

پیش‌بینی سری‌های زمانی قیمت فولاد با استفاده از معماری ترنسفورمر پیچ‌محور

فاطمه زارع مهرجردی^{۱*}، عباس نوروزی^۲

^۱ استادیار، دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه میبد، میبد، ایران

^۲ کارشناسی ارشد، دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه میبد، میبد، ایران

مقاله پژوهشی

چکیده

فولاد به عنوان یکی از مهم‌ترین مصالح در صنایع مختلف، نقشی حیاتی در اقتصاد جهانی ایفا می‌کند. اما نوسانات بالای قیمت، وابستگی‌های پیچیده و غیرخطی و عدم توانایی مدل‌های سستی در ثبت همبستگی‌های بلندمدت، پیش‌بینی دقیق آن را به یک چالش جدی تبدیل کرده است. این پژوهش برای غلبه بر این نارسایی‌ها، یک رویکرد جامع برای پیش‌بینی سری‌های زمانی قیمت فولاد با استفاده از مدل‌های یادگیری عمیق ارائه می‌دهد که شامل سه مرحله اصلی است: در مرحله اول، پیش‌پردازش داده‌ها از قبیل حذف مقادیر گم‌شده، نرمال‌سازی و ایجاد پنجره‌های زمانی ترتیبی انجام شده است؛ در مرحله دوم، مدل‌های یادگیری عمیق، شامل مدل‌های شبکه‌های عصبی بازگشتی، شبکه‌های ترنسفورمر (PatchTST) و شبکه‌های کانولوشنی آموزش داده‌شده‌اند و در نهایت، مرحله سوم ارزیابی عملکرد و مقایسه مدل‌ها با استفاده از معیارهای RMSE، MAE و R^2 صورت گرفته است. نتایج ارزیابی نشان می‌دهد که مدل PatchTST با بهره‌گیری از مکانیزم توجه و پردازش داده‌ها به صورت قطعات یا پیچ توانسته است وابستگی‌های پیچیده و بلندمدت را با دقت بالاتری نسبت به سایر مدل‌ها تشخیص دهد؛ به‌طور خاص، PatchTST در پیش‌بینی قیمت‌های فولاد توانست به دقت پیشرو R^2 با مقدار ۰/۹۸۱۵ و حداقل خطای RMSE با مقدار ۰/۰۳۰۴ دست یابد. این در حالی است که مدل‌های RNN به دلیل ماهیت ترتیبی و محدودیت‌های ساختاری خود، عملکرد ضعیف‌تری ارائه دادند. این یافته‌ها بر برتری مدل‌های مبتنی بر ترنسفورمر در پیش‌بینی دقیق سری‌های زمانی پیچیده تاکید می‌کند.

تاریخ دریافت:

۱۴۰۴/۰۵/۲۵

تاریخ پذیرش:

۱۴۰۴/۰۸/۱۳

کلیدواژه‌ها:

پیش‌بینی قیمت فولاد، یادگیری عمیق، سری زمانی، شبکه‌های عصبی بازگشتی، شبکه ترنسفورمر

نویسنده مسئول:

fzare@meybod.ac.ir

doi : 10.22034/ABMIR.2025.23536.1152

۱- مقدمه

همان‌طور که در پژوهش‌های اخیر اشاره شده، این مدل‌های سستی اغلب در ثبت الگوهای غیرخطی و پیچیده در سری‌های زمانی مالی با محدودیت‌هایی مواجه هستند.

در سال‌های اخیر با پیشرفت تکنیک‌های یادگیری عمیق، مدل‌های مبتنی بر شبکه‌های عصبی توانستند با یادگیری خودکار ویژگی‌ها و کشف الگوهای غیرخطی از داده‌ها در پیش‌بینی سری‌های زمانی مالی موفق ظاهر شوند [۴]. شبکه‌های عصبی بازگشتی^۵ به دلیل ساختار خود، توانایی ذاتی در پردازش داده‌های ترتیبی دارند. مدل‌های پیشرفته‌تر مانند شبکه‌های حافظه بلندمدت کوتاه^۶ (LSTM) و واحدهای بازگشتی دروازه‌ای^۶ (GRU)، با حل مشکل محو شدن گرادیان^۷ در RNN های ساده، توانستند وابستگی‌های بلندمدت در داده‌ها را به شکل بهتری مدل‌سازی کنند. به همین دلیل، LSTM و GRU به طور گسترده در پیش‌بینی قیمت کالاها و سهام مورد استفاده قرار گرفته‌اند [۵، ۶].

در سال‌های اخیر، معماری ترنسفورمر^۸ با معرفی مکانیزم توجه به خود^۹ انقلابی در پردازش توالی‌ها ایجاد کرد. این معماری به مدل اجازه می‌دهد تا به جای پردازش ترتیبی، به صورت موازی به تمام نقاط ورودی توجه کند و روابط پیچیده و غیرخطی بین آن‌ها را بدون از دست دادن اطلاعات، کشف کند. مدل PatchTST یک پیاده‌سازی نوین از ترنسفورمر است که به طور خاص برای سری‌های زمانی طراحی شده است و با تقسیم داده‌ها به قطعات^{۱۰}، توانایی ترنسفورمر را در یادگیری الگوهای محلی و بلندمدت به شکل چشم‌گیری بهبود می‌بخشد [۴، ۷].

با توجه به روش‌های مختلف پیش‌بینی سری زمانی مبتنی بر یادگیری عمیق، نیاز به یک تحلیل مقایسه‌ای جامع و کمی بین معماری‌های مختلف شبکه‌های عصبی بازگشتی، ترنسفورمرها و شبکه‌های کانولوشنی برای پیش‌بینی قیمت فولاد احساس می‌شود. تحلیل پژوهش‌های پیشین نشان می‌دهد که علی‌رغم اهمیت حیاتی

فولاد به عنوان یکی از مهم‌ترین و پرمصرف‌ترین مواد اولیه در جهان، نقشی حیاتی در صنایع گوناگون از جمله ساخت‌وساز، خودروسازی، زیرساخت‌ها و تولید ماشین‌آلات ایفا می‌کند. به دلیل اهمیت استراتژیک این محصول، نوسانات قیمت آن تأثیر مستقیمی بر اقتصاد جهانی و تصمیمات تجاری شرکت‌ها دارد. از این رو، پیش‌بینی دقیق و به موقع قیمت فولاد به منظور مدیریت ریسک، برنامه‌ریزی تولید و بهینه‌سازی زنجیره تأمین، از چالش‌های کلیدی در مدیریت صنعتی و مالی به شمار می‌رود. بر اساس مطالعات انجام شده، جهت پیش‌بینی سری‌های زمانی قیمت‌ها، دو رویکرد اصلی روش‌های سستی و روش‌های یادگیری عمیق وجود دارد.

در طول دهه‌های گذشته، روش‌های آماری و اقتصادسنجی متعددی برای پیش‌بینی قیمت‌ها در بازارهای مالی ارائه شده‌اند که به عنوان روش‌های سستی شناخته می‌شوند. مدل‌های ARIMA^۱ به عنوان یکی از اولین و پرکاربردترین رویکردهای سستی پیش‌بینی سری‌های زمانی است. این مدل ترکیبی از سه جز اصلی خودرگرسیون، تفاضل‌گیری و میانگین متحرک است. مدل ARIMA با استفاده از این سه جز، می‌تواند الگوهای خطی موجود در داده‌های سری زمانی را شناسایی کرده و برای پیش‌بینی مقادیر آینده استفاده کند. از آنجایی که قیمت فولاد تحت تأثیر عوامل پیچیده‌ای همچون تغییرات عرضه و تقاضا، شرایط کلان اقتصادی، سیاست‌های تجاری، و حتی تحولات ژئوپلیتیکی قرار دارد باعث شده پیش‌بینی آن با استفاده از روش‌های سستی آماری مانند ARIMA دشوار باشد [۱].

برای تحلیل دقیق‌تر نوسانات بازار، مدل‌های خانواده GARCH^۲ [۲] توسعه یافتند تا با در نظر گرفتن واریانس ناهمسان شرطی، پیش‌بینی ریسک را بهبود بخشند. هم‌چنین، رویکردهای چندمتغیره مانند VAR^۳ [۳] به محققان اجازه دادند تا روابط متقابل بین قیمت کالا و سایر متغیرهای کلان اقتصادی را بررسی کنند. با این حال،

^۵ Long Short-Term Memory (LSTM)

^۶ Gated Recurrent Unit (GRU)

^۷ Vanishing Gradient

^۸ Transformer

^۹ Self-Attention

^{۱۰} Patches

^۱ Autoregressive Integrated Moving Average (ARIMA)

^۲ Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

^۳ Vector Autoregression (VAR)

^۴ Recurrent Neural Network (RNN)

چهارم، نتایج حاصل از ارزیابی‌های کمی و تحلیل‌های مقایسه‌ای ارائه شده و برتری ساختاری مدل PatchTST مورد بحث قرار می‌گیرد. در نهایت، بخش پنجم به جمع‌بندی یافته‌های اصلی پژوهش و ارائه پیشنهادهایی برای تحقیقات آتی اختصاص دارد.

