریب توابع کاربرد فراوانی دارد. الگوریتم ماشین بردار پشتیبان، ابتدا تنها طبقه‌بندی دو کلاس را شامل می‌شد. این رویکرد در ادامه با استفاده از انواع تکنیک‌های ترکیب به طبقه‌بندی چند کلاس نیز تعمیم یافت (محمدی و سعیدی، ۱۳۹۰). داده‌های آموزشی به صورت  $(x_1, y_1), \dots, (x_m, y_m) \in N * R$  است که در آن  $X$  معرف داده ورودی است. در رگرسیون  $\varepsilon$ -SV هدف پیدا کردن تابع  $f(x)$  است که حداکثر به اندازه  $\varepsilon$  انحراف از مقدار واقعی  $y_1$  برای همه داده‌های آموزشی دارد. تابع خطی  $f(x)$  به صورت رابطه ۲ تعریف می‌شود:

$$f(x) = \langle w, x \rangle + b \text{ with } w \in N, b \in R \quad \text{رابطه ۲}$$

علامت  $\langle \rangle$  نشان‌دهنده ضرب داخلی است. هموارشدگی در رابطه ۱ به معنای  $w$  کوچک است. برای رسیدن به این منظور باید نرم اقلیدسی  $\|w\|^2$  کمینه شود. به این معنا که باید  $\frac{1}{2}\|w\|^2$  کمینه شود و داشته باشیم:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \quad \text{رابطه ۳}$$

مسئله بهینه‌سازی فوق، با مشخص بودن دقیق تابع  $f$  و تقریب همه جفت‌های  $(x_i, y_i)$  با  $\varepsilon$  دقیق حل می‌شود. با معرفی متغیرهای  $\xi_i^*$  و  $\xi_i$  برای مقابله با سایر محدودیت‌های غیرقابل حل، مسئله بهینه به صورت رابطه‌های ۴ و ۵ خواهد بود.

$$\text{Minimize } 1/2 \|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \quad \text{رابطه ۴}$$

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i > 0 \end{cases} \quad \text{رابطه ۵}$$

تابع خطای  $\varepsilon$ -insensitive به صورت رابطه ۶ تعریف می‌شود.

$$|\xi|_i = \begin{cases} 0 & \text{if } |\xi| < \varepsilon \\ |\xi| < \varepsilon & \text{otherwise} \end{cases} \quad \text{رابطه ۶}$$

فرمول دوگانه لاگرانژ، راه‌حلی به منظور توسعه ماشین بردار برای توابع غیرخطی فراهم می‌آورد. روش استاندارد دوگانه، از ضرایب لاگرانژ به صورت رابطه ۷ استفاده می‌کند.

$$\begin{aligned} L = \frac{1}{2\|w\|^2} &- \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ &- \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \\ &- \sum_{i=1}^l (\gamma_i \xi_i + \gamma_i^* \xi_i^*) + c \sum_{i=1}^l (\xi_i + \xi_i^*) \end{aligned} \quad \text{رابطه ۷}$$

ضرایب دوگانه رابطه ۵، باید با در نظر گرفتن اینکه همواره مثبت‌اند، تعیین شوند. برای حل رابطه بالا از  $L$  نسبت به متغیرهای  $w$ ،  $b$ ،  $\xi_i^*$  و  $\xi_i$  مشتق گرفته و برابر صفر قرار می‌دهیم تا مقادیر آنها مشخص شود (سولما و اسکولکاف، ۲۰۰۴).

### بهینه‌سازی ازدحام ذرات

در سال ۱۹۹۵ ابرهلت و کندی برای اولین بار الگوریتم بهینه‌سازی ازدحام ذرات<sup>۱</sup> را به‌عنوان روش جست‌وجوی غیرقطعی برای بهینه‌سازی تابع مطرح کردند. روش بهینه‌سازی ازدحام ذرات، نوعی الگوریتم جست‌وجوی سراسری با تابع هدف کمینه‌سازی است که به کمک آن می‌توان مسائلی را که جواب آنها یک نقطه یا سطح در فضای  $n$  بعدی است، از روی رفتار ذرات مدل‌سازی کرد (آبه، ۲۰۰۵).

در بهینه‌سازی ازدحام ذرات، ذرات در فضای جست‌وجو جاری می‌شوند و تغییر مکان ذرات در فضای جست‌وجو، تحت تأثیر سرعت و مسیر ابتدایی آنها، تجربه و دانش خودشان و همسایگانشان است. بر همین اساس در بهینه‌سازی ازدحام ذرات، در هر لحظه، هر ذره مکان خود را در فضای جست‌وجو با توجه به بهترین مکانی که تا کنون در آن قرار گرفته، تنظیم می‌کند و بر مبنای ملاک شایستگی به‌دست‌آمده، به سمت بهترین همسایگان خود می‌رود. در آغاز کار تعدادی ذره به‌صورت تصادفی در فضای مسئله به‌وجود می‌آیند و با به‌روز کردن نسل‌ها و توجه به هوش تجمعی و ملاک بهینه‌سازی، سعی در یافتن راه حل بهینه می‌نمایند. در هر تکرار، هر ذره با دو موقعیت بهینه به‌روز می‌شود. اولین مورد، بهترین موقعیتی است که تا کنون ذره موفق به رسیدن به آن شده است. دومین مورد، بهترین موقعیتی است که تا کنون توسط جمعیت ذرات به‌دست آمده است و به‌عنوان یک بهینه سراسری در این مرحله در نظر گرفته می‌شود.

$$v_j^i[t+1] = w * v_j^i[t] + c_1 * f_1(x_j^{i,best}[t] - x_j^i[t]) + c_2 * r_2(x_j^{g,best}[t] - x_j^i[t]) \quad \text{رابطه ۸}$$

$$x_j^i[t+1] = x_j^{i,best}[t] + v_j^i[t+1] \quad \text{رابطه ۹}$$

هر ذره از سه بردار  $d$  بعدی تشکیل شده است که  $d$  بعد فضای جست‌وجو است. ذره نام در زامین تکرار این سه بردار عبارت است از:  $x_j^i$  موقعیت فعلی ذره،  $v_j^i$  سرعت حرکت ذره و  $x_j^{i,best}$  بهترین موقعیتی که ذره تا تکرار تجربه کرده است (شای، ۲۰۰۴).

