

ارزیابی مدل درختی M5 و شبکه عصبی مصنوعی برای برآورد متوسط روزانه دمای هوا بر اساس داده‌های دمای سطح زمین سنجنده مودیس

سعید امامی فر^۱، علی رحیمی خوب^{۲*}، علی اکبر نوروزی^۳

۱. کارشناس ارشد گروه مهندسی آبیاری و زهکشی پردیس ابوریحان دانشگاه تهران

۲. دانشیار گروه مهندسی آبیاری و زهکشی پردیس ابوریحان دانشگاه تهران

۳. استادیار پژوهشکده حفاظت خاک و آبخیزداری

(تاریخ دریافت: ۱۳۹۳/۲/۳ - تاریخ تصویب: ۱۳۹۳/۶/۱۷)

چکیده

استفاده از داده‌های تصاویر ماهواره‌ای روشی مؤثر برای پهنه‌بندی دمای هواست. در این تحقیق مدل‌های شبکه عصبی و مدل درختی M5 برای تبدیل دمای سطح زمین در دو زمان روز و شب محصولات ماهواره ترا سنجنده مودیس برای استان خوزستان مقایسه شد. در کل ۳۶۵ تصویر از محصولات دمای سطح زمین در سال ۲۰۰۷، که منطقه مورد مطالعه را پوشش می‌داد، استفاده شد. داده‌های متوسط دمای هوای روزانه از ۲۹ ایستگاه هواشناسی سینوپتیک و کلیماتولوژی سال ۲۰۰۷ جمع‌آوری و به منزله داده‌های واقعی استفاده شدند. داده‌های ورودی مدل‌ها شامل دمای سطح زمین در دو زمان روز و شب و تابش بیرون زمینی بودند. نتایج نشان داد ضریب تعیین هر دو مدل بیش از ۰/۹۶ است. با این حال مدل شبکه عصبی با دقت بیشتری دمای هوا را برآورد می‌کند. جذر مربع میانگین خطا و ضریب تعیین مدل شبکه عصبی به ترتیب برابر ۱/۷ درجه سانتی‌گراد و ۰/۹۷ برآورد شد.

کلیدواژگان: دمای سطح زمین، دمای هوا، سنجنده مودیس، شبکه عصبی، مدل درختی M5

مقدمه

دمای هوا یکی از عناصر اساسی شناخت هوا و از پارامترهای مهم مورد نیاز برای مطالعات هیدرولوژیکی و هواشناسی است. بسیاری از فرایندهای سطح زمین، مانند فتوسنتز و تبخیر و تعرق، به کمک آن تنظیم می‌شوند. این پارامتر در ایستگاه‌های هواشناسی اندازه‌گیری می‌شود و معرف دمای هوای پیرامون ایستگاه است و نمی‌توان آن را به مناطق دور از ایستگاه تعمیم داد. زیرا دمای هوا تحت تأثیر خصوصیات سطح زمین است و این خصوصیات با توجه به زمان و مکان تغییر می‌کند. داده‌های ماهواره‌ای، به علت پیوستگی مکانی داده‌ها، نسبت به داده‌های ایستگاه‌های هواشناسی در مطالعات ناحیه‌ای کاربرد بهتری دارند (Rahimikhoob et al, 2008). تحقیقات مختلف نشان داد متغیرهای دمای سطح زمین و شاخص پوشش گیاهی، که با استفاده از تصاویر ماهواره به دست می‌آیند، بیشترین همبستگی را با دمای هوای نزدیک زمین دارند (Yao and Zhang, 2012). دو روش برای تعیین دمای هوا با استفاده از داده‌های تصاویر

ماهواره‌ای ارائه شده است. روش اول، به نام روش TVX^۱، مبتنی بر این فرض است که دمای سطح زمین در یک پوشش گیاهی کامل به دمای هوا نزدیک است. این روش در شمال ایالت کانزاس آمریکا ارزیابی شد (Prihodko and Goward, 1997). نتایج نشان داد همبستگی زیادی (ضریب همبستگی برابر ۰/۹۳) بین برآوردهای ماهواره‌ای و دمای هوای اندازه‌گیری شده در ایستگاه‌های هواشناسی با میانگین خطای ۲/۹۲ درجه سانتی‌گراد وجود دارد. اما بایاس ثابت مثبتی در برآوردهای ماهواره‌ای وجود داشت. در ایران نیز روش TVX برای تخمین دمای هوای حوضه آبریز سفیدرود بررسی شد و نتایج نشان داد این روش نسبت به تراکم پوشش گیاهی حساسیت دارد؛ طوری که در مناطق با تراکم کم پوشش گیاهی میزان خطا افزایش پیدا می‌کند (Parviz et al, 2011). روش دوم مبتنی بر استفاده از همبستگی آماری است. در این زمینه می‌توان به تحقیقی که در شمال آفریقا انجام شد اشاره کرد (Cresswell et al, 1999). در این مطالعه، از داده‌های دمای سطح زمین

1. Temperature-vegetation index

* نویسنده مسئول: akhob@ut.ac.ir

هدف این مطالعه مقایسه مدل شبکه عصبی و مدل درختی M5 در برآورد میانگین روزانه دمای هوا با استفاده از تصاویر دمای سطح زمین سنجنده مودیس است.