۲- مروری بر کارهای گذشته

پیش‌بینی دقیق قیمت کالاها، از جمله فولاد، سال‌هاست که موضوع پژوهش‌های فراوانی بوده است. با توجه به مطالعات انجام شده در این بخش مروری بر تعدادی مقالات انجام شده در پیش‌بینی سری‌های زمانی با روش‌های سنتی و یادگیری عمیق انجام شده است. لازم به ذکر است، از آنجایی که بازارهای کالایی مانند فولاد، نفت، طلا و رمزارزها همگی دارای ویژگی‌های دینامیکی مشترکی مانند نوسان‌پذیری^۲ بالا، وابستگی‌های غیرخطی پیچیده و ماهیت پرهیز و مرج هستند، متدولوژی‌های موفق در پیش‌بینی قیمت یک کالا معمولاً قابل‌تعمیم به کالاهای دیگر نیز می‌باشند. بنابراین، بررسی مطالعات انجام شده در حوزه‌های مشابه مانند طلا، نفت خام و بازارهای سهام، برای شناسایی کارآمدترین مدل‌سازی برای چنین دینامیک‌های پیچیده‌ای، ضروری و قابل قبول است.

۱-۲ پیش‌بینی سری‌های زمانی با روش‌های سنتی

گوها و همکارانش [۸] از رویکرد سنتی ARIMA برای پیش‌بینی قیمت طلا در کشور هند استفاده کرده‌اند. مطالعه آن‌ها بر اساس داده‌های ماهانه از نوامبر ۲۰۰۳ تا ژانویه ۲۰۱۴ بوده و پس از مقایسه چندین ترکیب مختلف پارامترها برای یافتن بهترین ساختار، به مدلی دست یافته‌اند که بهترین عملکرد را برای پیش‌بینی قیمت‌های آتی دارد. هرچند که این مطالعه دقت بالای مدل را با استفاده از معیارهای آماری تأیید می‌کند اما برای پیش‌بینی‌های کوتاه‌مدت مناسب هستند و در مواجهه با تغییرات ناگهانی و بزرگ (مانند تغییرات اقتصادی و سیاسی) که خطی نیستند، عملکرد ضعیفی از خود نشان می‌دهند.

در پژوهشی دیگر گریش [۹] بر پیش‌بینی قیمت‌های نقطه‌ای برق در بازار برق هند تمرکز کرده است. این مطالعه با استفاده از داده‌های ساعتی، عملکرد مدل‌های خودرگرسیون GARCH را

پیش‌بینی قیمت فولاد، سه نارسایی کلیدی در ادبیات وجود دارد که توجیهی برای پژوهش حاضر محسوب می‌شود: نخست، کمبود یک مقایسه جامع و نظام‌مند بین نسل‌های مختلف مدل‌های یادگیری عمیق نوین (از جمله مدل‌های بازگشتی، کانولوشنی و ترنسفورمر) در حوزه قیمت فولاد مشاهده می‌شود؛ تحقیقات پیشین اغلب بر روی یک یا دو مدل خاص متمرکز بوده‌اند. دوم، مدل‌های بازگشتی سنتی (RNN/LSTM)، که رایج‌ترین مدل‌های یادگیری عمیق در این حوزه هستند، در مواجهه با وابستگی‌های طولانی، غیرخطی و نوسانات شدید قیمت فولاد، دچار افت شدید دقت و پدیده محوشدگی گرادیان می‌شوند. سوم، علیرغم موفقیت‌های چشم‌گیر معماری‌های پیشرفته ترنسفورمر مانند PatchTST در سری‌های زمانی عمومی، کاربرد و ارزیابی عمیق این مدل‌های جدید در داده‌های تخصصی و پیچیده قیمت فولاد به‌طور قابل‌توجهی کم است. بنابراین، این پژوهش با هدف پر کردن این شکاف‌ها، یک ارزیابی مقایسه‌ای دقیق بین معماری‌های RNN، PatchTST و TCN را بر روی داده‌های واقعی قیمت فولاد انجام می‌دهد تا برترین معماری برای پیش‌بینی‌های استراتژیک این حوزه را مشخص سازد. بنابراین هدف از این پژوهش، فراتر از یک مقایسه ساده، ارزیابی دقیق عملکرد مدل‌های پیشرفته، به خصوص مدل PatchTST، در مسئله چالش‌برانگیز پیش‌بینی قیمت فولاد است. ما با استفاده از یک مجموعه داده واحد، نه تنها کارآمدترین رویکرد را شناسایی می‌کنیم، بلکه در پی ارائه یک بینش ساختاری هستیم تا مشخص شود که چگونه مزیت‌های معماری ترنسفورمر، مانند مکانیزم پیچ و توجه^۱، آن را قادر می‌سازد تا وابستگی‌های پیچیده و بلندمدت قیمت فولاد را با دقت بالاتری نسبت به مدل‌های بازگشتی سنتی مدل‌سازی کند و یک معیار عملکرد جدید در این حوزه تعیین نماید.

ساختار ادامه این مقاله به شرح زیر است. در بخش دوم، مروری بر کارهای پیشین انجام شده در حوزه پیش‌بینی سری زمانی قیمت‌ها با تمرکز بر مدل‌های یادگیری عمیق ارائه می‌گردد. بخش سوم به شرح جزئیات روش پیشنهادی می‌پردازد و مجموعه داده، مدل‌های انتخابی و تنظیمات آموزش را معرفی می‌کند. در بخش

² Volatility

¹ Patching and Attention

پیش‌بینی متغیرهایی با هم‌بستگی بالا عملکرد بهتری دارد. با این حال، برای متغیرهایی که هم‌بستگی کمتری با یکدیگر دارند، هر دو مدل VAR و ARIMA عملکرد تقریباً مشابهی داشته‌اند. این یافته‌ها نشان داد که قبل از انتخاب مدل پیش‌بینی، بررسی میزان هم‌بستگی بین متغیرها از اهمیت بالایی برخوردار است، چرا که مدل‌های چندمتغیره مانند VAR برای داده‌های با هم‌بستگی بالا کارایی بیشتری دارند.

۲-۲ پیش‌بینی سری‌های زمانی با روش‌های یادگیری

عمیق

پژوهش‌های اخیر نشان داده‌اند که رویکردهای یادگیری ماشین و یادگیری عمیق در مقایسه با مدل‌های سنتی، کارایی بالاتری در پیش‌بینی قیمت‌های غیرخطی و نوسانی دارند. به عنوان مثال، در مقاله [۱۳]، ویجه و همکارانش از دو تکنیک شبکه عصبی مصنوعی و جنگل تصادفی برای پیش‌بینی قیمت‌های پایانی سهام استفاده کردند. این تحقیق با استفاده از داده‌های مالی پنج شرکت، به این نتیجه رسید که مدل شبکه عصبی عملکرد پیش‌بینی بهتری را ارائه می‌دهد. در مقاله‌ای دیگر کوالکار و همکارانش [۱۴] از الگوریتم رگرسیون درخت تصمیم برای پیش‌بینی قیمت مسکن در شهر بمبئی استفاده کردند و توانستند به دقت ۸۹ درصدی دست یابند. این مطالعات تأیید می‌کنند که روش‌های یادگیری ماشین، به ویژه برای داده‌های غیرخطی و پیچیده مالی، می‌توانند ابزارهای قدرتمندی برای پیش‌بینی قیمت‌ها باشند. با در نظر گرفتن این پیش‌زمینه، مطالعات یادگیری عمیق در این حوزه را می‌توان به سه دسته‌بندی تحلیلی اصلی مدل‌های بازگشتی، مدل‌های ترنسفورمر و روش‌های ترکیبی تقسیم کرد. در ادامه برای این دسته‌بندی نمونه‌هایی آورده شده است.

۲-۲-۱ پژوهش‌های مبتنی بر مدل‌های بازگشتی

(RNN/LSTM/GRU)

در مقاله [۱۵] شاهی و همکارانش، عملکرد مدل‌های یادگیری عمیق بازگشتی LSTM و GRU برای پیش‌بینی قیمت سهام، تحت شرایط یکسان و کنترل‌شده را مورد بررسی و مقایسه قرار دادند. آن‌ها نشان دادند که هر دو مدل LSTM و GRU، به تنهایی و با استفاده از ویژگی‌های تاریخی سهام، در پیش‌بینی قیمت‌ها

برای این منظور ارزیابی کرده است. با این حال، این مدل‌ها در ثبت کامل پیچیدگی‌ها و الگوهای غیرخطی بازار که ممکن است تحت تأثیر عوامل مختلف قرار گیرد، محدودیت‌هایی دارند و اغلب برای تحلیل واریانس نسبت به پیش‌بینی دقیق قیمت‌ها در بازارهای بسیار پویا مناسب‌تر هستند.

مرابت و همکارانش [۱۰] یک روش ترکیبی جهت بررسی رفتار سری زمانی قیمت نفت، با هدف پیش‌بینی نوسانات آن استفاده کردند. آن‌ها نشان دادند که قیمت نفت، به‌ویژه در بازار مالی، دارای نوسانات بالایی است. آن‌ها با استفاده از داده‌های روزانه از سال ۲۰۱۰ تا ۲۰۱۹، ابتدا سری زمانی را به یک مدل ARIMA تبدیل کرده‌اند. سپس برای مدل‌سازی نوسانات واریانس، از مدل EGARCH استفاده کردند. پس از بررسی و تنظیم پارامترهای مختلف دریافتند که مدل ترکیبی ARIMA و EGARCH بهترین عملکرد را در پیش‌بینی قیمت‌های آتی نفت دارد. اگرچه این مطالعه دقت بالای مدل ترکیبی را تأیید می‌کند، اما اشاره به این نکته که تحقیقات کمی در مورد مدل‌های متقارن و نامتقارن برای نوسانات قیمت نفت وجود دارد، نشان‌دهنده نیاز به بررسی بیشتر در این حوزه است.

