

پیش‌بینی مکانی-زمانی غلظت آلاینده‌ی PM_{2.5} با استفاده از شبکه‌های بازگشتی عمیق: مطالعه موردی شهر تهران

مرجان فرجی^۱، سعید نادی^{۲*}، داود شجاعی^۳

^۱ دانشجوی کارشناسی ارشد مهندسی نقشه‌برداری، سنجش‌ازدور - دانشکده‌ی مهندسی عمران و حمل‌ونقل -

دانشگاه اصفهان

marjanfaraji729@yahoo.com

^۲ استادیار گروه مهندسی نقشه‌برداری - دانشکده‌ی مهندسی عمران و حمل‌ونقل - دانشگاه اصفهان

snadi@eng.ui.ac.ir

^۳ استادیار مرکز زیرساخت اطلاعات مکانی و مدیریت زمین - دانشکده مهندسی زیرساخت - دانشگاه ملبورن -

ویکتوریا - استرالیا

shojaeid@unimelb.edu.au

(تاریخ دریافت مرداد ۱۳۹۹، تاریخ تصویب آبان ۱۳۹۹)

چکیده

در سال‌های اخیر آلودگی هوا به یکی از چالش‌های مهم زیست‌محیطی در شهرهای بزرگ و صنعتی به ویژه شهر تهران تبدیل شده است. غلظت بالای ذرات معلق با قطر کمتر از ۲/۵ میکرومتر (PM_{2.5}) که به عنوان اصلی‌ترین عامل آلودگی شهر تهران شناخته شده است، با تأثیرات جبران ناپذیری بر سلامتی انسان همراه است. ارائه مدل مکانی-زمانی با دقت و سرعت بالا برای پیش‌بینی به عنوان یک روش موثر برای حفاظت از سلامت عمومی در برابر افزایش آلاینده‌های مضر هوا می‌باشد. رشد سریع فن‌آوری‌های محاسباتی و در دسترس بودن داده‌های مربوط به کیفیت هوا این فرصت را برای محققان فراهم کرده است تا مدل‌های پیچیده‌ای را در چارچوب یادگیری ماشین و یادگیری عمیق برای پیش‌بینی غلظت آلاینده‌های مختلف هوا ارائه دهند. در این پژوهش با هدف پیش‌بینی غلظت PM_{2.5} در فواصل زمانی مختلف، یک مدل ترکیبی مکانی-زمانی جدید مبتنی بر واحد گیت دار بازگشتی (GRU) با حفظ و استخراج وابستگی‌های زمانی و مکانی در داده‌های سری زمانی آلودگی هوا ارائه شده است و عملکرد آن با روش‌های رگرسیون ماشین بردار پشتیبان (SVR) و حافظه‌ی بلندمدت ماندگار (LSTM) مقایسه شده است. داده‌های مورد استفاده در این پژوهش شامل غلظت ساعتی آلاینده‌ی PM_{2.5} و پارامترهای هواشناسی ثبت شده به ترتیب توسط ۱۳ ایستگاه سنجش آلودگی و ۳ ایستگاه هواشناسی سینوپتیک در شهر تهران در بازه‌ی زمانی ۱۷ آذر ۱۳۹۵ تا ۸ اسفند ۱۳۹۷ می‌باشد. مدل ارائه شده در این مقاله با شاخص‌های ارزیابی $RMSE = 7.97 \mu g/m^3$ و $MAE = 5.35 \mu g/m^3$ بهترین نتیجه را برای پیش‌بینی آلودگی در مقایسه با روش‌های دیگر کسب کرده است. این مدل می‌تواند ۸۰ درصد ($R^2 = 80$) از تغییرات غلظت PM_{2.5} را تعیین و سطح آلودگی را پیش‌بینی کند. همچنین مدل پیشنهادی با استخراج ویژگی‌های زمانی، پیش‌بینی هم‌زمان برای تمام ایستگاه‌ها و در نظر گرفتن همبستگی‌های مکانی اثبات کرده است که می‌تواند برای پیش‌بینی و کنترل آلودگی هوا به طور مؤثر به کار گرفته شود.

واژگان کلیدی: آلودگی هوا، یادگیری عمیق، پیش‌بینی مکانی-زمانی، آلاینده PM_{2.5}، یادگیری ماشین

۱- مقدمه

امروزه آلودگی هوا به عنوان یک عامل مهم در تعیین کیفیت زندگی و سلامت عمومی در مناطق شهری و در مناطق پرجمعیت، شناخته شده است. شهرنشینی، رشد جمعیت، گسترش صنعتی شدن، افزایش مصرف سوخت‌های فسیلی در کنار کیفیت پایین سوخت، فقدان سیستم حمل و نقل کارآمد و ترافیک باعث شده است که روزانه مقادیر زیادی آلاینده وارد هوای شهر تهران شود که این امر با مکانیزم خود پالایی^۱ طبیعت ناسازگار است [۱]. در میان آلاینده‌های موجود در جو، آلاینده $PM_{2.5}$ به عنوان یکی از مهم‌ترین آلاینده‌های هوا، با اندازه‌ی کوچک خود به بخش‌های عمیق ریه نفوذ کرده و موجب بیماری‌های قلبی-عروقی، متابولیک و سرطان می‌شود [۲]. پیش‌بینی آلودگی هوا می‌تواند به عنوان یک ابزار قدرتمند به دولت‌ها کمک کند تا مردم در معرض خطر را از وضع آلودگی هوا مطلع سازد که علاوه بر اینکه راهکارهایی در جهت کاهش آلودگی هوا ارائه شود، از قرارگیری افراد در مکان‌های آلوده نیز جلوگیری شود [۳]. در طول دهه‌های گذشته، مطالعات متعددی برای پیش‌بینی غلظت مکانی - زمانی آلاینده‌ها در هوا انجام شده است. به‌طور کلی روش‌های پیش‌بینی آلودگی هوا به دو دسته روش‌های قطعی^۲ و روش‌های آماری^۳ تقسیم‌بندی می‌شوند [۳]. روش‌های قطعی شامل مدل‌سازی آلودگی هوا بر پایه شبیه‌سازی رفتار شیمیایی، فیزیکی، انتقال و پراکنش آلاینده‌ها در اتمسفر می‌باشد. نتایج حاصل از این روش‌ها به دلیل نیاز به دانستن اطلاعاتی نظیر چگونگی پراکنش، ضرایب انتشار آلاینده و تغییرپذیری ذاتی اتمسفر، با دقت پایینی مواجه است [۴]. برای غلبه بر محدودیت‌های روش‌های قطعی از مدل‌های آماری برای پیش‌بینی آلودگی هوا استفاده شده است. به طور کلی روش‌های آماری پیش‌بینی آلودگی هوا به دو دسته روش‌های مکانی بر پایه درون‌یابی و روش‌هایی بر پایه سری زمانی تقسیم‌بندی می‌شود. روش‌هایی بر پایه درون‌یابی تحت عنوان روش‌های زمین آماری برای مدل‌سازی آلودگی هوا مانند روش درون‌یابی معکوس

وزن‌دار^۴ [۵] روش کریجینگ^۵ [۶] و رگرسیون کاربری اراضی^۶ [۷] در تحقیقات گذشته برای تخمین مکانی آلودگی در سطح شهر در یک زمان مشخص به کار گرفته شده‌اند. دسته‌ی دوم روش‌های بر پایه سری زمانی با هدف پیش‌بینی آلودگی شامل روش‌های کلاسیک و روش‌های نوین می‌باشند. روش میانگین متحرک خود همبسته یکپارچه^۷ [۸] و رگرسیون خطی چندگانه [۹] از جمله روش‌های کلاسیک آماری پیش‌بینی محسوب می‌شوند. دقت به دست آمده در این روش‌ها به دلیل عدم توانایی این روش‌ها در مدل‌سازی داده‌های غیرخطی با محدودیت مواجه است.

