Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال سوم، شماره دوم، پاییز و زمستان ۱۴۰۴، صفحات ۱۰۵ - ۸۹

مدلی ترکیبی در پیش‌بینی داده محور مصرف انرژی در صنعت فولاد: رویکردی بر پایه LSTM و بهینه‌سازی هایپرپارامترها با الگوریتم Optuna

لیلا ذوالقدر[۱]، مجید شخصی نیائی[۲*]، یحیی زارع مهرجردی[۳]، محمدمهدی لطفی[۳]

[۱] دانشجوی دکتری مهندسی صنایع، گروه مهندسی صنایع، دانشگاه یزد، یزد، ایران

[۲] دانشیار، گروه مهندسی صنایع، دانشگاه یزد، یزد، ایران

[۳] استاد، گروه مهندسی صنایع، دانشگاه یزد، یزد، ایران

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۴/۰۵/۳۱

تاریخ پذیرش:
۱۴۰۴/۰۷/۲۸

کلیدواژه‌ها:
مدل‌سازی داده محور، LSTM-Optuna ، اعتبار سنجی گام به گام، مصرف انرژی الکتریکی، صنعت فولاد، RNN، SVM

نویسنده‌مسئول:
m.niaei@yazd.ac.ir

چکیده

در دنیای مدرن، انرژی الکتریکی یکی از اساسی‌ترین نیازهای جوامع، نقش حیاتی در بخش‌های صنعتی، اقتصادی و اجتماعی ایفا می‌کند. با توجه به عدم امکان ذخیره‌سازی در مقیاس بزرگ و نوسانات عرضه و تقاضا، پیش‌بینی دقیق مصرف بار الکتریکی ضروری است. مدل‌های سنتی سری‌های زمانی، اغلب در مواجهه با الگوهای غیرخطی و پیچیده داده‌های صنعتی دچار چالش می‌شوند. این پژوهش یک چارچوب ترکیبی نوین برای پیش‌بینی دقیق سری زمانی مصرف انرژی الکتریکی ارائه می‌دهد. رویکرد پیشنهادی بر پایه شبکه عصبی حافظه بلند-کوتاه مدت (LSTM) است که برای به حداکثر رساندن عملکرد مدل، هایپرپارامترهای شبکه LSTM با استفاده از الگوریتم بهینه‌سازی خودکار Optuna بهینه‌سازی شده‌اند. داده‌های مورد استفاده شامل سری‌های زمانی مصرف برق کارخانه فولاد بافت به همراه متغیرهای برونزا مانند میزان تولید فولاد، مدت زمان توقف تولید، شرایط آب و هوایی و تقویمی هستند. نتایج ارزیابی نشان می‌دهند که مدل ترکیبی LSTM-Optuna با ضریب تعیین (R2) ۰/۸۹و درصد خطای مطلق میانگین (MAPE)۱۸/۴۶٪، نه تنها از مدل پایه با ضریب تعیین ۰/۸۵ و درصد خطای مطلق میانگین با مقدار ۲۰/۴۴٪ عملکرد بهتری نشان می‌دهد، بلکه از سایر روش‌های یادگیری ماشین مانند شبکه‌های عصبی بازگشتی (RNN) و ماشین بردار پشتیبان (SVM).

: 10.22034/ABMIR.2025.23569.1157

E-ISSN: 2821-2037

۱- مقدمه

انرژی الکتریکی به‌عنوان یکی از اساسی‌ترین و حیاتی‌ترین نیازهای جوامع بشری، نقشی زیربنایی در توسعه صنعتی، اقتصادی و اجتماعی ایفا می‌کند و امکان ذخیره‌سازی این انرژی در ابعاد بزرگ با فناوری‌های موجود محدود است[۱]. از طرفی انرژی الکتریکی به‌عنوان نیروی محرکه اصلی در صنایع سنگین، نقشی زیربنایی در فرآیندهای تولید ایفا می‌کند. در صنعت فولاد، که فرآیندهای آن به شدت وابسته به کوره‌های قوس الکتریکی و ماشین‌آلات پرمصرف است، مصرف بهینه و پایدار انرژی، ستون فقرات تولید محسوب می‌شود. با توجه به اینکه هزینه‌های انرژی بخش قابل توجهی از هزینه‌های عملیاتی کارخانه‌ها را تشکیل می‌دهد، پیش‌بینی دقیق مصرف برق برای مدیران انرژی و تولید، یک ضرورت راهبردی است.

با این حال، دستیابی به پیش‌بینی دقیق در صنعت فولاد به دلیل وابستگی مصرف به عوامل پیچیده و غیرخطی، یک چالش بزرگ است. الگوهای مصرف انرژی نه تنها تحت تأثیر برنامه‌های تولیدی، بلکه به شدت به متغیرهای برون‌زا و نوسانات ناگهانی و شدید ناشی از عواملی چون قطعی‌های برق برنامه‌ریزی‌نشده و توقف‌های اضطراری نیز وابسته است. این شرایط، داده‌های صنعتی را به مراتب پیچیده‌تر از داده‌های عمومی یا خانگی مصرف برق می‌کند. مدل‌های سنتی سری‌های زمانی، که بر مفروضات خطی بنا شده‌اند، در مواجهه با این الگوهای غیرخطی و نوسانات ناگهانی، کارایی لازم را ندارند [۲]. این امر به‌ویژه در صنایع پرمصرف مانند صنعت فولاد، که مصرف انرژی آن تحت تأثیر متغیرهای برون‌زا قرار می‌گیرد، اهمیت بیشتری پیدا می‌کند[۳]. در سال‌های اخیر، با رشد یادگیری عمیق، رویکردهای نوین مبتنی بر شبکه‌های عصبی عمیق، به‌خصوص مدل حافظه بلند-کوتاه مدت(LSTM) ، به عنوان راهکارهای کارآمد برای پیش‌بینی سری‌های زمانی مطرح شده‌اند[٤].LSTM به دلیل توانایی منحصربه‌فرد خود در یادگیری وابستگی‌های طولانی‌مدت، به‌خوبی می‌تواند الگوهای پیچیده و غیرخطی را در داده‌های مصرف انرژی صنعتی شناسایی کند[٥]. اما چالش فراتر از این است؛ همان‌طور که در پژوهش‌های اخیر بر روی داده‌های صنعتی (مانند حوزه نفت و گاز) نشان داده شده، حتی به‌کارگیری چارچوب‌های پیشرفته یادگیری ماشین خودکار نیز می‌تواند به دلیل پیچیدگی داده‌ها با مشکلات جدی مانند بیش‌برازش مواجه شود و مدل‌هایی غیرقابل اعتماد تولید کند. مطالعه ژو و همکاران (۲۰۲۵) این امر نشان را می‌دهد که صرفاً استفاده از یک مدل پیشرفته کافی نیست و نیاز به یک رویکرد مستحکم و هوشمندانه برای ساخت و اعتبارسنجی مدل در شرایط واقعی صنعتی به شدت احساس می‌شود[٦].

با این حال، عملکرد بهینه این مدل‌ها به تنظیم دقیق هایپرپارامترهای آن‌ها وابسته است [۷]. تنظیم دستی این پارامترها فرآیندی زمان‌بر و غیربهینه است و تضمینی برای دستیابی به بهترین پیکربندی مدل نیز وجود ندارد[۸]. از طرفی با توجه به آن که نوسانات غیرقابل پیش‌بینی و شدید، یک چالش اساسی در مدیریت انرژی و پیش‌بینی دقیق مصرف برق به شمار می‌رود، در این پژوهش، تمرکز بر روی داده‌هایی است که علاوه بر الگوهای فصلی و روزانه، تحت تأثیر عواملی چون قطعی‌های برق برنامه‌ریزی‌نشده قرار دارند. چرا که این امر، دقت مدل‌های پیش‌بینی سنتی را به شدت کاهش می‌دهد و کاملاً تداعی‌کننده شرایط واقعی در صنعت فولاد و محدودیت‌های موجود در مصرف برق است.

این پژوهش در تلاش برای پاسخگویی به تمامی چالش‌های پیش‌بینی مصرف انرژی در صنایع سنگین، یک چارچوب داده‌محور جامع را معرفی می‌کند. این رویکرد پیشنهادی، بر قدرت شبکه عصبی LSTM در مدل‌سازی وابستگی‌های زمانی پیچیده تکیه دارد و آن را با یک الگوریتم بهینه‌سازی هوشمند هایپرپارامترها (Optuna) ترکیب می‌کند. نوآوری و هدف اصلی این تحقیق، نه صرفاً ترکیب دو تکنیک، بلکه اعتبارسنجی و به‌کارگیری این رویکرد پیشرفته بر روی داده‌های واقعی، پیچیده و پرنویز صنعت فولاد است تا نشان دهد آیا چنین مدلی می‌تواند با موفقیت از پس چالش‌های دنیای واقعی صنعت برآید و مدلی دقیق و در عین حال قابل تعمیم ارائه دهد. برای اثبات کارایی این چارچوب، عملکرد آن با مدل پایه و همچنین طیف وسیعی از روش‌های کلاسیک و مدرن پیش‌بینی سری زمانی شامل شبکه‌های عصبی بازگشتی (RNN) و ماشین بردار پشتیبان (SVM) مقایسه شده است.

۲- پیشینه تحقیق

ادبیات پژوهشی در زمینه پیش‌بینی مصرف برق یک مسیر تکاملی طولانی را طی کرده است. در ابتدا، روش‌های آماری و کلاسیک و رگرسیون، ابزارهای اصلی برای این کار بودند. بان و فارمر [۹]و موغرام و رحمان[۱۰]از اولین کسانی بودند که به مرور و مقایسه روش‌های پیش‌بینی کوتاه‌مدت در صنعت برق پرداختند. با این حال، همان‌طور که ژانگ و کی[۲]اشاره کردند، این مدل‌ها در مواجهه با روابط غیرخطی پیچیده با محدودیت روبرو هستند.