مواد و روش‌ها

محدوده مطالعه شده و منابع داده‌ها

استان خوزستان محدوده مطالعه این تحقیق بود که بین عرض جغرافیایی ۲۹/۹۵ و ۳۲/۹ درجه شمالی و طول جغرافیایی ۴۷/۶ و ۵۰/۶ درجه شرقی واقع شده است. این استان، بر اساس طبقه‌بندی اقلیمی کوپن، جزء طبقه خشک است. داده‌های اندازه‌گیری شده دمای بیشینه و کمینه هوا، به ترتیب T_n و T_x ، در سال ۲۰۰۷ از ۲۹ ایستگاه هواشناسی این استان (۱۲ ایستگاه کلیماتولوژی و ۱۷ ایستگاه سینوپتیک) جمع‌آوری شدند. شکل ۱ موقعیت این ایستگاه‌ها و توزیع آن‌ها را تقریباً در کل استان نشان می‌دهد. داده‌های میانگین روزانه دمای هوا (T_n) از متوسط‌گیری دمای بیشینه و کمینه هوا محاسبه شد و به مثابه داده‌های واقعی به کار رفت. مشخصات ایستگاه‌های استفاده شده در این تحقیق در جدول ۱ می‌آید.

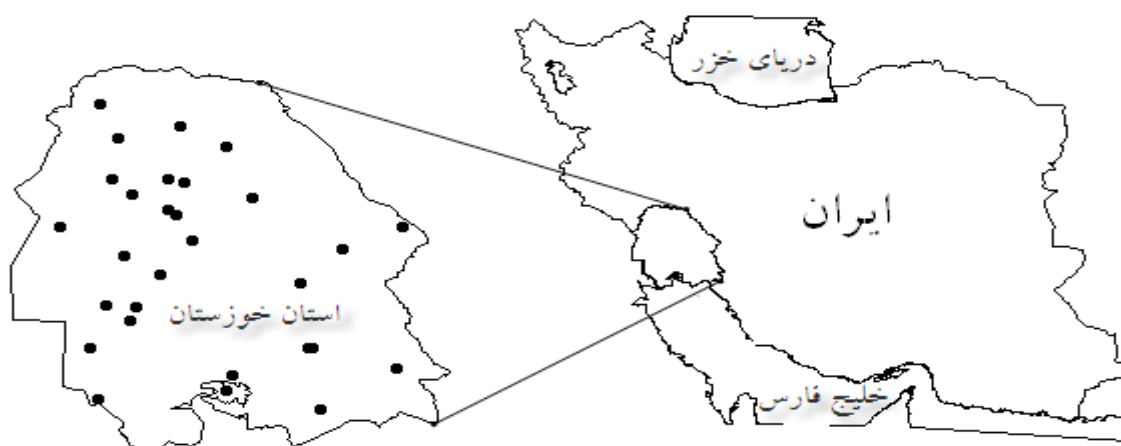
تصاویر مودیس^۲

در این تحقیق، داده‌های دمای سطح زمین در دو زمان روز و شب، حاصل تصاویر سنجنده مودیس، به منزله داده‌های ورودی مدل‌های شبکه عصبی و مدل درختی M5 استفاده شد. سنجنده مودیس، که بر دو ماهواره ترا و آکوا نصب می‌شود، تصاویر سطح زمین را در ۳۵ باند طیفی بین ۰/۶۲ و ۱۴/۳۸۵ میکرومتر با دقت مکانی ۲۵۰ متر، ۵۰۰ متر، و ۱ کیلومتر تهیه می‌کند. دقت مکانی این سنجنده در محدوده طیفی حرارتی ۱ کیلومتر است. ماهواره‌های ترا و آکوا به ترتیب در تاریخ‌های ۱۸ دسامبر ۱۹۹۹ و ۴ می ۲۰۰۲ در مدار قرار گرفتند. این ماهواره‌ها خورشیدآهنگ‌اند و در دو مسیر مخالف روزی دو بار دور زمین می‌گردند. ماهواره ترا از شمال به جنوب حرکت می‌کند و زمان‌های عبور آن از استوا حدود ۱۰:۳۰ صبح و شب به وقت محلی در هر روز است؛ حال آنکه ماهواره آکوا، در مسیری مخالف ماهواره ترا، از جنوب به شمال حرکت می‌کند و زمان‌های عبور آن از استوا حدود ۱:۳۰ صبح و بعد از ظهر به وقت محلی در هر روز است.