### مدل GJR

مدل GJR برای اولین بار توسط گلسنن و همکارانش در سال ۱۹۹۳، به‌عنوان نسخه نامتقارنی از مدل گارچ معرفی شد. این مدل، طبیعت نامتقارن رفتار سرمایه‌گذار را در پاسخ به بازده سهام یا شاخص‌های بازار و عدم تقارن اثر شوک‌های مثبت و منفی بر واریانس‌های شرطی، در نظر می‌گیرد. منطق مدل GJR بر این واقعیت تجربی استوار است که شوک‌های منفی در زمان  $t - 1$  تأثیر قوی‌تری در واریانس در زمان  $t$  نسبت به شوک‌های مثبت در همان زمان دارند. این عدم تقارن که اثر اهرمی نامیده می‌شود، خود ریسک را افزایش می‌دهد و منشأ آن شوک‌های منفی است (دوان، گوتیر و سیموناتو، ۲۰۰۶).

$$\sigma_t^2 = w + \sum_{i=1}^p (\alpha_i + \gamma_i I_{t-i}) \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad \text{رابطه ۱۰}$$

### پیشینه تجربی

نتایج مطالعات گذشته نشان می‌دهد استفاده از ابزارها و روش‌های کلاسیک پیش‌بینی، خطای زیادی داشته و در مقایسه با روش‌های جدیدتر و مدل‌های غیرخطی، عملکرد ضعیف‌تری دارند. با پیشرفت‌های اخیر به‌ویژه در حوزه هوش مصنوعی، روش‌های جدیدی برای پیش‌بینی ارائه شده است که نسبت به روش‌های سنتی، از دقت بالاتری برخوردارند (فلاح‌پور، گل‌عزیزی و فتوره‌چیان، ۱۳۹۲). در ادامه به برخی از پژوهش‌هایی که در این حوزه انجام شده است، اشاره کوتاهی می‌شود.

پژوهشگران با استفاده از روش ترکیبی شبکه عصبی و تبدیل موجک به پیش‌بینی شاخص قیمت سهام شانگهای پرداختند. در این مقاله دو روش ترکیبی شبکه عصبی و تبدیل موجک با شبکه عصبی مقایسه شد و نتایج نشان داد کاهش نویز و اختلال در داده‌ها، سبب بهبود پیش‌بینی و کاهش خطا می‌شود (ژو ژی وانگ، ژیان ژو وانگ، ژانگ و گو، ۲۰۱۲). در پژوهشی دیگر، محققان یک مدل پیش‌بینی با استفاده از روش تجزیه حالت تجربی EMD و گروه تجربی حالت تجزیه EEMD و شبکه عصبی مصنوعی ANN برای پیش‌بینی کوتاه‌مدت شاخص سهام ارائه دادند. در مرحله اول، با استفاده از دو روش EMD و EEMD داده‌ها به مجموعه‌های کوچک‌تر تجزیه شده و در مرحله بعد، به‌عنوان ورودی مدل شبکه عصبی، برای آموزش و پیش‌بینی قیمت سهام در آینده استفاده شدند. نتایج نشان داد مدل پیش‌بینی با استفاده از روش ارائه شده نسبت به مدل‌های شبکه عصبی و رگرسیون بردار پشتیبان خطای کمتری دارد

(جویتمانی، شانکار و یاداو، ۲۰۱۶). در پژوهشی، برای پیش‌بینی شاخص قیمت‌های آتی چین (CSI)، از مجموعه‌های فازی و روش سری زمانی فازی چندمتغیره استفاده شده است. در این مقاله محققان از ۳۰۰ شاخص آتی به‌عنوان ورودی مدل استفاده کردند و ایجاد یک مدل چندمتغیره به بهبود دقت و صحت محاسبات منجر شد (سون و همکاران، ۲۰۱۵). پاتل، شاه، تاکار و کوتچه (۲۰۱۵) نیز با استفاده از چهار مدل شبکه عصبی مصنوعی، ماشین بردار پشتیبان، جنگل تصادفی و شبکه‌های بیزین، به پیش‌بینی مسیر حرکت سهام و قیمت سهام شاخص بازارهای سهام هند پرداختند. این پژوهش با دو روش ورودی به مقایسه مدل‌های یاد شده می‌پردازد. در روش اول داده‌های ورودی شامل ۱۰ پارامتر تکنیکال محاسبه‌شده با استفاده از داده معاملات سهام (باز، بالا، پایین و قیمت نزدیک) است؛ ولی در روش دوم پارامترهای تکنیکال دارای روند قطعی هستند. نتایج پژوهش نشان داد برای روش اول، جنگل تصادفی در کل بهتر از سه مدل پیش‌بینی دیگر عمل می‌کند. همچنین عملکرد تمام مدل‌های پیش‌بینی هنگامی که پارامترهای تکنیکال دارای روند هستند، بهبود داده شده است.