به طور کلی روش‌های شبکه عصبی، یادگیری ماشین و یادگیری عمیق در دسته روش‌های نوین پیش‌بینی آلودگی هوا قرار می‌گیرند [۱۰]. رگرسیون بردار پشتیبان و شبکه عصبی از دیگر روش‌های رایج در پیش‌بینی آلودگی هوا می‌باشند. پرز و تریر^۸ (۲۰۰۱) نشان دادند که مدل شبکه عصبی مصنوعی (ANN) برای پیش‌بینی غلظت مونوکسید نیتروژن (NO) و دی‌اکسید نیتروژن (NO₂) بهتر از مدل رگرسیون چند متغیره عمل می‌کند [۱۱]. امروزه یادگیری عمیق به یکی از مهم‌ترین روش‌ها در حوزه یادگیری ماشین تبدیل شده است که با موفقیت در حوزه‌های بینایی ماشین، طبقه‌بندی تصویر، شناسایی گفتار و پیش‌بینی سری زمانی استفاده شده است [۱۲]. علاوه بر آن در حوزه‌ی پیش‌بینی کیفیت هوا نیز به طور گسترده مورد استفاده قرار گرفته است [۱۳-۱۵]. ردی^۹ و همکاران (۲۰۱۷) در پژوهش انجام‌شده بر روی شهر پکن کشور چین، به پیش‌بینی غلظت آلاینده $PM_{2.5}$ در ۵ تا ۱۰ ساعت آینده با به‌کارگیری داده‌ها در فاصله‌ی زمانی ۲۰۱۵ تا ۲۰۱۷ پرداخته‌اند. در مدل ارائه‌شده در آن پژوهش از شبکه بازگشتی عمیق، با معماری LSTM به همراه داده‌های هواشناسی و غلظت آلاینده در ایستگاه‌های سنجش آلودگی استفاده شده است. تحلیل نتایج حاصله نشان داده است که مدل LSTM پیش‌بینی کننده‌ی بهتری برای داده‌های سری زمانی نسبت به مدل‌های

۴ Inverse distance weighted (IDW)

۵ Kriging

۶ Land use regression (LUR)

۷ Autoregressive integrated moving average (ARIMA)

۸ Perez and Trier

۹ Reddy

۱ Self-purification

۲ Deterministic method

۳ Statistical method

بنابراین لازم است که هنگام بررسی کیفیت هوا همبستگی‌های مکانی و زمانی ایستگاه‌ها در نظر گرفته شود [۱۰]. در این پژوهش به پیش‌بینی مکانی-زمانی غلظت آلاینده $PM_{2.5}$ در چندین گام جلوتر (تا ۲۴ ساعت آینده)، با استفاده از گسترشی بر روش یادگیری عمیق با استفاده از شبکه GRU پرداخته شده است.

۲- مواد و روش‌ها

در این بخش به معرفی منطقه‌ی مورد مطالعه و داده‌های استفاده شده در پژوهش پرداخته می‌شود. همچنین مبانی نظری روش‌های مطرح شده در این پژوهش تشریح خواهند شد.

۲-۱- منطقه مورد مطالعه

شهر تهران به عنوان پایتخت ایران در عرض جغرافیایی ۵۱ درجه شرقی و طول جغرافیایی ۳۵ درجه شمالی واقع شده است. شهر تهران دارای هوای نیمه خشک و سرد با میانگین سالانه رطوبت نسبی ۴۶ درصد و دمای ۵- تا ۳۸ درجه در ارتفاع ۱۱۱۷ تا ۱۷۱۲ از سطح دریا قرار دارد. تهران در دامنه جنوبی کوه البرز قرار دارد که تأثیر بسزایی در آب و هوای منطقه دارد. جهت باد تهران از سمت شمال غربی به جنوب شرقی است. بنابراین باد آلودگی هوا را از نقاط مختلف شهر و حتی از شهرهای مستقر در شمال غربی تهران (مانند کرج) به جنوب و جنوب شرقی پایتخت منتقل می‌کند.

وسعت و جمعیت شهر تهران طی چند دهه اخیر رو به افزایش بوده است. بر اساس سرشماری ملی انجام شده در سال ۲۰۱۶، جمعیت شهر تهران ۸/۶۹ میلیون نفر برآورد شده است که ۱۰/۸ درصد از کل جمعیت کشور را تشکیل داده است. آلاینده $PM_{2.5}$ با غلظتی سالانه ۸۶/۸ میکروگرم بر مترمکعب (بر اساس ۴ سال مشاهدات، از سال ۲۰۱۵ تا ۲۰۱۸) به طور قابل توجهی از مقادیر دستورالعمل سازمان بهداشت جهانی^۶ فراتر رفته است [۲۲]. وسایل نقلیه بنزینی و دیزلی، انتشار گازهای گلخانه‌ای صنعتی و طوفان‌های گرد و غبار مهم‌ترین دلایل غلظت بالای $PM_{2.5}$ در تهران هستند [۲۱].

غیرخطی دیگر نظیر رگرسیون بردار پشتیبان می‌باشد [۱۶]. آتیرا^۱ و همکاران (۲۰۱۸) در پژوهشی دیگر، از طریق سه ساختار یادگیری عمیق شامل مدل شبکه عصبی بازگشتی، شبکه LSTM و شبکه GRU به پیش‌بینی آلاینده‌ی $PM_{2.5}$ پرداخته‌اند. داده‌های استفاده شده در آن تحقیق شامل داده‌های هواشناسی و غلظت آلاینده‌های اندازه‌گیری شده توسط ۱۴۹۸ ایستگاه سنجش آلودگی در کشور چین و در بازه‌ی زمانی سال‌های ۲۰۱۵ تا ۲۰۱۷ می‌باشد. درصد میانگین مطلق خطا^۲ برابر با مقدار عددی ۰/۴۳ بیان‌گر عملکرد بهتر مدل GRU نسبت به مدل‌های دیگر می‌باشد [۱۵].

پیش‌بینی آلودگی هوا در کشور ایران و به ویژه شهر تهران نیز مورد توجه پژوهشگران متعددی بوده است. کمالی و همکاران (۲۰۱۵) با ارائه مدل ترکیبی از فیلتر پایین گذر کلموگوروف-زوربنکو^۳ و شبکه عصبی، به پیش‌بینی آلاینده‌های گازی و PM_{10} در شهر تهران پرداختند. آن‌ها با اعمال یک فیلتر در داده‌های ورودی شبکه عصبی به ضریب تعیین ۰/۹ در یک ایستگاه سنجش آلودگی رسیدند [۱۷]. در پژوهشی دیگر معماریان فرد و حاتمی (۲۰۱۷) یک مدل پیش‌بینی روزانه آلاینده $PM_{2.5}$ با استفاده از یک شبکه عصبی پیش‌خور سه لایه^۴ و داده‌های روزانه هواشناسی ارائه دادند [۱۸]. جمال و نبی زاده (۲۰۱۷) با به‌کارگیری مدلی ترکیبی از درخت تصمیم^۵ و شبکه عصبی به توسعه‌ای در پیش‌بینی آلاینده $PM_{2.5}$ با شاخص ارزیابی RMSE به مقدار عددی $21/6 \mu g/m^3$ دست یافتند [۱۹]. از دیگر پژوهش‌های انجام شده در حوزه پیش‌بینی آلودگی هوا در شهر تهران با به‌کارگیری مدل‌های یادگیری ماشین و یادگیری عمیق، می‌توان به پژوهش‌های کائیمیان و همکاران (۲۰۱۹) و زمانی و همکاران (۲۰۱۹) اشاره کرد [۲۰، ۲۱].

مطابق با پژوهش‌های پیشین، انتشار و پراکنش آلاینده‌های هوا دارای ارتباطات مکانی می‌باشند [۱۰]. اگر آلاینده‌های هوا فقط در حوزه زمانی تجزیه و تحلیل شوند، تأثیرات و روابط بین ایستگاه‌ها ممکن است نادیده گرفته شود. در مقابل، اگر فقط روابط مکانی در نظر گرفته شود، انتشار آلاینده‌ها با توجه به زمان نادیده گرفته می‌شود.

^۱ Athira

^۲ Mean absolute percentage error

^۳ Kolmogorov-Zurbenko

^۴ Feedforward neural network (FNN)

^۵ Decision tree

^۶ World Health Organization (WHO)

۲-۲- داده غلظت آلاینده PM_{2.5} و پارامترهای هواشناسی

در این پژوهش از داده‌های غلظت آلاینده از ۱۳ ایستگاه سنجش آلودگی تهیه شده توسط سازمان کنترل کیفیت هوا در شهر تهران استفاده شده است. داده‌های جمع‌آوری شده به صورت ساعتی و در بازه زمانی ۱۷ آذر ۱۳۹۵ تا ۸ اسفند ۱۳۹۷ می‌باشد. به منظور بهبود عملکرد پیش‌بینی با قرار دادن یک حد آستانه مناسب داده‌های پرت حذف شده است. اطلاعات آماری میانگین، انحراف معیار، بیشینه و کمینه‌ی پارامتر غلظت آلاینده از ۱۳ ایستگاه سنجش آلودگی و پارامترهای هواشناسی از سه ایستگاه سینوپتیک استفاده شده در پژوهش به صورت جدول (۱) بیان شده است.

