

## پیش‌بینی کوتاه مدت بار الکتریکی از طریق یادگیری ویژگی‌های بار با استفاده از شبکه‌های عمیق کانولوشن و بازگشتی

حسین اسکندری<sup>۱</sup>، مریم ایمانی\*<sup>۲</sup>، محسن پارسا مقدم<sup>۳</sup>

<sup>۱</sup> کارشناسی ارشد مهندسی برق مخابرات، دانشگاه تربیت مدرس

<sup>۲\*</sup> استادیار گروه مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، maryam.imani@modares.ac.ir

<sup>۳</sup> استاد گروه قدرت، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس

### چکیده

پیش‌بینی بار الکتریکی مصرفی برای عملکرد مطمئن سیستم‌های قدرت و همچنین برای برنامه‌ریزی مدیریت تقاضا و ذخیره بار حیاتی است. در این مقاله پیش‌بینی بار با استخراج ویژگی‌های تاریخچه بار مصرفی با استفاده از شبکه‌های عمیق انجام شده است. شبکه‌های عصبی بازگشتی، به ویژه شبکه‌های عصبی بازگشتی بهبود یافته مانند LSTM و GRU قادر به نگه‌داشتن حافظه کوتاه مدت و طولانی هستند تا روابط بین مقادیر بار را از سری زمانی استخراج کنند. از طرف دیگر، شبکه‌های عصبی کانولوشنی قادر به یادگیری خودکار ویژگی‌ها هستند و می‌توانند مستقیماً یک بردار را برای پیش‌بینی تولید کنند. روش پیشنهادی این مقاله، استخراج ویژگی‌های بار با استفاده از شبکه‌های عصبی کانولوشنی و سپس استخراج اطلاعات توالی زمانی بار با بهره‌گیری از لایه‌های بازگشتی GRU است. نتایج آزمایش‌ها بر روی سه مجموعه داده ISO-NE و Toronto و North American Utility نشان دهنده کاهش خطای پیش‌بینی روش پیشنهادی در مقایسه با سایر روش‌های رقیب است.

### کلیدواژه

پیش‌بینی بار الکتریکی، یادگیری عمیق، شبکه‌های عصبی کانولوشنی، شبکه‌های بازگشتی بهبود یافته، خطای پیش‌بینی.

### مقدمه

عصبی مصنوعی<sup>۱</sup> است، زیرا آن‌ها می‌توانند روابط پیچیده و غیر خطی بین بار و عوامل خارجی را مدل کنند [۲]. در مواردی که رفتار بار الکتریکی بسیار پیچیده می‌شود، دقت پیش‌بینی در شبکه‌های عصبی مصنوعی سطحی رضایت بخش نیست. بنابراین از روش یادگیری عمیق<sup>۲</sup> استفاده می‌شود. اخیراً برخی از روش‌های یادگیری عمیق مانند شبکه‌های باور عمیق<sup>۳</sup> (DBNs)، شبکه‌های عصبی بازگشتی<sup>۴</sup> (RNNs) و شبکه‌های عصبی کانولوشنی<sup>۵</sup> (CNNs) را برای بهبود دقت پیش‌بینی بسیار کوتاه مدت و کوتاه مدت از طریق استخراج و تبدیل ویژگی‌ها مورد استفاده قرار داده‌اند [۳]-[۵].

با اینکه شبکه‌های عصبی بازگشتی و به ویژه LSTM<sup>۶</sup> و GRU<sup>۷</sup> قادر به نگه‌داشتن حافظه کوتاه مدت و طولانی هستند با این حال، RNNها در هنگام پردازش مجموعه‌های بار با گستردگی زمانی زیاد، به دلیل افزایش غیر خطی پیچیدگی-های آموزشی و از بین رفتن گرادیان که در LSTM وجود دارد،

پیش‌بینی بار الکتریکی<sup>۱</sup> (ELF) برای عملکرد و برنامه‌ریزی یک شرکت برق ضروری است. به طور کلی، روش‌های پیش‌بینی بار الکتریکی را می‌توان به چهار دسته از دیدگاه زمانی تقسیم کرد: پیش‌بینی بار بسیار کوتاه مدت<sup>۲</sup> (VSTLF)، پیش‌بینی بار کوتاه مدت<sup>۳</sup> (STLF)، پیش‌بینی بار میان مدت<sup>۴</sup> (MTLF) و پیش-بینی بار درازمدت<sup>۵</sup> (LTLF). افق قطع زمان برای این چهار دسته به ترتیب یک روز، دو هفته و سه سال است [۱]. بار الکتریکی مصرفی تحت تاثیر بسیاری از عوامل خارجی مانند آب و هوا، روزهای هفته، قیمت برق و غیره است. پیش‌بینی بار با استفاده از این عوامل زمانیکه آن‌ها به صورت دقیق پیش‌بینی نشده‌اند منجر به افزایش خطای پیش‌بینی می‌شود. بنابراین، مطالعه و استفاده از روش‌های پیش‌بینی که به این عوامل بیرونی بستگی ندارند اهمیت دو چندان پیدا می‌کند. بیشتر روش‌های پیش‌بینی بار الکتریکی موجود، براساس شبکه‌های

<sup>۱</sup> artificial neural networks  
<sup>۲</sup> Deep learning  
<sup>۳</sup> Deep Belief Networks  
<sup>۴</sup> Recurrent Neural Networks  
<sup>۵</sup> Convolutional Neural Networks  
<sup>۶</sup> Long Short Term Memory  
<sup>۷</sup> Gated Recurrent Unit

<sup>۱</sup> Electrical Load Forecasting  
<sup>۲</sup> very short-term load forecasting  
<sup>۳</sup> short-term load forecasting  
<sup>۴</sup> medium-term load forecasting  
<sup>۵</sup> long-term load forecasting

دستیابی به اطلاعات از نقاط همسایه (نمونه‌های زمانی قبلی و بعدی) از نقطه بار در نظر گرفته معرفی شده است. بازنمایی مشترک بردار ویژگی مرتبط با هر نقطه بار، شامل اطلاعات محلی ارزشمند در مورد نقاط بار مجاور است. لی و همکاران منحنی‌های بار را با استفاده از یک مدل شبکه عصبی کانولوشنی عمیق به چند دسته تقسیم کردند [۸]. یک شبکه عصبی مصنوعی با سه لایه پنهان برای پیش‌بینی بار کوتاه مدت، با توجه به عوامل خارجی دیگر، به عنوان مثال دما، رطوبت، سرعت باد و غیره مورد استفاده قرار گرفت. همچنین، پانگ و همکاران [۹] یک پیش‌بینی کننده مبتنی بر شبکه عصبی را برای پیش‌بینی بار بسیار کوتاه مدت پیشنهاد داده اند که تنها مقادیر بار مراحل فعلی و قبلی را به عنوان ورودی می‌گیرد تا مقدار بار را در مرحله زمان آینده پیش‌بینی کند. Amarasinghe و همکاران [۱۰] رویکردهای شبکه‌های عصبی عمیق، از جمله شبکه‌های عصبی کانولوشنی، LSTM و شبکه‌های باور عمیق را برای پیش‌بینی بار بسیار کوتاه مدت به کار بردند. در [۱۱] chen و همکاران با بهره‌گیری از یک شبکه عصبی residual عمیق و با استفاده از بار الکتریکی و دمای هفته‌های قبل، به پیش‌بینی بار در ۲۴ ساعت آینده پرداخته‌اند. همچنین در [۱۲] ترکیبی از LSTM و رگرسیون بردار پشتیبان برای پیش‌بینی بار الکتریکی کوتاه مدت ارائه شده است روش پیشنهادی در [۱۲] به این صورت است که از ویژگی‌های استخراج شده توسط LSTM به عنوان ورودی رگرسیون بردار پشتیبان برای پیش‌بینی استفاده شده است. در مقاله [۱۳] توالی بار الکتریکی ابتدا به تصاویر بار دو بُعدی تبدیل شده است و سپس از شبکه‌های عصبی کانولوشنی برای پیش‌بینی بار استفاده شده است. از آنجایی که سری زمانی بار الکتریکی دارای ذات غیر ایستاد است و ممکن است اطلاعات ارزشمندی در توالی بار الکتریکی موجود باشد که در حوزه زمان، باز نیست بنابراین در مقاله [۱۴] فضای ویژگی جدیدی از توالی بار الکتریکی تشکیل شده است. حوزه ارائه شده شامل اطلاعات تکمیلی در مورد شکل و تغییرات توالی بار الکتریکی است. ویژگی‌های بار به دست آمده با مقادیر اصلی بار الکتریکی در دامنه زمانی یکپارچه می‌شود تا یک منبع غنی برای پیش‌بینی کننده فراهم شود. در نهایت از یک تکنیک یادگیری عمیق از خانواده شبکه‌های عصبی بازگشتی، به نام LSTM، برای یادگیری تقاضای بار الکتریکی از ویژگی‌های ارائه شده در حوزه‌های تکی و هیبریدی استفاده می‌شود. دامنه‌های