پژوهشگران داخلی نیز از الگوریتم‌های هوش مصنوعی و به‌طور مشخص از شبکه عصبی برای پیش‌بینی سری‌های زمانی بهره برده‌اند که در ادامه بخشی از این تحقیقات مرور می‌شود. در پژوهشی، شناسایی و رتبه‌بندی عوامل مالی مؤثر بر شاخص بورس اوراق بهادار تهران با استفاده از روش تاپسیس انجام گرفته است. نتایج به‌دست آمده از روش تاپسیس نشان می‌دهد به ترتیب نسبت قیمت بر درآمد، سیر تاریخی قیمت سهام، عایدی هر سهم و میزان بازدهی دارایی‌ها بیشترین تأثیر را بر شاخص قیمت سهام دارد (پاکدین امیری، پاکدین امیری، پاکدین امیری، ۱۳۸۸). رابطه میان شاخص صنعت و مالی دارای شدت مولتی فراکتالیته بیشتری است و نیز، آثار دم‌های تابع توزیع احتمال در شاخص‌های صنعت و مالی بر ساختارهای همبستگی آنها با شاخص قیمت اهمیت دارد. پژوهشگران با استفاده از الگوریتم پرواز پرندگان به پیش‌بینی شاخص کل قیمت سهام در بازه زمانی بین سال‌های ۱۳۷۸ تا ۱۳۸۷ پرداختند. آنها به‌منظور ارزیابی، الگوریتم پرندگان را با الگوهای سنتی مانند نمو هموار ساده، هلت وینترز، اتو رگرسیون، میانگین متحرک و آریما مقایسه کردند که خطای برآورد نشان‌دهنده برتری بر الگوهای سنتی بود (دموری، فرید و اشهر، ۱۳۹۰). در پژوهشی دیگر بر مبنای مدل‌های ترکیب خبرگان، ساختار مدل GLGNN معرفی شده و با استفاده از آن، مدلی برای پیش‌بینی شاخص بورس تهران ارائه شده است. مدل ترکیب خبرگان، سیستمی شامل مجموعه‌ای از خبرگان است که در آن، هر خبره یک شبکه عصبی پیش‌خور محسوب می‌شود. نتایج نشان داد با وجود شکل نامتعارف توزیع داده‌ها، مدل ارائه‌شده کارایی مناسبی در مدل‌سازی و پیش‌بینی شاخص بورس تهران دارد



(یگانی و چینی‌پرداز، ۱۳۹۰). در پژوهشی، تهرانی، نمکی و هدایتی‌فرد (۱۳۹۱) با بررسی ساختار همبستگی میان شاخص‌های بورس اوراق بهادار تهران در قالب یک شاخص قیمت و دو شاخص شاخص مالی و صنعت به نتایجی در خصوص وابستگی میان تغییرات هر شاخص به تغییرات مقادیر گذشته شاخص‌های دیگر دست یافتند. مشبکی و کردانی (۱۳۹۲) نیز به بررسی پیش‌بینی‌پذیری شاخص قیمت و بازده نقدی بورس اوراق بهادار تهران به کمک انفیس<sup>۱</sup> و یافتن مدل مناسب برای پیش‌بینی این شاخص پرداختند. بدین منظور، نخست سه متغیر کلان اقتصادی همراه با مقادیر تاریخی شاخص به‌عنوان ورودی‌های مدل انتخاب شدند؛ سپس ساختارهای گوناگون انفیس و شبکه عصبی مصنوعی پس‌انتشار خطا برای بررسی پیش‌بینی‌پذیری و شناسایی مدل مناسب انتخاب شد. نتایج به‌دست‌آمده در بازه زمانی ۱۳۷۹ تا ۱۳۹۱، نشان می‌دهد مدل‌های ساده‌تر در پیش‌بینی شاخص، دقت و عملکرد بهتری دارند و پیش‌بینی انفیس نسبت به شبکه عصبی پس‌انتشار خطا دقیق‌تر است. محققان دیگری برای پیش‌بینی شاخص هفتگی قیمت سهام در بورس تهران در سال‌های ۸۷-۱۳۸۳، از روش سری‌های زمانی تک‌متغیره و شبکه عصبی مصنوعی بهره بردند. هدف از انجام این پژوهش، معرفی بهترین روش برای پیش‌بینی شاخص کل بورس و مقایسه آن توان آزمون‌های خطی و غیرخطی در پیش‌بینی بازارهای مالی بود (خسروی‌نژاد و شعبانی، ۱۳۹۳). فلاح‌پور و علی‌پور (۱۳۹۳)، به پیش‌بینی شاخص کل سهام بورس اوراق بهادار تهران با استفاده از شبکه‌های عصبی موجکی پرداختند. ابتدا از تبدیل موجک گسسته برای نویززدایی داده‌ها در سری زمانی استفاده کردند؛ سپس به کمک شبکه‌های عصبی به پیش‌بینی شاخص سهام پرداختند. بر اساس نتایج، عملکرد شبکه عصبی موجکی سطح خطای کمتری نسبت به شبکه عصبی معمولی در پیش‌بینی شاخص سهام داشت.

### روش‌شناسی پژوهش

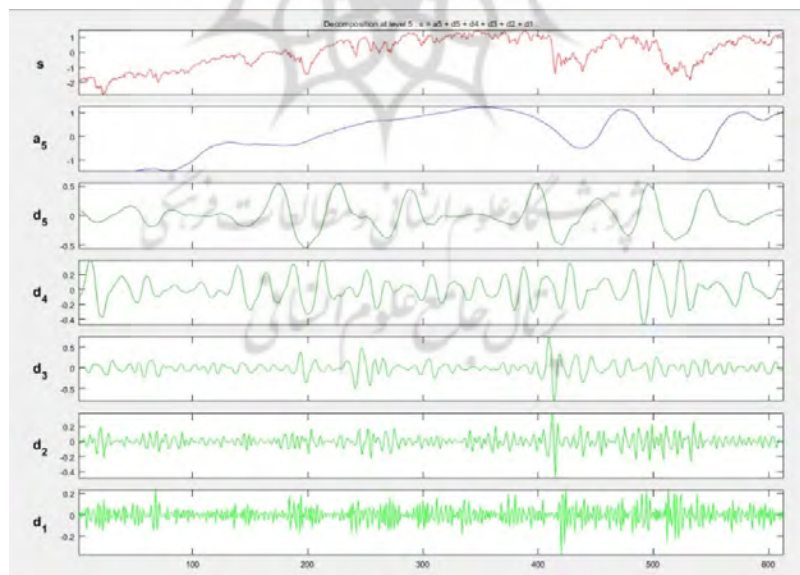
این پژوهش از نظر رویکرد کاربردی بوده و مبتنی بر پژوهش‌های میدانی است. جامعه آماری آن، شاخص کل سهام بورس اوراق بهادار تهران از ابتدای سال ۱۳۸۷ تا پایان سال ۱۳۹۴ است. برای ایجاد مدل‌سازی و پیش‌بینی شاخص کل، نمونه‌ای از اطلاعات، مشتمل بر ۶۱۳ مورد داده از معاملات روزانه بورس از طریق نرم‌افزار سازمان بورس استخراج شد که به دو سری داده آموزش و آزمون دسته‌بندی شدند. این پژوهش درصدد است که مدل پیش‌بینی جدیدی برای پیش‌بینی شاخص سهام بازار ارائه کند؛ به‌طوری که با تفکیک سری داده‌ها به چند بخش و پیش‌بینی آن