با پیشرفت علوم رایانه، مدل‌های مبتنی بر هوش مصنوعی و شبکه‌های عصبی مطرح شدند. این مدل‌ها، به دلیل توانایی در یادگیری الگوهای پیچیده و غیرخطی از داده‌ها، عملکرد بهتری نسبت به روش‌های سنتی نشان داده‌اند. شبکه عصبی مصنوعی (ANN)و پس از آن، مدل‌های پیشرفته‌تری نظیر شبکه‌های عصبی پیچشی (CNN) و شبکه‌های عصبی بازگشتی (RNN) پا به عرصه گذاشتند [۱۱]. مطالعات مختلفی، کارایی شبکه‌های عصبی مصنوعی (ANN) را در پیش‌بینی مصرف برق و سایر متغیرهای صنعتی نشان داده‌اند[۱۲].

در ادامه ماشین بردار پشتیبان (SVM) با معرفی نسخه رگرسیون آن (SVR)، به ابزاری قدرتمند برای پیش‌بینی مقادیر پیوسته تبدیل شد. این مدل به دلیل توانایی در مدیریت فضاهای با ابعاد بالا و مقاومت در برابر بیش‌برازش از طریق تکنیک هسته، در پیش‌بینی سری‌های زمانی، به ویژه در حوزه‌های مالی و انرژی، کاربرد فراوانی دارد [۱۳].

برای حل مشکل محوشدگی گرادیان در RNN ها، هوخرایتر و شمیدهوبر [۱۴]مدل حافظه طولانی-کوتاه‌مدت (LSTM) را معرفی کردند. ژنگ و همکاران[۱۵]و پی و همکاران[۱۶]نیز از LSTM برای پیش‌بینی کوتاه‌مدت با نتایج قابل توجهی استفاده کرده‌اند که به دلیل ساختار حافظه‌دار خود، برای پیش‌بینی سری‌های زمانی بسیار مناسب هستند[۱۷]. تحقیقات متعددی اثبات کرده‌اند که LSTM عملکرد قابل توجهی در پیش‌بینی مصرف انرژی[۱۸]و سایر سری‌های زمانی پیچیده دارد[۱۹،۲۰]. برخی مطالعات به مقایسه LSTM با مدل‌های آماری پرداخته و برتری آن را در شرایط خاص نشان داده‌اند[۲۱].

با وجود کارایی بالای مدل‌های یادگیری عمیق، محققان به این نتیجه رسیده‌اند که یک مدل تکی به‌تنهایی نمی‌تواند تمام پیچیدگی‌های سری‌های زمانی را مدل‌سازی کند. از این رو، رویکردهای ترکیبی توسعه یافتند. یکی از این تکنیک‌ها، تجزیه سیگنال است. هوانگ و همکاران[۲۲]روش تجزیه مد تجربی (EMD) و دراگومیرتسکی و زوسو[۲۳]روش تجزیه مد تغییراتی (VMD) را معرفی کردند. روان و همکاران[۲۴]از یک مدل ترکیبی VMD-LSTM استفاده کردند که نتایج بهتری نسبت به مدل‌های تکی به دست آورد.

همچنین پژوهش‌هایی مانند اکوندایو (۲۰۲۰) و قَرایی و همکاران (۲۰۲۵)، از چارچوب Optuna برای بهینه‌سازی مدل‌های ترکیبی مانند CNN-LSTM در پیش‌بینی مصرف برق استفاده کردند [۲۵،۲۶]. به طور مشابه، یودین و همکاران (۲۰۲۵) با بهینه‌سازی هایپرپارامترهای یک مدل LSTM به خطای بسیار پایینی دست یافتند[۲۷]. هی و همکاران[۷] از بهینه‌سازی بیزی برای بهینه‌سازی LSTMاستفاده کردند که به بهبود قابل توجهی در دقت پیش‌بینی منجر شد.

با نگاهی به پیشینه تحقیق، مشخص می‌شود که تحقیقات به سمت مدل‌های ترکیبی و بهینه‌سازی‌شده حرکت کرده‌اند. مطالعاتی مانند بوکتیف و همکاران [۲۸]از بهینه‌سازی‌های فراابتکاری استفاده کرده‌اند که اغلب از بهینه‌سازی خودکار و هوشمند هایپرپارامترها غفلت کرده‌اند و اکثر تحقیقات مذکور برروی داده‌های عمومی یا خانگی مصرف برق متمرکز شده‌اند و کمتر به چالش‌های منحصربه‌فرد داده‌ها همان‌طور که در پژوهش ژو و همکاران (۲۰۲۵) در حوزه صنعت نفت و گاز نشان داده شد، پرداخته شده است[۶]. چرا که به‌کارگیری مدل‌های یادگیری ماشین پیشرفته برروی داده‌های صنعتی می‌تواند با چالش‌هایی جدی مانند بیش‌برازش همراه باشد.

این پژوهش، با اعتبارسنجی یک چارچوب ترکیبی که به‌طور هم‌زمان از قابلیت‌های یادگیری عمیق LSTM برای مدل‌سازی وابستگی‌های بلندمدت و از یک الگوریتم بهینه‌سازی هوشمند و خودکار مانند Optuna برای تنظیم دقیق هایپرپارامترها در حوزه پیش‌بینی مصرف انرژی بر روی داده‌های پیچیده و پرنویز صنعت فولاد استفاده می‌کند، به دنبال پر کردن این شکاف پژوهشی است.

۳- مفاهیم پایه پژوهش

۳-۱ شبکه‌های عصبی عمیق، مدل حافظه کوتاه‌مدت طولانی LSTM برای سری‌های زمانی

حافظه کوتاه‌مدت طولانی (LSTM) به دلیل توانایی منحصر به فرد خود در یادگیری وابستگی‌های طولانی‌مدت، به یکی از محبوب‌ترین مدل‌ها برای سری‌های زمانی تبدیل شده است [۲۹]. LSTM یک نوع خاص از شبکه عصبی بازگشتی (RNN) است که توسط سپ هوخرایتر و یورگن شمیدهوبر در سال ۱۹۹۷ معرفی شد[۳۰]. یک سلول LSTM، به جای یک تابع فعال‌سازی ساده، دارای ساختار داخلی پیچیده‌تری است که شامل سه دروازه و یک حالت سلول است[۲۹].

۱- حالت سلول: این خط اصلی حافظه سلول است که اطلاعات را در طول توالی منتقل می‌کند.

۲- دروازه فراموشی: این دروازه تصمیم می‌گیرد که چه اطلاعاتی از حالت سلول قبلی (C_{t-1})باید فراموش شوند یا کنار گذاشته شوند. این دروازه یک لایه سیگموئید است که خروجی بین ۰ و ۱ تولید می‌کند. عدد ۱ به معنای نگه‌داشتن کامل و ۰ به معنای کاملاً فراموش کردن است که در آن دروازه فراموشی از رابطه شماره ۱ محاسبه می‌شود:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (۱)$$

که در آن x_tورودی فعلی (مشاهده در زمان t)، h_{t-1} خروجی پنهان از گام زمانی قبلی، W_f ماتریس وزن‌های دروازه فراموشی، b_f بردار بایاس دروازه فراموشی و σ تابع فعال‌سازی سیگموئید است.

۳- دروازه ورودی: این دروازه تصمیم می‌گیرد که چه اطلاعات جدیدی به حالت سلول فعلی (C_t) اضافه شوند. این شامل دو بخش است که یک لایه سیگموئید دیگر که تصمیم می‌گیرد کدام مقادیر آپدیت شوند (دروازهi_t) و یک لایه tanh که یک بردار از کاندیداهای مقادیر جدید $(\tilde{C}_t)$ را ایجاد می‌کند که می‌توانند به حالت سلول اضافه شوند که با توابع نمایش داده شده در رابطه ۲ محاسبه می‌شوند.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (۲)$$
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

که در آن W_i, W_C ماتریس‌های وزن‌ها، b_i, b_C بردارهای بایاس و tanh تابع فعال‌سازی تانژانت هایپربولیک است.

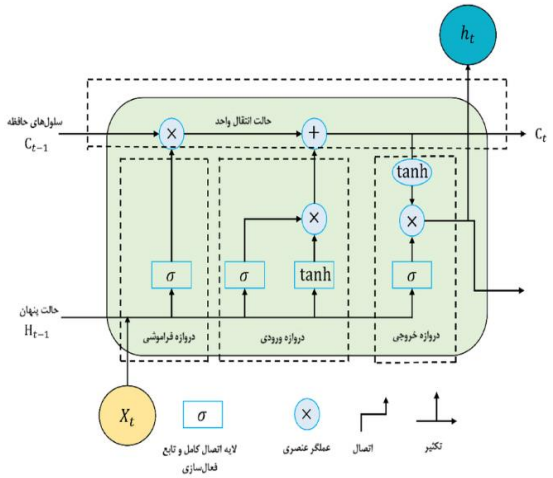
٤- به‌روزرسانی حالت سلول: در این مرحله، حالت سلول قبلی C_{t-1}با ترکیب اطلاعات فراموشی و اطلاعات ورودی جدید، به حالت سلول فعلی C_tبه‌روز می‌شود و در رابطه شماره ۳ ارائه شده است.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (۳)$$

۵- دروازه خروجی: این دروازه تصمیم می‌گیرد که چه بخشی از حالت سلول فعلی C_tبه عنوان خروجی پنهان در این گام زمانی ارائه شود. این خروجی به گام زمانی بعدی منتقل شده و همچنین می‌تواند به عنوان پیش‌بینی مدل استفاده شود. دروازه خروجی o_t و خروجی پنهان h_t با روابط نمایش داده در رابطه ٤ محاسبه می‌شود.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (٤)$$
$$h_t = o_t * \tan h(C_t)$$

که در آن W_o ماتریس وزن‌های دروازه خروجی و b_oبردار بایاس دروازه خروجی است. LSTM ها قادرند وابستگی‌های طولانی‌مدت را یاد بگیرند. این قابلیت کنترلی، به LSTM امکان می‌دهد تا اطلاعات مرتبط را برای مدت‌های طولانی حفظ کند و اطلاعات نامربوط را به سرعت فراموش کند [۲۹،۳۰]. در شکل ۱ معماری داخلی مدل LSTM نمایش داده شده است.