موثر نیستند. به همین منظور در [۲] روشی تحت عنوان C-LSTM پیشنهاد شده است که یک ترتیب جدید کوتاه‌تر از سری بار اولیه را بازسازی می‌کند. این روش محدودیت‌های LSTM را در گام‌های زمان طولانی برطرف می‌کند، در حالی که بیشترین اطلاعات اولیه را حفظ می‌نماید [۲].

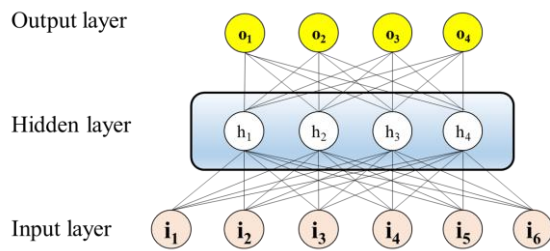
انواع مختلفی از مدل‌های کانولوشنی وجود دارد که می‌توانند برای هر نوع مسئله خاص پیش‌بینی سری زمانی استفاده شوند. اگرچه شبکه‌های کانولوشنی برای داده‌های تصویر دو بُعدی توسعه یافته‌اند، اما می‌توان از این شبکه‌ها در مسائل مربوط به پیش‌بینی سری‌های زمانی نیز استفاده کرد. شبکه عصبی کانولوشنی یک بُعدی نوعی از شبکه‌های کانولوشنی است که بر روی توالی یک بُعدی اعمال می‌شود. استفاده از لایه‌های کانولوشنی ما قبل لایه‌های بازگشتی برای به دست آوردن الگوهای زمانی و محلی می‌تواند در موارد خاص بسیار مفید واقع شود. ادامه مقاله به شرح زیر تدوین شده است. در بخش دوم به مروری بر کارهای گذشته در زمینه پیش‌بینی بار می‌پردازیم. در بخش سوم شبکه‌های عصبی به کار رفته در روش پیشنهادی این مقاله شرح داده شده و سپس در بخش چهارم نتایج شبیه‌سازی بحث گردیده است. نتیجه‌گیری نیز در بخش پنجم آمده است.

### مروری بر کارهای گذشته

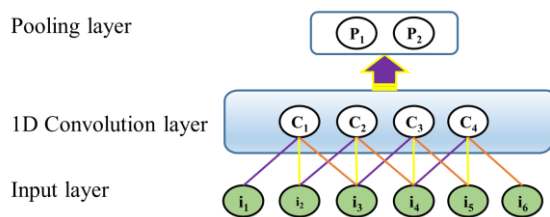
در زمینه پیش‌بینی بار کوتاه مدت، پژوهش‌های زیادی انجام شده است. کائو و همکاران [۶] مدل پیش‌بینی ARIMA<sup>۱۳</sup> و روش روز مشابه را برای پیش‌بینی بار درون روز اتخاذ کردند. مکانیسم روش روز مشابه آن‌ها این است که روز هدف را با روزهای مشابه هواشناسی در تاریخ گروه‌بندی کرده و بار را بر اساس تقاضای متوسط آن روزها پیش‌بینی کنند. در این کار نشان داده شده است که در روزهای عادی، ARIMA بهتر عمل می‌کند در حالی که روش روز مشابه در روزهای غیرعادی عملکرد بهتری از خود نشان می‌دهد. در [۷] از تبدیل موجک و تبدیل بازنمایی مشترک<sup>۱۴</sup> برای پیش‌بینی بار الکتریکی در سطح خانگی استفاده شده است. در مقاله [۷]، یک بردار متغیر بار تاخیر یافته به هر نقطه از منحنی بار اختصاص داده می‌شود. سپس برای حذف جزئیات<sup>۱۵</sup> اضافی و استفاده از مولفه تقریب<sup>۱۶</sup> بردار ویژگی، تجزیه موجک روی آن اعمال می‌شود. علاوه بر این، یک نسخه جدید از بازنمایی مشترک برای

<sup>۱۳</sup> Integrated Moving Average Autoregressive  
<sup>۱۴</sup> Collaborative Representation  
<sup>۱۵</sup> details  
<sup>۱۶</sup> approximate

کاهش یافته است. فرایند pooling یک روش مهم برای استخراج ویژگی‌های کانولوشنی مهم است [۱۶].



شکل ۱. پرسپترون چندلایه



شکل ۲. کانولوشن یک بُعدی

### شبکه LSTM

شبکه‌های عصبی بازگشتی با حافظه کوتاه مدت طولانی (LSTM) به عنوان یک روش قابل اعتماد در مدل‌سازی، پیش‌بینی و تجزیه و تحلیل توالی از داده‌ها محسوب می‌شوند [۱۷]. LSTM معمولاً در حل مسائل در برنامه‌های مربوط به داده‌های متوالی مانند صدا و زبان استفاده می‌شود. LSTM‌ها در به دست آوردن وابستگی‌های زمانی طولانی مدت بدون مواجهه با چالش‌های بهینه‌سازی پیش روی شبکه بازگشتی سنتی، مؤثر هستند [۱۸]. کلید معماری LSTM یک سلول حافظه است که حافظه خود را با گذشت زمان حفظ می‌کند. واحدهای گیت غیر خطی، جریان اطلاعات در سلول را تنظیم می‌کنند. طرزکار گیت، LSTM را قادر می‌سازد تا در مقابله با چالش ناپدید شدن گرادینت<sup>۲۲</sup> (که توسط شبکه‌های بازگشتی سنتی تجربه می‌شود و باعث می‌شود فرایند به روز رسانی وزن‌ها یا اصطلاحاً پس انتشار در طول زمان<sup>۲۳</sup> در شبکه‌های بازگشتی استاندارد به خوبی انجام نشود) موفق عمل کند. ساختار بلوک واحد LSTM، در شکل (۳) این دروازه‌های غیرخطی را نشان می‌دهد. هنگامی که چند لایه LSTM پشت سرهم قرار می‌گیرد، شبکه عصبی بازگشتی LSTM عمیق حاصل می‌شود [۱۹].

فرکانسی، سرم<sup>۱۷</sup>، استروئید طیفی<sup>۱۸</sup>، غلتیدن طیفی<sup>۱۹</sup>، شار طیفی، انرژی، اختلاف زمان، اختلاف فرکانس، گابور و بازنمایی مشترک<sup>۲۰</sup> در این کار مورد بررسی قرار گرفته است.

### استخراج ویژگی‌های بار با یادگیری عمیق

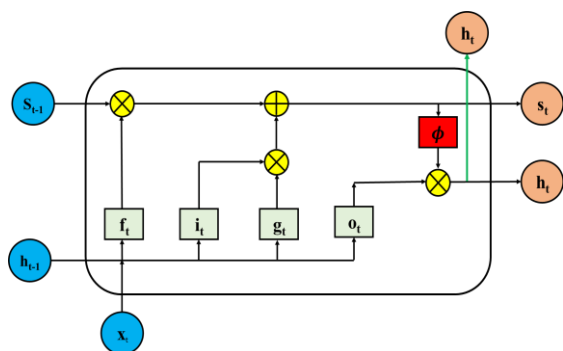
شبکه عصبی مصنوعی با الهام از شبکه عصبی بیولوژیکی، یک سیستم محاسباتی با توانایی بالا در مدل‌کردن است که در بخش یادگیری ماشین بسیار محبوب است. معماری عمومی شبکه عصبی مصنوعی شامل نورون‌ها، وزن‌ها و بایاس است. در این بخش توضیح مختصری در مورد چندین شبکه عصبی مصنوعی از جمله پرسپترون چند لایه، شبکه عصبی کانولوشنی و شبکه‌های بازگشتی LSTM و GRU ارائه شده است.