اجزا به وسیله روش های متناسب با ویژگی آنها، نتیجه مطلوب تری ارائه دهد. مدل ترکیبی مدنظر با استفاده از تبدیل موجک، سری زمانی را به شش سری مجزا تفکیک می کند که هر یک از این سری ها، ویژگی های غیرخطی و متلاطم شاخص مدنظر را نمایندگی می کنند؛ سپس با استفاده از ترکیب الگوریتم های بهینه سازی ازدحام ذرات و ماشین بردار پشتیبان، سه سری مربوط به ویژگی های غیرخطی را پیش بینی می کند. همچنین سه سری مرتبط با ویژگی های متلاطم با استفاده از مدل GJR پیش بینی می شوند.

در ادامه کار با بهره مندی از ترکیب خطی نتایج به دست آمده از این شش پیش بینی، سری زمانی مربوط به شاخص را تخمین می زند. با ترکیب این روش های مختلف تلاش شده است که جنبه های مختلف روند تغییرات شاخص کل قیمت به تصویر کشیده شود. مراحل اجرای این پژوهش به شرح زیر است:

مرحله ۱) سری زمانی شاخص کل قیمت با استفاده از روش موجک، به سری های زمانی جزئیات و تقریب، بسته به سطح تجزیه تبدیل می شود. شاخص کل قیمت به صورت رابطه ۱۱ به دست می آید.

$$X(t) = \sum_{i=1}^6 x_i(t) \quad \text{رابطه ۱۱}$$



شکل ۱. سری زمانی داده های واقعی و ویژگی های استخراج شده از تبدیل موجک

مرحله ۲) سری‌های زمانی به‌دست آمده از مرحله اول را می‌توان به‌صورت دو مؤلفه‌ای متشکل از ویژگی‌های غیرخطی و متلاطم در نظر گرفت. از روش موجک برای استخراج ویژگی‌های غیرخطی و متلاطم شاخص کل استفاده می‌شود. رابطه ۱۲، مؤلفه اول رفتار غیرخطی و مؤلفه دوم رفتار متلاطم را به تصویر می‌کشد.

$$\sum_{i=1}^6 x_i(t) = \sum_{i=1}^m N_i(t) + \sum_{m+1}^6 V_i(t) \quad (\text{رابطه ۱۲})$$

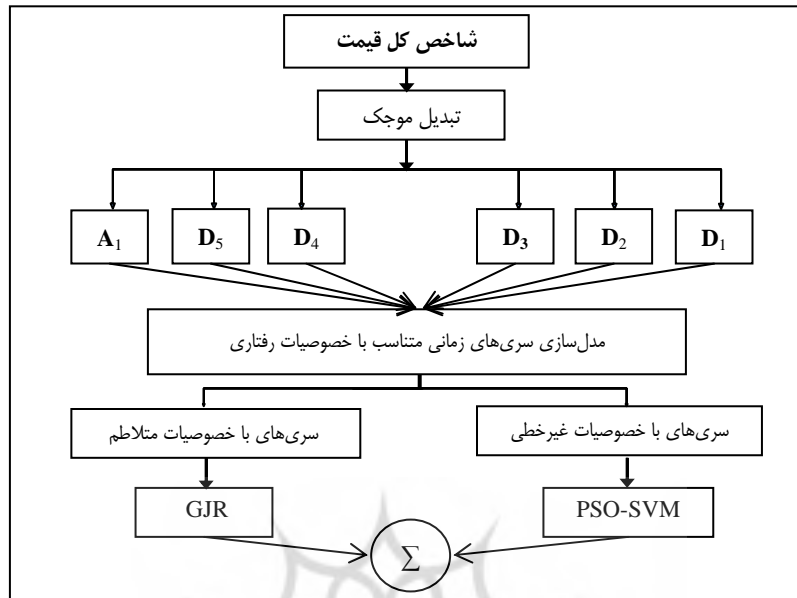
مرحله ۳) با توجه به شکل ۲، رفتار متلاطم شاخص کل قیمت را می‌توان در سری‌های زمانی  $d_1, d_2, d_3$  مشاهده کرد. برای مدل‌سازی و پیش‌بینی مؤلفه متلاطم از مدل GJR استفاده می‌شود. مقدار پیش‌بینی به‌دست آمده از مدل GJR برابر با  $\hat{v}_i(t)$  است.

مرحله ۴) با توجه به شکل ۲، سری‌های زمانی  $d_4, d_5$  و دارای ویژگی غیرخطی هستند که برای مدل‌سازی و پیش‌بینی آنها، الگوریتم‌های ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات به کار گرفته شده است. برای تنظیم پارامترهای الگوریتم ماشین بردار پشتیبان، از جمله پارامتر ضریب جریمه در تابع هدف و مقدار واریانس در تابع کرنل، از الگوریتم بهینه‌سازی ازدحام ذرات استفاده می‌کنیم. مقدار پیش‌بینی به‌دست آمده از الگوریتم ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات برابر با  $\widehat{N}_i(t)$  است.