شکل (۱): معماری داخلی مدل حافظه کوتاه‌مدت طولانی LSTM [۳۱]

۳-۲ بهینه‌سازی هایپرپارامترها با استفاده از Optuna

عملکرد مدل‌های یادگیری عمیق، از جمله LSTM، به شدت به تنظیم دقیق هایپرپارامترها وابسته است. هایپرپارامترها، پارامترهایی هستند که قبل از فرآیند آموزش مدل تعیین می‌شوند و بر ساختار و فرآیند یادگیری مدل تأثیر می‌گذارند [۸]. انتخاب نادرست این پارامترها می‌تواند منجر به عملکرد ضعیف مدل، بیش‌برازش یا کم‌برازش شود. هایپرپارامترهای معمول در مدل‌های LSTM شامل مواردی همچون تعداد لایه‌ها و نورون‌ها، نرخ یادگیری[1]، اندازه دسته‌ای[2]، تعداد دوره‌های آموزش[3] و نرخ حذف تصادفی[4] برای جلوگیری از بیش‌برازش، می‌باشند.

روش‌های سنتی برای تنظیم این هایپرپارامترها، مانند جستجوی شبکه‌ای[5] یا جستجوی تصادفی[6]، اغلب ناکارآمد و زمان‌بر هستند، به‌ویژه زمانی که فضای جستجو بسیار بزرگ است. در پاسخ به این چالش، الگوریتم‌های بهینه‌سازی خودکار هایپرپارامترها توسعه یافته‌اند. Optuna یک فریم‌ورک متن‌باز و مدرن برای بهینه‌سازی هایپرپارامترها است که با استفاده از الگوریتم‌های هوشمند و پویا، بهترین ترکیب از پارامترها را پیدا می‌کند [۸،۳۲]. این امر باعث می‌شود که فرآیند بهینه‌سازی سریع‌تر و کارآمدتر انجام شود[۸]. انتخاب Optuna به جای الگوریتم‌های فرابتکاری سنتی (مانند الگوریتم ژنتیک یا ازدحام ذرات) به دلیل کارایی بالاتر آن در فضاهای جستجوی پیچیده است. Optuna از بهینه‌سازی بیزی مبتنی بر مدل جانشین[7] استفاده می‌کند که به آن اجازه می‌دهد تا با هر آزمایش، درک بهتری از تابع هدف پیدا کرده و نواحی امیدوارکننده‌تر را به صورت هوشمندانه کاوش کند. این رویکرد، برخلاف جستجوی عمدتاً تصادفی در روش‌های فرابتکاری، تعداد آزمایش‌های مورد نیاز برای رسیدن به یک راه‌حل بهینه را به شدت کاهش می‌دهد و فرآیند بهینه‌سازی را سریع‌تر و کارآمدتر می‌سازد [۸، ۳۲]. مراحل کار با Optuna به صورت خلاصه به شرح زیر است:

۱- تعریف تابع هدف: تابعی که هایپرپارامترها را به عنوان ورودی دریافت و معیار ارزیابی را به عنوان خروجی بازمی‌گرداند. هدف Optuna کمینه‌سازی یا بیشینه‌سازی این مقدار است.

۲- ایجاد مطالعه: یک شیء Study در Optuna ایجاد می‌شود که فرآیند بهینه‌سازی را مدیریت می‌کند.

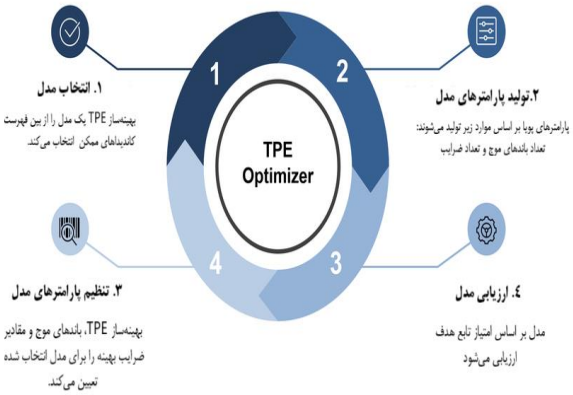
۳- بهینه‌سازی: Optuna به صورت تکراری تابع هدف را با ترکیبات مختلفی از هایپرپارامترها اجرا می‌کند و با استفاده از الگوریتم‌های نمونه‌برداری هوشمند خود، بهترین ترکیب بعدی را پیشنهاد می‌دهد.

۴- تحلیل نتایج: پس از اتمام بهینه‌سازی، Optuna بهترین مجموعه هایپرپارامترها را به همراه بهترین عملکرد مدل ارائه می‌دهد.

Optuna با استفاده از الگوریتم درخت پارزن[7] (TPE) به عنوان مدل جانشین، فضای جستجو را کاوش می‌کند. در هر مرحله، Optuna بهترین نقطه بعدی را برای جستجو با توجه به تعادل بین اکتشاف[8] (جستجوی نواحی جدید و ناشناخته) و بهره‌برداری[9] (تمرکز بر نواحی امیدوارکننده) انتخاب می‌کند. این رویکرد هوشمندانه باعث می‌شود که Optuna نسبت به روش‌های سنتی، سریع‌تر و کارآمدتر عمل کرده و نتایج بهتری را در بهینه‌سازی مدل‌های پیچیده مانند LSTM ارائه دهد.

این الگوریتم، داده‌های تریال‌های قبلی را به دو گروه خوب (دارای عملکرد بالا) و بد (دارای عملکرد پایین) تقسیم کرده و دو مدل چگالی احتمالی[10] را برای هر گروه تخمین می‌زند. این مدل‌ها به ترتیب با $l(x)$ برای گروه خوب و $g(x)$ برای گروه بد نشان داده می‌شوند. در هر مرحله، TPE به دنبال یافتن هایپرپارامترهای جدید x است که تابع اکتساب (E.I.) که در رابطه ۵ ارائه شده است را به حداکثر برسانند. این تابع، تعادل بین اکتشاف و بهره‌برداری را برقرار می‌کند [۳۲].

[1] Learning Rate
[2] Batch Size
[3] Epochs
[4] Dropout
[5] Grid Search
[6] Random Search
[7] Surrogate Model
[8] Exploration
[9] Exploitation
[10] Probability Density Function



شکل (۲): شماتیک عملکرد الگوریتم درخت پارزن (TPE) در بهینه‌سازی هایپرپارامترها با استفاده از Optuna [۳۳]

$$E.I.(x) = \frac{g(x)}{l(x)} \qquad (۵)$$

با استفاده از Optuna ، مدل LSTM قادر خواهد بود به صورت خودکار و بهینه، بهترین پیکربندی را برای پیش‌بینی مصرف انرژی در صنعت فولاد پیدا کند، که این امر به بهبود قابل توجهی در دقت و کارایی مدل منجر خواهد شد [۷۸].

همان‌طور که در شکل ۲ نشان داده شده است، الگوریتم بهینه‌ساز TPE یک فرآیند چرخشی را برای یافتن بهترین هایپرپارامترها دنبال می‌کند. این فرآیند از چهار گام اصلی تشکیل شده است که در گام اول، انتخاب مدل که در آن بهینه‌ساز از میان کاندیداهای ممکن، یک مدل را انتخاب می‌کند. گام دوم تولید پارامترهای مدل که در آن پارامترهای پویا بر اساس توزیع‌های احتمالی قبلی تولید می‌شوند. گام سوم، تنظیم پارامترهای مدل که در آن الگوریتم به صورت هوشمندانه بهترین مقادیر را برای پارامترهای مدل انتخاب می‌کند. در نهایت، ارزیابی مدل انجام می‌شود که در آن مدل بر اساس امتیاز توابع هدف که در اینجا R2 و MAPE است، ارزیابی می‌گردد. این چرخه تا زمانی که بهترین پیکربندی ممکن یافت شود یا تعداد تریال‌ها به پایان برسد، تکرار می‌شود.

۳-۳ معیارهای ارزیابی عملکرد مدل‌ها

۳-۳-۱ ضریب تعیین[1] (R^2)

برای ارزیابی دقت مدل‌های پیش‌بینی، از معیار آماری استاندارد استفاده می‌شود. این معیار میزان برازش مدل به داده‌های واقعی را اندازه‌گیری می‌کنند و امکان مقایسه عادلانه بین مدل‌ها را فراهم می‌کند.R2 یک معیار آماری است که در رابطه شماره ٦ نمایش داده شده است و نسبت واریانس در متغیر وابسته را که از متغیرهای مستقل قابل پیش‌بینی است، نشان می‌دهد. مقدار آن از ۰ تا ۱۰۰٪ (یا ۰ تا ۱) متغیر است، که مقدار بالاتر نشان‌دهنده برازش بهتر مدل به داده‌ها است [۳۳].

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad (٦)$$

که در آن SSresمجموع مربع‌های باقیمانده و SStotمجموع کل مربع‌ها است.

برای ارزیابی دقیق عملکرد مدل‌های پیش‌بینی، از معیارهای ارزیابی کمی استاندارد استفاده می‌شود. این معیارها امکان مقایسه عینی مدل‌ها و سنجش دقت پیش‌بینی آن‌ها را فراهم می‌آورند [۱۲،۳٤].

در این پژوهش، چهار معیار اصلی مورد استفاده قرار گرفته است.