جدول ۱- مشخصات آماری داده‌های ورودی مدل				
نام پارامتر	انحراف	میانگین	بیشینه	کمینه
معیار				
آلاینده PM _{2.5}	۲۰	۳۲	۲۷۷	۱
جهت باد (dd)	۱۱۶/۳	۱۳۹	۳۶۰	۰
سرعت باد (ff)	۱/۵۵	۲	۱۵	۰
دما (T)	۱۰/۳۹	۱۶	۴۰	-۷
نقطه‌ی شبنم (Td)	۵/۴۴	-۰/۴۱	۹/۵	-۲۷
فشار ایستگاه (p)	۷/۲۸	۸۶۲	۹۱۲	۸۴۷
فشار سطح دریا (P ₀)	۴/۲۱	۸۴۶/۹	۹۱۱/۹۳	۸۶۱/۸۹
رطوبت نسبی (U)	۲۰۵۲	۳۸	۹۸	۳

پارامترهای هواشناسی یک عامل تأثیرگذار در غلظت آلاینده‌های هوا به شمار می‌آیند. برای مثال در فصل زمستان با کاهش دمای هوا پدیده‌ی وارونگی هوا اتفاق می‌افتد و موجب افزایش غلظت آلاینده‌ها می‌شود از طرفی افزایش باد و رطوبت کاهش آلودگی را در پی دارد. بنابراین در این پژوهش از ۷ پارامتر هواشناسی شامل دما، رطوبت نسبی، فشار هوای ایستگاه و دریا، سرعت و جهت باد و نقطه‌ی شبنم تهیه شده توسط ۳ ایستگاه هواشناسی سینوپتیک شامل ایستگاه مهرآباد، شمیران و ژئوفیزیک در بازه زمانی ۳ ساعته استفاده شده است. طریقه‌ی به کارگیری داده‌های هواشناسی بدین گونه است که برای هر ایستگاه سنجش آلودگی از داده‌های نزدیک‌ترین ایستگاه سینوپتیک به ایستگاه مربوطه استفاده شده است. موقعیت

ایستگاه‌های سنجش آلودگی به همراه ایستگاه‌های سینوپتیک در شکل (۱) مشاهده می‌شود.

۳-۲- آنالیز وابستگی بین ایستگاه‌ها

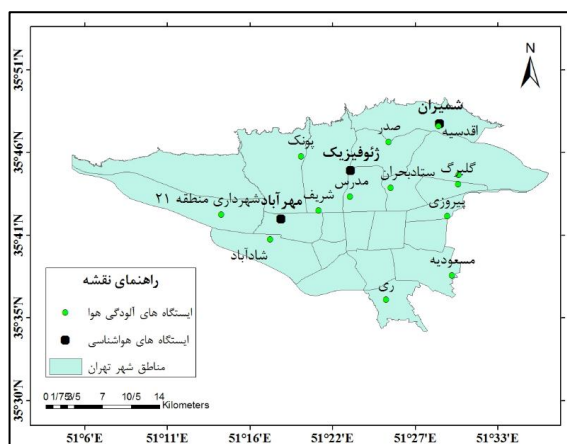
در این بخش ارتباط مکانی غلظت آلاینده PM_{2.5} را در بین ۱۳ ایستگاه مورد بررسی قرار داده شد. برای اندازه‌گیری همبستگی مکانی بین دو ایستگاه از ضریب همبستگی پیرسون مطابق با رابطه‌ی (۱) استفاده شده است [۲۳].

$$r_{ij} = \frac{\sum_t (S_i - \bar{S}_i)(S_j - \bar{S}_j)}{\sqrt{\sum_{i=1}^n S_i - \bar{S}_i)^2} \sqrt{\sum_{j=1}^n S_j - \bar{S}_j)^2}} \quad (1)$$

که در آن S_i سری زمانی ایستگاه i ، $\bar{S}_i$ سری زمانی ایستگاه i و $\bar{S}_j$ و $\bar{S}_i$ میانگین مقادیر سری زمانی ایستگاه‌های i و j می‌باشند.

همبستگی بین ایستگاه‌ها در جدول (۲) نشان داده شده است. نکته قابل توجه این است که تمام مقادیر همبستگی بالاتر از ۰/۴ ($P \text{ value} < 0.05$) می‌باشد که بیان‌کننده ارتباط مؤثر بین ایستگاه‌های سنجش آلودگی است. این ارتباط مکانی بیان‌کننده‌ی لزوم استفاده از یک مدل برای پیش‌بینی هم‌زمان در ایستگاه‌ها به جای استفاده از مدل‌های مجزا برای هر ایستگاه را به اثبات می‌رساند [۲۴].

علاوه بر آن استفاده از ورودی‌های مرتبط در مدل موجب بهبود عملکرد پیش‌بینی می‌شود [۲۵]. در این پژوهش برای پیش‌بینی هر ایستگاه از سه ایستگاه دارای بیش‌ترین همبستگی با ایستگاه هدف استفاده شده است.



شکل ۱- موقعیت مکانی ایستگاه‌های سنجش آلودگی و ایستگاه‌های هواشناسی شهر تهران

جدول ۲- ضریب همبستگی پیرسون بین ایستگاه‌ها

ایستگاه	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
S1	۱	۰/۷۱	۰/۶۱	۰/۶۱	۰/۶۸	۰/۵۱	۰/۸۲	۰/۷۵	۰/۸۷	۰/۸۱	۰/۷۱	۰/۶۸	۰/۵۳
S2	۰/۷۱	۱	۰/۴۶	۰/۵۳	۰/۵۷	۰/۴۱	۰/۶۶	۰/۶۷	۰/۷۱	۰/۶۱	۰/۶۶	۰/۵۱	۰/۵
S3	۰/۶۱	۰/۴۶	۱	۰/۴۶	۰/۴۸	۰/۵	۰/۵۶	۰/۵	۰/۵۵	۰/۵۷	۰/۴۸	۰/۵	۰/۴۴
S4	۰/۶۱	۰/۵۳	۰/۴۶	۱	۰/۶۲	۰/۵۷	۰/۶۲	۰/۵۴	۰/۵۹	۰/۵۸	۰/۵۳	۰/۵۲	۰/۵۲
S5	۰/۶۸	۰/۵۷	۰/۴۸	۰/۶۲	۱	۰/۵۳	۰/۶۷	۰/۵۹	۰/۶۴	۰/۶۸	۰/۵۹	۰/۶۲	۰/۴۹
S6	۰/۵۱	۰/۴۱	۰/۵	۰/۵۷	۰/۵۳	۱	۰/۵۷	۰/۴۷	۰/۵۳	۰/۵۴	۰/۴۶	۰/۵۲	۰/۴۸
S7	۰/۸۲	۰/۶۷	۰/۵۶	۰/۶۲	۰/۶۷	۰/۵۷	۱	۰/۶۷	۰/۷۸	۰/۷۸	۰/۶۵	۰/۶۵	۰/۵۴
S8	۰/۷۵	۰/۶۷	۰/۵۴	۰/۵۹	۰/۴۷	۰/۶۷	۱	۰/۷۷	۰/۷۷	۰/۶۶	۰/۷۵	۰/۵۹	۰/۵
S9	۰/۶۷	۰/۷۱	۰/۵۵	۰/۵۹	۰/۶۴	۰/۵۳	۰/۷۸	۰/۷۷	۱	۰/۷۵	۰/۷۳	۰/۶۶	۰/۵۸
S10	۰/۸۱	۰/۶۱	۰/۵۷	۰/۵۸	۰/۶۸	۰/۵۴	۰/۷۸	۰/۶۶	۰/۷۵	۱	۰/۶۱	۰/۶۴	۰/۴۵
S11	۰/۷۱	۰/۶۶	۰/۴۸	۰/۵۳	۰/۵۹	۰/۴۶	۰/۶۵	۰/۷۵	۰/۷۳	۰/۶۱	۱	۰/۵۷	۰/۴۷
S12	۰/۶	۰/۵۱	۰/۵	۰/۵۲	۰/۶۲	۰/۵۲	۰/۶۵	۰/۵۹	۰/۶۶	۰/۶۴	۰/۵۷	۱	۰/۴۶
S13	۰/۵۳	۰/۵	۰/۴۴	۰/۵۳	۰/۴۹	۰/۴۸	۰/۵۴	۰/۵	۰/۵۸	۰/۴۵	۰/۴۷	۰/۴۶	۱

در ادامه به تشریح مبانی تئوری روش‌های به کار رفته در این مقاله همانند روش LSTM، SVR و مدل پیشنهادی بر مبنای GRU پرداخته خواهد شد.

۲-۴- شبکه حافظه کوتاه مدت ماندگار (LSTM)

روش‌های یادگیری عمیق می‌توانند از حجم بالای داده استفاده کرده و همبستگی‌های پیچیده بین داده‌ها را از طریق آموزش مؤثر شناسایی کنند [۲۶]. شبکه LSTM نوع خاصی از شبکه‌های بازگشتی از دسته روش‌های با نظارت یادگیری عمیق هستند که قادر به حفظ وابستگی‌های طولانی مدت با صرف نظر کردن از بخشی از اطلاعات غیر ضروری به عنوان یک راه‌حل برای حل مشکل محوشدگی گرادیان می‌باشد [۲۷]. شبکه LSTM از ۳ گیت یا دروازه تشکیل شده است که در روابط (۲) تا (۵) بیان شده اند [۲۸]. وزن‌های شبکه برای پیش‌بینی گام‌های زمانی بعدی به‌روزرسانی می‌شوند که این کار مستلزم حفظ اطلاعات زمانی از گام‌های ابتدایی می‌باشد.