مرحله ۵) مجموع مقادیر پیش‌بینی شده توسط الگوریتم ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات برای سری زمانی با خاصیت غیرخطی را با مقادیر پیش‌بینی شده توسط مدل GJR برای سری‌های زمانی با خاصیت متلاطم به‌دست می‌آوریم و به‌عنوان مقدار پیش‌بینی شاخص کل قیمت سهام در نظر می‌گیریم که به‌صورت رابطه ۱۳ تعریف می‌شود.

$$\sum_{i=1}^6 \hat{x}_i(t) = \sum_{i=1}^m \hat{v}_i(t) + \sum_{m+1}^6 \widehat{N}_i(t) \quad (\text{رابطه ۱۳})$$

برای مقایسه عملکرد نتایج پیش‌بینی الگوریتم ترکیبی ارائه‌شده، از الگوریتم‌های ترکیبی ماشین بردار پشتیبان - موجک، موجک - شبکه عصبی و موجک - GJR استفاده شده است. در هر سه روش، ابتدا تبدیل موجک را اعمال می‌کنیم؛ سپس هر شش سری زمانی تفکیک‌شده را با استفاده از یک روش پیش‌بینی کرده و نتایج را ترکیب می‌نماییم. اکنون نتایج به‌دست آمده از اجرای مدل ترکیبی را با سه مدل پایه مقایسه می‌کنیم.



شکل ۲. الگوریتم پیش بینی شاخص کل قیمت توسط مدل هیبریدی ارائه شده

پس از طراحی مدل ها و پیش بینی نتایج با روش مدل ترکیبی ارائه شده و سه روش دیگر، به محاسبه میزان خطای روش ها و مقایسه دقت پیش بینی این روش ها پرداخته شده است. برای تحلیل نتایج و مقایسه این روش ها از معیار میانگین قدر مطلق خطاها، ریشه میانگین مجذور خطاها و میانگین قدر مطلق درصد خطاها استفاده می کنیم. شکل ۲ مراحل کلی روش ارائه شده در این مقاله را نمایش می دهد.

فرضیه پژوهش به این صورت تعریف شده است: روش ترکیبی ارائه شده در این پژوهش نسبت به الگوریتم های ترکیبی ماشین بردار پشتیبان - موجک، موجک - شبکه عصبی و موجک - GJR، عملکردی بهتری در پیش بینی شاخص کل سهام بورس اوراق بهادار تهران دارد.

### یافته های پژوهش

ارزیابی نتایج پیش بینی و در واقع گزینش معیار ارزیابی مناسب از نکات حائز اهمیت در مسئله پیش بینی است. در ادبیات این حوزه، عمدتاً از سه معیار میانگین قدر مطلق خطاها؛ ریشه میانگین مجذور خطاها و میانگین قدر مطلق درصد خطاها استفاده شده است. دو معیار نخست کاربرد بیشتری در زمینه سنجش میزان دقت پیش بینی دارند. معیار دوم از خانواده محاسبات میانگین

ارائه روش هیبریدی نوین برای پیش‌بینی شاخص کل قیمت .... ۶۲۵

خطای استاندارد و معیار اول و سوم در مورد خطای مطلق است. در این رابطه‌ها  $Q_t$  مقدار واقعی شاخص کل قیمت و  $\hat{Q}_t$  برابر با مقدار پیش‌بینی شده است.

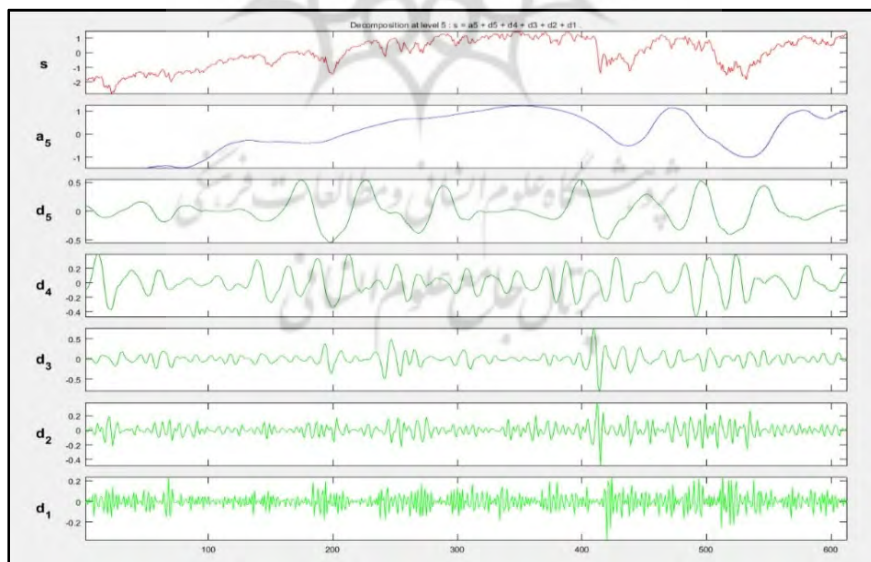
$$MAE = 1/T \sum_{t=1}^T |Q_t - \hat{Q}_t| \quad \text{رابطه ۱۴}$$

$$RMSE = \sqrt{1/T \sum_{t=1}^T (Q_t - \hat{Q}_t)^2} \quad \text{رابطه ۱۵}$$

$$MAPE = 1/T \sum_{t=1}^T \frac{|Q_t - \hat{Q}_t|}{Q'_t} \quad \text{رابطه ۱۶}$$

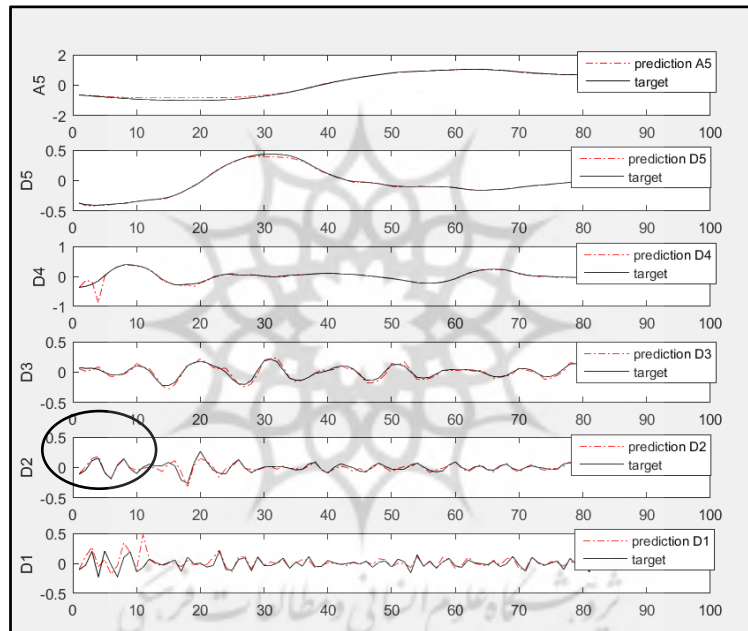
$$Q'_t = 1/T \sum_{t=1}^T Q_t \quad \text{رابطه ۱۷}$$