رهنمون و همکارانش [۱۱] از یک مدل خودرگرسیون برداری (VAR) برای پیش‌بینی قیمت‌های ماهانه فولاد در ایران استفاده کردند. آن‌ها در مطالعه خود، ضمن شناسایی متغیرهای مؤثر بر قیمت فولاد بر روی داده‌های دوره ۱۳۹۸ تا ۱۴۰۲، نشان دادند که شوک‌های وارد شده از متغیرهای نرخ ارز غیررسمی و شاخص قیمت تولیدکننده صنعت، بیشترین تأثیر را بر نوسانات قیمت فولاد داشته است. آن‌ها با کسب دقت بالای مدل، بر مزیت مدل‌های VAR به دلیل امکان بررسی تأثیر هم‌زمان چند متغیر، نسبت به مدل‌های تک‌متغیره مانند شبکه‌های عصبی یا الگوریتم ژنتیک، تأکید کردند.

در پژوهشی خان و همکارانش [۱۲] عملکرد مدل‌های VAR و ARIMA در پیش‌بینی متغیرهای اقتصادی بنگلادش را مورد مقایسه قرار دادند. هدف اصلی آن‌ها، پیش‌بینی هم‌زمان یک گروه از متغیرها و بهره‌گیری از هم‌بستگی بین آن‌ها بوده است. نتایج این مطالعه نشان داد که مدل VAR در مقایسه با مدل ARIMA، در

همین موفقیت باعث شده تا مدل‌های مبتنی بر ترنسفورمر و انواع مختلف آن به حوزه سری‌های زمانی نیز راه یابند. زو و همکارانش [۱۸] با معرفی مدل Informer کارایی این معماری را در پیش‌بینی سری‌های زمانی بلندمدت به اثبات رساندند. در همین راستا، مدل PatchTST [۱۹] به عنوان یک پیاده‌سازی پیشرفته و بهینه از ترنسفورمر برای سری‌های زمانی، با تقسیم داده‌ها به قطعات، توانایی چشم‌گیر مدل در یادگیری الگوهای محلی و بلندمدت را نشان می‌دهد.

۲-۲-۳ پژوهش‌های مبتنی بر روش‌های ترکیبی (مدل‌های کانولوشنی، بازگشتی و ترنسفورمر)

با پیشرفت فناوری، مدل‌های ترکیبی^۲ و مکانیزم‌های جدیدی برای بهبود عملکرد مدل‌های یادگیری عمیق در پیش‌بینی سری‌های زمانی معرفی شدند. به عنوان مثال، پاتنایک و همکارانش [۲۰] با استفاده از یک مدل ترکیبی شبکه‌های عصبی کانولوشنال، LSTM و الگوریتم تکاملی گرگ خاکستری، به پیش‌بینی قیمت سهام پرداختند. آن‌ها نشان دادند که ترکیب این معماری‌ها می‌تواند ویژگی‌های مهم‌تری را از داده‌ها استخراج کرده و دقت پیش‌بینی را افزایش دهد. در پژوهشی دیگر، جیالینو همکارانش [۲۱] از ترکیب یک مدل CNN-LSTM با مکانیزم توجه^۳ برای پیش‌بینی قیمت سهام استفاده کردند و نشان دادند که مکانیزم توجه می‌تواند به مدل در شناسایی نقاط کلیدی داده‌ها کمک کند و عملکرد آن را بهبود بخشد. هم‌چنین، ونگ [۲۲] یک روش جدید با ترکیب Bi-LSTM و ترنسفورمر برای بهبود دقت و پایداری پیش‌بینی قیمت سهام معرفی کردند. این مدل ترکیبی از Bi-LSTM (برای جذب اطلاعات دوطرفه در توالی‌ها) و یک نسخه بهبودیافته از ترنسفورمر تشکیل شده است. نتایج این مطالعه نشان می‌دهد که این روش جدید، عملکرد بهتری نسبت به روش‌های موجود دارد و با کاهش خطا و افزایش دقت، قابلیت تعمیم بالایی در پیش‌بینی قیمت سهام دارد. این مطالعات تأیید می‌کنند که ادغام معماری‌های مختلف یادگیری عمیق، رویکردی مؤثر برای افزایش دقت در پیش‌بینی سری‌های زمانی پیچیده است.

عملکردی وابسته به شرایط دارند. با این حال، هنگامی که داده‌های مربوط به احساسات^۱ اخبار مالی به ویژگی‌های ورودی مدل‌ها اضافه می‌شود، عملکرد پیش‌بینی قیمت سهام به طور قابل توجهی بهبود می‌یابد. نتایج این تحقیق نشان داد که هر دو مدل ترکیبی LSTM و GRU با داده‌های احساسات توانسته‌اند عملکرد یکسانی را در پیش‌بینی بهتر قیمت سهام ارائه دهند. آن‌ها بر اهمیت گنجاندن داده‌های احساسات در کنار داده‌های تاریخی برای افزایش دقت پیش‌بینی در بازارهای مالی غیرخطی و نوسانی تأکید کردند. گونارتو و همکارانش [۱۶] به مقایسه عملکرد دو مدل یادگیری عمیق RNN و LSTM برای پیش‌بینی قیمت رمزارزها پرداخته‌اند. آن‌ها داده‌های قیمت روزانه بیت‌کوین و اتریوم را از سال ۲۰۱۵ تا ۲۰۲۱ جمع‌آوری و برای پیش‌بینی قیمت‌های آینده استفاده کرده‌اند. نتایج تحقیق نشان می‌دهد که مدل LSTM عملکرد بهتری نسبت به مدل RNN داشته است. این یافته نشان می‌دهد که مدل LSTM ابزار مؤثرتری برای پیش‌بینی قیمت رمزارزها است و می‌تواند به سرمایه‌گذاران در اتخاذ استراتژی‌های مناسب و مدیریت ریسک کمک کند.

۲-۲-۲ پژوهش‌های مبتنی بر مدل‌های ترنسفورمر

مدل ترنسفورمر برای اولین بار برای پیش‌بینی قیمت سهام در بورس داکا بنگلادش به کار گرفته شد [۱۷]. این مدل با استفاده از یک مکانیزم پیشرفته برای نمایش ویژگی‌های سری زمانی، قادر به تحلیل مؤثر داده‌ها است. محققان این مدل را بر روی داده‌های روزانه و هفتگی هشت شرکت منتخب آزمایش کردند و نتایج، عملکرد قابل قبولی را در اکثر سهام‌ها نشان داد. این مطالعه بر شکاف تحقیقاتی موجود در استفاده از مدل‌های مبتنی بر توجه برای پیش‌بینی قیمت سهام در بازار بورس بنگلادش تأکید کرده و با ارائه نتایج امیدوارکننده، پتانسیل بالای این مدل‌ها را در تحلیل و پیش‌بینی سری‌های زمانی پیچیده اثبات می‌کند.

نقطه قوت معماری ترنسفورمر و مکانیزم توجه به خود، امکان پردازش موازی و درک وابستگی‌های بلندمدت بدون محدودیت‌های ذاتی شبکه‌های عصبی بازگشتی RNN است.

³ Attention

¹ Sentiment

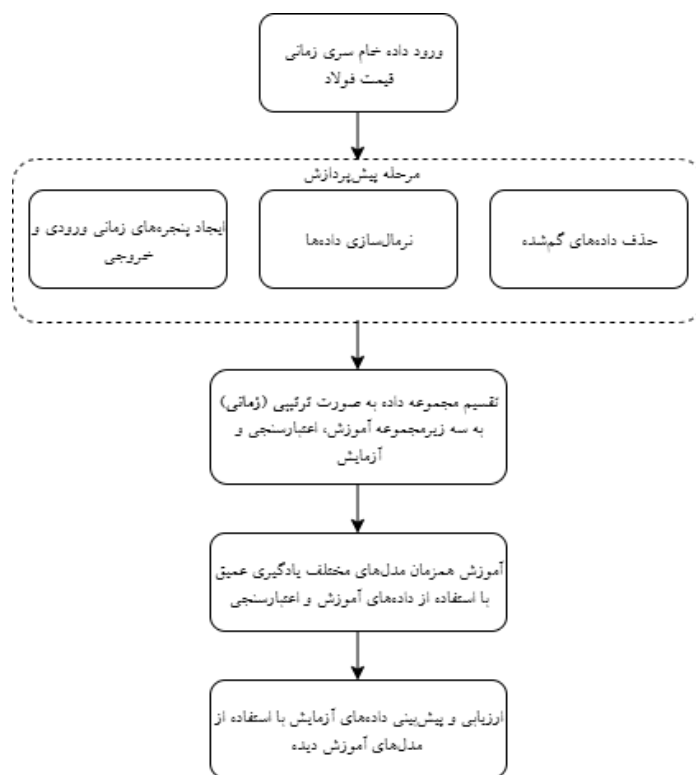
² Hybrid

بر روی یک سری زمانی تکی است، این پژوهش به صورت تک‌متغیره انجام شده تا مقایسه‌ای دقیق از کارایی ذاتی هر مدل ارائه گردد و کارآمدترین رویکرد برای پیش‌بینی قیمت فولاد شناسایی شود.

۳- روش پیشنهادی

در این بخش، رویکرد پژوهش برای پیش‌بینی سری‌های زمانی قیمت فولاد که شامل مراحل اصلی آماده‌سازی و پیش‌پردازش مجموعه داده، معرفی معماری مدل‌های یادگیری عمیق مورد استفاده و تشریح معیارهای ارزیابی و تنظیمات پارامترهای آموزش است، با جزئیات کامل شرح داده شده است. خلاصه روش پیشنهادی در شکل (۱) آورده شده است.