### شبکه کانولوشن

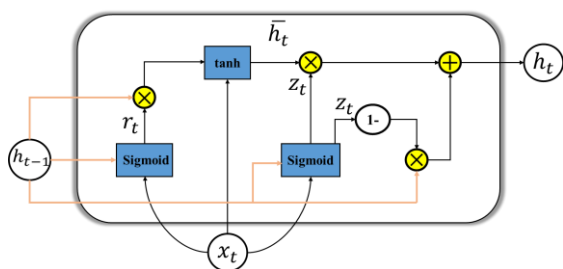
از میان معماری‌های مختلف شبکه عصبی مصنوعی که در مسائل یادگیری ماشین به کار گرفته می‌شوند، پرسپترون چند لایه<sup>۲۱</sup> (MLP) متداول‌ترین مدلی است که مورد استفاده قرار گرفته است [۱۵]. همانطور که در شکل (۱) نشان داده شده است پرسپترون چند لایه در اصل یک معماری شبکه عصبی با لایه‌های تمام متصل است. در پرسپترون چند لایه، تمام نورون‌های لایه قبلی به صورت تمام متصل به نورون‌های لایه بعدی متصل هستند. در شکل (۱)،  $i_1, i_2, \dots, i_6$  نورون‌های لایه ورودی،  $h_1, h_2, h_3, h_4$  نورون‌های لایه پنهان و  $o_1, o_2, o_3, o_4$  نورون‌های لایه خروجی هستند. برخلاف ساختار ساده‌ای که در پرسپترون چند لایه مشاهده می‌شود، استفاده از این معماری نتایج خوبی در بسیاری از کاربردها به همراه دارد. اگرچه پرسپترون چند لایه در مدل‌کردن و شناسایی الگو به خوبی عمل می‌کند اما شبکه‌های عصبی کانولوشنی که از مفهوم به اشتراک گذاری وزن استفاده می‌کنند، دقت بهتری در بسیاری از مسائل غیرخطی همانند پیش‌بینی بار از خود نشان می‌دهند. لایه کانولوشنی یک بُعدی و همچنین لایه pooling در شکل (۲) نشان داده شده است. خطوط با رنگ یکسان، وزن مشترک را نشان می‌دهد و مجموعه‌ای از وزن‌های مشترک به عنوان هسته در نظر گرفته می‌شوند [۱۶]. بعد از فرایند کانولوشن ورودی‌های  $i_1, i_2, \dots, i_6$  به نگاشت ویژگی‌های  $c_1, c_2, c_3, c_4$  تبدیل می‌شوند. مرحله بعد در شکل (۲)، pooling است که در آن نگاشت ویژگی مربوط به لایه کانولوشنی نمونه برداری می‌شود و ابعاد آن کاهش می‌یابد. به عنوان مثال، ۴ بُعد موجود در نگاشت ویژگی در شکل (۲)، بعد از فرایند pooling به ۲ بُعد

<sup>۲۲</sup> vanishing gradient  
<sup>۲۳</sup> backpropagation through time

<sup>۱۷</sup> cepstrum  
<sup>۱۸</sup> spectral centroid  
<sup>۱۹</sup> spectral roll-off  
<sup>۲۰</sup> collaborative representation  
<sup>۲۱</sup> Multilayer Perceptron



شکل ۳. ساختار LSTM



شکل ۴. ساختار GRU

یک سلول GRU از دو دروازه به روزرسانی  $z$  و دروازه بازنشانی  $r$  ساخته شده است. در روابط (۹) تا (۱۲)،  $z_t$  خروجی دروازه به روزرسانی در لحظه  $t$ ،  $r_t$  خروجی دروازه بازنشانی در لحظه  $t$ ، نشان‌دهنده ماتریس وزن‌ها و  $h_t$  حالت را در لحظه  $t$  نشان می‌دهد.

### پیش‌بینی بار الکتریکی با استفاده از مدل پیشنهادی

از آنجایی که بار الکتریکی ساعتی خصوصیات مشابهی اعم از همبستگی بالای مولفه‌های بار الکتریکی و همچنین رفتار پریودیک در روزها و هفته‌های متوالی دارد، بنابراین می‌توان این خصوصیات و ویژگی‌های مشابه در سری زمانی بار الکتریکی را با استفاده از شبکه‌های کانولوشنی استخراج کرد. از طرفی می‌دانیم که شبکه‌های برگشتی بهبود یافته در زمینه استخراج اطلاعات توالی زمانی قدرتمند عمل می‌کنند و کمتر با پدیده محو گرادیان مواجه می‌شوند. بنابراین روش پیشنهادی در این مقاله همانطور که در شکل (۵) نشان داده شده است به این صورت است که ابتدا ویژگی‌های موجود در سری زمانی بار الکتریکی را با استفاده از لایه‌های کانولوشنی استخراج کنیم. سپس اطلاعات موجود در این ویژگی‌های استخراج شده را با استفاده از لایه GRU استخراج می‌نماییم. در روش پیشنهادی

در  $\{x_1, x_2, \dots, x_t\}$  را به عنوان یک دنباله ورودی برای LSTM در نظر بگیریم که  $x_t$  نشان‌دهنده مقدار حقیقی توالی در گام زمانی  $t$  است. LSTM دارای سه دروازه برای کنترل مقدار سلول حالت است. LSTM به ترتیب دروازه فراموشی ( $f_t$ )، دروازه ورودی ( $i_t$ )، گره ورودی ( $g_t$ ) و دروازه خروجی ( $o_t$ ) را با استفاده از روابط (۱) تا (۴) مشخص می‌کند. معادلات (۵) و (۶) به ترتیب سلول حالت حافظه ( $s_t$ ) و حالت ( $h_t$ ) را در گام زمانی  $t$  مشخص می‌کنند.

$$f_t = \sigma(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \quad (۱)$$

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \quad (۲)$$

$$g_t = \phi(w_{gx}x_t + w_{gh}h_{t-1} + b_g) \quad (۳)$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \quad (۴)$$

$$s_t = g_t \odot i_t + s_{t-1} \odot f_t \quad (۵)$$

$$h_t = \phi(s_t) \odot o_t \quad (۶)$$

که  $w_{oh}$ ،  $w_{ox}$ ،  $w_{gh}$ ،  $w_{gx}$ ،  $w_{ih}$ ،  $w_{ix}$ ،  $w_{fh}$ ،  $w_{fx}$  وزن برای ورودی‌های مربوط به توابع فعال ساز شبکه و  $b_g$ ،  $b_i$ ،  $b_f$  و  $b_o$  بایاس‌های مربوط به هر کدام از آن‌ها هستند و  $\odot$  ضرب نقطه‌ای را نشان می‌دهد. توابع  $\sigma$  و  $\phi$  که در روابط (۷) و (۸) نشان داده شده‌اند به ترتیب نمایش توابع سیگموئید و تانژانت هایپربولیک هستند.

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (۷)$$

$$\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (۸)$$

### شبکه GRU

معماری GRU در سال ۲۰۱۴ توسط Cho و همکاران معرفی شد [۲۰]. این معماری که ساختار آن در شکل (۴) نشان داده شده است به منظور برطرف کردن کاستی‌های شبکه عصبی بازگشتی سنتی (RNN) نظیر مشکل محو شدگی گرادیان و همچنین کاهش سربار موجود در معماری LSTM ارائه شده است. GRU مبتنی بر LSTM بهبود یافته است و عموماً به عنوان نسخه‌ای تغییر یافته از LSTM در نظر گرفته می‌شود. در GRU، دروازه فراموشی و دروازه ورودی موجود در LSTM، در یک دروازه به روزرسانی (update gate) ترکیب شده‌اند. کل فرایند را می‌توان با روابط (۹) تا (۱۲) بیان کرد:

$$z_t = \text{Sigmoid}(W_z \times [h_{t-1}, x_t]) \quad (۹)$$

$$r_t = \text{Sigmoid}(W_r \times [h_{t-1}, x_t]) \quad (۱۰)$$

$$\bar{h}_t = \tanh(W \times [r_t \times h_{t-1}, x_t]) \quad (۱۱)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \bar{h}_t \quad (۱۲)$$

<sup>۱۱</sup> Update gate  
<sup>۱۲</sup> Reset gate

باعث بهبود چشمگیر عملکرد شبکه‌های عصبی عمیق شده است [۲۱].