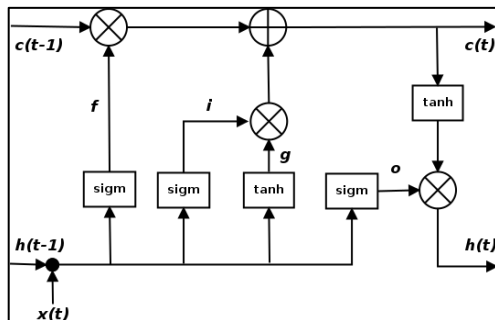
$$f_t = \text{sigm}(U_f x_t + W_f h_{t-1} + b_f) \quad (2)$$

$$i_t = \text{sigm}(U_i x_t + W_i h_{t-1} + b_i) \quad (3)$$

$$c_t = \text{tanh}(U_c x_t + W_c h_{t-1} + b_c) \quad (4)$$

$$o_t = \text{tanh}(U_o x_t + W_o h_{t-1} + b_o) \quad (5)$$

در معادلات فوق، f_t بیانگر گیت فراموشی، i_t گیت ورودی، o_t گیت خروجی، c_t سلول حالت، U و W وزن‌ها و b بایاس می‌باشند. ساختار یک شبکه LSTM در شکل (۲) مشاهده می‌شود. عنصر اصلی شبکه LSTM سلول حالت است که به صورت یک خط افقی با ساختار ساده و تغییرات کم در شکل (۲) مشاهده می‌شود. LSTM قادر است اطلاعات جدیدی را به سلول حالت اضافه یا حذف کند که این کار از طریق گیت‌ها انجام می‌پذیرد.



شکل ۲- ساختار یک بلوک LSTM [۲۸]

خروجی گیت فراموشی f به صورت یک عدد بین ۰ و ۱ حاصل از لایه تابع فعال‌سازی سیگموئید می‌باشد. برای مثال اگر $f = 0$ باشد همه اطلاعات در سلول c_t فراموش می‌شود و اگر $f = 1$ باشد تمام اطلاعات حفظ می‌شود. به بیان دیگر این گیت وظیفه کنترل اطلاعات از گام زمانی قبلی را دارد که اطلاعات حافظه از گام زمانی قبل مورد استفاده قرار بگیرند یا خیر. گیت ورودی i تصمیم‌گیری می‌کند که چه اطلاعاتی در این وضعیت از گام زمانی قبلی به‌روزرسانی شوند. در واقع این گیت

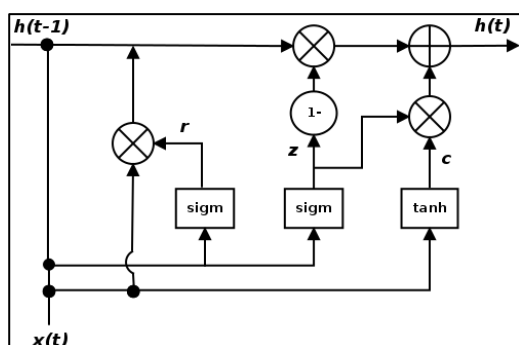
یک بردارند که تعیین می‌کنند چه اطلاعاتی به خروجی منتقل شود [۳۱].

$$r_t = \text{sigm}(U_r x_t + W_r h_{t-1} + b_r) \quad (8)$$

$$z_t = \text{sigm}(U_z x_t + W_z h_{t-1} + b_z) \quad (9)$$

$$\hat{h}_t = \tanh(U_h x_t + W_h (r_t \cdot h_{t-1}) + b_h) \quad (10)$$

در معادلات (۸) و (۹) پارامترهای r_t و z_t به ترتیب بیان‌کننده‌ی لایه گیت بازنشانی و لایه‌ی گیت به‌روزرسانی می‌باشند. در صورتی‌که وضعیت پنهان منتخب تنها بر مبنای ورودی x_t فعلی باشد، تمامی عناصر گیت بازنشانی به صفر نزدیک شده و اطلاعات از گام‌های زمانی h_{t-1} کنار گذاشته می‌شود. ساختار یک بلوک GRU به صورت شکل (۳) نمایش داده می‌شود.



شکل ۳- ساختار یک بلوک GRU [۱۵]

در انتها وضعیت پنهان جدید با روشی مشابه محاسبه ی سلول حالت LSTM به دست می‌آید با این تفاوت که در شبکه GRU گیت به روز رسانی مشابه گیت فراموشی در LSTM عمل کرده و در واقع در GRU گیت فراموشی و گیت به روز رسانی ادغام شده‌اند و خروجی شبکه به صورت معادله (۱۱) بیان می‌شود [۳۱].

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \hat{h}_t \quad (11)$$

که در رابطه فوق z_t نشان‌دهنده‌ی گیت به روز رسانی، h_{t-1} حالت پنهان پیشین و $\hat{h}_t$ حالت پنهان به روز رسانی شده می‌باشد. در واقع h_t در شبکه GRU در زمان t حاصل از یک تعامل خطی بین وضعیت پنهان پیشین و وضعیت منتخب می‌باشد [۳۲].

مشخص می‌کند که آیا در گام زمانی فعلی باید از اطلاعات جدید استفاده شود یا خیر. مطابق با رابطه (۶) ترکیب خروجی گیت فراموشی در سلول حالت پیشین و خروجی گیت ورودی در سلول حالت فعلی، سلول حالت را به‌روزرسانی می‌کند [۲۸].

$$c_t^l = f \cdot c_{t-1}^l + i \cdot c_t \quad (6)$$

که در آن f گیت فراموشی، c_{t-1}^l سلول حالت پیشین، i گیت ورودی c_t سلول حالت فعلی می‌باشد. بدین ترتیب اطلاعات مهم با ورودی‌های اضافه‌شده جایگزین نمی‌شود و می‌توان آن را برای یک دوره‌ی طولانی‌تر حفظ کرد. گیت خروجی o مشخص می‌کند که چه میزان اطلاعات در سلول حافظه حفظ شود و چه میزان از اطلاعات گام زمانی قبل با اطلاعات گام زمانی فعلی، به گام زمانی بعد منتقل شود. خروجی وضعیت به صورت معادله ۷ نشان داده‌شده است [۲۸].

$$h_t^l = o_t \cdot \tanh(c_t^l) \quad (7)$$

که در رابطه فوق o_t گیت خروجی و c_t^l سلول حالت به روز رسانی شده می‌باشد. در نهایت h_t^l از حافظه‌ی به‌روزرسانی محاسبه و در گیت خروجی o_t ضرب می‌شود و بردار خروجی به عنوان یک تابع از وضعیت پنهان جدید محاسبه می‌شود.

۲-۵- شبکه واحد گیت دار بازگشتی (GRU)

این شبکه معمولاً به صورت نسخه تغییر یافته LSTM در نظر گرفته می‌شود با این تفاوت که در این شبکه سلول حالت و وضعیت پنهان ادغام شده است. نتایج تحقیقات گذشته نشان داده است که در داده‌های یکسان این شبکه نتایج بهتری به نسبت LSTM ارائه می‌دهد [۲۹، ۳۰]. در شبکه GRU برخلاف شبکه LSTM سلول حالت به صورت مجزا و گیت خروجی وجود ندارد و تنها از دو گیت به روز رسانی^۱ و بازنشانی^۲ به صورت معادلات (۸) و (۹) تشکیل شده است. همچنین وضعیت پنهان به صورت معادله‌ی (۱۰) نشان داده می‌شود. این گیت‌ها در اصل

^۱ Update gate

^۲ Reset gate

۲-۶- ماشین بردار پشتیبان

به طور کلی مدل ماشین بردار پشتیبان به دو گروه کلی طبقه بندی و رگرسیون ماشین بردار پشتیبان تقسیم شده است و در مدل های پیش بینی از روش های رگرسیونی بردار پشتیبان استفاده می شود. ماشین بردار پشتیبان توسط وینیک و همکاران (۱۹۹۵) ارائه شد [۳۳]. در ماشین های بردار پشتیبان رگرسیونی، مسئله اولیه به صورت رابطه ۱۲ بیان می شود [۳۴].

$$f(x) = (w * \phi(x)) + b \quad (12)$$

که در آن $f(x)$ بیانگر مقدار پیش بینی شده، $\phi(x)$ تابع غیر خطی در فضای ویژگی (تابع کرنل)، پارامتر w و b بردار وزن و بایاس را نشان می دهند. در این مدل هدف تعیین تابع $f(x)$ از طریق مدل رگرسیونی است. منظور از رگرسیون به دست آوردن یک ابر صفحه است که بر داده های مورد نظر برازش داده می شود. فاصله ی هر نقطه از ابر صفحه بیان کننده خطای نقطه ی مورد نظر است. بهترین روشی که تاکنون برای برازش رگرسیون خطی پیشنهاد شده است روش کمترین مربعات است که در حضور داده های پرت عملکرد ضعیفی از خود نشان می دهد. بنابراین به یک برآوردگر نیرومند نیاز است که نسبت به تغییرات کوچک حساس نباشد و خطاها را در یک فاصله ی معین از مقادیر مشاهداتی نادیده بگیرد. از این رو یک تابع جریمه به صورت رابطه (۱۳) تعریف شده است [۳۴].

$$L(x, y, f) = \begin{cases} 0 & \text{if } |y_i - f(x)| \leq \varepsilon \\ |y_i - f(x)| - \varepsilon & \text{else} \end{cases} \quad (13)$$

هنگامی که خطای داده ها از مقدار ε تجاوز نماید تغییری تحت عنوان متغیر کمبود تعریف می شود. مطابق با تابع جریمه ارائه شده به منظور بهینه سازی محدوده خطا، تابع خطا به همراه قیود به صورت رابطه (۱۴) تعریف می شود [۳۴].