در شکل ۳ ویژگی‌های استخراج شده توسط تبدیل موجک نمایش داده شده است که  $d_1$ ،  $d_2$  و  $d_3$ ، گویای ویژگی متلاطم شاخص کل است و سری‌های  $d_4$ ،  $d_5$  و  $a_5$  ویژگی غیرخطی این شاخص را نشان می‌دهند.



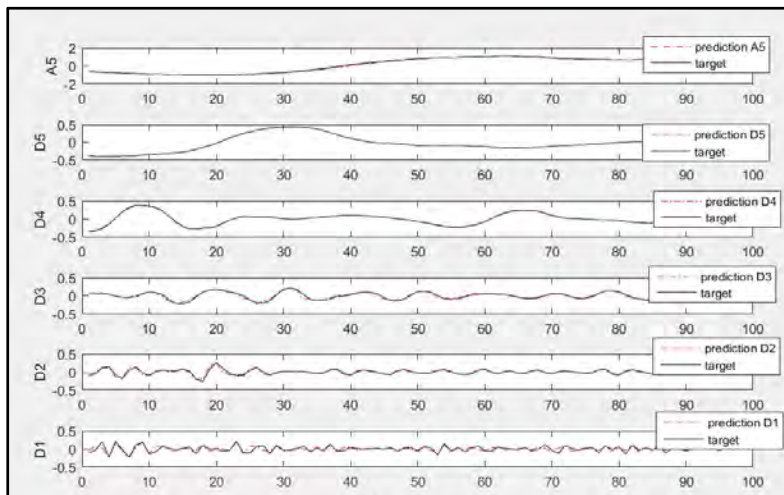
شکل ۳. سری زمانی داده‌های واقعی و ویژگی‌های استخراج شده از تبدیل موجک

شکل ۴ نتایج پیش‌بینی شش سری استخراج شده تابع موجک با استفاده از شبکه عصبی نشان داده شده است. همان‌گونه که می‌بینیم در ویژگی‌های متلاطم سری زمانی، به‌ویژه در فواصل زمانی که تلاطم شدت یافته و سرعت تغییرات قیمت بسیار بالاست، شبکه عصبی توانایی پیش‌بینی نوسانات شدید را ندارد و نمی‌تواند تغییرات شدید را به خوبی تخمین بزند. اما در سری‌های زمانی که شدت نوسانات کمتر شده و بیشتر رفتار غیرخطی وجود دارد، شبکه عصبی مدل‌سازی بهتری انجام داده و قابلیت پیش‌بینی بهتری دارد. ضعف شبکه عصبی در پیش‌بینی تلاطم سری زمانی در  $d_1$  با دایره سیاه نشان داده شده است.



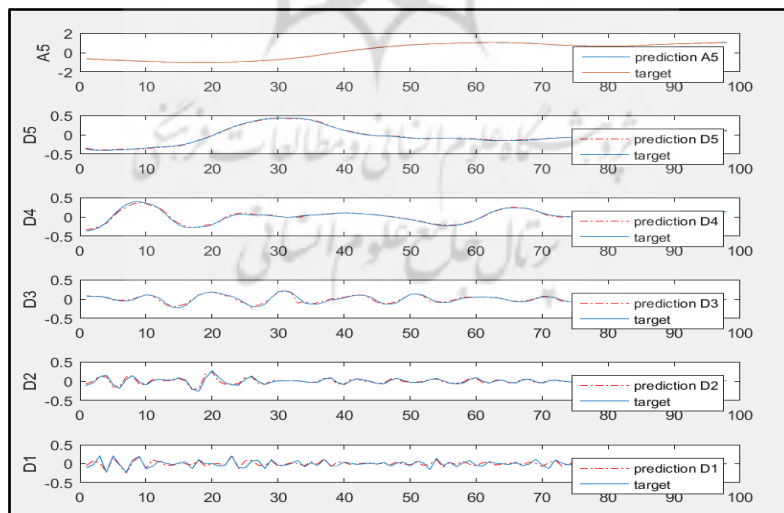
شکل ۴. پیش‌بینی شاخص کل قیمت توسط مدل شبکه عصبی

شکل ۵ سری‌های زمانی تفکیک‌شده که با روش GJR پیش‌بینی شده‌اند را نمایش می‌دهد. در سری  $d_3$ ، مشاهده می‌شود که مقدار تلاطم داده‌ها به خوبی پیش‌بینی شده است؛ سری‌های  $d_1$  و  $d_2$  به نسبت تلاطم بیشتری از سری زمانی در خود نشان می‌دهند و به‌طور طبیعی اختلاف مقادیر پیش‌بینی‌شده با مقادیر واقعی بیشتر خواهد بود. نوسان داده و شکست ساختاری عوامل مهمی در حساسیت پیش‌بینی سری‌های زمانی مالی هستند که سری‌های متلاطم این ویژگی‌ها را دارند؛ مشاهده می‌شود که مدل موجک-GJR توانایی بهتری در پیش‌بینی این نوع سری‌ها از خود نشان داده است.