با توجه به بررسی مطالعات انجام شده در پیش‌بینی قیمت فولاد، یک خلأ پژوهشی بزرگی مانند فقدان یک مقایسه جامع بین معماری‌های نوین یادگیری عمیق (به ویژه مدل‌های ترنسفورمر) در حوزه تخصصی قیمت فولاد به چشم می‌خورد. لذا، این پژوهش مستقیماً با هدف پر کردن این خلأ، کارآمدترین معماری را برای مدل‌سازی دینامیک‌های پیچیده قیمت فولاد شناسایی می‌کند. هم‌چنین با توجه به محدودیت‌های روش‌های سنتی در تحلیل الگوهای غیرخطی و پیچیده سری‌های زمانی و با الهام از موفقیت چشم‌گیر مدل‌های یادگیری عمیق در این زمینه، در پژوهش حاضر تلاش شده است تا با استفاده از مدل‌های یادگیری عمیق مانند RNN، LSTM، GRU، TCN، Bi-LSTM، RCNN، PatchTST و Transformer به پیش‌بینی قیمت فولاد پرداخته شود. از آنجا که هدف این تحقیق ارزیابی عملکرد این معماری‌ها



شکل (۱): خلاصه روش پیشنهادی

۱-۳ آماده‌سازی و پیش‌پردازش مجموعه داده

مجموعه داده مورد استفاده در این پژوهش با عنوان "TATASTEEL.NS_stock_data.csv" از پلتفرم معتبر Kaggle تهیه شده است. این داده‌ها شامل اطلاعات تاریخی عملکرد سهام شرکت چندملیتی هندی در صنعت فولاد است. این مجموعه داده به تحلیلگران و پژوهشگران امکان می‌دهد تا روندهای گذشته و نوسانات قیمت سهام را در یک بازه زمانی مشخص بررسی کنند. این اطلاعات می‌تواند برای تحلیل سرمایه‌گذاری، تحقیقات مالی، و پیش‌بینی روندهای بازار به کار گرفته شود. مجموعه داده مذکور، اطلاعات روزانه قیمت سهام را در بازه زمانی ۳ جولای ۲۰۱۵ تا ۲ جولای ۲۰۲۱ ارائه می‌دهد. جدول (۱) ستون‌های کلیدی موجود در مجموعه داده را نشان می‌دهد.

جدول (۱): ستون‌های موجود در مجموعه داده مورد استفاده

مفهوم ستون	نام ستون
تاریخ روزانه	Date
قیمت باز شدن فولاد	Open
بالترین قیمت در طول روز	High
پایین‌ترین قیمت در طول روز	Low
قیمت بسته شدن فولاد	*Close
حجم معادلات روزانه	Volume

با توجه به جدول (۱) قیمت بسته شدن فولاد به عنوان متغیر اصلی و هدف پیش‌بینی در این پژوهش مورد استفاده قرار گرفته است. مجموعه داده معرفی شده به عنوان یک سری زمانی مالی پرنوسان مرتبط با فولاد شناخته شده است که دلیل اصلی جهت انتخاب مجموعه داده اصلی برای این پژوهش است. لازم به ذکر است پژوهش جاری جهت تعمیم‌پذیری بر روی دو مجموعه داده مستقل دیگر نیز اجرا و ارزیابی شده است، که شامل قیمت آتی کالای جهانی فولاد و یک شاخص کلان ماهانه قیمت فولاد هستند. پیش‌پردازش داده‌ها به منظور آماده‌سازی آن‌ها برای مدل‌های یادگیری عمیق، در سه مرحله انجام شده است:

- حذف داده‌های گم‌شده: ابتدا، تمام ردیف‌هایی که شامل مقادیر گم‌شده هستند، از مجموعه داده حذف شدند تا از یکپارچگی داده‌ها اطمینان حاصل شود.
- نرمال‌سازی داده‌ها: برای کاهش حساسیت مدل‌ها نسبت به مقیاس داده‌ها و بهبود همگرایی در طول آموزش، از روش نرمال‌سازی MinMaxScaler استفاده شد. این روش مقادیر قیمت را به بازه [۰،۱] نگاشت می‌کند. نحوه نرمال‌سازی در رابطه (۱) آورده شده است:

$$x'_{\text{MinMax}} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \quad (1)$$

- ایجاد پنجره‌های زمانی: در این مرحله داده‌های سری زمانی به پنجره‌های ورودی و خروجی تبدیل شدند. یک پنجره با طول seq_len برابر با ۳۰ (قیمت ۳۰ روز گذشته) به عنوان ورودی و یک پنجره با طول pred_len برابر با ۷ (قیمت ۷ روز آینده) به عنوان خروجی تعریف شده است. لازم به ذکر است انتخاب این اعداد با استفاده از مفهوم اعتبارسنجی متقابل^۲ انجام شده است. درنهایت، این داده‌ها به سه مجموعه آموزش (۷۰٪)، اعتبارسنجی (۱۵٪) و آزمون (۱۵٪) تقسیم شده‌اند.

۲-۳ معرفی معماری مدل‌های یادگیری عمیق مورد

استفاده

انتخاب مدل‌های یادگیری عمیق در این پژوهش بر اساس توانایی ساختاری و عملکردی آن‌ها در مواجهه با چالش‌های اساسی سری‌های زمانی مالی پرنوسان قیمت فولاد (یا سهام مرتبط با آن) صورت گرفته است. این چالش‌ها عمدتاً شامل نویز بالا، الگوهای غیرخطی و حضور وابستگی‌های زمانی با طول‌های متفاوت است. برای پوشش‌دهی این طیف از ویژگی‌ها، سه خانواده اصلی مدل‌ها برای مقایسه عملکرد انتخاب شدند: مدل‌های بازگشتی برای ارزیابی و مدل‌سازی وابستگی‌های کوتاه‌مدت و میان‌مدت و تأثیر توالی مستقیم داده‌ها؛ مدل‌های کانولوشنی به دلیل توانایی در

¹<https://www.kaggle.com/datasets/nitirajkulkarni/tata-steel-ns-stock-performance>

² Cross Validation

مدت‌های طولانی است. LSTM با استفاده از سه دروازه اصلی، جریان اطلاعات را به دقت کنترل می‌کند: دروازه فراموشی^۱، دروازه ورودی^۲ و دروازه خروجی^۳، این مکانیزم به LSTM اجازه می‌دهد تا اطلاعات مهم را برای مدت‌های طولانی به خاطر بسپارد و وابستگی‌های پیچیده را به خوبی مدل‌سازی کند.

ج. GRU: این مدل که توسط Cho معرفی شد، نسخه‌ای ساده‌تر و کارآمدتر از LSTM است [۶]. این مدل با ترکیب دو دروازه (دروازه ورودی و فراموشی) به یک دروازه به‌روزرسانی^۴، پیچیدگی محاسباتی کمتری نسبت به LSTM دارد. GRU با داشتن تنها دو دروازه (به‌روزرسانی و ریست^۵) سریع‌تر از LSTM آموزش می‌بیند و پارامترهای کمتری دارد، اما در اکثر موارد، عملکردی مشابه با LSTM ارائه می‌دهد. این ویژگی، GRU را به یک انتخاب محبوب در مواقعی که سرعت و کارایی محاسباتی اهمیت دارد، تبدیل کرده است.

د. LSTM دو طرفه^۶: این مدل یک تغییر ساختاری از LSTM است که توالی داده‌ها را در هر دو جهت (به جلو و عقب) پردازش می‌کند [۲۰]. این مدل از دو لایه LSTM استفاده می‌کند: یکی که توالی را از ابتدا تا انتها می‌خواند و دیگری که آن را از انتها به ابتدا می‌خواند. Bi-LSTM با ترکیب خروجی‌های این دو لایه، به یک درک جامع و کامل از هر گام زمانی دست پیدا می‌کند، زیرا اطلاعات را هم از گذشته و هم از آینده توالی (بخش‌های بعدی) دریافت می‌کند. این ویژگی می‌تواند در مسائلی که وابستگی‌های دوطرفه در داده‌ها وجود دارد، دقت پیش‌بینی را به شکل چشم‌گیری بهبود بخشد.

مدل‌های ترنسفورمر:

ترنسفورمرها، که در ابتدا توسط Vaswani برای پردازش زبان طبیعی توسعه یافتند، با استفاده از مکانیزم توجه به خود، انقلابی در پردازش داده‌های ترتیبی ایجاد کردند. برخلاف RNN که داده‌ها را به صورت متوالی پردازش می‌کند، ترنسفورمر می‌تواند همه نقاط داده را به صورت موازی تحلیل کرده و وابستگی بین آن‌ها را فارغ از فاصله‌شان درک کند. این معماری به دلیل کارایی بالا و توانایی

استخراج ویژگی‌های محلی موضعی (نظیر نوسانات کوچک) و استفاده از Dilated Convolutions برای گسترش مؤثر میدان دید در توالی‌های زمانی؛ و در نهایت، مدل‌های مبتنی بر ترنسفورمر که هدف اصلی پژوهش را تشکیل می‌دهند، برای بررسی کارایی مکانیزم توجه به خود در کشف و مدل‌سازی همبستگی‌های بسیار بلندمدت و غیرخطی داده‌های قیمت فولاد انتخاب گردیدند. به‌طور ویژه، PatchTST با رویکرد پیچ‌سازی، باعث مدل‌سازی مؤثرتر الگوهای محلی و بلندمدت در سری‌های زمانی با طول توالی زیاد می‌شود. در این بخش، به بررسی دقیق‌تر معماری مدل‌های یادگیری عمیقی پرداخته شده است که برای تحلیل و پیش‌بینی سری‌های زمانی قیمت فولاد مورد استفاده قرار گرفته‌اند. این مدل‌ها به دلیل توانایی در یادگیری الگوهای پیچیده و وابستگی‌های زمانی، از مدل‌های سنتی عملکرد بهتری دارند.