رابطه تابع فعال‌ساز ReLU به صورت زیر است:

$$\text{ReLU}(Z_i) = \max(0, Z_i) \quad (13)$$

که  $Z_i$  فعال‌ساز خطی از گره  $i$ ام یک لایه است. مشکلی که تابع فعال‌ساز ReLU دارد این است که اگر یک واحد با هیچ‌کدام از ورودی‌هایی که در یک پایگاه داده است نتواند فعال شود، الگوریتم بهینه‌سازی مبتنی بر گرادیان نمی‌تواند وزن‌های آن واحد را به روز رسانی کند به همین دلیل آن واحد دیگر هرگز فعال نمی‌شود. علاوه بر این، اگر بخش بزرگی از واحدهای پنهان، گرادیان ثابت صفر تولید کنند، آموزش شبکه بسیار سخت می‌شود [۲۲]. این مشکل با اضافه کردن شیب به نیمه منفی محور ReLU قابل حل است. با یک تغییر ساده در فرمول ReLU در نیمه منفی محور، PReLU را بدست می‌آوریم [۲۳]. رابطه تابع فعال‌ساز PReLU به این صورت است:

$$\text{PReLU}(Z_i) = \begin{cases} Z_i & \text{if } Z_i > 0 \\ \theta_i Z_i & \text{if } Z_i < 0 \end{cases} \quad (14)$$

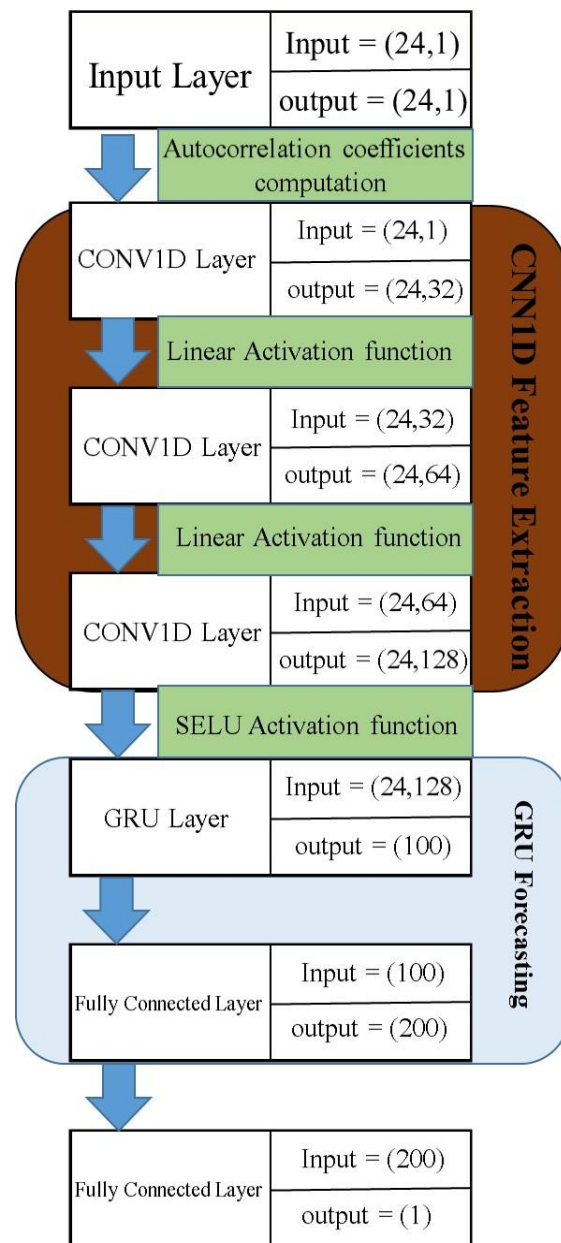
که  $\theta_i$  ضریب کنترل شیب  $\theta_i Z_i$  است برای  $Z_i < 0$ . تغییر دیگری در ReLU که باعث ایجاد خاصیت self normalizing می‌شود در [۲۴] آورده شده است. رابطه تابع فعال‌ساز SELU به صورت زیر می‌باشد:

$$\text{SELU}(Z_i) = \Gamma \begin{cases} Z_i & \text{if } Z_i > 0 \\ \beta e^{Z_i} - \beta & \text{if } Z_i < 0 \end{cases} \quad (15)$$

که  $\Gamma$  و  $\beta$  دو پارامتر قابل تنظیم هستند. در مقاله [۲۴] نشان داده شده است که اگر  $\Gamma \approx 1/0.577$  و  $\beta \approx 1/67.33$  باشد، خروجی یک لایه تمام متصل، زمانیکه ورودی آن از توزیع نرمال پیروی کند، به سمت توزیع نرمال میل می‌کند.

همچنین می‌دانیم که به طور گسترده‌ای در زمینه یادگیری ماشین اذعان می‌شود که یک مجموعه از مدل‌های مختلف از قابلیت تعمیم بالاتری [۲۵] نسبت به مدل‌های تکی برخوردار هستند. در [۲۶]، نویسندگان تجزیه و تحلیل مجموعه‌های شبکه عصبی برای پیش‌بینی کوتاه مدت بار الکتریکی ساختمان‌های اداری ارائه می‌دهند. نتایج نشان می‌دهد که یک مجموعه از شبکه‌های عصبی، واریانس عملکرد را کاهش می‌دهد. استراتژی گروهی (دسته جمعی)<sup>۳۰</sup> مورد استفاده در این مقاله در شکل (۶) نشان داده شده است. همانطور که از شکل (۶) مشخص است در مدل پیشنهادی، ما مدل مورد نظر را ۵

با استفاده از بار الکتریکی ۲۴ ساعت قبل، به دنبال پیش‌بینی بار الکتریکی در ساعت آینده هستیم. همچنین برای انتخاب اندازه فیلتر در لایه‌های کانولوشنی از ضرایب خودهمبستگی بار الکتریکی استفاده کرده‌ایم به این صورت که با محاسبه ضرایب خود همبستگی بار، بارهای ساعتی که بیشترین ضرایب خودهمبستگی را دارند در یک فیلتر (هسته<sup>۳۶</sup>) قرار می‌گیرند.



شکل ۵. ساختار مدل پیشنهادی

در مدل پیشنهادی از تابع فعال‌ساز خطی<sup>۳۷</sup> و تابع فعال‌ساز SELU<sup>۳۸</sup> استفاده شده است. استفاده از تابع فعال‌ساز ReLU<sup>۳۹</sup>

<sup>۳۰</sup> ensemble strategy

<sup>۳۶</sup> Kernel  
<sup>۳۷</sup> linear  
<sup>۳۸</sup> scaled exponential linear units  
<sup>۳۹</sup> Rectified Linear Unit

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|x_t - x_{pt}|}{x_t} \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - x_{pt})^2} \quad (19)$$

که در روابط (۱۶) تا (۱۹)،  $N$  تعداد نمونه‌های آزمایشی،  $x_t$  بار الکتریکی واقعی و  $x_{pt}$  بار الکتریکی پیش‌بینی شده است. به منظور مقایسه دقیق‌تر و منصفانه‌تر، تمام مدل‌های شبکه عصبی با ۵ وزن رندوم اولیه (همانطور که پیش‌تر در توضیحات مدل پیشنهادی گفته شد) آموزش داده شده‌اند و از خروجی مدل که با استفاده از این ۵ وزن رندوم اولیه آموزش داده شده است، میانگین گرفته شده و پیش‌بینی انجام شده است.