۳-۳-۲ میانگین مربعات خطا[2]

میانگین مربع تفاوت بین مقادیر واقعی Y_iو پیش‌بینی شده $\hat{Y}_i$است که به خطاهای بزرگتر وزن بیشتری می‌دهد و طبق رابطه شماره ۷ محاسبه می‌شود.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \qquad (۷)$$

۳-۳-۳ میانگین خطای مطلق[3]

MAE میانگین تفاوت‌های مطلق بین مقادیر پیش‌بینی شده و واقعی را محاسبه می‌کند و طبق رابطه ۸ محاسبه می‌شود. این معیار خطای پیش‌بینی را در واحدهای داده اصلی ارائه می‌دهد و نسبت به داده‌های پرت کمتر حساس است.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | Yi - \hat{Y}_i | \qquad (۸)$$

۳-۳-٤ ریشه میانگین مربع خطا[4]

RMSE ریشه دوم میانگین مربع خطاهای پیش‌بینی را محاسبه می‌کند. این معیار به خطاهای بزرگ‌تر وزن بیشتری می‌دهد و یک

[1] Coefficient of Determination
[2] Mean Squared Error
[3] Mean Absolute Error
[4] Root Mean Squared Error

دیدگاه جامع از دقت مدل ارائه می‌کند و در رابطه شماره ۹ نمایش داده شده است.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{Y}_t)^2} \quad (۹)$$

۳-۳-۵ میانگین خطای مطلق درصدی[1]

MAPE میانگین خطای مطلق را به صورت درصدی از مقادیر واقعی بیان می‌کند. این معیار امکان مقایسه عملکرد مدل را در میان مجموعه‌های داده مختلف با واحدهای متفاوت فراهم می‌کند و به راحتی قابل تفسیر است و با رابطه شماره ۱۰ محاسبه می‌شود.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \quad (۱۰)$$

۳-٤ اعتبارسنجی مدل

برای ارزیابی عملکرد یک مدل یادگیری ماشین و اطمینان از قابلیت تعمیم‌پذیری آن به داده‌های دیده نشده، از روش‌های اعتبارسنجی متقاطع استفاده می‌شود. این روش‌ها با تقسیم مجموعه داده به زیرمجموعه‌های آموزش و اعتبارسنجی، عملکرد مدل را به صورت پایدار و قابل اعتماد می‌سنجند. رایج‌ترین نوع این تکنیک، اعتبارسنجی متقاطع K-Fold است.

در روش K-Fold، مجموعه داده D به K زیرمجموعه مساوی و تصادفی (F1,F2,...,FK) تقسیم می‌شود، به طوری که هیچ زیرمجموعه‌ای شامل داده‌های قبل یا بعد از زیرمجموعه دیگر نیست. در هر تکرار (k∈{1,...,K})، مجموعه آموزشی (Dtrain,k) و اعتبارسنجی (Dvalid,k) به صورت رابطه‌های ۱۱ و ۱۲ تعریف می‌شود:

$$D_{valid,k} = F_k \quad (۱۱)$$

$$D_{train,k} = D / F_k \quad (۱۲)$$

با این حال، استفاده از K-Fold برای داده‌های سری زمانی کاملاً نامناسب است؛ زیرا ترتیب زمانی داده‌ها را نادیده گرفته و به داده‌های آینده اجازه می‌دهد تا به صورت تصادفی وارد مجموعه آموزش شوند. این پدیده که به عنوان نشت داده[2] شناخته می‌شود، منجر به ارزیابی بیش از حد خوش‌بینانه از عملکرد مدل می‌گردد.

به همین دلیل، در این پژوهش از روش تخصصی اعتبارسنجی متقاطع گام به گام[3] که در کتابخانه‌های پایتون تحت عنوان TimeSeriesSplit پیاده‌سازی شده، استفاده گردید. این روش با حفظ کامل توالی زمانی داده‌ها، یک چارچوب واقع‌گرایانه برای ارزیابی مدل فراهم می‌کند. در روش TimeSeriesSplit، داده‌ها به صورت پیوسته و بدون تداخل زمانی تقسیم می‌شوند. اگر مجموعه داده کامل را به صورت D={x1,x2,...,xN} در نظر بگیریم، در هر تکرار (k∈{1,...,K})، مجموعه آموزشی و اعتبارسنجی به صورت رابطه‌های ۱۳ و ۱٤ تعریف می‌شود:

$$D_{train,k} = \{x_1, x_2, \dots, x_{t_k}\} \quad (۱۳)$$

$$D_{valid,k} = \{x_{t_{k+1}}, x_{t_{k+2}}, \dots, x_{t_{k+N}}\} \quad (۱٤)$$

که در آن tkنقطه پایانی مجموعه آموزشی در تکرار k ام است. این رویکرد تضمین می‌کند که مدل تنها بر روی داده‌های گذشته آموزش دیده و بر روی داده‌های آینده ارزیابی می‌شود. این روش، ارزیابی پایداری مدل را در طول زمان امکان‌پذیر می‌سازد و برای بهینه‌سازی هایپرپارامترها با ابزارهایی مانند Optuna بسیار ضروری است تا از انتخاب پارامترهایی که به صورت تصادفی روی یک برش زمانی خاص عملکرد خوبی داشته‌اند، جلوگیری شود و بهترین مدل به صورت پایدار پیدا شود [۳۵]که این امر اعتبار و قابلیت تعمیم‌پذیری نتایج را در شرایط عملیاتی به‌شدت افزایش می‌دهد.

٤- روش‌شناسی پژوهش

٤-۱ محدوده پژوهش و رویکرد داده‌کاوی

پژوهش حاضر بر پیش‌بینی مصرف انرژی در کارخانه احیای مستقیم فولاد بافت در استان کرمان، شهرستان بافت، تمرکز دارد. این کارخانه با ظرفیت تولید ۸۰۰ هزار تن آهن اسفنجی در سال، از فناوری پرد[4] استفاده می‌کند. این تحقیق از منظر هدف، کاربردی و توسعه‌ای محسوب می‌شود. از منظر زمان جمع‌آوری داده‌ها، پیمایشی است. از بعد ماهیت داده‌ها، کمی بوده و نهایتاً در روش جمع‌آوری داده، از نوع میدانی به شمار می‌رود.

[1] Mean Absolute Percentage Error
[2] Data Leakage
[3] Walk-Forward Cross-Validation
[4] PERED

٤-٢ داده‌ها و منبع آن

پژوهش حاضر بر اساس داده‌های ۵ ساله مصرف انرژی برق کارخانه فولاد بافت صورت گرفته است. این داده‌ها به صورت روزانه از تاریخ ۱ فروردین ۱۳۹۸ تا ۲۹ اسفند ۱۴۰۲ (معادل ۲۱ مارس ۲۰۱۹ تا ۱۹ مارس ۲۰۲٤) در بازه زمانی پنج سال گذشته جمع‌آوری شده‌اند و در مجموع شامل ۱۸۲۷ رکورد می‌باشند. متغیرهای مورد استفاده در این تحلیل شامل مصرف برق به عنوان متغیر هدف و متغیرهای تولید فولاد، ساعات توقف، میانگین، حداقل و حداکثر دما، فشار هوا، رطوبت، تعطیلات رسمی ایران، ویژگی روز هفته به عنوان متغیر برون‌زا در نظر گرفته شده‌اند. داده‌های مربوط به دما، فشار و رطوبت، که به عنوان متغیرهای برون‌زا در مدل‌سازی تأثیرگذار هستند، از سایت‌های جهانی معتبر هواشناسی و اقلیم‌شناسی احصا شده و به مجموعه داده اصلی اضافه گردیده‌اند. این رویکرد داده‌محور، امکان بررسی جامع تأثیر عوامل مختلف بر مصرف انرژی را فراهم می‌کند.

٤-٣ پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها گامی حیاتی در مدل‌سازی سری‌های زمانی، به ویژه در کاربردهای صنعتی، محسوب می‌شود، چرا که کیفیت داده‌های ورودی تأثیر مستقیمی بر دقت و قابلیت اطمینان مدل‌های پیش‌بینی دارد. در این پژوهش، مراحل پیش‌پردازش داده‌ها به شرح زیر انجام شده است:

۱- تبدیل تاریخ‌ها: ابتدا تاریخ‌های شمسی به میلادی تبدیل و به عنوان ایندکس زمانی دیتافریم تنظیم شده‌اند.

۲- رسیدگی به مقادیر گمشده[1]: مقادیر گمشده در مجموعه داده با استفاده از روش میانگین‌گیری[2] پر شده‌اند.

۳- ایجاد ویژگی‌های جدید[3]: برای مدل‌سازی بهتر تأثیر عوامل زمانی، ویژگی‌های روز هفته (مانند دوشنبه، سه‌شنبه و غیره) با استفاده از روش One-Hot Encoding ایجاد شده‌اند [۱۱].

٤- اعمال تبدیل لگاریتمی و نرمال‌سازی: برای کاهش نوسانات شدید و پایدارسازی سری زمانی، تبدیل لگاریتمی بر روی متغیر هدف (مصرف برق) اعمال شد. سپس، تمامی متغیرها با استفاده از روش نرمال‌سازیMin-Max به مقیاسی بین ۰ و ۱ درآورده شدند تا عملکرد مدل‌های یادگیری عمیق بهبود یابد.

۵- تقسیم داده‌ها[4]: داده‌ها به دو مجموعه آموزش[5]، و آزمون[6] تقسیم شده‌اند و در نهایت ارزیابی نهایی عملکرد مدل بر روی داده‌های دیده نشده (مجموعه آزمون) را فراهم می‌آورد [۱۸].

٤-٤ ابزارهای نرم‌افزاری

پیاده‌سازی این پژوهش با استفاده از زبان برنامه‌نویسی پایتون و ابزارهای متنوعی صورت گرفته است. برای مدیریت داده‌ها و عملیات پیش‌پردازش از کتابخانه Pandas و برای محاسبات عددی از NumPy استفاده شد. مدل‌سازی شبکه‌های عصبی عمیق با استفاده از فریم‌ورک TensorFlow/Keras انجام گردید. همچنین، بهینه‌سازی هایپرپارامترها با استفاده از کتابخانه Optuna و بصری‌سازی نتایج با Matplotlib و Seaborn صورت پذیرفت.

٤-٥ مدل LSTM و تنظیمات آن

در این پژوهش مدل LSTM یک شبکه عصبی متوالی با سه لایه LSTM با ۵۰ نرون[7] در هر لایه است و سه لایه Dropout با نرخ ۰,۲ برای جلوگیری از بیش‌برازش[8] قرار داده شده است. لایه خروجی یک لایه Dense با ۱ واحد برای پیش‌بینی مقدار مصرف انرژی است. این مدل با بهینه‌ساز 'adam' و تابع هزینه 'mean_squared_error'کامپایل شده و برای ۱۵۰ دوره آموزش دیده است. ازEarly Stopping با صبر ۱۰ دوره برای جلوگیری از بیش‌برازش و بازیابی بهترین وزن‌ها استفاده شده است. در جدول ۱ به تشریح تمامی تنظیمات بیان شده است.