مدل‌های شبکه عصبی بازگشتی:

شبکه‌های عصبی بازگشتی (RNN) دسته‌ای از شبکه‌های عصبی هستند که به دلیل ماهیت ترتیبی خود، برای تحلیل داده‌های سری زمانی مناسب هستند. این مدل‌ها با استفاده از یک حلقه بازگشتی، اطلاعات را از گام‌های زمانی قبلی به گام‌های بعدی منتقل می‌کنند. الف. RNN ساده: این مدل به عنوان پایه‌ای‌ترین مدل در این دسته، اطلاعات را از گام زمانی قبلی به گام زمانی فعلی منتقل می‌کند. این مدل یک حلقه بازگشتی دارد که به آن اجازه می‌دهد تا اطلاعات را در حافظه خود نگه دارد. RNN در مدل‌سازی وابستگی‌های کوتاه مدت عملکرد خوبی دارد. با این حال، مهم‌ترین ضعف آن، پدیده محو شدن گرادیان است. این پدیده باعث می‌شود که مدل نتواند وابستگی‌های بلندمدت را به خوبی یاد بگیرد، زیرا تأثیر اطلاعات دوردست در طول زمان از بین می‌رود [۵]. به همین دلیل، در مسائل پیچیده سری زمانی، از مدل‌های پیشرفته‌تر استفاده می‌شود.

ب. LSTM: این مدل که توسط Hochreiter و Schmidhuber معرفی شد، یک راه‌حل کارآمد برای مشکل محو شدن گرادیان در RNN است [۶]. این مدل از یک ساختار پیچیده‌تر به نام سلول حافظه استفاده می‌کند که قادر به ذخیره و بازیابی اطلاعات برای

⁴ Update Gate

⁵ Reset Gate

⁶ Bi-LSTM

¹ Forget Gate

² Input Gate

³ Output Gate

در این روش، به جای اینکه مدل هر نقطه زمانی را به صورت جداگانه پردازش کند، سری زمانی به قطعات کوچکی با طول ثابت تقسیم می‌شود. این پچ‌ها سپس به صورت موازی به مدل ترنسفورمر داده می‌شوند. این کار سه مزیت کلیدی دارد:

- ۱- استخراج ویژگی‌های محلی: هر پچ به مدل اجازه می‌دهد تا الگوهای محلی (مانند تغییرات کوچک یا روندهای کوتاه) را به صورت مؤثرتر استخراج کند، مشابه کاری که شبکه‌های عصبی کانولوشنال در پردازش تصویر انجام می‌دهند.
- ۲- کاهش پیچیدگی محاسباتی: با کاهش طول توالی از تعداد کل نقاط داده به تعداد پچ‌ها، پیچیدگی محاسباتی مکانیزم توجه به خود ترنسفورمر به شکل چشم‌گیری کاهش می‌یابد و سرعت آموزش و پیش‌بینی افزایش پیدا می‌کند.
- ۳- مدل‌سازی وابستگی‌های بلندمدت: مکانیزم توجه به خود به مدل اجازه می‌دهد تا وزن هر پچ را نسبت به پچ‌های دیگر در کل توالی مشخص کند و وابستگی‌های بلندمدت را بدون از دست دادن اطلاعات، کشف نماید.

مدل PatchTST هم‌چنین از یک لایه RevIN برای نرمال‌سازی درونی استفاده می‌کند که پایداری آموزش را بهبود می‌بخشد و حساسیت مدل به تغییرات مقیاس داده‌ها را کاهش می‌دهد. این مدل با ترکیب قدرت ترنسفورمر و رویکرد نوآورانه تقسیم‌بندی به قطعات، عملکردی برتر در پیش‌بینی دقیق سری‌های زمانی، به‌ویژه در مقایسه با مدل‌های قدیمی‌تر، ارائه می‌دهد [۲۰-۲۲].

مدل‌های مبتنی بر شبکه‌های کانولوشنی

شبکه‌های عصبی کانولوشنال برای سری‌های زمانی TCN^۶: این مدل یک معماری است که از شبکه‌های عصبی کانولوشنال برای پردازش داده‌های سری زمانی استفاده می‌کند [۲۳]. ویژگی کلیدی TCN، استفاده از پیچش‌های سببی متسع‌شده^۹ است که به آن اجازه می‌دهد تا میدان دید خود را برای دربرگرفتن وابستگی‌های بلندمدت گسترش دهد. این معماری می‌تواند در شرایطی که الگوهای محلی در طول زمان تکرار می‌شوند، مفید باشد. با این

در مدل‌سازی وابستگی‌های بلندمدت، به سرعت به حوزه سری‌های زمانی راه یافت.

مدل‌های ترنسفورمر از دو بخش اصلی رمزگذار^۱ و رمزگشا^۲ تشکیل شده‌اند. بلوک رمزگذار مسئول پردازش ورودی‌ها و تولید بازنمایی‌های داخلی است. این بلوک شامل لایه‌های توجه چندسر^۳ و شبکه‌های پیش‌خور موضعی است. بلوک رمزگشا مسئول تولید خروجی‌ها بر اساس بازنمایی‌های داخلی تولیدشده توسط رمزگذار است [۱۷].

مفهوم توجه چندسر، خود از تعدادی لایه توجه به خود تشکیل شده است. هر یک از این لایه‌ها میزان توجه و ارزش‌گذاری بر روی قسمت‌های مختلف ورودی را با استفاده از سه ماتریس پرسش^۴، کلید^۵ و ارزش^۶ متفاوت انجام داده و یک بازنمایی منحصر به فردی از ورودی را ایجاد می‌کنند. بازنمایی‌های مختلف به دست آمده از لایه‌های خود توجه با هم ترکیب شده تا بازنمایی قوی‌تر از ویژگی‌ها به دست آید. در ادامه با اضافه کردن لایه‌های پیش‌خور، بازنمایی‌های تولیدشده به خروجی مناسب نگاشت می‌شوند. نحوه عملکرد مفهوم توجه به خود در رابطه (۲) آورده شده است:

$$\begin{aligned} \text{head}_i &= \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \\ &= \text{Softmax} \left[\frac{QW_i^Q (KW_i^K)^T}{\sqrt{d_k}} \right] VW_i^V \quad (2) \\ \text{MultiHead}(Q, K, V) &= \\ &= \text{concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W_O \end{aligned}$$

این فرمول نشان می‌دهد که چگونه وابستگی بین ورودی‌ها محاسبه می‌شود. پس از آن، لایه توجه چندسر، چندین بار این عملیات را به صورت موازی انجام می‌دهد و خروجی‌ها را به هم می‌چسباند تا یک نمایش کامل‌تر از ورودی به دست آید.

مدل PatchTST^۷: این مدل یک رویکرد پیشرفته و قدرتمند در حوزه پیش‌بینی سری‌های زمانی است که معماری انقلابی ترنسفورمر را برای این نوع داده‌ها بهینه کرده است. نقطه قوت و نوآوری اصلی این مدل، استفاده از رویکرد قطعات یا پچینگ است.

⁶ Value

⁷ Patch-wise Time Series Transformer (PatchTST)

⁸ Temporal Convolutional Networks (TCN)

⁹ Dilated Causal Convolutions

¹ Encoder

² Decoder

³ Multi-Head Attention

⁴ Query

⁵ Key

جدول (۲): پارامترهای مدل‌های یادگیری عمیق استفاده‌شده

پارامتر	مقدار
بهینه‌ساز	Adam
نرخ یادگیری	۰/۰۰۰۱
تابع هزینه	MSE
تعداد تکرارهای آموزش	۱۵
اندازه دسته	۳۲

معیارهای ارزیابی:

عملکرد مدل‌ها در مجموعه داده آزمون با سه معیار اصلی ارزیابی شده است. این معیارها کمک می‌کنند تا دقت و توانایی مدل در پیش‌بینی قیمت‌های آینده از زوایای مختلف سنجیده شوند.

خطای میانگین مربعات ریشه^۲ (RMSE)

این معیار، میانگین اختلاف بین مقادیر واقعی و مقادیر پیش‌بینی شده را محاسبه می‌کند. از آنجا که خطاهای بزرگ‌تر را به صورت درجه دوم جریمه می‌کند، نسبت به خطاهای پرت^۳ حساسیت بیشتری دارد. هرچه مقدار RMSE کمتر باشد، عملکرد مدل بهتر است. رابطه (۳) نحوه محاسبه این معیار را نمایش داده است.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

خطای میانگین مطلق (MAE)

این معیار، میانگین قدر مطلق اختلاف بین مقادیر واقعی و پیش‌بینی شده را نشان می‌دهد. تفسیر MAE ساده‌تر از RMSE است، زیرا هر خطایی را به صورت خطی محاسبه می‌کند MAE. نشان‌دهنده میانگین اندازه خطای پیش‌بینی بدون در نظر گرفتن جهت آن است. رابطه (۴) نحوه محاسبه این معیار را نمایش داده است.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4)$$

حال، در مسائل پیچیده‌ای که وابستگی‌های غیرخطی و دوردست غالب هستند، ممکن است به اندازه مدل‌های مبتنی بر توجه مانند ترنسفورمرها عملکرد قوی‌ای ارائه ندهد.