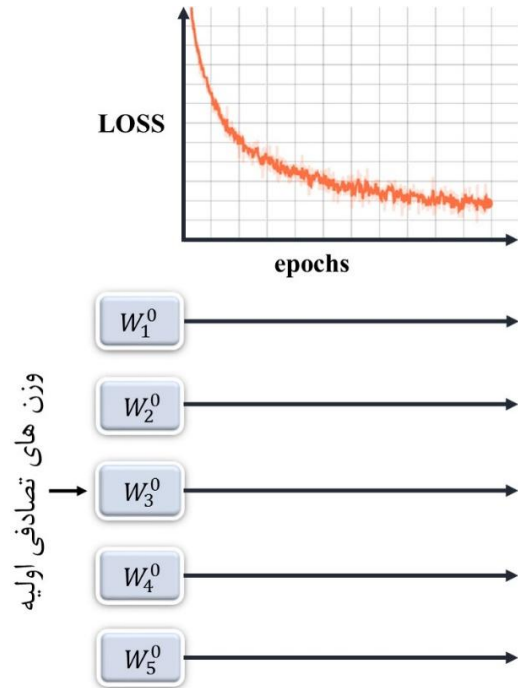
همچنین در تمام مدل‌های طراحی شده توسط نویسندگان از بهینه‌ساز<sup>۳۵</sup> ADAM و تابع خطا<sup>۳۶</sup> میانگین خطای مطلق (MAE) برای آموزش مدل‌ها استفاده شده است. تمام مدل‌ها با استفاده از keras 2.2.4 تحت بستر tensorflow 1.13.1 و CUDA 8.0 و CuDNN 7.0 در محیط python 3.7 اجرا شده‌اند. یک لپ‌تاپ با مشخصات intel@core™ i7-6700 HQ cpus برای آموزش مدل‌ها استفاده شده است.

اولین پایگاه داده مورد استفاده در این مقاله، داده‌های مربوط به بار مصرفی تورنتو [۲۸] در طول هفت ماه از سال ۲۰۱۶ با رزولوشن یک ساعت است که در شکل (۷) به نمایش در آمده است. پایگاه داده دوم، پایگاه داده North American Utility [۲۹] است که شامل بار الکتریکی ساعتی از ژانویه سال ۱۹۸۵ تا ۱۲ اکتبر ۱۹۹۲ است که در شکل (۸) نشان داده شده است. پایگاه داده سوم که در شکل (۹) نمایش داده شده است، پایگاه داده ISO-NE [۳۰] است که شامل بار الکتریکی ساعتی ایالت New-England است که از مارچ سال ۲۰۰۳ تا دسامبر ۲۰۱۴ جمع‌آوری شده است (پایگاه‌های داده North American Utility و ISO-NE از [۳۱] استخراج شده است).

مرتباً با استفاده از ۵ مجموعه وزن تصادفی اولیه ( $W_1^0$  تا  $W_5^0$ ) آموزش می‌دهیم پس از اتمام آموزش، از خروجی میانگین می‌گیریم و پیش‌بینی نهایی را به دست می‌آوریم.

## نتایج آزمایش‌ها

در این مقاله عملکرد روش پیشنهادی با ۵ شبکه عصبی و یک



شکل ۶. نمایش مدل گروه استفاده شده در این مقاله. مدل مورد نظر با ۵ وزن تصادفی اولیه  $W_1^0$  تا  $W_5^0$  آموزش داده می‌شود

روش رگرسیون بردار پشتیبان<sup>۳۱</sup> (SVR) که توسط نویسندگان طراحی شده و چند مقاله دیگر که آزمایشات آن‌ها با پایگاه‌های داده یکسان با این مقاله انجام شده است، مورد بررسی و مقایسه قرار گرفته است. جزئیات و مشخصات لایه‌ها و همچنین تعداد لایه‌های روش‌های طراحی شده توسط نویسندگان در جدول (۱) آورده شده است.

همانند [۲۷] از معیارهای ارزیابی میانگین خطای مطلق<sup>۳۲</sup> (MAE)، میانگین درصد خطای مطلق<sup>۳۳</sup> (MAPE) و جذر میانگین مربعات خطا<sup>۳۴</sup> (RMSE) که در روابط (۱۶) تا (۱۹) به ترتیب نشان داده شده‌اند، برای ارزیابی مدل‌ها استفاده کرده‌ایم:

$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |x_t - x_{pt}| \quad (16)$$

<sup>۳۵</sup> optimizer  
<sup>۳۶</sup> loss

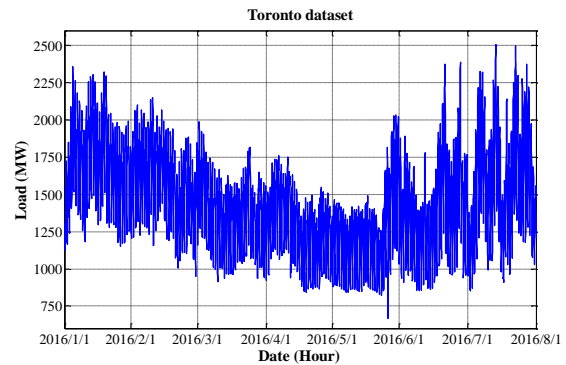
<sup>۳۱</sup> Support Vector Regression  
<sup>۳۲</sup> Mean Absolute Error  
<sup>۳۳</sup> Mean Absolute Percentage Error  
<sup>۳۴</sup> Root Mean Squared Error

این مقاله در تعیین اندازه فیلتر در لایه‌های کانولوشنی از این شباهت استفاده شده است و اندازه فیلتر برابر ۳ در نظر گرفته شده است تا بارهای الکتریکی که در یک فیلتر (هسته) با هم قرار می‌گیرند بیشترین شباهت را به هم داشته باشند.

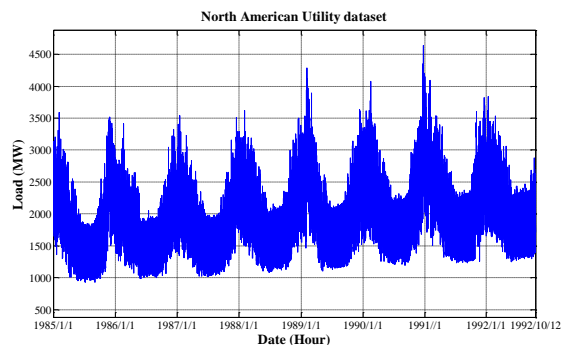
پارامترها	لایه‌ها	الگوریتم
[filters , kernel_size] = [32 , 3]	Conv1D	<b>Proposed method</b>
[filters , kernel_size] = [64 , 3]	Conv1D	
[filters , kernel_size] = [128 , 3]	Conv1D	
n <sub>recurrent</sub> = 100	GRU	
200 output nodes	Dense	
1 output nodes	Dense	
[filters , kernel_size] = [32 , 3]	Conv1D	<b>CNN1D</b>
[filters , kernel_size] = [64 , 3]	Conv1D	
[filters , kernel_size] = [128 , 3]	Conv1D	
	Flatten	
200 output nodes	Dense	
1 output nodes	Dense	
50 output nodes	Dense	<b>F-LSTM</b>
100 output nodes	Dense	
n <sub>recurrent</sub> = 50	LSTM	
dropout rate = 20%	Dropout	
n <sub>recurrent</sub> = 100	LSTM	
dropout rate = 20%	Dropout	
1 output nodes	Dense	
n <sub>recurrent</sub> = 50	LSTM	<b>LSTM</b>
dropout rate = 20%	Dropout	
n <sub>recurrent</sub> = 100	LSTM	
dropout rate = 20%	Dropout	
1 output nodes	Dense	<b>MLP</b>
200 output nodes	Dense	
100 output nodes	Dense	
50 output nodes	Dense	
dropout rate = 20%	Dropout	
1 output nodes	Dense	<b>Linear SVR</b>
Epsilon parameter = 0.01		
loss = epsilon_insensitive		
Regularization parameter = 1		

جدول ۱. تنظیمات مدل‌های استفاده شده

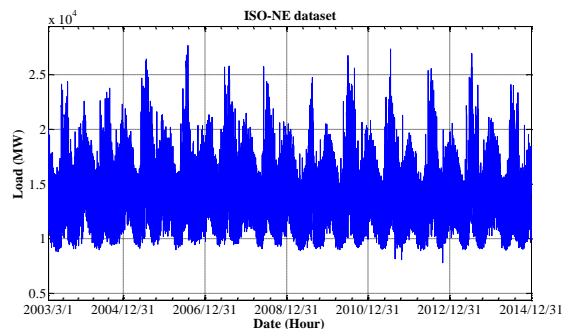
<sup>۱</sup> تنظیمات مدل GRU نیز دقیقاً همانند مدل LSTM است و تنها لایه‌های LSTM به GRU تبدیل شده است.