[1] Missing Values Imputation
[2] Mean Imputation
[3] Feature Engineering
[4] Data Splitting
[5] Training Set
[6] Test Set
[7] neurons
[8] Overfitting

جدول (۱): تنظیمات مدل شبکه عصبی عمیق LSTM و مدل بهینه Optuna

شاخص	مقدار
بخش داده و مدل‌سازی	
تعداد کل رکوردها	۱۸۲۷
تقسیم داده	۸۰٪ (۱۴۶۱ رکورد) برای آموزش ۲۰٪ (۳۶۶ رکورد) برای آزمون
مدل‌سازی	
نوع مدل	شبکه عصبی LSTM
تابع هزینه	MSE
الگوریتم بهینه‌سازی	Adam
تعداد دوره‌های آموزش	۱۵۰
اندازه دسته	۳۲
بخش بهینه‌سازی(Optuna)	
نوع بهینه‌سازی	بهینه‌سازی دو هدفه
تعداد اهداف	۱. حداکثرسازی ضریب تعیین (R^2) ۲. حداقل‌سازی درصد خطای مطلق میانگین
تعداد تریال‌ها	۵۰
الگوریتم نمونه‌برداری	TPE

ابعاد ورودی این لایه با لایه‌های قبلی متفاوت است. این نشان می‌دهد که لایه آخر LSTM (lstm_14)فقط آخرین خروجی را به لایهDense فرستاده است.

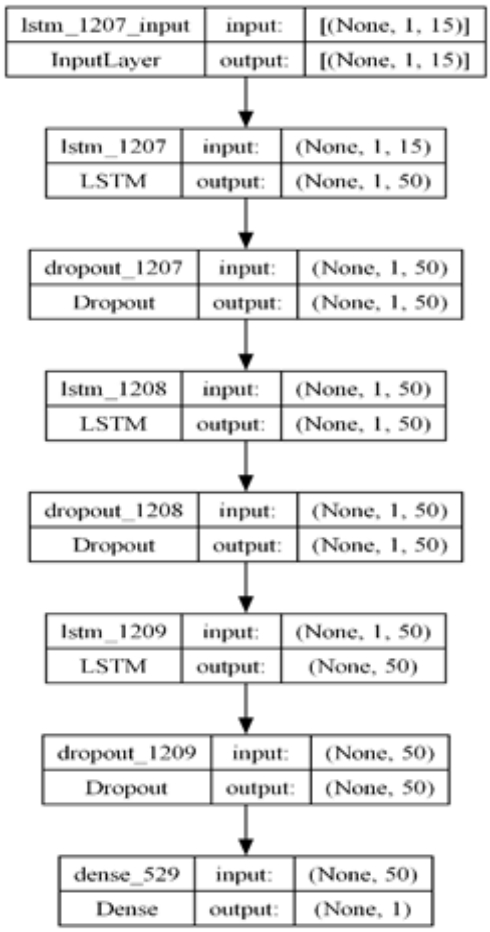
۵- تحلیل نتایج

در این بخش، نتایج حاصل از ارزیابی دو مدل LSTM پایه و LSTM بهینه‌سازی‌شده با Optuna ، براساس معیارهای کمی و بصری تحلیل می‌شوند. مقایسه جامع عملکرد مدل‌ها در دو مرحله پیش از بهینه‌سازی و پس از آن، ارزش رویکرد پیشنهادی را به وضوح نشان می‌دهد.

۵-۱ تحلیل بهینه‌سازی هایپرپارامترها و نمودارهای Loss

مقایسه هایپرپارامترهای به‌دست‌آمده از Optuna با هایپرپارامترهای مدل پایه نشان می‌دهد که بهینه‌ساز تغییراتی استراتژیک و هدفمند را اعمال کرده است. مقادیر مدل پایه و مدل بهینه‌سازی شده در جدول ۲ نمایش داده شده است.

همان‌طور که در جدول مشاهده می‌شود، الگوریتم Optuna با وجود حفظ تعداد لایه‌ها و واحدهای لایه اول، تغییرات کلیدی را در هایپرپارامترهای دیگر ایجاد کرده که به بهبود قابل توجهی در عملکرد مدل منجر شده است. مهم‌ترین تغییری که Optuna ایجاد کرده، افزایش نرخ یادگیری از ۰/۰۰۱ به ۰/۰۰۲۳۷ است. این امر به مدل اجازه داده است تا وزن‌ها را با سرعت بیشتری به‌روز کند و به یک نقطه بهینه در فضای پارامترها دست یابد. تعداد واحدهای لایه سوم را از ۵۰ به ۶۴ افزایش داده و هم‌زمان، نرخ Dropout را در لایه‌های میانی و انتهایی کاهش داده است.



شکل (۳): معماری و اندازه تنسورهای ورودی و خروجی هر لایه شبکه عصبی

این تغییرات نشان می‌دهد که بهینه‌ساز دریافته است که مدل به ظرفیت یادگیری بیشتری در لایه‌های عمیق‌تر نیاز دارد و با کاهش نرخ Dropout، به مدل اجازه داده است تا وابستگی‌های پیچیده‌تر را به طور کامل‌تری یاد بگیرد، بدون اینکه دچار بیش‌برازش شود. این مقایسه به وضوح نشان می‌دهد که بهینه‌سازی خودکار، به جای

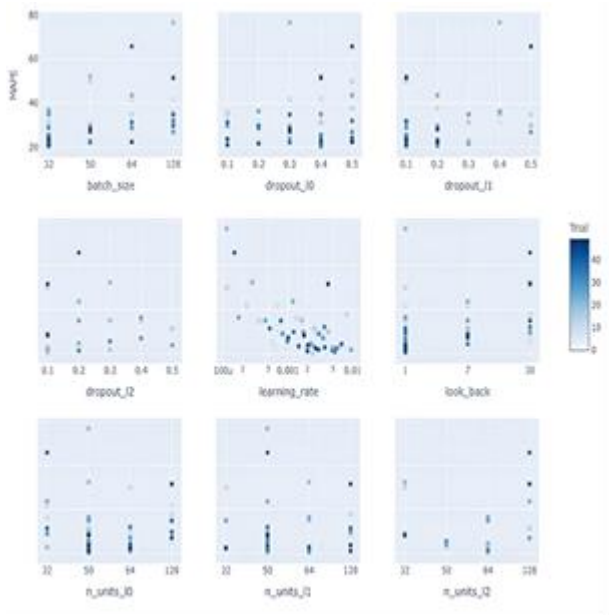
تغییرات کوچک و تصادفی، تغییراتی استراتژیک و هدفمند را در هایپرپارامترها اعمال کرده است. در ادامه نمودار Slice Plot در شکل ٤ به تفکیک هر مورد نمایش داده شده است و تأثیر هر یک از هایپرپارامترها را بر روی معیار درصد خطای مطلق میانگین (MAPE)به صورت جداگانه نمایش می‌دهد.

جدول (۲): مقایسه هایپرپارامترها در مدل شبکه عصبی عمیق LSTM و مدل بهینه Optuna

هایپرپارامتر	مدل پایه	مدل بهینه‌شده باOptuna	تحلیل تغییرات
تعداد لایه‌ها(n_layers)	۳	۳	بدون تغییر
تعداد واحدها در لایه اول (n_units_l0)	۵۰	۵۰	بدون تغییر
نرخ Dropout در لایه اول(dropout_l0)	۰/۲	۰/٤	افزایش یافته
تعداد واحدها در لایه دوم (n_units_l1)	۵۰	۵۰	بدون تغییر
نرخ Dropout در لایه دوم(dropout_l1)	۰/۲	۰/۱	کاهش یافته
تعداد واحدها در لایه سوم (n_units_l2)	۵۰	٦٤	افزایش یافته
نرخ Dropout در لایه سوم(dropout_l2)	۰/۲	۰/۱	کاهش یافته
نرخ یادگیری	۰/۰۰۱	۰/۰۰۲۳۷	افزایش یافته
اندازه دسته	۳۲	۳۲	بدون تغییر

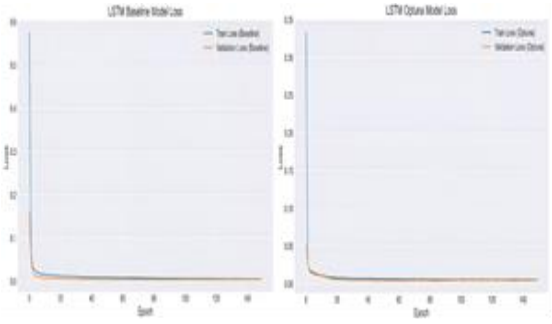
هر نقطه در این نمودار نشان‌دهنده یک تریال بهینه‌سازی است و رنگ نقاط، بیانگر ترتیب زمانی تریال‌ها است. نمودار نرخ یادگیری به وضوح نشان می‌دهد که مقادیر پایین MAPE عمدتاً در یک بازه مشخص از نرخ یادگیری (تقریباً ۰/۰۰۱ تا ۰/۰۰۵) متمرکز شده‌اند. این امر تأیید می‌کند که نرخ یادگیری یکی از حیاتی‌ترین هایپرپارامترها برای دستیابی به دقت بالا در این مدل بوده است. طول پنجره زمانی بهترین نتایج (پایین‌ترین MAPE) به طور مشخص در پنجره زمانی برابر با ۱ به دست آمده‌اند.

این یافته نشان می‌دهد که برای این مجموعه داده خاص، نگاه به یک گام زمانی گذشته، برای پیش‌بینی مقدار بعدی کافی بوده و استفاده از پنجره‌های طولانی‌تر منجر به بهبود عملکرد نشده است. در مجموع، این نمودار به صورت بصری تأیید می‌کند که Optuna به جای جستجوی تصادفی، به صورت هوشمندانه فضای پارامترها را کاوش کرده و توانسته است به ناحیه‌ای از هایپرپارامترها با عملکرد بهینه همگرا شود.



شکل (٤): نمودار Slice Plot و نمایش تأثیر هایپرپارامترها بر معیار درصد خطای مطلق میانگین (MAPE)

در ادامه در بررسی شکل ۵، نمودار روند کاهش خطای مدل (Loss) را در طول فرآیند آموزش برای هر دو مدل نشان می‌دهند.