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & \text{subject to } \begin{cases} y_i - w \cdot \phi(x_i) - b = \varepsilon + \xi_i \\ w \cdot \phi(x_i) + b - y_i = \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (14)$$

متغیرهای ξ_i^* و ξ_i معرف متغیرهای کمبود می باشند که حد بالا و پایین خطا با توجه به انحراف ε را تعیین می کند. همچنین C معیار جریمه برای جمله ی خطا می باشد. در نهایت حل مسئله بهینه سازی از طریق استفاده از ضرایب لاگرانژ به صورت رابطه (۱۵) خواهد بود [۳۵].

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \cdot \phi(x_i)^T \cdot \phi(x) + b \quad (15)$$

در رابطه فوق $\bar{\alpha}_i$ و $\phi(x)$ بیانگر میانگین دو ضریب لاگرانژ و کرنل می باشند. برای پیش بینی مسائل غیر خطی از تابع کرنل مشخص برای نگاشت داده ی ورودی به ابعاد بالاتر استفاده می شود. به این ترتیب که در فضای بالاتر مسئله غیر خطی به خطی تبدیل می شود. یکی از توابع کرنل متداول در مدل SVR تابع کرنل شعاعی^۱ است که به صورت رابطه (۱۶) نمایش داده می شود [۳۵].

$$k(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\delta^2}\right), \frac{1}{2\delta^2} = \gamma \quad (16)$$

در این تابع پارامتر γ ضریب کرنل است و هر چه مقدار بزرگتری داشته باشد موجب بیش برازش مدل و افزایش خطا می شود [۳۵].

۳- پیاده سازی

در این بخش، به بیان معماری ارائه شده بر مبنای روش GRU و پیاده سازی و مقایسه ی آن با روش های LSTM و SVR پرداخته خواهد شد.

۳-۱- معماری توسعه داده شده بر مبنای شبکه

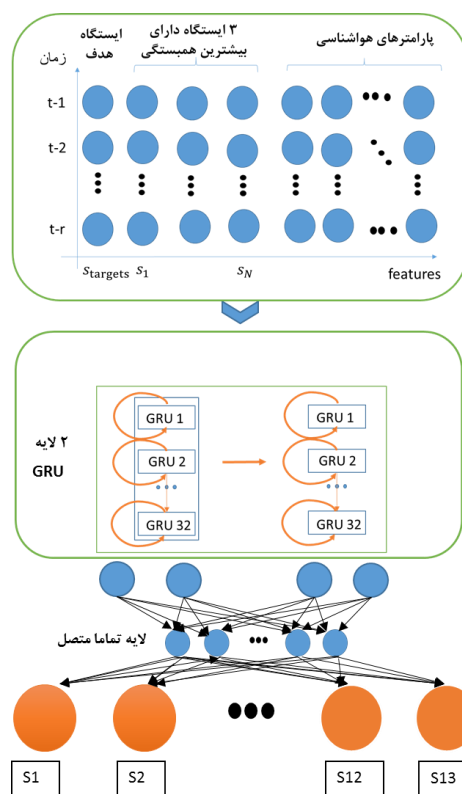
GRU و روش های مقایسه ای

ساختار معماری کلی مدل ارائه شده در شکل (۴) نشان داده شده است. ورودی شبکه داده های غلظت آلاینده نرمال شده در بازه ی [0,1] ایستگاه هدف به همراه ۳ ایستگاه دارای بیشترین همبستگی با ایستگاه هدف در کنار پارامترهای هواشناسی نرمال شده آن ایستگاه در تأخیر زمانی تعیین شده می باشد. پس از ساخت ماتریس ورودی دو لایه GRU با هدف استخراج ویژگی های زمانی از اطلاعات سری زمانی قرار داده شده است. در این بخش

^۱ Raial baiss function (RBF)

استخراج خودکار ویژگی‌های زمانی و مکانی با در نظر گرفتن وابستگی‌های مکانی بین ایستگاه‌ها با حفظ وابستگی‌های طولانی مدت صورت می‌گیرد. پس از آن یک لایه تماماً متصل قرار گرفته است. در نهایت پیش‌بینی غلظت آلاینده برای ۱۳ ایستگاه سنجش آلودگی به طور هم‌زمان برای زمان‌های $t+1, t+2, \dots, t+24$ انجام می‌شود.

تأخیر زمانی^۱ یک پارامتر مهم در مدل پیش‌بینی محسوب می‌شود. تأخیر زمانی طولانی موجب افزایش پارامترهای غیر مرتبط به مدل شده و در مقابل تأخیر زمانی کوتاه نمی‌تواند ورودی مناسبی برای مدل پیش‌بینی باشد بدین منظور با توجه به تحقیقات [۳، ۲۴]، تأخیر زمانی بر روی عدد ۸ ساعت تنظیم شده است.



شکل ۴- ساختار کلی مدل پیشنهادی

در این پژوهش از تکنیک حذف تصادفی استفاده شده است که در آن معماری یک شبکه با حذف تصادفی نورون‌ها و اتصالات پنهان از شبکه، در هر تکرار اصلاح می‌شود. در این روش، به دلیل اینکه یک نورون نمی‌تواند به حضور نورون‌های دیگر متکی باشد، انتظار می‌رود یک نورون ویژگی‌های مقاوم‌تری را یاد بگیرد که این فرایند

موجب بهبود عملکرد شبکه می‌شود [۳۶]. توقف زودهنگام یکی دیگر از تکنیک‌هایی است که از بیش برازش شبکه جلوگیری می‌کند و در این پژوهش مورد استفاده قرار گرفته است. اگر عملکرد یک مدل در مجموعه داده‌های اعتبارسنجی نتواند بعد از تعداد اختیاری از دوره‌ها بهبود یابد، این روش آموزش را متوقف می‌کند. روش به هنگام سازی وزن‌ها که به معنی کاهش اختلاف بین مقادیر واقعی و پیش‌بینی شده است، توسط الگوریتم بهینه‌ساز انجام می‌شود. تاکنون الگوریتم‌های بسیاری همانند Adadelata، Adam، Adagrad، RMSprop معرفی شده‌اند که انتخاب الگوریتم مناسب موجب افزایش سرعت یادگیری شبکه می‌شود [۳۷].

یکی از روش‌های محبوب برای تعیین فرامترهای شبکه، الگوریتم جست‌وجوی شبکه‌ای نام دارد. سازوکار این روش به این ترتیب است که مجموعه‌ای از پارامترهای موردنظر همانند سائز دسته‌ها، الگوریتم‌های بهینه‌ساز را دریافت کرده و مدل را با ترکیب‌های متفاوت از این پارامترها آموزش می‌دهد تا در نهایت پارامتر بهینه با کمتر شدن تابع خطا مشخص شود. این روش کامل‌ترین روش محاسباتی برای تعیین پارامترهای بهینه‌ی شبکه است [۳۸]. در این پژوهش به منظور تعیین بهینه‌ی پارامترهای آموزش مدل شامل الگوریتم‌های بهینه‌سازی و تعداد سائز دسته‌ها از الگوریتم جست‌وجوی شبکه‌ای استفاده شده است. لازم به ذکر است که محدوده‌ی جست و جو به منظور محاسبه مقادیر بهینه بر اساس تحقیقات [۳، ۲۰، ۳۹] تعیین شده است. تابع هزینه مشخص می‌کند که چه پاسخی به جواب بهینه نزدیک‌تر است که در این پژوهش از تابع هزینه MAE استفاده شده است. جزئیات فرا پارامترهای مدل ارائه شده در جدول (۳) نمایش داده شده است.

به منظور ارزیابی عملکرد مدل پیشنهادی، روش‌های مقایسه‌ای LSTM و SVR مورد استفاده قرار گرفته‌اند. ۸۰ درصد داده‌ها در بازه‌ی زمانی ۲ سال کامل به عنوان داده‌ی آموزشی و ۲۰ درصد داده‌ها به عنوان داده‌ی آزمایشی تعیین شده است. بازه‌ی داده‌ی آموزشی و آزمایشی در سه روش ارائه شده به منظور مقایسه عملکرد روش‌ها، یکسان در نظر گرفته شده است. لازم به ذکر است که هر آزمایش ۳ بار تکرار شده و نتایج ارزیابی به صورت میانگین آزمایش‌ها در نظر گرفته شده است.

^۱ Time lag

مختلف کرنل گوسی شعاعی^۱ برای هدف پیش‌بینی، عملکرد بهتری دارد [۴۱]. بر این اساس در این مدل از کرنل گوسی شعاعی استفاده شده است.