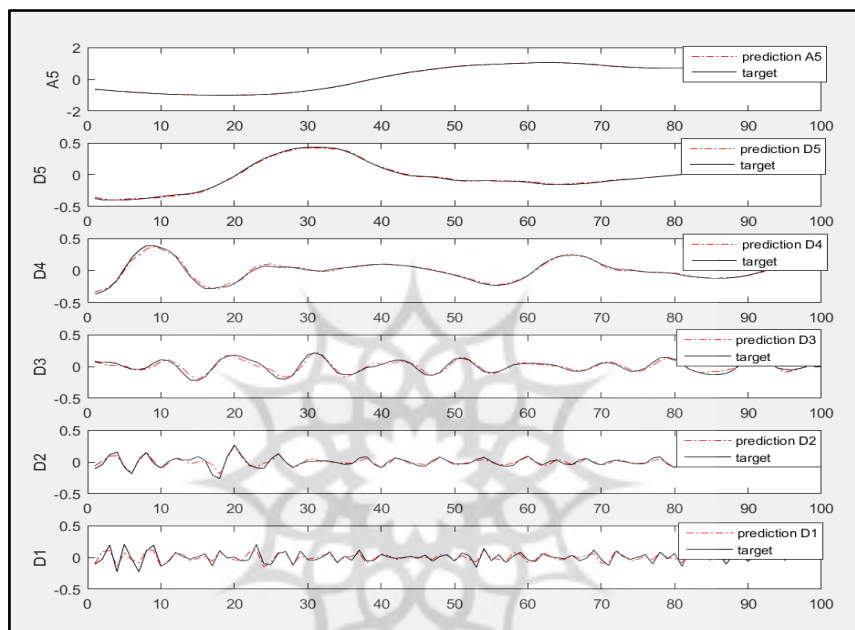
شکل ۵. پیش‌بینی شاخص کل قیمت توسط روش GJR

در شکل ۶ همان‌طور که مشاهده می‌شود، سری‌های پیش‌بینی شده با ویژگی‌های غیرخطی دقت زیادی دارند و این نتیجه قابل تصور است که ماشین بردار پشتیبان نسبت به شبکه عصبی، تئوری بنیادی قوی‌تری داشته و به دلیل ساختار برنامه‌ای درجه دوم به جواب جهانی می‌رسد؛ نسبت به بیش‌برازش مقاوم‌تر است و به تفسیر هندسی دست پیدا می‌کند. به همین دلیل، انتظار می‌رود نسبت به مدل شبکه عصبی پیش‌بینی بهتری انجام دهد.



شکل ۶. پیش‌بینی شاخص کل قیمت توسط روش SVM-PSO

در شکل ۷ نتایج پیش‌بینی روش هیبریدی نشان داده شده است؛ همان‌طور که انتظار می‌رود ویژگی‌های غیرخطی که توسط روش SVM پیش‌بینی شده‌اند، دقت زیادی دارند و ویژگی‌های متلاطم که توسط الگوریتم GJR پیش‌بینی شده‌اند، تغییرات را بیشتر پوشش داده‌اند. در نتیجه ترکیب این شش سری، پیش‌بینی دقیق‌تری از داده‌ها را ارائه می‌کند.



شکل ۷. پیش‌بینی شاخص کل قیمت توسط روش هیبریدی

بنابراین، با مقایسه نتایج هر یک از الگوریتم‌ها با داده‌های واقعی، سه شاخص مد نظر محاسبه شدند که نتایج این محاسبات در جدول ۱ آمده است. ارزیابی قدرت پیش‌بینی مدل هیبریدی و مقایسه آن با سه مدل دیگر حاکی از تفاوت در میزان صحت پیش‌بینی این الگو نسبت به سایر الگوهاست. با توجه به نتایج مندرج در جدول ۱، می‌توان پی برد که مدل هیبریدی ضمن عملکرد بهتر، در پیش‌بینی شاخص قیمت کمترین خطا را دارد. در مقایسه قدرت پیش‌بینی، شبکه عصبی - موجک در هر سه شاخص عملکرد ضعیف‌تری نسبت به سه مدل دیگر دارد. دو مدل موجک - GJR و ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات - موجک با اختلاف کمی نسبت به مدل هیبریدی، پیش‌بینی مناسبی انجام داده‌اند.



نتایج ارائه شده در جدول ۱ نشان می‌دهد که روش هیبریدی پیشنهاد شده، عملکرد بهتری برای مدل‌سازی سیستم از خود نشان داده است. سایر روش‌های مندرج در این جدول، یافته‌های تحقیقات دیگر است و روش‌های آن مقالات در این مقاله پیاده‌سازی شده است. برای مثال، روش شبکه عصبی - موجک در کار تحقیقی وانگ و همکاران (۲۰۱۱) و فلاح‌پور و علی‌پور (۱۳۹۳) بررسی شده و در تحقیق فلاح‌پور و همکارانش (۱۳۹۲) روشی بر پایه الگوریتم ماشین بردار پشتیبان پیاده‌سازی شده است.

در جدول ۲ نیز، مقایسه نتیجه الگوریتم هیبریدی پیشنهادی با نتایج برخی از الگوریتم‌های پیاده‌سازی شده در ادبیات موضوع در این حوزه نمایش داده شده است.

جدول ۱. نتایج خطای پیش‌بینی مدل‌های

| MAPE     | RMSE     | MAE      | مدل   | معیار |
|----------|----------|----------|---|-------|
| ۰/۲۰۱۲۰۳ | ۰/۰۵۳۴۵۸ | ۰/۰۳۳۸۴۶ | مدل هیبریدی   |       |
| ۰/۳۳۶۹۷۱ | ۰/۱۰۸۲۰۷ | ۰/۰۵۶۶۸۴ | شبکه عصبی - موجک                                    |       |
| ۰/۲۳۸۸۲۶ | ۰/۰۶۵۳۵۲ | ۰/۰۴۰۱۷۴ | ماشین بردار پشتیبان و بهینه‌سازی ازدحام ذرات - موجک |       |
| ۰/۲۴۹۶۶۷ | ۰/۰۶۷۷۲۴ | ۰/۰۴۱۹۹۸ | موجک - GJR  |       |