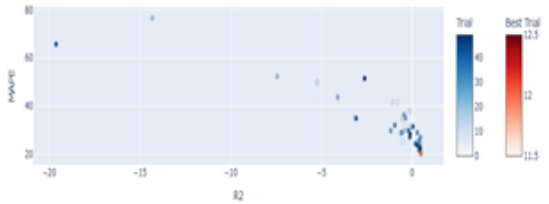


شکل (۵): نمودار تابع هزینه (Loss) مدل پایه در طول فرآیند آموزش

هر دو نمودار همگرایی مناسبی را نشان می‌دهند، اما مدل بهینه‌سازی‌شده از همان ابتدای آموزش، با یک کاهش سریع‌تر و پایدارتر در خطا مواجه شده و به یک سطح خطای بسیار پایین‌تر همگرا می‌شود. این امر نشان می‌دهد که هایپرپارامترهای بهینه شده،

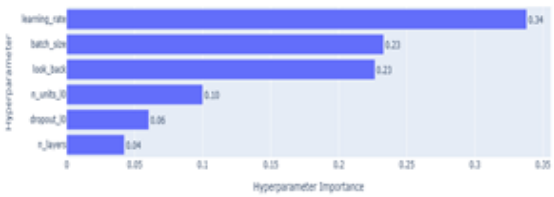
فرآیند یادگیری را کارآمدتر کرده و از نوسانات غیرضروری جلوگیری کرده‌اند.

در تحلیل فرآیند بهینه‌سازی نمودارهای مرز پارتو در شکل ٦، اهمیت هایپرپارامترها و تاریخچه بهینه‌سازی مورد بررسی قرار گرفته است. در نمودار مرز پارتو توازن بین دو هدف بهینه‌سازی (حداکثرسازی R2 و حداقل‌سازی MAPE) را نشان می‌دهد. بهترین تریال‌ها نمایش داده شده که نشان‌دهنده پیدا شدن یک نقطه تعادل بهینه بین دقت و خطای پیش‌بینی است.



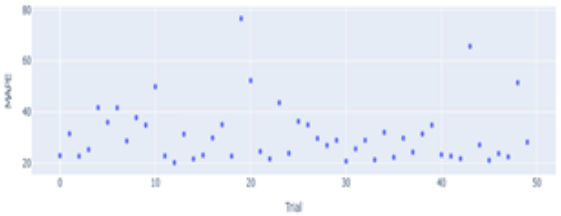
شکل (٦): نمودار مرز پارتو در ایجاد توازن بین دو هدف بهینه‌سازی

نمودار اهمیت هایپرپارامترها[1] تأثیر هر هایپرپارامتر را بر دقت مدل در شکل ۷ نشان می‌دهند. هایپرپارامترهای learning_rate (نرخ یادگیری) و look_back (تعداد گام‌های زمانی گذشته) بیشترین تأثیر را بر روی هر دو معیار R2 و MAPE داشته‌اند. این یافته اهمیت تنظیم دقیق این دو پارامتر را در مدل‌سازی سری‌های زمانی با LSTM برجسته می‌کند.



شکل (۷): نمودار اهمیت هایپرپارامترها و بررسی تأثیر هر هایپرپارامتر بر دقت مدل

نمودار تاریخچه بهینه‌سازی[2] در شکل ۸ روند تغییرات MAPE را در طول ۵۰ تریال بهینه‌سازی نشان می‌دهند. با پیشروی تریال‌ها، الگوریتم Optuna به سمت نواحی با MAPE پایین‌تر حرکت کرده، که نشان‌دهنده کارایی الگوریتم در کاوش هوشمندانه فضای پارامترها است.



شکل (۸): نمودار تاریخچه بهینه‌سازی و بررسی روند تغییرات MAPE در طول ۵۰ تریال

۵-۲ تحلیل عملکرد مدل‌های LSTM و LSTM-Optuna براساس معیارهای کمی

نتایج اعتبارسنجی اولیه مدل پایه با استفاده از روش TimeSeriesSplit، میانگین عملکرد آن را با R2 برابر ۰/٤۸ و MAPEبرابر ۱۹/۵۸٪ نشان داد. این نتایج به عنوان معیاری برای شروع فرآیند بهینه‌سازی در نظر گرفته شدند. پس از آموزش مدل‌ها بر روی کل مجموعه داده آموزش و ارزیابی نهایی در مجموعه آزمون، نتایج به شرح جدول ۳ به دست آمد.

جدول (۳): مقایسه معیارهای ارزیابی در مدل شبکه عصبی عمیق LSTM و مدل بهینه Optuna

معیار ارزیابی	مدل پایه (Baseline)	مدل بهینه‌سازی‌شده با Optuna	بهبود
ضریب تعیین (R^2)	۰/۸۵	۰/۸۹	٪٤/۷۰
درصد خطای مطلق میانگین (MAPE)	٪۲۰/٤٤	٪۱۸/٤٦	٪۹/۷۰
ریشه میانگین مربعات خطا (RMSE)	۹۳۲/۵۵ ,۳۵	۸۲٤/٦۵ ,۳۰	٪۱٤/۱۰
میانگین مربعات خطا (MSE)	۱, ۲۹۱ ,۱٤۷ ۹۵۵/۹۷	,۱۵۸ ,۹۵۰ ۷۵۵/۹۱	٪۲٦/٤۱
میانگین خطای مطلق (MAE)	۸۸۵/۰۲ ,۲٤	۸۵٦/٤۰ ,۲۱	٪۱۲/۱۷

[1] Hyperparameter Importances

[2] Optimization History

همان‌طور که در جدول ۳ مشاهده می‌شود، مدل LSTM-Optuna در تمامی معیارهای ارزیابی به طور قابل توجهی بهتر از مدل پایه عمل کرده است. افزایش ضریب تعیین از ۰/۸۵ به ۰/۸۹ و کاهش ۱/۹۸ درصدی در MAPE نشان‌دهنده توانایی تبیینی بالاتر و دقت بیشتر مدل بهینه‌سازی‌شده است.

در نهایت با مقایسه بصری پیش‌بینی‌ها در شکل ۹، مقایسه بصری عملکرد دو مدل را بر روی مجموعه آزمون نهایی نشان می‌دهد. همان‌طور که در نمودار مشاهده می‌شود، پیش‌بینی‌های مدل بهینه‌سازی‌شده با Optuna (خط قرمز) به طور چشمگیری انطباق نزدیک‌تری با مقادیر واقعی (خط آبی) دارند. به خصوص در دوره‌هایی که نوسانات شدید یا افت ناگهانی در مصرف انرژی وجود دارد، مدل بهینه‌سازی‌شده دقت بالاتری در دنبال کردن الگوهای واقعی نشان می‌دهد. این مقایسه بصری، برتری مدل ترکیبی را در مواجهه با داده‌های پیچیده و غیرخطی صنعتی تأیید می‌کند.

شکل (۹): مقایسه بصری پیش‌بینی مدل شبکه عصبی عمیق LSTM و مدل بهینه Optuna

۵-۳ مقایسه عملکرد با مدل‌های دیگر

برای ارزیابی جامع و اثبات برتری مدل پیشنهادی، عملکرد آن با چندین مدل شناخته‌شده دیگر در حوزه پیش‌بینی سری زمانی مقایسه گردید. این مدل‌ها شامل شبکه عصبی بازگشتی ساده، ماشین بردار پشتیبان ساده و بهینه‌شده بودند. تمامی مدل‌ها بر روی یک مجموعه داده یکسان آموزش دیده و ارزیابی شدند. نتایج کمی این مقایسه در جدول ۵ ارائه شده است.

جدول (٤): مقایسه عملکرد مدل پیشنهادی با سایر مدل‌های پیش‌بینی

معیار ارزیابی	RNN	SVR	SVR بهینه شده	LSTM	LSTM-Optuna
ضریب تعیین (R^2)	۰/۸	۰/٤٤	۰/۸۵	۰/۸۵	۰/۸۹
درصد خطای مطلق میانگین (MAPE)	٪۲۱/۰۸	٤۳/۰۲	٪۲٤/٤۹	٪۲۰/٤٤	٪۱۷/٦٤
ریشه میانگین مربعات خطا (RMSE)	۸٦٦/٤٤/۳۳	٥۷۰/۲۹۰/۲	۲۱٤/۹۸/۳۸	۹۳۲/٥٥/۳٥	۸۲٤/٦٥/٤۷

نتایج جدول ٤ به وضوح نشان می‌دهد که مدل پیشنهادی LSTM-Optuna در تمامی معیارهای کلیدی عملکرد بهتری نسبت به سایر مدل‌ها داشته است. مدل RNN به دلیل ساختار حافظه‌دار خود عملکردی منطقی از خود نشان داد، اما نتوانست با نسخه پیشرفته‌تر خود یعنی LSTM رقابت کند. مدل SVR بهینه شده نیز اگرچه توانست بخشی از الگوها را یاد بگیرد، اما در مواجهه با نوسانات شدید و وابستگی‌های زمانی پیچیده داده‌های صنعتی، با محدودیت مواجه شد. این مقایسه جامع، ارزش ترکیب معماری قدرتمند LSTM با یک فرآیند بهینه‌سازی هوشمند هایپرپارامترها را برای دستیابی به بالاترین سطح از دقت در پیش‌بینی‌های صنعتی تأیید می‌کند.

٦- نتیجه‌گیری

پژوهش حاضر، با موفقیت یک چارچوب ترکیبی داده‌محور بر پایه LSTM و بهینه‌سازی چندهفته هدفه با Optuna را برای پیش‌بینی دقیق مصرف انرژی در صنعت فولاد معرفی و ارزیابی کرد. نتایج به وضوح برتری قابل توجه این مدل را نسبت به مدل پایه نشان داد و تأکید کرد که تنها استفاده از شبکه‌های عصبی عمیق کافی نیست و بهینه‌سازی هایپرپارامترها، عاملی حیاتی در رسیدن به عملکرد بهینه است.