ثبت شده سنجنده متوست^۱ در هنگام شب همراه پارامترهای زاویه زنیت خورشیدی جهت برآورد مینیمم دمای هوا استفاده شد. نتایج نشان داد انحراف معیار برآوردها کمتر از ۳ درجه سانتی‌گراد است. در روش‌های فوق پیش‌بینی دمای هوا به میزان دقت برآورد شاخص پوشش گیاهی و دمای سطح زمین بستگی دارد. مقدار خطای هر دو روش TVX و همبستگی آماری برای برآورد دمای هوا حدود ۳ درجه سانتی‌گراد است (Yan et al, 2012).

در سال‌های اخیر روش‌های یادگیری ماشینی، مانند شبکه‌های عصبی و بردار پشتیبان ماشین و مدل‌های درختی M5، به طور موفقیت‌آمیزی در دامنه‌های مختلف علوم مهندسی استفاده شد. این روش‌ها برای مدل‌سازی فرایندهای فاقد راه‌حل دقیق کاربرد دارد. برای پهنه‌بندی مکانی دمای هوا در یکی از زیرحوزه‌های هندوستان از مدل‌های شبکه عصبی، به منظور تبدیل داده‌های دمای سطح زمین، و تصاویر لندست و روزشمار سال (روز ژولوسوس) همراه داده‌های اقلیمی حوزه، مثل رطوبت نسبی و تبخیر از تشت و تابش رسیده به زمین و سرعت باد، استفاده شد (Aher et al, 2011). در این تحقیق شاخص‌های آماری ضریب تعیین (R^2)، میانگین خطای مطلق (MAE)، و جذر میانگین مربعات خطا (RMSE) به ترتیب ۰/۹۸ و ۰/۳۴ و ۰/۳۶ درجه سانتی‌گراد برآورد شدند. در پژوهشی دیگر از داده‌های پنج باند ماهواره نوا، ارتفاع سطح زمین، زاویه سمت‌الارض خورشید، و روزشمار سال به منزله داده‌های ورودی مدل شبکه عصبی برای برآورد دمای هوا استفاده شد (Jang et al, 2004). نتایج نشان داد دقت برآورد دمای هوا حدود ۲ درجه سانتی‌گراد و ضریب همبستگی برابر ۰/۹۳ است. دقت برآورد دمای بیشینه هوا با استفاده از داده‌های ماهواره نوا و مدل شبکه عصبی در منطقه خوزستان حدود ۱/۷ درجه سانتی‌گراد تعیین شده است (Rahimikhoob et al, 2008).

مدل شبکه درختی M5 برای مدل‌سازی دمای هوا، با استفاده از تصاویر ماهواره‌ای، موفق ارزیابی شده است (Emamifar et al, 2013). مزیت شبکه درختی M5 نسبت به مدل شبکه عصبی این است که این مدل روابط جبری ساده‌ای بر اساس شروط «اگر-آن‌گاه» ارائه می‌دهد و شناخت پدیده را آسان‌تر می‌کند. کاربرد مدل شبکه درختی M5 در پیش‌بینی سیلاب (Solomatin and Xue, 2004)، روابط دبی، اشل (Bhattacharya and Solomatin, 2005)، و مدل سازی بارش-رواناب (Solomatin and Dulal, 2003) موفق بوده است.