شکل ۷. بار ساعتی تورنتو از ژانویه ۲۰۱۶ تا جولای ۲۰۱۶



شکل ۸. بار ساعتی در پایگاه داده North American Utility



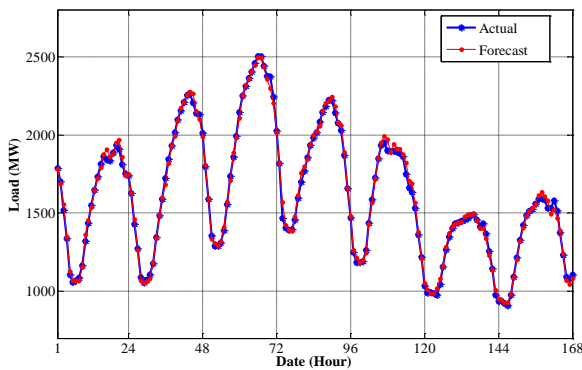
شکل ۹. بار ساعتی مربوط به پایگاه داده ISO-NE

به طور شهودی بارهای الکتریکی در یک بازه کوتاه به دلیل تغییر اندک عوامل وابستگی شهر، یعنی تولید ناخالص داخلی شهر با یکدیگر همبستگی بالایی دارند [۲]. بنابراین ما می‌توانیم با محاسبه ضرایب خودهمبستگی سری بار الکتریکی ارتباط خطی بین بار الکتریکی و همسایگان آن را ارزیابی کنیم. شکل (۱۰) تا (۱۲) به وضوح نشان می‌دهد که بار الکتریکی ساعتی با بار مصرفی ساعت قبل و بعد و همچنین بار الکتریکی مصرفی در روز بعد همان ساعت و ساعت قبل و بعدش (یعنی ۲۴ ساعت بعد) همبستگی بالایی دارد. همانطور که از شکل-های (۱۰) تا (۱۲) مشخص است ضریب خود همبستگی بار الکتریکی هر ساعت با دو ساعت بعد خود بیشتر از ۰/۸ است. بنابراین شباهت بسیار زیادی به هم دارند. به همین دلیل در

Toronto در جدول (۲) گردآوری شده است. همانطور که از جدول (۲) مشخص است روش پیشنهادی در هر سه معیار گفته شده بهترین عملکرد را داشته است و بعد از آن شبکه CNN1D بهترین عملکرد را از خود نشان داده است که نشان‌دهنده قدرت شبکه‌های کانولوشنی در زمینه پیش‌بینی بار است. همچنین از نظر معیار MAPE، روش آماری رگرسیون خطی بردار پشتیبان بدترین عملکرد را داشته است. در شکل (۱۳) عملکرد روش پیشنهادی نشان داده شده است. همانطور که گفته شد ما هر مدل را ۵ بار آموزش می‌دهیم و از Loss به دست آمده میانگین می‌گیریم. نمودار Loss مربوط به روند آموزش مدل پیشنهادی با استفاده از پایگاه داده Toronto در شکل (۱۴) نشان داده شده است که ناحیه هاشورخورده ناحیه بین میانگین Loss و یک انحراف معیار بالاتر و پایین‌تر از میانگین Loss را نشان می‌دهد.

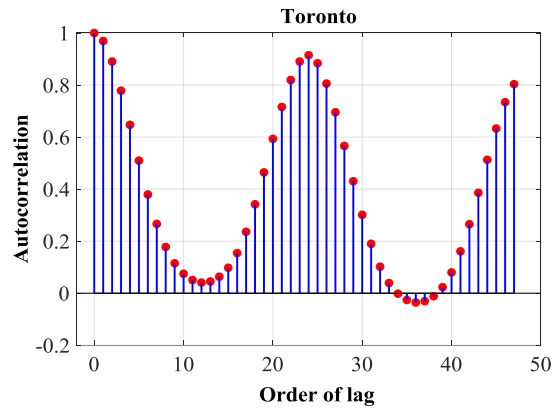
### نتایج آزمایش مربوط به پایگاه داده North American Utility

در این پایگاه داده برای مقایسه با مقالات دیگر، از بارهای الکتریکی ساعتی ۱۹۸۸ تا ۱۲ اکتبر ۱۹۹۰ به عنوان داده‌های آموزشی و از بار الکتریکی ۱۲ اکتبر ۱۹۹۰ تا ۱۲ اکتبر ۱۹۹۲ به عنوان داده‌های آزمایشی استفاده کرده‌ایم.

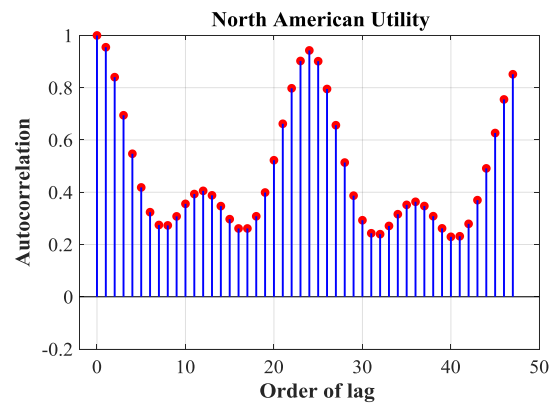


شکل ۱۳. مقایسه بار الکتریکی واقعی و پیش‌بینی شده در مدل پیشنهادی در پایگاه داده Toronto در مدت یک هفته

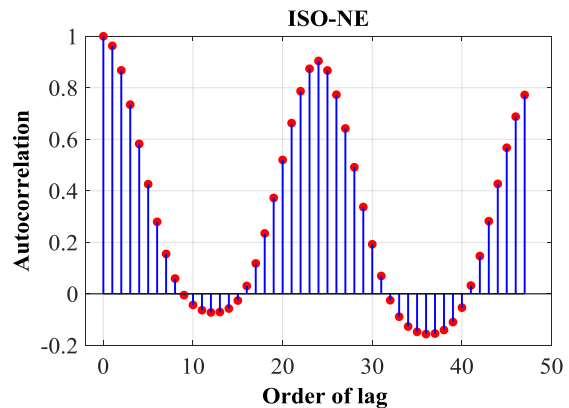
عملکرد مدل پیشنهادی در دو بازه یک هفته‌ای متفاوت در شکل‌های (۱۵) و (۱۶) نشان داده شده است. در شکل (۱۵) عملکرد پیش‌بینی مدل پیشنهادی در هفته اول سپتامبر ۱۹۹۱ و در شکل (۱۶) عملکرد آن در هفته آخر فوریه ۱۹۹۲ نمایش داده شده است. برای مقایسه عملکرد مدل پیشنهادی با سایر مدل‌ها و مقالات دیگری که از این پایگاه داده در آزمایشات خود استفاده کرده‌اند جدول (۳) گردآوری شده است. لازم به ذکر است که نتایج [۲۷] ML-BRNN و [۳۵] CCRN از طریق شبیه‌سازی به دست آمده است.