۴- بحث و نتایج

برای ارزیابی عملکرد مدل‌های ارائه شده از شاخص‌های ارزیابی جذر میانگین مربعات خطا^۲ و میانگین مطلق خطا^۳ که نزدیکی مقادیر پیش‌بینی شده به مقادیر مشاهداتی را اندازه‌گیری می‌کند، به صورت روابط (۱۷) و (۱۸) استفاده شده است [۳]. همچنین برای تجزیه و تحلیل قدرت پیش‌بینی مدل‌های مختلف، ضریب تعیین و ضریب همبستگی نیز مطابق با روابط (۱۹) و (۲۰) به کار گرفته شده است [۴۲].

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \right]^{1/2} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (18)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n [P_i - O_i]^2}{\sum_{i=1}^n [P_i - \bar{O}]^2} \quad (19)$$

$$r = \frac{\sum_t (P_i - \hat{p}_i)(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (P_i - \bar{p})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (20)$$

جهت اعتبار سنجی مدل‌های پیش‌بینی در روابط فوق N بیان‌کننده تعداد داده‌ها، O_i مقادیر مشاهداتی، P_i مقادیر برآورد شده و $\bar{O}$ میانگین مقادیر مشاهداتی می‌باشد. جدول (۴) نتایج پیش‌بینی غلظت آلاینده بر اساس شاخص‌های ارزیابی ضریب تعیین و ضریب همبستگی برای یک ساعت آینده در ۱۳ ایستگاه سنجش آلودگی در مدل پیشنهادی را نمایش می‌دهد.

برای پیش‌بینی در یک ساعت آینده مطابق با جدول (۴) ایستگاه ستاد بحران با ضریب تعیین ۰٫۹۶ و ضریب همبستگی ۰٫۹۸ بهترین نتیجه و ایستگاه گلبرگ با ۰٫۶۲ و ضریب همبستگی ۰٫۸ ضعیف‌ترین نتیجه را در بین ۱۳ ایستگاه سنجش آلودگی کسب کرده است. مقادیر

در روش LSTM هر ایستگاه به همراه اطلاعات هواشناسی مربوطه به عنوان ویژگی وارد شبکه می‌شود و تأخیر زمانی همانند روش GRU، مقدار ۸ ساعت در نظر گرفته شده است. تعداد ویژگی‌ها در مدل برابر عدد ۸ بیانگر مقادیر آلاینده ایستگاه مورد نظر به همراه ۷ پارامتر هواشناسی می‌باشد. معماری این شبکه با سباز دسته ۱۶، تابع هزینه MAE و بهینه‌ساز Adagrad از دو لایه LSTM با ۱۶ نورون و تابع فعال ساز Tanh به همراه یک لایه Dropout و لایه تماماً متصل تشکیل شده است. برای تعیین پارامترهای بهینه‌ساز و سباز دسته‌ها از الگوریتم جست و جوی شبکه‌ای استفاده شده است. همچنین تعداد واحدهای هر لایه LSTM و تابع فعال ساز از طریق روش آزمون و خطا تعیین شده است.

جدول ۳- تنظیمات مدل ارائه شده بر اساس GRU

مقدار	پارامتر
۲	تعداد لایه GRU
۳۲-۳۲	تعداد بلوک GRU
۲	تعداد لایه تماماً متصل
۱-۲۴	زمان پیش‌بینی بر حسب ساعت
tanh	تابع فعالیت
MAE	تابع هزینه
۸۰۰	تعداد دوره
۳۲	سباز دسته‌ها
۱۰٪	حذف تصادفی
Adagrad	تابع بهینه‌ساز

ورودی مدل SVR شامل غلظت آلاینده ایستگاه هدف به همراه پارامترهای هواشناسی می‌باشد. همچنین مشابه روش LSTM پیش‌بینی برای هر ایستگاه جداگانه انجام می‌شود. پارامترهای این مدل شامل ضریب کرنل، پارامتر جریمه‌ی جمله‌ی خطا (c) و پارامتر دقت (ε) به ترتیب عدد ۱۰۰، ۰٫۰۰۰۴ و ۰٫۱ در نظر گرفته شده است. جست و جوی پارامترها از طریق اعتبار سنجی متقابل با ۱۰ بار تکرار انجام شده است. لازم به ذکر است که برای بهینه‌سازی پارامترها ابتدا مقادیر آن در یک بازه‌ی بزرگ انتخاب و پس از پیدا کردن محدوده تقریبی جواب‌ها، بازه‌ی کوچک‌تری تعیین می‌شود [۴۰]. در مدل رگرسیون ماشین بردار پشتیبان از کرنل‌های مختلفی همانند کرنل خطی، کرنل شعاعی درجه دوم، کرنل گوسی و چندجمله‌ای استفاده می‌شود که در این بین در تحقیقات

^۱ Radial basis function

^۲ Root mean square error (RMSE)

^۳ Mean absolute error

پیش‌بینی شده در مقابل مقادیر مشاهداتی در یک بخش از داده‌های آزمایشی در ایستگاه ستاد بحران و گلبرگ به صورت شکل (۵) و (۶) ارائه گردیده است.

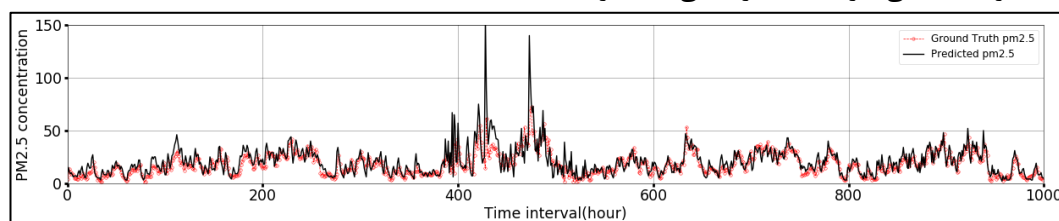
جدول ۴- نتایج مدل پیش‌بینی در روش پیشنهادی بر اساس GRU

نام ایستگاه	R-Squared	R-coefficient
شهر ری	۰/۸۹	۰/۹۵
شهرداری منطقه ۴	۰/۷۸	۰/۹۱
پیروزی	۰/۸۲	۰/۹۱
پونک	۰/۷۷	۰/۹
گلبرگ	۰/۶۲	۰/۸
ستاد بحران	۰/۹۶	۰/۹۸
شهرداری منطقه ۲۱	۰/۷۵	۰/۷۸
مسعودیه	۰/۶۹	۰/۸۴
شادآباد	۰/۷۷	۰/۸۸
شریف	۰/۸۳	۰/۹۱
مدرس	۰/۸	۰/۹
صدر	۰/۸۳	۰/۹۳
اقدسیه	۰/۸۷	۰/۹۱

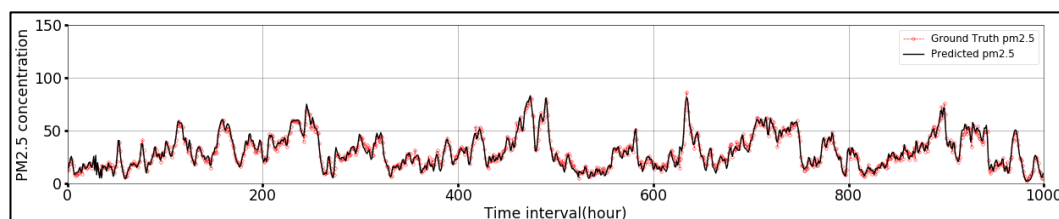
مطابق با اشکال (۵) و (۶) نزدیکی مقادیر پیش‌بینی شده به مقادیر مشاهداتی در داده‌ی آزمایشی، عملکرد

خوب مدل پیشنهادی را در ثبت تغییرات غلظت آلاینده نشان می‌دهد. اما باید به این نکته توجه داشت که مدل پیش‌بینی در فراز و فرودهای ناگهانی در غلظت بالای آلودگی قدرت پیش‌بینی را از دست داده است و علت پایین بودن دقت در ایستگاه گلبرگ را می‌توان به این عامل نسبت داد اما با این وجود روند تغییرات غلظت آلاینده را به درستی پیش‌بینی کرده است. اما در داده‌ی آزمایشی ایستگاه ستاد بحران، پیش‌بینی با دقت بسیار بالایی انجام شده و تغییرات بسیار شدید و ناگهانی مشاهده نشده است.

عملکرد مدل‌های GRU، LSTM و SVR در پیش‌بینی غلظت آلاینده از ۱ تا ۲۴ ساعت آینده با شاخص‌های ارزیابی ضریب تعیین، RMSE و MAE به صورت جدول (۵) مشاهده می‌شود. لازم به ذکر است که در روش‌های LSTM و SVR که پیش‌بینی بر روی هر ایستگاه جداگانه انجام شده است، معیارهای ارزیابی داده‌های هر ایستگاه محاسبه شده و میانگین آن در جدول مذکور ارائه شده است.