با توجه به نتایج حاصل، در حالی که محققانی مانند ژنگ و همکاران [۱۵]و پی و همکاران[۱٦]ارزش مدل‌های LSTM را در

پیش‌بینی سری‌های زمانی اثبات کردند، پژوهش حاضر نشان داد که ترکیب این مدل با یک رویکرد بهینه‌سازی هوشمند، دقت را به سطحی جدید ارتقا می‌دهد و به R2 برابر ۰/۸۹ و MAPE با مقدار ۱۸/۴۶٪ می‌رسد. نتایج ارزیابی نشان داد که مدل بهینه‌سازی‌شده LSTM-Optuna ، از مدل پایه LSTM و همچنین مدل‌های دیگری مانند RNN، SVR و SVR بهینه شده عملکرد بهتری دارد. پژوهش حاضر، با تأیید یافته‌های هی و همکاران[۷]و گائو و همکاران[۸]در خصوص اهمیت بهینه‌سازی هایپرپارامترها، این رویکرد را با استفاده از فریم‌ورک مدرن Optuna در یک کاربرد صنعتی خاص به کار گرفت. مقایسه هایپرپارامترهای بهینه‌شده با مدل پایه نشان داد که Optuna به جای تغییرات کوچک، تغییراتی استراتژیک در نرخ یادگیری و ظرفیت مدل (تعداد واحدها و نرخ Dropout) ایجاد کرده که منجر به همگرایی سریع‌تر و دقت بالاتر شده است. این تحقیق نشان داد که گنجاندن متغیرهای برون‌زای صنعتی و اقلیمی، به همراه داده‌های مربوط به قطعی‌های برنامه‌ریزی‌نشده، برای مدل‌سازی دقیق نوسانات شدید مصرف انرژی در صنعت فولاد ضروری است. این امر تأییدکننده یافته‌های پاشایی و دهخوارقانی[۳]در مورد اهمیت عوامل مؤثر بر مصرف انرژی در صنایع پرمصرف است.

تحلیل مقایسه‌ای در چشم‌انداز پژوهش‌های اخیر برای ارزیابی دقیق‌تر جایگاه این دستاورد، نتایج پژوهش حاضر با یافته‌های جدیدترین مقالات در حوزه پیش‌بینی مصرف انرژی در جدول شماره ۵ مقایسه گردید. پژوهش‌های اخیر مانند یودین و همکاران (۲۰۲۵) و قرایی و همکاران (۲۰۲۵)، با به‌کارگیری مدل‌های LSTM بهینه‌سازی‌شده بر روی داده‌های استاندارد مصرف برق، به دقت‌های بسیار بالایی (به ترتیب R^2=۰/۹۴ و R^2=۰/۹۹) دست یافته‌اند [۲۷،۲۶]. هرچند در نگاه اول، این ضرایب تبیین قدرت کامل مدل پیش‌بینی را به نمایش می‌گذارد، اما تفاوت ماهوی در پیچیدگی و ماهیت مجموعه داده‌ها، نقطه کلیدی تحلیل است. مقالات مذکور بر روی داده‌های عمومی یا خانگی کار کرده‌اند که الگوهای فصلی و روزانه مشخصی دارند. در مقابل، قدرت اصلی مدل حاضر در عملکرد موفق بر روی یک مجموعه داده واقعی، صنعتی و بسیار پرچالش نهفته است. داده‌های مورد استفاده در این پژوهش، شامل نوسانات شدید ناشی از فرآیندهای تولید و به‌ویژه قطعی‌های برق برنامه‌ریزی‌نشده بود که مدل‌سازی را به مراتب دشوارتر می‌کند.

جدول (۵): مقایسه تحلیلی نتایج پژوهش با مطالعات برجسته اخیر

پژوهش	مدل اصلی	حوزه داده	R^2 (آزمون)	MAPE (آزمون)	نکات کلیدی تحلیل
پژوهش حاضر	LSTM + Optuna (دو هدفه)	صنعت فولاد (با قطعی برق)	۰/۸۹	٪۱۸/۴۶	عملکرد قوی و پایدار بر روی داده‌های صنعتی بسیار پرنویز
Uddin et al. (2025)	LSTM + HPO	مصرف برق عمومی	۰/۹۴	٪۲/۶۳	دقت بالا بر روی داده‌های استاندارد و کم‌نویز
Gharai et al. (2025)	LSTM + Optuna	مصرف برق خانگی	۰/۹۹	گزارش نشده	دقت بسیار بالا بر روی داده‌های استاندارد و قابل پیش‌بینی
Zhu et al. (2025)	AutoML	صنعت نفت و گاز	۰/۲۲	گزارش نشده	نمونه‌ای از چالش بیش‌برازش در داده صنعتی

اهمیت این موضوع زمانی برجسته‌تر می‌شود که نتایج پژوهش ژو و همکاران (۲۰۲۵) را در نظر بگیریم. ایشان با استفاده از یک چارچوب پیشرفته یادگیری ماشین خودکار بر روی داده‌های صنعتی نفت و گاز، با مشکل جدی بیش‌برازش مواجه شدند، به طوری که R^2 مدل از ۰/۷۹ بر روی داده‌های آموزش به ۰/۲۳ بر روی داده‌های آزمون سقوط کرد. در مقابل، مدل پیشنهادی ما با دستیابی به R^2 پایدار ۰/۸۹ بر روی داده‌های آزمون، نشان‌دهنده استحکام۱ و قدرت تعمیم‌پذیری۲ بالای آن در شرایط واقعی و

[1] Robustness

[2] Generalization

پرنویز صنعتی است [٦]. علاوه بر این، رویکرد بهینه‌سازی در این پژوهش با در نظر گرفتن هم‌زمان دو هدف (کاهش MAPE و افزایش R^2)، به دنبال یافتن یک مدل متوازن و جامع بود؛ رویکردی که مشابه آن در پژوهش بیجو و پیلای (۲۰۲۳) برای یافتن توازن بهینه بین «دقت» و «تفسیرپذیری» به کار رفته و نشان‌دهنده همسویی روش‌شناسی ما با جدیدترین روندهای علمی است[۳٦].

در نهایت، این پژوهش با ارائه یک مدل دقیق و قابل اعتماد، به مدیران انرژی در صنعت فولاد کمک می‌کند تا یک استراتژی مدیریت انرژی فعال اتخاذ کنند. پیش‌بینی‌های دقیق، امکان برنامه‌ریزی بهینه تولید، مذاکره هوشمندانه‌تر در خصوص قراردادهای تأمین انرژی و جلوگیری از جریمه‌های ناشی از تجاوز از پیک مصرف را فراهم می‌آورد. همچنین، این مدل به مدیران کمک می‌کند تا تأثیر عوامل مختلف بر مصرف انرژی را بهتر درک کرده و برای حفظ پایداری عملیاتی در برابر نوسانات بازار و شرایط محیطی، آمادگی داشته باشند.

علیرغم دستاوردهای مهم این پژوهش، محدودیت‌هایی وجود دارد که می‌تواند به عنوان فرصت‌هایی ارزشمند برای تحقیقات آینده در نظر گرفته شود:

۱- مدل‌های ترکیبی پیشرفته‌تر: پژوهش‌های آتی می‌توانند رویکردی فراتر از ترکیب ساده LSTM با بهینه‌سازی را دنبال کنند. به عنوان مثال، کاوش در مدل‌های ترکیبی مبتنی بر تجزیه سیگنال مانند VMD-LSTM یا ترکیب مدل‌های Encoder-Decoder با مکانیزم توجه[1] می‌تواند به بهبود دقت پیش‌بینی، به ویژه در الگوهای پیچیده‌تر، منجر شود.

۲- بهینه‌سازی پویا و خودکار: در این مطالعه، از Optuna برای بهینه‌سازی پارامترهای اصلی مدل استفاده شد. تحقیقات آینده می‌توانند بهینه‌سازی‌های پیشرفته‌تری را بررسی کنند که به صورت خودکار معماری مدل را نیز تعیین می‌کنند یا از الگوریتم‌های بهینه‌سازی دیگری مانند Hyperopt برای مقایسه عملکرد استفاده نمایند.

۳- افزایش تفسیرپذیری: مدل‌های یادگیری عمیق اغلب به عنوان "جعبه سیاه "شناخته می‌شوند. برای افزایش اعتماد مدیران به نتایج، پژوهش‌های آتی می‌توانند از روش‌های تفسیرپذیری مدل‌های هوش مصنوعی (XAI) مانند SHAP یا LIME استفاده کنند تا دلایل پشت هر پیش‌بینی، به صورت شفاف مشخص شود.

٤- پیش‌بینی چند گام جلوتر و عدم قطعیت: این مطالعه بر پیش‌بینی کوتاه‌مدت متمرکز بود. توسعه چارچوب‌های مدل‌سازی که به طور هم‌زمان پیش‌بینی چند گام جلوتر را با دقت بالا و همراه با تخمین عدم قطعیت ارائه می‌دهند، می‌تواند کاربردهای عملیاتی گسترده‌تری داشته باشد.

سپاسگزاری

این پژوهش با حمایت مالی ارزشمند دانشگاه یزد و شرکت توسعه فناوری و نوآوری فولاد مبارکه اصفهان به سرانجام رسید. نویسندگان از همکاری و پشتیبانی این نهادها در قالب گرنت پارسا در اجرای پروژه پایان‌نامه دکتری با محوریت حل یک چالش کلیدی صنعتی، صمیمانه سپاسگزاری می‌کنند.

References

[1] S. Ghasaee and R. Ravanmehr, “Short-Term Prediction of Electrical Load Consumption Using Deep Neural Networks, CNN, and LSTM,” J. Qual. Product. Iran Electr. Ind., vol. 10, no. 1, pp. 35–51, 2021 [In Persian].

[2] G. Zhang and M. Qi, “Neural Network Forecasting for Seasonal and Trend Time Series,” Eur. J. Oper. Res., vol. 160, no. 2, pp. 501-514, 2005, doi: 10.1016/j.ejor.2003.08.037.

[3] H. Pashaei and M. Dehkharghani, “Predictive Modeling of Energy Consumption in the Steel Industry Using CatBoost Regression: A Data-Driven Approach for Sustainable Energy Management,” J. Clean. Prod., vol. 401, p. 136706, 2023, doi: 10.1016/j.jclepro.2023.136706.