شکل ۱۰. ضرایب خودهمبستگی بار در پایگاه داده Toronto



شکل ۱۱. ضرایب خودهمبستگی بار در پایگاه داده North American Utility

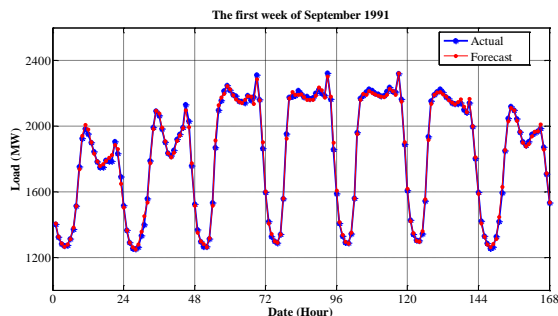


شکل ۱۲. ضرایب خودهمبستگی بار در پایگاه داده ISO-NE

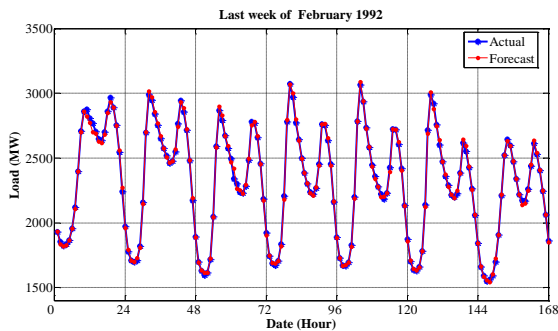
### نتایج آزمایش مربوط به پایگاه داده Toronto

در این پایگاه داده، در تمامی مدل‌های جدول (۲) از ۹۰ درصد داده‌ها برای آموزش و ۱۰ درصد داده‌ها برای آزمایش مدل‌ها استفاده کرده‌ایم. همچنین ۵ درصد داده‌های آموزشی به عنوان داده‌های اعتبارسنجی<sup>۳۷</sup> استفاده شده است. نتایج مربوط به پیش‌بینی بار الکتریکی در ساعت آینده برای پایگاه داده

با توجه به جدول (۳)، مدل پیشنهادی در مقایسه با [۳۴] ۲۴/۳٪، در مقایسه با [۲۷] ۲۰/۳٪ و در مقایسه با [۳۵] ۱۷/۷٪ در معیار میانگین درصد خطای مطلق (MAPE) بهتر عمل کرده است.



شکل ۱۵. عملکرد پیش‌بینی مدل پیشنهادی در پایگاه داده North American Utility در هفته اول سپتامبر ۱۹۹۱



شکل ۱۶. عملکرد پیش‌بینی مدل پیشنهادی در پایگاه داده North American Utility در هفته آخر فوریه ۱۹۹۲

### نتایج آزمایش مربوط به پایگاه داده ISO-NE

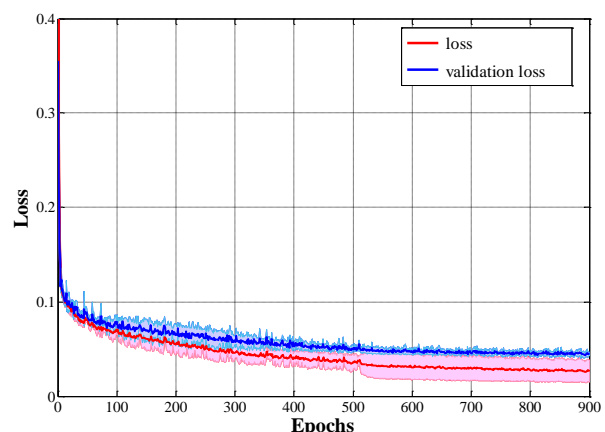
در پایگاه داده ISO-NE برای مقایسه با مقالات دیگر، از بار الکتریکی ۱ ژانویه ۲۰۰۷ تا ۳۱ دسامبر ۲۰۰۷ برای آموزش مدل استفاده کرده‌ایم. از بار الکتریکی ۱ ژانویه ۲۰۰۸ تا ۳۰ ژوئن ۲۰۰۸ به عنوان داده‌های اعتبار سنجی و از بار الکتریکی ۱ جولای ۲۰۰۸ تا ۳۱ دسامبر ۲۰۰۹ برای ارزیابی عملکرد مدل استفاده کرده‌ایم. همانند آنچه در [۳۶] مطرح شده است بازه پیش‌بینی برای مقایسه عملکرد مدل‌ها از ۱ جولای ۲۰۰۸ تا ۳۱ جولای ۲۰۰۸ در نظر گرفته شده است. در شکل‌های (۱۷) و (۱۸) عملکرد پیش‌بینی مدل پیشنهادی در دو بازه یک هفته‌ای متفاوت از داده‌های آزمایشی نشان داده شده است و به منظور مقایسه عملکرد مدل پیشنهادی با سایر مدل‌ها و مقالات دیگر جدول (۴) گردآوری شده است. لازم به ذکر است که

Algorithm	MAPE (%)
Linear SVR	28.288
F-LSTM	1.043
LSTM	1.020
GRU	0.967
CNN1D	0.933
MLP	1.291
Application of echo state networks in STLF [32]	1.140
Feature extraction via multiresolution analysis for STLF [33]	1.100
STLF of power systems by combination of WT and neuro-evolutionary algorithm [34]	0.990
ML-BRNN[27]	0.939
CCRN [35]	0.910
<b>Proposed method</b>	<b>0.749</b>

جدول ۲. مقایسه عملکرد مدل‌ها در پایگاه داده Toronto

Algorithm	MAE (MW)	RMSE (MW)	MAPE (%)
Linear SVR	24.405	31.916	28.821
F-LSTM	25.751	33.952	1.541
LSTM	36.673	46.409	2.352
GRU	35.355	44.857	2.301
CNN1D	21.654	28.021	1.347
MLP	23.893	30.634	1.516
<b>Proposed method</b>	<b>17.74</b>	<b>24.103</b>	<b>1.105</b>

جدول ۳. مقایسه عملکرد مدل‌ها در پایگاه داده North American Utility



شکل ۱۴. نمودار Loss مربوط به روند آموزش مدل پیشنهادی بر روی پایگاه داده Toronto. خطوط پررنگ میانگین Loss حاصل از آموزش مدل با ۵ وزن تصادفی اولیه در داده‌های آموزشی و داده‌های اعتبار سنجی را نشان می‌دهد و نواحی هاشور خورده یک انحراف معیار بالاتر از میانگین Loss و یک انحراف معیار پایین‌تر از میانگین Loss را نشان می‌دهد.

دهنده قابلیت تعمیم روش پیشنهادی در این مقاله باشد. مدل پیشنهادی به ترتیب  $1.18/6\%$ ،  $5.07/7\%$ ،  $11/3\%$ ،  $15/3\%$  و  $5/9\%$  در مقایسه با [۳۶]، [۳۷]، [۳۱]، [۲۷] و [۳۵] در معیار میانگین درصد خطای مطلق (MAPE) بهتر عمل کرده است. همچنین در معیار میانگین خطای مطلق (MAE) نیز در مقایسه با [۳۶]  $1.18\%$ ، در مقایسه با [۲۷]  $12/8\%$  و در مقایسه با [۳۵]  $5/3\%$  بهتر عمل کرده است.

### نتیجه‌گیری و پیشنهادات

استفاده از شبکه‌های عصبی در پیش‌بینی بار الکتریکی به ویژه شبکه‌های بازگشتی بهبود یافته امری رایج و مرسوم است در حالی که شبکه‌های کانولوشنی برای تصویر توسعه داده شده‌اند. ولی نتایج آزمایش‌ها در این مقاله نشان می‌دهد که استفاده از شبکه‌های کانولوشنی به تنهایی، در زمینه پیش‌بینی بار به نتایج مطلوبی منجر می‌شود. روش پیشنهادی در این مقاله یک مدل ترکیب شده از لایه‌های کانولوشنی و لایه برگشتی GRU بود که همانطور که از نتایج مربوط به سه پایگاه داده مورد آزمایش مشخص شد عملکرد به مراتب بهتری نسبت مدل‌های رقیب خود داشت.