جدول ۲. مقایسه نتایج الگوریتم هیبریدی با الگوریتم پیاده‌سازی شده در ادبیات موضوعی

| MAPE     | RMSE     | سال  | مقاله             | مدل         | معیار |
|----------|----------|------|-------------------|-------------|-------|
| ۰/۲۰۱۲۰۳ | ۰/۰۵۳۴۵۸ | ---  | ---               | مدل هیبریدی |       |
| ---      | ۰/۱۵۳۴۵۸ | ۲۰۰۹ | پاکدامن و همکاران | شبکه عصبی   |       |
| ۰/۰۷۰۷۲۴ | ---      | ۲۰۰۹ | دموری و همکاران   | پرندهگان    |       |
| ---      | ۰/۱۳۴۳   | ۲۰۱۳ | مشیکی و همکاران   | ANFIS       |       |

## نتیجه گیری

شاخص کل قیمت از عوامل گوناگون و گاه متناقضی تأثیر می‌پذیرد که تحلیل این رفتار غیرخطی و پر نوسان، مستلزم استفاده از الگوریتم‌های پیچیده است. با توجه به کارایی بالای الگوریتم‌های هوش مصنوعی و مدل‌های اقتصادسنجی در پیش‌بینی، مدل هیبریدی ارائه شده هر دو مفهوم ممکن را در خود جای داده است. مقایسه خطای کل پیش‌بینی مدل هیبریدی ارائه شده با مدل‌های دیگر، نشان داد مدل ترکیبی قوی‌تر از مدل‌های دیگر عمل کرده و پیش‌بینی به‌دست آمده از آن خطای کمتری داشته است.

موضوعات زیر برای پژوهش‌های آتی پیشنهاد می‌شود:

۱. ترکیب SVM با الگوریتم‌های بهینه‌سازی دیگری همچون الگوریتم مورچگان، الگوریتم ژنتیک و غیره بررسی شود.
۲. متغیرهای تأثیرگذار بر شاخص قیمت انتخاب شده و به‌وسیله آنها پیش‌بینی انجام گیرد.
۳. از مدل‌های دیگری برای پیش‌بینی ویژگی غیرخطی و متلاطم شاخص استفاده شود.

## References

- Abe, S. (2005). *Support vector machines for pattern classification* (Vol. 53). London: Springer.
- Debnath, L. & Shah, F. A. (2002). *Wavelet transforms and their applications*. Boston: Birkhäuser.
- Dmory, D., Farid, D., Ashhr, M. (2011). Forecasting Tehran Stock Exchange index using Swarm comparison with the traditional patterns. *Journal of Accounting*, 2 (5), 7-30. (in Persian)
- Duan, J., Gauthier, G., Simonato, J. & Sasseville, C. (2006). Approximating the GJR-GARCH and EGARCH option pricing models analytically. *Journal of Computational Finance*, 9(3), 41.
- Engle, R. F. & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1 (2), 237-245.
- Fallahpour, R., Golarzi, Gh. & Fatourechian, N. (2014). Predicting the trend of stock prices using support vector machine based on genetic algorithms in Tehran Stock Exchange. *Scientific-Research Journal of Finance*, 15 (2), 269-288. (in Persian)

- Fallahpour, S., Alipur Reykandeh, C. (2014). Predicting the stock index using wavelet neural networks in the Tehran Stock Exchange. *Financial management strategy*, 2 (4), 15-31. (in Persian)
- Jothimani, D., Shankar, R. & Yadav, S.S. (2016). A Comparative Study of Ensemble-based Forecasting Models for Stock Index Prediction. *MWAIS 2016 Proceedings*. Paper 5
- Khosravinezhad, A., Shabani, M. (2014). Evaluation of linear and nonlinear models to predict stock price index at Tehran Stock Exchange. *Journal of Economic Sciences*, 38(8), 51-64. (in Persian)
- Moshabaki, A. & Kordnaeij, A. (2013). Predicting Tehran's Stock Market Index With Adaptive Network-Based Fuzzy Inference System (ANFIS). *Asset Management & Financing*, 1(1), 27-44. (in Persian)
- Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y. & Ngo, D. C. L. (2015). Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications*, 42(1), 306-324.
- Pakdin Amiri, M., Pakdin Amiri, M. & Pakdin Amiri, A. (2009). Prioritize financial factors affecting the price index at Tehran Stock Exchange using TOPSIS. *Financial Research*, 10(26), 61-71. (in Persian)
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268.
- Saeedi, H. & Mohammadi, S. (2012). Prediction market return volatility using GARCH-ANN hybrid models. *Journal of Stock Exchange*, 4 (16), 153-174. (in Persian)
- Shi, Y. (2004). Particle swarm optimization. *IEEE Connections*, 2(1), 8-13.
- Smola, A. J. & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.
- Sun, B., Guo, H., Karimi, H. R., Ge, Y. & Xiong, S. (2015). Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing*, 151, 1528-1536.
- Tehrani, R., Namaki, A., Hedayatifar, L. (2013). Cross-correlation indexes Tehran Stock Exchange using the multifractal analysis of correlations without trend (MF-DXA). *Financial Research*, 14(1), 55-61. (in Persian)

- Tian, G. A. O. (2015). Research on Stock Price Prediction Based on Optimal Wavelet Packet Transformation and ARIMA-SVR Mixed Model. *Journal of Guizhou University of Finance and Economics*, 6(6), 57-69.
- Wang, J. Z., Wang, J. J., Zhang, Z. G., & Guo, S. P. (2011). Stock index forecasting based on a hybrid model. *Omega*, 40(6), 758-766.
- Wang, J. J., Wang, J. Z., Zhang, Z. G., & Guo, S. P. (2012). Stock index forecasting based on a hybrid model. *Omega*, 40(6), 758-766.
- Yegani, D. R., Chini Pardaz, R. (2011). Predicting the price index of Tehran Stock Exchange using a combination of Experts. *Quantitative Economics*, 8 (3), 53-73. (in Persian)