شبکه‌های عصبی بازگشتی-کانولوشنال^۱ CRNN:

CRNN یک مدل ترکیبی است که از ترکیب دو معماری، شبکه‌های عصبی کانولوشنال و شبکه‌های عصبی بازگشتی ایجاد شده است [۲۴]. در این معماری، لایه‌های CNN در ابتدا قرار می‌گیرند تا به صورت خودکار ویژگی‌های محلی را از داده‌های ورودی استخراج کنند. پس از این مرحله، خروجی لایه‌های CNN به عنوان ورودی به لایه‌های RNN (مانند LSTM یا GRU) داده می‌شود. لایه‌های RNN مسئول یادگیری وابستگی‌های زمانی و الگوهای ترتیبی از ویژگی‌های استخراج شده هستند. این رویکرد دو مرحله‌ای به CRNN اجازه می‌دهد تا هم ویژگی‌های محلی را تشخیص دهد و هم وابستگی‌های زمانی را مدل‌سازی کند. با این حال، در برخی موارد، پیچیدگی این مدل ترکیبی می‌تواند منجر به کاهش کارایی در مقایسه با مدل‌های ساده‌تر و هدفمندتر شود.

۳-۳ معیارهای ارزیابی و تنظیمات پارامترهای

آموزش

تمامی مدل‌ها با استفاده از مجموعه‌ای یکسان از تنظیمات آموزش و معیارهای ارزیابی، مورد مقایسه قرار گرفتند تا نتایج به صورت منصفانه و دقیق قابل تحلیل باشند. این رویکرد، امکان مقایسه مستقیم عملکرد مدل‌ها را فراهم می‌کند. جدول (۲) پارامترها و مقادیر آن‌ها را نمایش می‌دهد. توجه به این نکته ضروری است که تنظیمات ساده و یکسان آموزش به طور عمدی انتخاب شدند تا یک مقایسه کنترل‌شده بین توانایی‌های ذاتی معماری مدل‌ها فراهم شده و از هرگونه مزیت ناشی از بهینه‌سازی‌های اختصاصی برای یک مدل خاص جلوگیری شود.

³ Outliers

¹ Convolutional Recurrent Neural Networks (CRNN)

² Root Mean Square Error (RMSE)

ضریب تعیین (R^2)

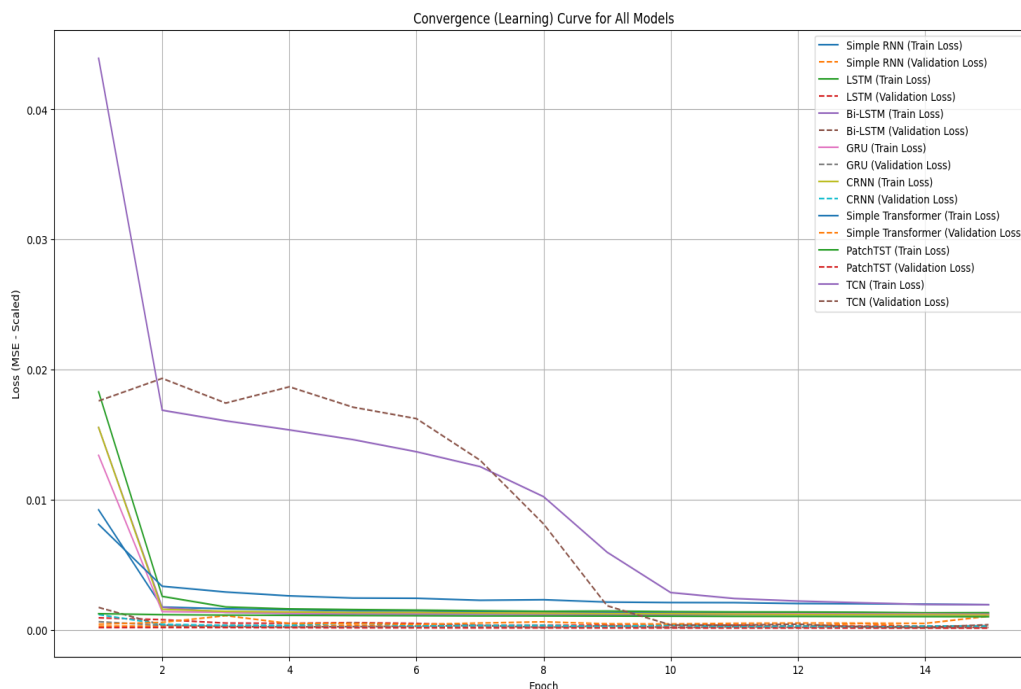
این معیار، نشان می‌دهد که چه نسبتی از واریانس داده‌های واقعی توسط مدل قابل توضیح است. R^2 به عنوان یک معیار نرمال‌شده، مقایسه عملکرد مدل‌ها را آسان‌تر می‌کند. مقداری نزدیک به ۱ نشان‌دهنده عملکرد عالی مدل است که تقریباً تمام واریانس داده‌ها را توضیح می‌دهد. مقدار ۰ به این معنی است که مدل هیچ بهبودی نسبت به یک مدل خط پایه ساده (پیش‌بینی میانگین) ندارد، و مقادیر منفی نشان‌دهنده عملکردی بدتر از خط پایه است. هرچه مقدار R^2 به ۱ نزدیک‌تر باشد، مدل از تناسب بهتری با داده‌ها برخوردار است. رابطه (۵) نحوه محاسبه این معیار را نمایش داده است.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (5)$$

۴- نتایج ارزیابی

در این پژوهش، برای پیش‌بینی قیمت سهام فولاد، عملکرد چندین مدل یادگیری عمیق از جمله مدل‌های بازگشتی، مدل‌های مبتنی بر ترنسفورمر و شبکه‌های کانولوشنی مورد بررسی و مقایسه قرار گرفته است. برای حفظ شفافیت، قابلیت تکرارپذیری و تضمین

پیش‌بینی واقع‌بینانه، داده‌های سری زمانی پس از پیش‌پردازش شامل حذف داده‌های گم‌شده، نرمال‌سازی با MinMaxScaler و ایجاد پنجره‌های زمانی به صورت کاملاً ترتیبی (زمانی) به سه زیرمجموعه آموزش، اعتبارسنجی و آزمون با نرخ ۱۵/۱۵/۷۰ تقسیم شدند تا از نشت اطلاعات به‌طور کامل جلوگیری شود. تمامی مدل‌ها با استفاده از تنظیمات یکسانی ارزیابی شدند: طول توالی ورودی برابر با ۳۰ روز و طول پیش‌بینی خروجی برابر با ۷ روز در نظر گرفته شد؛ هایپرپارامترهای بهینه‌سازی شامل بهینه‌ساز Adam، نرخ یادگیری $1e-4$ ، اندازه دسته ۶۴ و ۵۰ دوره آموزش بودند. برای اطمینان از ثبات نتایج، هر آزمایش ۳ بار به صورت مستقل با مقادیر بذر تصادفی متفاوت تکرار شد. در نهایت، ارزیابی عملکرد مدل‌ها بر روی داده‌های آزمون با استفاده از سه معیار اصلی RMSE، MAE و R^2 انجام شده است. برای ارزیابی همگرایی مدل‌ها و اطمینان از عدم وقوع بیش‌برازش، نمودار منحنی یادگیری تمام مدل‌های یادگیری عمیق در شکل (۲) ترسیم شده است. این نمودار نشان می‌دهد که تابع هزینه فاز آموزش و اعتبارسنجی هم‌زمان کاهش یافته و در طول تکرارها تثبیت شده‌اند، که نشان‌دهنده فرایند یادگیری پایدار و بهینه است.



شکل (۲): نمودار همگرایی تابع هزینه فاز آموزش و اعتبارسنجی تمام مدل‌های مورد استفاده

ترنسفورمر است که با استفاده از مکانیزم توجه به خود، به مدل اجازه می‌دهد تا روابط پیچیده و وابستگی‌های بلندمدت را در سری زمانی کشف کند، بدون اینکه با مشکلات مدل‌های ترتیبی مواجه شود. هم‌چنین، رویکرد تقسیم داده‌ها به قطعات، به این مدل کمک کرده تا الگوهای محلی را با دقت بیشتری یاد بگیرد. در میان مدل‌های بازگشتی، GRU با RMSE معادل ۰/۰۵۶۱ و R^2 برابر ۰/۹۳۷۱، بهترین عملکرد را به نمایش گذاشته است. این نتایج نشان می‌دهد که مدل GRU با وجود سادگی ساختاری، در مقایسه با LSTM از لحاظ دقت و کارایی محاسباتی، برتری دارد. مدل RNN با ضعیف‌ترین عملکرد R^2 به مقدار ۰/۷۸۶۸ به عنوان یک خط مبنا، به خوبی نشان می‌دهد که مدل‌های پیشرفته‌تر تا چه اندازه توانسته‌اند مشکلات مدل‌های ساده‌تر، به خصوص پدیده محو شدن گرادیان، را حل کرده و دقت پیش‌بینی را به طور چشمگیری افزایش دهند. مدل TCN در بین مدل‌های مبتنی بر شبکه کانولوشن عملکرد متوسطی دارد اما به دلیل استفاده از مفهوم کانولوشن بیشترین زمان آموزش را از بین تمام مدل‌ها به خود اختصاص داده است.

برای درک بهتر عملکرد مدل‌ها، شکل‌های (۳) و (۴) عملکرد دو مدل برتر را ترسیم کرده است.

جدول (۳) خلاصه‌ای از میانگین عملکرد حاصل از این ۳ تکرار اجرای مدل‌ها را نشان می‌دهد.