[4] A.Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network,” Physica D: Nonlinear Phenomena, vol. 404, p. 132306, 2020, doi: 10.1016/j.physd.2020.132306.

[5] S. Arslan, “A Hybrid Forecasting Model Using LSTM and Prophet for Energy Consumption with

[1] Attention

Decomposition of Time Series Data," PeerJ Comput. Sci., vol. 8, p. e1001, 2022, doi: 10.7717/peerj-cs.1001.
[6] R. Zhu, N. Li, Y. Duan, G. Li, G. Liu, F. Qu, et al., "Well-Production Forecasting Using Machine Learning with Feature Selection and Automatic Hyperparameter Optimization," Energies, vol. 18, no. 1, p. 99, dic. 2024, doi: 10.3390/en18010099.
[7] F. He, J. Zhou, Z.-k. Feng, G. Liu, and Y. Yang, "A Hybrid Short-Term Load Forecasting Model Based on Variational Mode Decomposition and Long Short-Term Memory Networks, Considering Relevant Factors with a Bayesian Optimization Algorithm," Appl. Energy, vol. 237, pp. 108–117, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.055.
[8] T. Gao, Y. Sun, and C. Li, "LSTM Network Hyperparameter Optimization for Stock Price Prediction Using the Optuna Framework," J. Phys.: Conf. Ser., vol. 1785, no. 1, p. 012061, 2021, doi: 10.1088/1742-6596/1785/1/012061.
[9] D. Bunn and E. Farmer, "Review of Short-Term Forecasting Methods in the Electric Power Industry," in Comparative Models for Electrical Load Forecasting, D. Bunn and E. Farmer, Eds. Berlin: Springer, 1985, pp. 13-30.
[10] Moghram and S. Rahman, "Analysis and Evaluation of Five Short-Term Load Forecasting Techniques," IEEE Trans. Power Syst., vol. 4, no. 4, pp. 1484-1491, Nov. 1989.
[11] K. Amasyali and N. M. El-Gohary, "A Review of Data-Driven Building Energy Consumption Prediction Studies," Renew. Sustain. Energy Rev., vol. 81, pp. 1192-1205, Jan. 2018.
[12] M. H. L. Lee, Y. C. Ser, G. Selvachandran, P. H. Thong, L. Cuong, L. H. Son, et al., "A Comparative Study of Forecasting Electricity Consumption Using Machine Learning Models," Mathematics, vol. 10, no. 1329, pp. 1-23, 2022.
[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2017.
[14] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
[15] Y. Zheng, J. Zhang, and Y. Wang, "Short-Term Load Forecasting Based on a Long Short-Term Memory Neural Network," J. Mod. Power Syst. Clean Energy, vol. 5, no. 2, pp. 296-302, Mar. 2017.
[16] S. Pei, H. Qin, L. Yao, Y. Liu, C. Wang, and J. Zhou, "Multi-Step Ahead Short-Term Load Forecasting Using Hybrid Feature Selection and Improved Long Short-Term Memory Network," Energies, vol. 13, no. 16, p. 4121, Aug. 2020, doi: 10.3390/en13164121.
[17] M. Zohaki Raht and H. Sadeghi Saghadol, "Modeling and Forecasting of Iran's Short-Term Electricity Consumption Using Neural Network and TPE Algorithm," Journal of Energy Economics Studies, vol. 20, no. 83, pp. 159-182, 2024 [In Persian].
[18] J. Zhu, Y. Wang, C. Lai, and X. Zhou, "Deep Learning-Based Energy Consumption Prediction Model for Green Industrial Parks," Appl. Artif. Intell., vol. 39, no. 1, p. 2462375, 2021.
[19] L. Shen, W. Chen, and J. Kwok, "Deep Learning for Short-Term Energy Consumption Forecasting of Smart Buildings Considering Microgrid and External Factors," Mathematics, vol. 10, no. 1329, pp. 1-23, 2022.
[20] S. Sayadinejad, A. Esmailzadeh Maghari, and M. R. Rostami, "Presenting a Bitcoin Return Prediction Model Using a Hybrid Deep Learning Signal Decomposition Algorithm (CEEMD-DL)," Quarterly Journal of Financial Economics, vol. 17, no. 62, pp. 217-238, 2022 [In Persian].
[21] J. Kim, H. Kim, H. Kim, D. Lee, and S. Yoon, "A Comprehensive Survey of Deep Learning for Time Series Forecasting: Architectural Diversity and Open Challenges," Artif. Intell. Rev., vol. 58, p. 216, Apr. 2025.
[22] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, et al., "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis," Proc. R. Soc. Lond. A, vol. 454, no. 1971, pp. 903-995, Mar. 1998.
[23] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531-544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.
[24] Y. Ruan, G. Wang, H. Meng, and F. Qian, "A Hybrid Model for Power Consumption Forecasting Using VMD-Based Long Short-Term Memory Neural Network," Front. Energy Res., vol. 9, p. 772508, Feb. 2022, doi: 10.3389/fenrg.2021.772508.
[25] I. O. Ekundayo, "Optuna optimization-based CNN-LSTM model for predicting electric power energy consumption," M.S. thesis, School of Comput., Nat. College of Ireland, Dublin, 2020.
[26] C. Gharai, S. K. Mohapatra, S. Parida, C. Dora, R. K. Mohanta, and S. Chakravarty, "Optimized Deep Learning Model for Power Consumption Prediction Using Hyperparameter Tuning Techniques," in Int. Conf. Intell. Cloud Comput.

(ICoICC), Bhubaneswar, India, 2025, pp. 1-6, doi: 10.1109/ICoICC64033.2025.11052111.
[27] R. Uddin, M. R. Hazari, S. Ahmad, C. A. Hossain, M. S. Rahman Zishan, and A. Ahmed, "Short-Term Load Forecasting Using Deep Learning Algorithms with Hyperparameter Optimization," in 4th Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST), Dhaka, Bangladesh, 2025, pp. 366-371, doi: 10.1109/ICREST63960.2025.10914446.
[28] S. Bouktif, R. F'Haim, and M. Lazaar, "A Novel Hybrid Model for Short-Term Load Forecasting Using LSTM and Metaheuristic Optimization Algorithms," Energy, vol. 196, p. 117042, Apr. 2020.
[29] J. Brownlee, Deep Learning for Time Series Forecasting: Theory to Practice. San Juan, PR: Machine Learning Mastery, 2018.
[30] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.
[31] J. Zhao, G. Nie, and Y. Wen, "Monthly Precipitation Prediction in Luoyang City Based on EEMD-LSTM-ARIMA Model," Water Sci. Technol., vol. 87, no. 1, pp. 318–335, Jan. 2023, doi: 10.2166/wst.2022.425.
[32] Y. Ren, Z. Yan, D. Liu, J. Hou, and W. Zhou, "Optimized EWT-Seq2Seq-LSTM with Attention Mechanism for Insulator Fault Prediction," Energies, vol. 15, no. 2, p. 528, Jan. 2022, doi: 10.3390/en15020528.
[33] A. Alghamdi and A. Desuqi, "Predicting Electricity Consumption Using Machine Learning Techniques," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 11, pp. 384-391, 2020.
[34] M. Alimohammadi Ardakani, M. H. Karimi-Zarchi, and D. Shishebori, "A Hybrid Forecasting Model for Accurate Prediction of Building Energy Consumption Based on Multi-Stage Decomposition and Optimized Deep Learning," Energy, vol. 305, p. 126955, 2024.
[35] R. J. Hyndman and G. Athanasopoulos. (2021). Forecasting: Principles and Practice (3rd ed.) [Online]. Available: https://otexts.com/fpp3/
[36] B. G. M. and G. N. Pillai, "Hyperparameter Optimization of Long Short Term Memory Models for Interpretable Electrical Fault Classification," IEEE Access, vol. 11, pp. 123688-123704, Nov. 2023, doi: 10.1109/ACCESS.2023.3330056.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 3, No. 2, Autumn & Winter 2025, pp. 89 - 105

A Hybrid Model for Data-Driven Energy Consumption Forecasting in the Steel Industry: An Approach Based on LSTM and Hyperparameter Optimization with the Optuna Algorithm

Leila Zolqadr[1], Majid Shakhsi-Niaei [2*], Yahia Zare Mehrjerdid [3], Mohammad Mehdi Lotfi[3]

[1]Ph.D. Student in Industrial Engineering, Department of Industrial Engineering, Yazd University, Yazd, Iran

[2]Associate Professor, Department of Industrial Engineering, Yazd University, Yazd, Iran

[3]Professor, Department of Industrial Engineering, Yazd University, Yazd, Iran

Article Information

Original Research Paper

Received:
2025 August 22

Accepted:
2025 October 20

Keywords:
Data-Driven Modeling, STM-Optuna, Walk-Forward Validation, Electrical Energy Consumption, Steel Industry, RNN, SVM

Corresponding Author*:

m.niaei@yazd.ac.ir

Abstract

In the modern world, electrical energy, as one of the most fundamental needs of societies, plays a vital role in the industrial, economic, and social sectors. Given the impossibility of large-scale storage and fluctuations in supply and demand, accurate forecasting of electrical load consumption is essential. Traditional time-series models often struggle to handle the nonlinear and complex patterns of industrial data. This research presents a novel hybrid framework for accurately forecasting electrical energy consumption time series. The proposed approach is based on a Long Short-Term Memory (LSTM) neural network; to maximize the model's performance, the LSTM network's hyperparameters have been optimized using the Optuna automatic optimization algorithm. The data used includes time series of electricity consumption from the Baft steel factory, along with exogenous variables such as steel production volume, production downtime, and weather and calendar conditions. The evaluation results indicate that the hybrid LSTM-Optuna model, with a coefficient of determination (R^2) of 0.89 and a Mean Absolute Percentage Error (MAPE) of 18.46%, not only demonstrates better performance than the baseline model (with an R^2 of 0.85 and a MAPE of 20.44%) but also outperforms other machine learning methods such as Recurrent Neural Networks (RNN) and both simple and optimized Support Vector Machines (SVM).

: 10.22034/ABMIR.2025.23569.1157

E-ISSN: 2821-2037