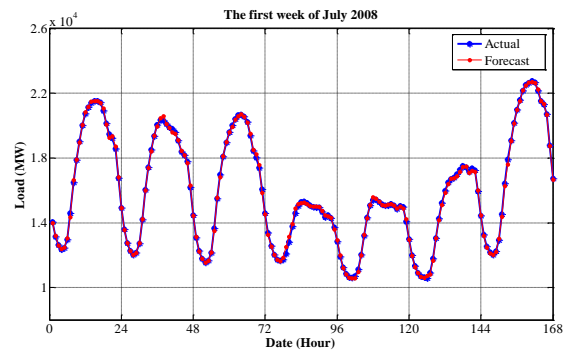
به عنوان پیشنهاد برای بهبود عملکرد این روش می‌توان از رگولایزر<sup>۳۸</sup> و همچنین الگوریتم‌هایی همانند BatchNormalization استفاده کرد. همچنین می‌دانیم که با افزایش عمق شبکه می‌توانیم به کارایی بهتر و پیش‌بینی دقیق‌تری دست پیدا کنیم اما در عمل افزایش تعداد لایه‌ها تا یک حدی منجر به افزایش کارایی شبکه می‌شود و اضافه کردن لایه بیشتر نه تنها عملکرد شبکه را بهبود نمی‌بخشد بلکه باعث کاهش دقت آموزش می‌شود بنابراین می‌توانیم از ساختار شبکه‌های ResNet و استفاده از نگاشت همانی برای لایه‌هایی که عملکرد شبکه را بهبود نمی‌بخشند، به منظور افزایش لایه‌های بیشتر و اصطلاحاً افزایش عمق شبکه و در نتیجه، بهبود عملکرد شبکه جهت افزایش دقت پیش‌بینی بار الکتریکی استفاده کنیم.

### مراجع

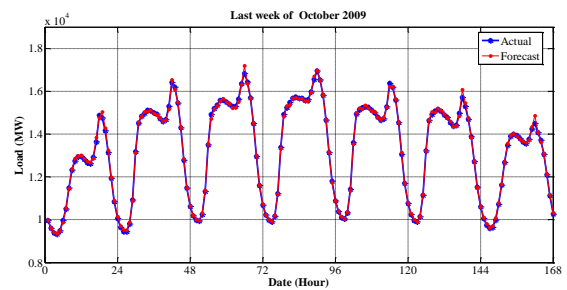
- [1] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," International Journal of Forecasting, vol. 32, pp. 914-938, 2016.
- [2] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou, and L. Shu, "Enhanced deep networks for short-term and medium-term load forecasting," IEEE Access, vol. 7, pp. 4045-4055, 2018.

<sup>۳۸</sup>regularizer

نتایج [۲۷] ML-BRNN و [۳۵] CCRN از طریق شبیه‌سازی به دست آمده است.



شکل ۱۷. عملکرد پیش‌بینی مدل پیشنهادی در پایگاه داده ISO-NE در هفته اول جولای ۲۰۰۸



شکل ۱۸. عملکرد پیش‌بینی مدل پیشنهادی در پایگاه داده ISO-NE در هفته آخر اکتبر ۲۰۰۹

Algorithm	MAE (MW)	MAPE (%)
Linear SVR	158.528	24.616
F-LSTM	136.442	0.753
LSTM	166.946	0.955
GRU	179.369	0.995
CNN1D	143.884	0.824
MLP	203.403	1.154
WNN [36]	83.540	0.490
ISO-NE [37]	138.330	0.810
EA-ELM [38]	70.220	0.410
ML-BRNN [27]	78.481	0.471
CCRN [35]	72.274	0.424
WT-Evolutionary ELM [31]	74.410	0.450
<b>Proposed method</b>	<b>68.467</b>	<b>0.399</b>

جدول ۴. مقایسه عملکرد مدل‌ها در پایگاه داده ISO-NE

همانطور که از جدول (۴) قابل مشاهده است روش پیشنهادی در این پایگاه داده نیز همانند دو پایگاه داده قبلی عملکرد بهتری نسبت به سایر روش‌ها داشته است و این می‌تواند نشان-

- Iranian Conference on Electrical Engineering (ICEE)*, Yazd, Iran, pp. 1362-1366, 2019.
- [14] M. Imani. "Deep Learning Based Electricity Demand Forecasting in Different Domains". *Iranian (Iranica) Journal of Energy & Environment*, vol. 11, no. 1, 33-39, 2020.
- [15] B. White, "Principles of neurodynamics: Perceptrons and the theory of brain mechanisms," ed: JSTOR, 1963.
- [16] P.-H. Kuo and C.-J. Huang, "An electricity price forecasting model by hybrid structured deep neural networks," *Sustainability*, vol. 10, p. 1280, 2018.
- [17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735-1780, 1997.
- [18] A. Narayan and K. W. Hipel, "Long short term memory networks for short-term electric load forecasting," in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2573-2578, 2017.
- [19] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms," *IEEE Access*, vol. 7, pp. 82584-82598, 2019.
- [20] Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555, 2014.
- [21] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, pp. 8609-8613, 2013.
- [22] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the International Conference on Machine Learning*, vol. 30, no. 1, p. 3, 2013.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026-1034, 2015.
- [24] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self normalizing neural networks," in *Advances in Neural Information Processing Systems*, pp. 972-981, 2017.
- [25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [26] M. De Felice and X. Yao, "Short-term load forecasting with neural network ensembles: A comparative study [application notes]," *IEEE*
- [3] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi and M. Khazaei, "Deep-Based Conditional Probability Density Function Forecasting of Residential Loads," *IEEE Transactions on Smart Grid*, 2020.
- [4] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85-117, 2015.
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.
- [6] X. Cao, S. Dong, Z. Wu, and Y. Jing, "A data-driven hybrid optimization model for short-term residential load forecasting," in 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 283-287, 2015.
- [7] M. Imani and H. Ghassemian, "Residential load forecasting using wavelet and collaborative representation transforms," *Applied Energy*, vol. 253, p. 113505, Nov. 2019.
- [8] L. Li, K. Ota, and M. Dong, "Everything is image: CNN-based short-term electrical load forecasting for smart grid," in 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), pp. 344-351, 2017.
- [9] P. Qingle and Z. Min, "Very short-term load forecasting based on neural network and rough set," in 2010 International Conference on Intelligent Computation Technology and Automation, pp. 1132-1135, 2010.
- [10] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), pp. 1483-1488, 2017.
- [11] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, 2018.
- [12] M. Imani, "Long Short-Term Memory Network and Support Vector Regression for Electrical Load Forecasting," 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), Istanbul, Turkey, pp. 1-6, 2019.
- [13] M. Imani and H. Ghassemian, "Sequence to Image Transform Based Convolutional Neural Network for Load Forecasting," *2019 27th*

- of wavelet transform and neuro-evolutionary algorithm", *Energy*, vol. 34, no. 1, p.p. 46-57, 2009.
- [35] H. Eskandari, M. Imani and M. P. Moghadam, "Correlation based Convolutional Recurrent Network for Load Forecasting," *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 2020.
- [36] C. Guan, P. B. Luh, L. D. Michel, Y. Wang and P. B. Friedland, "Very Short-Term Load Forecasting: Wavelet Neural Networks With Data Pre-Filtering," in *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30-41, Feb. 2013.
- [37] P. Shamsollahi, K. W. Cheung, Q. Chen and E. H. Germain, "A neural network based very short term load forecaster for the interim ISO New England electricity market system," *PICA 2001. Innovative Computing for Power - Electric Energy Meets the Market. 22nd IEEE Power Engineering Society. International Conference on Power Industry Computer Applications (Cat. No.01CH37195)*, Sydney, NSW, Australia, pp. 217-222, 2001.
- [38] S. Li, L. Goel and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22 – 29, 2016.
- Comput. Intell. Mag., vol. 6, no. 3, pp. 47–56, Jul. 2011.
- [27] X. Tang, Y. Dai, T. Wang and Y. Chen, "Short-term power load forecasting based on multi-layer bidirectional recurrent neural network," in *IET Generation, Transmission & Distribution*, vol. 13, no. 17, pp. 3847-3854, 2019.
- [28] Available at <https://www.torontohydro.com>
- [29] <http://sites.google.com/site/fkeynia/loaddata>
- [30] <http://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info>
- [31] S. Li, P. Wang, and L. Goel, "Short-term load forecasting by wavelet transform and evolutionary extreme learning machine," *Electric Power Systems Research*, vol. 122, p. 96-103, May. 2015.
- [32] A. Deihimi, H. Showkati, "Application of echo state networks in short-term electric load forecasting," *Energy*, vol. 39, no. 1, p.p. 327–340, 2012.
- [33] A.J.R. Reis, A.P.A. da Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 1, p.p. 189–198, 2005.
- [34] N. Amjady, F. Keynia, "Short-term load forecasting of power systems by combination