جدول (۳): نتایج ارزیابی میانگین مدل‌های مختلف

مدل	RMSE	MAE	R^2	زمان آموزش (ثانیه)
CRNN	۰/۱۰۱۲	۰/۰۸۰۱	۰/۷۹۷۵	۱۴/۳۳
TCN	۰/۰۸۸۱	۰/۰۶۸۱	۰/۸۴۵۳	۱۹۶/۴۷
RNN	۰/۱۰۳۴	۰/۰۸۱۴	۰/۷۸۶۸	۱۱/۲۹
LSTM	۰/۰۶۰۴	۰/۰۴۶۹	۰/۹۲۷۲	۱۲/۳۲
Bi-LSTM	۰/۰۸۳۹	۰/۰۶۶۸	۰/۸۵۹۶	۱۵/۸۸
GRU	۰/۰۵۶۱	۰/۰۴۳۴	۰/۹۳۷۱	۱۱/۰۱
Transformer	۰/۰۶۲۸	۰/۰۴۸۰	۰/۹۲۱۲	۳۴/۸۸
PatchTST	۰/۰۳۰۴	۰/۰۲۲۱	۰/۹۸۱۵	۳۷/۱۵

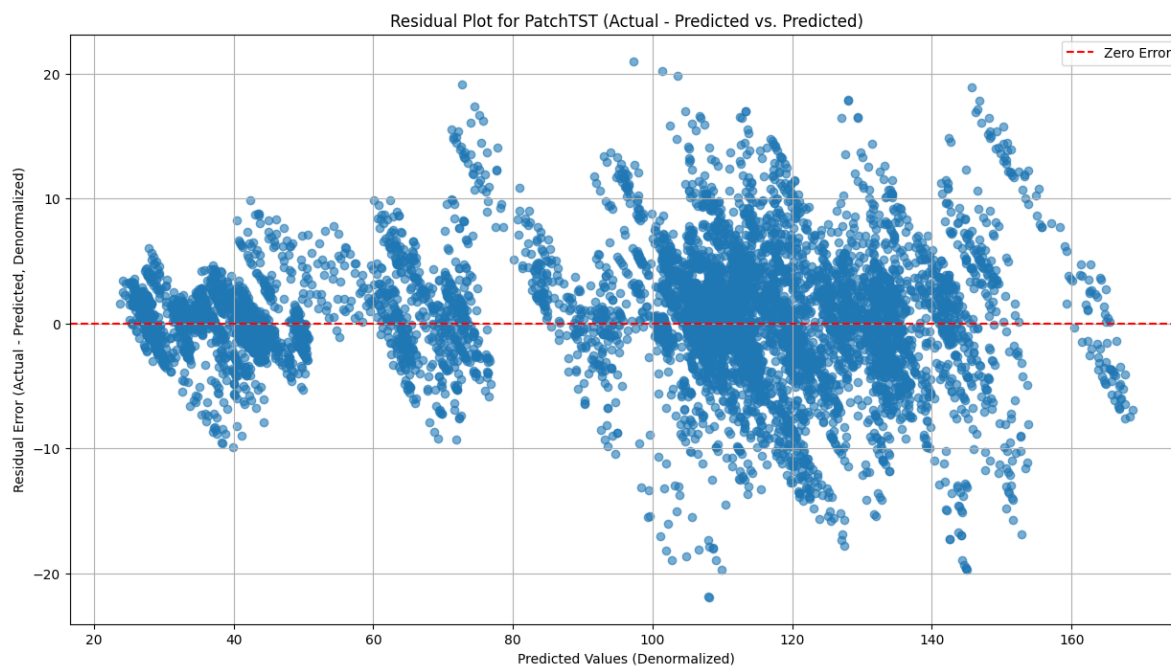
همان‌طور که از جدول (۳) مشخص است، نتایج به وضوح برتری برخی مدل‌ها را نسبت به دیگری نشان می‌دهد. مدل PatchTST با کسب بهترین امتیاز در تمامی معیارها، عملکردی کاملاً برتر را ارائه داد. با RMSE بسیار پایین (۰/۰۳۰۴) و R^2 فوق‌العاده بالا (۰/۹۸۱۵)، این مدل توانسته است بیش از ۹۸٪ از تغییرات قیمت فولاد را توضیح دهد. این برتری به دلیل معماری مبتنی بر



شکل (۳): مقایسه مقادیر واقعی و پیش‌بینی شده با روش PatchTST



شکل (۴): مقایسه مقادیر واقعی و پیش‌بینی شده با روش GRU



شکل (۵): نمودار خطای باقی‌مانده مقدار واقعی و پیش‌بینی شده برای مدل PatchTST

نتایج نهایی موجود در جدول (۴) نشان می‌دهد که در هر سه مجموعه داده متفاوت، مدل PatchTST به‌طور قاطع بهترین عملکرد را ثبت کرده است. این برتری، ریشه در معماری PatchTST دارد که با تقسیم سری زمانی به قطعات و استفاده از مکانیزم توجه، می‌تواند همبستگی‌های بلندمدت و دوردست را به صورت موازی و مؤثر، بدون مشکل محوشدگی گرادیان که محدودیت مدل‌های RNN5/LSTM است، استخراج و پردازش کند. این مقایسه قاطع، فرضیه اصلی پژوهش مبنی بر برتری مدل‌های مبتنی بر ترنسفورمر را بر مدل‌های RNN در پیش‌بینی سری‌های زمانی قیمت فولاد، به دلیل توانایی بالاتر در پردازش موازی و مدیریت وابستگی‌های دوردست، به‌طور کامل اثبات می‌کند.

۵- نتیجه‌گیری

پیش‌بینی دقیق قیمت فولاد به دلیل اهمیت استراتژیک این ماده در صنایع گوناگون، از اهمیت حیاتی برخوردار است. این پژوهش با هدف ارزیابی و مقایسه عملکرد مدل‌های پیشرفته یادگیری عمیق در پیش‌بینی قیمت سری زمانی فولاد انجام شد. در این تحقیق، مجموعه‌ای از مدل‌های بازگشتی، ترنسفورمر و کانولوشنی بر روی داده‌های تاریخی قیمت سهام فولاد مورد آزمایش قرار گرفتند و نتایج به وضوح نشان داد که مدل PatchTST با کسب بهترین امتیاز در تمام معیارهای ارزیابی عملکردی فراتر از سایر مدل‌ها ارائه داده است. این برتری قابل توجه، بیش از بازگویی نتایج عددی است و دلیل تحلیلی عمیقی دارد. رویکرد نوآورانه PatchTST در تقسیم سری زمانی به قطعات و استفاده از مکانیزم توجه به خود، به‌طور مؤثر چالش‌های مدل‌های بازگشتی (مانند محوشدگی گرادیان) را حل کرده است؛ این مکانیزم به مدل اجازه می‌دهد که وابستگی‌های پیچیده و بلندمدت را با دقت بالا مدل‌سازی کند، قابلیت‌هایی که مدل‌های RNN به دلیل ماهیت ترتیبی و محدودیت‌های ساختاری خود فاقد آن بودند. با وجود این نتایج موفقیت‌آمیز، این پژوهش با چالش تک‌متغیره بودن داده‌ها روبرو است که یک

این شکل‌ها قیمت‌های پیش‌بینی‌شده را با قیمت‌های واقعی در مجموعه داده آزمون مقایسه کرده است. این نمودارها به صورت بصری نشان می‌دهند که چگونه نمودار پیش‌بینی مدل PatchTST با کمترین انحراف، نزدیک‌ترین تطابق را با داده‌های واقعی داشته است. به منظور بررسی سیستماتیک و سلامت آماری مدل برتر، نمودار خطای باقی‌مانده برای مدل PatchTST در شکل (۵) آورده شده است. توزیع تصادفی خطاها حول خط صفر، شواهدی بر عدم وجود الگوی خطای ساختاریافته و توانایی مدل در ثبت وابستگی‌های غیرخطی بدون سوگیری است.

۴-۱ مقایسه روش پیشنهادی با دیگر پایگاه داده‌ها

برای اطمینان از تعمیم‌پذیری و مقاومت روش پیشنهادی، دو پایگاه داده مستقل مرتبط با قیمت واقعی فولاد، علاوه بر داده‌های اصلی سهام فولاد (TATASTEEL.NS_stock_data)، مورد ارزیابی قرار گرفتند. این اقدام به طور خاص جهت پاسخ به دغدغه‌های مربوط به نوع داده و اعتبار علمی پژوهش انجام شده است. مجموعه داده‌های اعتبارسنجی شامل شاخص رسمی قیمت ماهانه ورق فولاد (FRED WPU101707)^۱ است که الگوهای کلان را در یک بازه طولانی (از ژوئن ۱۹۸۲ تا می ۲۰۲۴) پوشش می‌دهد و برای این مجموعه داده، پارامترهای بلندمدت ورودی (Seq_len) با مقدار ۱۲ ماه و پارامتر طول خروجی (Pred_len) برابر با ۳ ماه تنظیم شده است. دومین مجموعه داده، قیمت آتی روزانه کلاف نورد گرم (Futures HRC)^۲ است که قیمت خام فولاد را با فرکانس روزانه در بازه (از سپتامبر ۲۰۱۵ تا اکتبر ۲۰۲۵) بازتاب می‌دهد و با پارامترهای طول ورودی ۳۰ روزه و طول خروجی ۷ روزه مدل‌سازی شده است. تمام مدل‌ها بر روی هر سه مجموعه داده با ۱۵ تکرار آموزش داده شدند تا مقایسه عملکرد کاملاً منصفانه و مستند باشد. جدول (۴) نتایج مقایسه‌ای جامع مدل‌ها را بر اساس سه معیار اصلی RMSE، MAE و R^2 در سه مجموعه داده با فرکانس‌های را گزارش می‌دهد.

¹ <https://fred.stlouisfed.org/series/WPU101707>

² <https://www.investing.com/commodities/us-steel-coil-futures-historical-data>

گزارش‌های مالی و ترکیب آن‌ها با داده‌های سری زمانی می‌تواند دقت پیش‌بینی را بهبود بخشد. هم چنین، انجام یک فرآیند بهینه‌سازی سیستماتیک برای یافتن بهترین هایپرپارامترهای هر مدل و آزمایش با طول‌های ورودی و خروجی متفاوت نیز پیشنهاد می‌گردد.