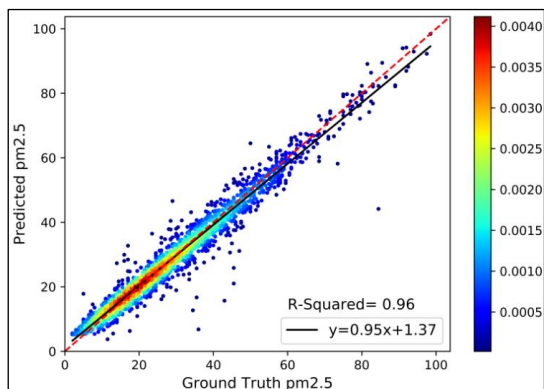
شکل ۵- مقایسه‌ی غلظت آلاینده‌ی پیش‌بینی شده توسط پیشنهادی بر اساس شبکه GRU در مقابل داده‌ی مشاهداتی ایستگاه گلبرگ



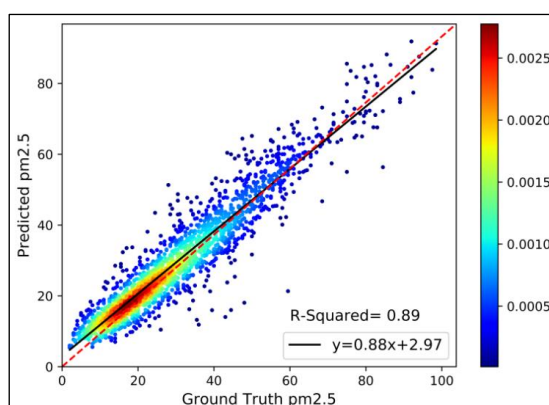
شکل ۶- مقایسه‌ی غلظت آلاینده‌ی پیش‌بینی شده توسط پیشنهادی بر اساس شبکه GRU در مقابل داده‌ی مشاهداتی ایستگاه ستاد بحران

جدول ۵- مقایسه مدل‌های پیش‌بینی آلودگی GRU، LSTM و SVR در گام‌های زمانی پیش‌بینی مختلف

Prediction Time	RMSE ($\mu\text{g}/\text{m}^3$)			MAE ($\mu\text{g}/\text{m}^3$)			R Squared		
	GRU	LSTM	SVR	GRU	LSTM	SVR	GRU	LSTM	SVR
1 h	۷/۹۷	۸/۵۹	۹/۱۸	۵/۳۵	۵/۹۳	۷/۱۰	۰/۸	۰/۷۶	۰/۷
2 h	۱۰/۳۵	۱۱/۳۶	۱۱/۴۵	۷/۲۷	۸/۱	۸/۹۲	۰/۶۶	۰/۵۹	۰/۵۳
3 h	۱۲/۰۲	۱۲/۶۶	۱۲/۶۱	۸/۴۹	۸/۹۱	۹/۸۶	۰/۵۵	۰/۵۲	۰/۴۵
4 h	۱۲/۸	۱۳/۹۷	۱۳/۵۰	۹/۷۳	۹/۷۵	۱۰/۵۹	۰/۵	۰/۴۲	۰/۳۷
5 h	۱۳/۴۲	۱۴/۵۸	۱۴/۰۹	۹/۶۲	۱۰/۲۲	۱۱/۱۱	۰/۴۴	۰/۳۷	۰/۳۲
6 h	۱۴/۰۱	۱۴/۸۶	۱۵/۴۷	۱۰/۰۱	۱۰/۵۳	۱۱/۳۹	۰/۳۸	۰/۳۴	۰/۲۸
12 h	۱۵/۵۵	۱۵/۸۱	۱۶/۶۷	۱۱/۳۷	۱۱/۵۴	۱۲/۴	۰/۲۵	۰/۲۳	۰/۱۷
24 h	۱۶/۶۸	۱۶/۷۳	۱۷/۱۲	۱۲/۴۵	۱۲/۶۵	۱۳/۷۸	۰/۱۷	۰/۱۲	۰/۰۲



شکل ۷- نمودار پراکندگی مقادیر پیش‌بینی شده و مقادیر مشاهداتی (روش پیشنهادی بر اساس شبکه GRU)



شکل ۸- نمودار پراکندگی مقادیر پیش‌بینی شده و مقادیر مشاهداتی (روش LSTM)

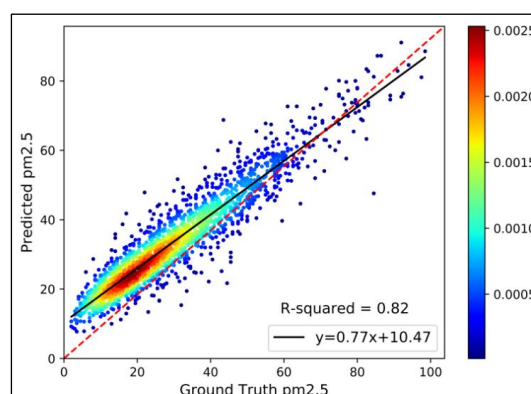
۵- نتیجه‌گیری

در سال‌های اخیر، محققان مدل‌های پیشرفته‌ای را برای پیش‌بینی غلظت آلاینده هوا پیشنهاد داده‌اند. در این پژوهش مدل‌های یادگیری عمیق و یادگیری ماشین همانند LSTM و SVR با یک مدل پیشنهادی بر مبنای GRU مورد مقایسه واقع شده‌اند. روش پیشنهادی با حفظ وابستگی‌های طولانی مدت می‌تواند همبستگی‌های مکانی و زمانی را در داده‌های غلظت آلاینده هوا به طور خودکار و مؤثر استخراج کند.

نتایج به دست آمده از این پژوهش حاکی از برتری روش پیشنهادی مکانی-زمانی GRU با شاخص ارزیابی RMSE و MAE برابر مقدار $7.97 \mu\text{g}/\text{m}^3$ و $5.35 \mu\text{g}/\text{m}^3$ می‌باشد. علاوه بر این، مدل پیشنهادی با ضریب تعیین (۰.۸) می‌تواند ۸۰ درصد از سطح آلودگی هوا را بر اساس غلظت $\text{PM}_{2.5}$ پیش‌بینی کرده و در واقع مدل پیشنهادی با لحاظ کردن وابستگی‌های مکانی-زمانی، در

همان‌طور که در جدول (۵) مشاهده می‌شود، برای پیش‌بینی در یک ساعت آینده برای روش پیشنهادی در مقایسه با روش‌های دیگر با ضریب تعیین ۰.۸، شاخص MAE و RMSE با مقادیر $5.35 \mu\text{g}/\text{m}^3$ و $7.97 \mu\text{g}/\text{m}^3$ با اختلاف کمترین خطا، بیش‌ترین ضریب تعیین و ضریب همبستگی را در برداشته است. پس از آن شاخص‌های ارزیابی ضریب تعیین، MAE و RMSE در روش LSTM دارای مقادیر ۰.۷۶، $5.93 \mu\text{g}/\text{m}^3$ و $8.59 \mu\text{g}/\text{m}^3$ و در روش SVR دارای مقادیر ۰.۷، $7.10 \mu\text{g}/\text{m}^3$ و $9.18 \mu\text{g}/\text{m}^3$ می‌باشد. به طور کلی روش‌های یادگیری عمیق نتایج بهتری نسبت به روش SVR به دست آورده است.

اشکال (۷) تا (۹) نمودارهای پراکندگی مقادیر پیش‌بینی شده در مقابل داده‌های مشاهداتی در داده‌های آزمایشی ایستگاه ستاد بحران برای دو روش LSTM، SVR و روش پیشنهادی بر اساس GRU را نمایش می‌دهند.



شکل ۹- نمودار پراکندگی مقادیر پیش‌بینی شده و مقادیر مشاهداتی (روش SVR)

این نمودارها ارتباط بین مقادیر پیش‌بینی شده و مقادیر واقعی را به صورت یک معادله خطی درجه اول نمایش داده است. در نمودار اشکال ۷ تا ۹ خط‌چین قرمز بیانگر خط $x=y$ نیم ساز ربع اول می‌باشد. هرچه میزان انحراف خط رگرسیون نسبت به خط نیم ساز ربع اول کمتر باشد در آن صورت توانایی مدل در پیش‌بینی آلودگی هوا بیشتر می‌شود و به دنبال آن مقادیر شاخص ارزیابی ضریب تعیین و ضریب همبستگی نزدیک عدد ۱ خواهند شد. همان‌طور که در شکل (۷) مشاهده می‌شود تراکم نقاط در نمودار پراکندگی حول خط نیم ساز ربع اول است که توانایی بالای مدل پیشنهادی در پیش‌بینی غلظت آلاینده در ایستگاه ستاد بحران را نمایش می‌دهد.

سری زمانی و استخراج ویژگی‌های پیچیده قادر به شناسایی الگوی تغییرات غلظت آلاینده می‌باشد.

در مدل ارائه شده توسط زمانی و همکاران (۲۰۱۹) به منظور پیش‌بینی غلظت آلاینده PM_{2.5} با به‌کارگیری مدل LSTM بر اساس شاخص ارزیابی RMSE و MAE مقدار $13,58 \mu\text{g}/\text{m}^3$ و $9,93 \mu\text{g}/\text{m}^3$ گزارش شده است [۲۱]. همچنین نتایج مدل ارائه شده توسط کائیمیان و همکاران (۲۰۱۹) جهت پیش‌بینی غلظت آلاینده در ۹ ایستگاه سنجش آلودگی شهر تهران با استفاده از مدل LSTM بر مبنای شاخص‌های ارزیابی RMSE و MAE برابر مقدار $8,91 \mu\text{g}/\text{m}^3$ و $6,21 \mu\text{g}/\text{m}^3$ گزارش شده

است [۲۰]. که نشان دهنده برتری نتایج مدل پیشنهادی در این پژوهش بر اساس شبکه بازگشتی GRU در مقایسه با پژوهش‌های [۲۱] و [۲۰] می‌باشد.

مدل ارائه شده در این پژوهش از داده‌های قبلی برای پیش‌بینی مقادیر آلاینده استفاده می‌کند و فقدان داده در سری زمانی موجب به هم خوردن توالی و ترتیب داده‌ها و کاهش قدرت پیش‌بینی در بلندمدت شده است. در پژوهش‌های آ آینده با افزودن پارامترهای تأثیرگذار بر آلودگی هوا به عنوان داده‌های کمکی همانند پارامترهای ترافیک و عمق نوری هواویز می‌توان دقت مدل پیش‌بینی را افزایش داد.

مراجع

- [1] A. Shamsoddini, M. Aboodi, and J. Karami, "Tehran air pollution prediction based on random forest feature selection method," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, 2017.
- [2] J. Ma, Y. Ding, J. C. Cheng, F. Jiang, V. J. Gan, and Z. Xu, "A Lag-FLSTM deep learning network based on Bayesian Optimization for multi-sequential-variant PM_{2.5} prediction," *Sustainable Cities and Society*, vol. 60, p. 102237, 2020.
- [3] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, and T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Science of The Total Environment*, vol. 654, pp. 1091-1099, 2019.
- [4] L. Bai, J. Wang, X. Ma, and H. Lu, "Air Pollution Forecasts: An Overview," *International journal of environmental research and public health*, vol. 15, no. 4, p. 780, 2018.
- [5] F. Tuna and M. Buluc, "Analysis of PM₁₀ pollutant in Istanbul by using Kriging and IDW methods: between 2003 and 2012," *Int J Comput Inf Technol*, vol. 4, no. 1, pp. 170-5, 2015.
- [6] A. M. Vicedo-Cabrera, A. Biggeri, L. Grisotto, F. Barbone, and D. Catelan, "A Bayesian kriging model for estimating residential exposure to air pollution of children living in a high-risk area in Italy," *Geospatial health*, pp. 87-95, 2013.
- [7] J.-H. Lee, C.-F. Wu, G. Hoek, K. de Hoogh, R. Beelen, B. Brunekreef, and C.-C. Chan, "LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction," *Science of the Total Environment*, vol. 514, pp. 178-184, 2015.
- [8] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [9] C. Li, N. C. Hsu, and S.-C. Tsay, "A study on the potential applications of satellite data in air quality monitoring and forecasting," *Atmospheric Environment*, vol. 45, no. 22, pp. 3663-3675, 2011.
- [10] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak, and C. Pak, "Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: A case study of Beijing, China," *Science of The Total Environment*, vol. 699, p. 133561, 2020.
- [11] P. Perez and A. Trier, "Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, Chile," *Atmospheric Environment*, vol. 35, no. 10, pp. 1783-1789, 2001.
- [12] T. Li, M. Hua, and X. Wu, "A hybrid CNN-LSTM model for forecasting particulate matter (PM_{2.5})," *IEEE Access*, vol. 8, pp. 26933-26940, 2020.
- [13] C.-J. Huang and P.-H. Kuo, "A deep cnn-lstm model for particulate matter (PM_{2.5}) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, 2018.
- [14] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework," *arXiv preprint arXiv:1812.04783*, 2018.
- [15] V. Athira, P. Geetha, R. Vinayakumar, and K. Soman, "DeepAirNet: Applying Recurrent Networks for Air Quality Prediction," *Procedia Computer Science*, vol. 132, pp. 1394-1403, 2018.

- [16] V. Reddy, P. Yedavalli, S. Mohanty, and U. Nakhat, "Deep Air: Forecasting Air Pollution in Beijing, China," pp. 1-8, 2017.
- [17] N. Kamali, M. Zare Shahne, and M. Arhami, "Implementing spectral decomposition of time series data in artificial neural networks to predict air pollutant concentrations," *Environmental Engineering Science*, vol. 32, no. 5, pp. 379-388, 2015.
- [18] M. Memarianfard and A. Hatami, "Artificial neural network forecast application for fine particulate matter concentration using meteorological data," 2017.
- [19] A. Jamal and R. N. Nodehi, "Predicting air quality index based on meteorological data: A comparison of regression analysis, artificial neural networks and decision tree," *Journal of Air Pollution And Health*, vol. 2, no. 1, 2017.
- [20] H. Kaimian, Q. Li, C. Wu, Y. Qi, Y. Mo, G. Chen, X. Zhang, and S. Sachdeva, "Evaluation of different machine learning approaches to forecasting PM2. 5 mass concentrations," *Aerosol and Air Quality Research*, vol. 19, no. 6, pp. 1400-1410, 2019.
- [21] M. Zamani Joharestani, C. Cao, X. Ni, B. Bashir, and S. Talebiesfandarani, "PM2. 5 Prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, no. 7, p. 373, 2019.
- [22] A. H. Al Hanai, D. S. Antkiewicz, J. D. Hemming, M. M. Shafer, A. M. Lai, M. Arhami, V. Hosseini, and J. J. Schauer, "Seasonal variations in the oxidative stress and inflammatory potential of PM2. 5 in Tehran using an alveolar macrophage model; The role of chemical composition and sources," *Environment international*, vol. 123, pp. 417-427, 2019.
- [23] K. Pearson, "VII. Note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347-352, pp. 240-242, 1895.
- [24] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, and T. Chi, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environmental Pollution*, vol. 231, pp. 997-1004, 2017.
- [25] S. Li, G. Xie, J. Ren, L. Guo, Y. Yang, and X. Xu, "Urban PM 2.5 Concentration Prediction via Attention-Based CNN-LSTM," *Applied Sciences*, vol. 10, no. 6, p. 1953, 2020.
- [26] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85-117, 2015.
- [27] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM 2.5 Concentration," *IEEE Access*, vol. 7, pp. 20050-20059, 2019.
- [28] L. Finsveen, "Time-series predictions with Recurrent Neural Networks-Studying Recurrent Neural Networks predictions and comparing to state-of-the-art Sequential Monte Carlo methods," *NTNU*, 2018.
- [29] V. Reddy, P. Yedavalli, S. Mohanty, and U. Nakhat, "Deep Air: Forecasting Air Pollution in Beijing, China," ed, 2018.
- [30] Y. A. Ayturan, Z. C. Ayturan, H. O. Altun, C. Kongoli, F. D. Tunccez, S. Dursun, and A. Ozturk, "Short-term prediction of PM2. 5 pollution with deep learning methods," *GLOBAL NEST JOURNAL*, vol. 22, no. 1, pp. 126-131, 2020.
- [31] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [33] V. N. Vapnik, "The nature of statistical learning theory," 1995.
- [34] A. Ardajani and V. R. Kohestani, "Prediction of Lateral Bearing Capacity of Pile in Clay Using Support Vector Machine," (in Persian), *Journal of Civil and Environmental Engineering*, vol. 47, no. 2, pp. 1-10, 2017.
- [35] A. Ng, "Cs229 lecture notes, part v: Support vector machines," *Technical report*, 2012.
- [36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
- [37] M. Farnaghi and H. Rahimi, "Spatio-Temporal Prediction of Monthly Rainfall using Deep Neural Network: A Case Study in North-west Iran," 2017.
- [38] M. Claesen and B. De Moor, "Hyperparameter search in machine learning," *arXiv preprint arXiv:1502.02127*, 2015.
- [39] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory," *Science of the Total Environment*, vol. 664, pp. 1-10, 2019.

- [40] M. Kashipazan Qomi and B. Shirgir, "Short-term Prediction of Particulate Matter Related to Traffic Flow Using Support Vector Machine (Case Study: Tehran City)," (in Persian), Modares Civil Engineering journal, vol. 18, no. 4 ,pp. 201-210, 2018.
- [41] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," Neural networks, vol. 17, no. 1, pp. 113-126, 2004.
- [42] N. Mohammadi, K. Zoroufchi , M. Shakerkhatibi, E. Fatehifar, A. Behrooz Sarand , A. Mahmoudian, and F. Sheikholeslami, "Forecasting Concentrations of Gaseous Air Pollutants Using Artificial Neural Networks in Tabriz," (in Persian), Journal of Civil and Environmental Engineering, vol. 46, no. 2, 2016.