محدودیت تحلیلی محسوب می‌شود، چرا که عوامل کلان اقتصادی، عرضه و تقاضا، اخبار جهانی و سیاست‌های دولت نیز تأثیر بسزایی در نوسانات قیمت فولاد دارند. از این رو، برای کارهای آینده پیشنهاد می‌شود که از داده‌های چندمتغیره شامل حجم معاملات و شاخص‌های اقتصادی مرتبط استفاده شود، همچنین بررسی امکان استفاده از تکنیک‌های پردازش زبان طبیعی برای تحلیل اخبار و

جدول (۴): مقایسه روش پیشنهادی با استفاده از سه مجموعه داده مختلف

معیار ارزیابی	مدل	TATA STEEL	HRC Futures	FRED WPU101707
R ²	PatchTST	۰/۹۸۱۵	۰/۸۸۶۸	۰/۷۵۲۹
	GRU	۰/۹۳۷۱	۰/۸۵۶۴	-۱/۴۸۳۵
	LSTM	۰/۹۲۷۲	۰/۶۱۵۷	-۲/۴۷۵۰
	Bi-LSTM	۰/۷۸۶۸	۰/۸۵۹۶	-۲/۰۴۷۷
	Simple Transformer	۰/۹۲۱۲	۰/۷۷۵۵	۰/۱۰۸۸
	RNN	۰/۸۵۹۶	۰/۷۵۹۱	۰/۳۲۱۵
	TCN	۰/۸۴۵۳	-۱/۴۹۰۲	-۲/۶۵۳۴
	CRNN	۰/۷۶۷۵	۰/۷۹۷۴	-۰/۶۵۹۱
RMSE	PatchTST	۰/۰۳۰۴	۰/۰۲۰۳	۰/۱۱۳۱
	GRU	۰/۰۵۶۱	۰/۰۲۲۸	۰/۳۵۸۵
	LSTM	۰/۰۶۰۴	۰/۰۳۷۴	۰/۴۲۴۱
	Bi-LSTM	۰/۱۰۳۴	۰/۰۲۲۶	۰/۳۹۷۱
	Simple Transformer	۰/۰۶۲۸	۰/۰۲۸۵	۰/۲۱۴۸
	RNN	۰/۰۸۳۹	۰/۰۲۹۶	۰/۱۸۷۴
	TCN	۰/۰۸۸۱	۰/۰۹۵۱	۰/۴۳۴۸
	CRNN	۰/۱۰۷۹	۰/۰۲۷۱	۰/۲۹۳۰
MAE	PatchTST	۰/۰۲۲۱	۰/۰۱۴۳	۰/۰۷۴۲
	GRU	۰/۰۴۳۴	۰/۰۱۵۰	۰/۲۹۶۱
	LSTM	۰/۰۴۶۹	۰/۰۲۶۰	۰/۳۶۱۲
	Bi-LSTM	۰/۰۸۱۴	۰/۰۱۵۶	۰/۳۳۷۱
	Simple Transformer	۰/۰۴۸۰	۰/۰۱۹۳	۰/۱۳۳۳
	RNN	۰/۰۶۶۸	۰/۰۲۴۸	۰/۱۱۵۱
	TCN	۰/۰۶۸۱	۰/۰۷۹۶	۰/۳۷۲۱
	CRNN	۰/۰۸۸۱	۰/۰۱۷۷	۰/۲۲۱۴

References

- [1] G. Dooley, and H. Lenihan, "An assessment of time series methods in metal price forecasting," Resources Policy, Vol. 30, No. 3, pp. 208-217, 2005.

- [2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of econometrics, Vol. 31, No. 3, pp. 307-327, 1986.

- [3] C. A. Sims, "Macroeconomics and reality," *Econometrica: journal of the Econometric Society*, pp. 1-48, 1980.
- [4] [4] C. Zhang, N. N. A. Sjarif, and R. Ibrahim, "Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 14, No. 1, 2024.
- [5] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [6] C. Ozdemir, K. Buluş, and K. Zor, "Medium-to long-term nickel price forecasting using LSTM and networks," *Resources Policy*, Vol. 78, p. 102906, 2022.
- [7] L'Heureux, K. Grolinger, and M. A. Capretz, "Transformer-based model for electrical load forecasting," *Energies*, Vol. 15, No. 14, p. 4993, 2022.
- [8] Guha, and G. Bandyopadhyay, "Gold price forecasting using ARIMA model," *Journal of Advanced Management Science*, Vol. 4, No. 2, 2016.
- [9] G. P. Girish, "Spot electricity price forecasting in Indian electricity market using autoregressive-GARCH models," *Energy Strategy Reviews*, Vol. 11, pp. 52-57, 2016.
- [10] F. Merabet, H. Zeghdoudi, R. H. Yahia, and I. Saba, "Modelling of oil price volatility using ARIMA-GARCH models," *Adv. Mathematics*, Vol. 10, pp. 2361-2380, 2021.
- [11] T. Rahnemoun Piruj, S. S. Akbar Mousavi, and M. Asgari, "Modeling and Monthly Price Forecasting of Steel in Iran," *Stable Economy Journal*, Vol. 4, No. 4, pp. 60-95, 2024.
- [12] M. S. Khan, and U. Khan, "Comparison of forecasting performance with VAR vs. ARIMA models using economic variables of Bangladesh," *Asian Journal of Probability and Statistics*, Vol. 10, No. 2, pp. 33-47, 2020.
- [13] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia computer science*, Vol. 167, pp. 599-606, 2020.
- [14] Kuvalekar, S. Manchewar, S. Mahadik, and S. Jawale, "House price forecasting using machine learning," In *Proceedings of the 3rd international conference on advances in science & technology (ICAST)*, 2020.
- [15] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, Vol. 8, No. 9, p. 1441, 2020.
- [16] M. Gunarto, S. Sa'adah, and D. Q. Utama, "Predicting cryptocurrency price using rnn and lstm method," *Journal Sisfokom (Sistem Informasi dan Komputer)*, Vol. 12, No. 1, pp. 1-8, 2023.
- [17] T. Muhammad, A. B. Aftab, M. Ibrahim, M. M. Ahsan, M. M. Muhu, S. I. Khan, and M. S. Alam, "Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market," *International Journal of Computational Intelligence and Applications*, Vol. 22, No. 3, p. 2350013, 2023.
- [18] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, No. 12, pp. 11106-11115, 2021.
- [19] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.
- [20] Patnaik, N. J. Rao, B. Padhiari, and S. Patnaik, "Optimised hybrid CNN-LSTM model for stock price prediction," *International Journal of Management and Decision Making*, Vol. 23, No. 4, pp. 438-460, 2024.
- [21] L. Jialin, Q. Shanwen, Z. Zhikai, L. Keyao, M. Jiayong, and T. T. Toe, "Cnn-lstm model stock forecasting based on an integrated attention mechanism," In *2022 3rd International conference on pattern recognition and machine learning (PRML)*, pp. 403-408, 2022.
- [22] S. Wang, "A stock price prediction method based on BiLSTM and improved transformer," *IEEE Access*, Vol. 11, pp. 104211-104223, 2023.
- [23] Y. Liu, H. Dong, X. Wang, and S. Han, "Time series prediction based on temporal convolutional network," In *2019 IEEE/ACIS 18th International conference on computer and information science (ICIS)*, pp. 300-305, 2019.
- [24] J. Shi, R. Myana, V. Stebliankin, A. Shirali, and G. Narasimhan, "Explainable parallel rcnn with novel feature representation for time series forecasting," In *International Workshop on Advanced Analytics and Learning on Temporal Data*, pp. 56-75, 2023.

Steel Price Time Series Forecasting Using a Patch-based Transformer Architecture

Fatemeh Zare Mehrjardi^{1*}, Abbas Norouzi²

¹Assistant Professor Computer Engineering Department, Meybod University, Meybod, Iran

²Ms Student Computer Engineering Department, Meybod University, Meybod, Iran

Article Information

Original Research Paper

Received:

2025 August 17

Accepted:

2025 November 04

Keywords:

Steel price forecasting, deep learning, time series, recurrent neural networks, transformer network

Corresponding Author*:

fzare@meybod.ac.ir

Abstract

Steel, as one of the most critical materials across various industries, plays a vital role in the global economy. However, its high price volatility, complex and non-linear dependencies, and the inability of traditional models to capture long-term correlations have made its accurate forecasting a serious challenge. To overcome these shortcomings, this research proposes a comprehensive deep learning approach for steel price time series forecasting, structured in three main phases. First, the data undergoes preprocessing, including handling missing values, normalization, and creating time window inputs. Second, various deep learning models, including Recurrent Neural Networks (RNNs), Transformer networks (PatchTST), and Convolutional Networks, are trained and optimized. Finally, the third phase involves evaluating and comparing the models' performance using the R^2 , RMSE, and MAE metrics. Evaluation results demonstrate that the PatchTST model, by utilizing its attention mechanism and processing data in segments or "patches," successfully identified complex, long-term dependencies with superior accuracy compared to other models. Specifically, PatchTST achieved a leading R^2 value of 0.9815 and the lowest RMSE of 0.0304 in steel price prediction. Conversely, standard RNN models exhibited significantly weaker performance due to their sequential nature and inherent structural limitations. These findings underscore the definitive superiority of Transformer-based models for accurate forecasting of complex time series data.

 : 10.22034/ABMIR.2025.23536.1152

E-ISSN: [2821-2037](#) /© 2025. Published by Yazd University This is an open access article under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