شکل ۱. موقعیت ایستگاه‌های هواشناسی در استان خوزستان

جدول ۱. مشخصات ایستگاه‌های استفاده شده

شماره ایستگاه	نام ایستگاه	کد ایستگاه	طول جغرافیایی (°E)	عرض جغرافیایی (°N)	بیشینه دمای هوا (°C)	کمینه دمای هوا (°C)
۱	باغ‌ملک	BM	۴۹,۸۸	۳۱,۵۳	۲۹,۱	۱۷,۷
۲	بندر	BA	۴۹,۱۵	۳۰,۵۵	۳۲,۴	۱۸,۶
۳	اهواز	AH	۴۸,۶۷	۳۱,۳۳	۳۳,۳	۱۹,۲
۴	امبدیه	OM	۴۹,۶۷	۳۰,۷۷	۳۵,۱	۱۸,۳
۵	کشت	KE	۴۸,۷۷	۳۱,۸۰	۳۲,۳	۱۷,۲
۶	هفت‌تپه	HA	۴۸,۳۵	۳۲,۰۸	۳۲,۱	۱۷,۱
۷	دعبل‌خزایی	DE	۴۸,۵۰	۳۱,۰۸	۳۴,۱	۱۹,۳
۸	میناب	MI	۴۸,۴۸	۳۱,۹۷	۳۱,۹	۱۷,۳
۹	مسجد سلیمان	MA	۴۹,۲۸	۳۱,۹۳	۳۲,۱	۱۹,۶
۱۰	شوشر	SH	۴۸,۸۳	۳۲,۰۵	۳۳,۰	۲۰,۳
۱۱	سلمان‌فارسی	SA	۴۸,۴۷	۳۰,۹۸	۳۴,۲	۱۹,۴
۱۲	آبادان	AB	۴۸,۲۵	۳۰,۳۷	۳۳,۸	۱۸,۹
۱۳	ملاثانی	MO	۴۸,۸۸	۳۱,۶۰	۳۲,۱	۱۸,۱
۱۴	امیرکبیر	AM	۴۸,۳۰	۳۱,۱۰	۳۴,۳	۱۹,۵
۱۵	بهبهان	BE	۵۰,۲۳	۳۰,۶۰	۳۲,۷	۱۷,۳
۱۶	بوستان	BO	۴۸,۰۰	۳۱,۷۲	۳۳,۳	۱۶,۵
۱۷	دریاچه	DA	۴۹,۱۰	۳۰,۴۳	۳۲,۴	۱۸,۶
۱۸	دهدز	DH	۵۰,۲۷	۳۱,۷۲	۲۸,۳	۱۷,۱
۱۹	دزفول	DZ	۴۸,۳۸	۳۲,۴۰	۳۲,۱	۱۶,۳
۲۰	حمیدیه	HA	۴۸,۴۳	۳۱,۴۸	۳۳,۰	۱۸,۹
۲۱	هندیجان	HE	۴۹,۷۳	۲۸,۲۸	۳۵,۷	۱۹,۳
۲۲	حسینیه	HO	۴۸,۲۷	۳۲,۶۷	۳۲,۷	۱۶,۰
۲۳	ایزه	IZ	۴۸,۷۲	۳۱,۸۵	۲۸,۴	۱۳,۹
۲۴	کشت‌وصنعت	KS	۴۸,۷۲	۳۲,۰۸	۳۲,۸	۱۹,۸
۲۵	لالی	LA	۴۹,۱۰	۳۲,۳۳	۳۰,۱	۱۷,۵
۲۶	میرزا	MR	۴۸,۲۰	۷۷,۳۰	۳۴,۷	۱۹,۷
۲۷	امبدیه	OI	۴۹,۶۵	۷۷,۳۰	۳۴,۷	۱۸,۷
۲۸	رامهرمز	RA	۴۹,۶۰	۳۱,۲۷	۳۲,۷	۱۹,۷
۲۹	سردشت	SA	۴۸,۸۰	۳۲,۵۰	۲۹,۱	۱۶,۵

Frank, 2005). مدل رگرسیون برای کل فضای داده‌ها یک معادله رگرسیون ارائه می‌دهد. ولی در مدل رگرسیون درختی محدوده داده‌ها به زیرناحیه‌هایی که برگ نامیده می‌شوند تقسیم و به هر برگ یک برجسب عددی داده می‌شود. جایگزینی معادله رگرسیون خطی به جای برجسب در گره‌ها شیوه‌ای است که در مدل M5 اجرا می‌شود و می‌تواند متغیرهای عددی پیوسته را پیش‌بینی یا برآورد کند (Quinlan, 1986). ساختار درخت تصمیم‌گیری شبیه درختی است که از ریشه^۶، شاخه‌ها^۷، گره‌ها^۸، و برگ‌ها^۹ تشکیل شده است. درخت تصمیم از بالا به پایین ترسیم می‌شود. ریشه، به منزله اولین گره، در بالا قرار می‌گیرد و زنجیره‌ای از شاخه‌ها و گره‌ها به برگ‌ها ختم می‌شود. هر گره مربوط به یک متغیر پیش‌بینی‌کننده است و به وسیله شاخه‌ها عمل انشعاب در گره انجام می‌شود. شاخه‌ها شامل بازه‌ای عددی است که از گره والد^{۱۰} منشعب می‌شود و به گره فرزند^{۱۱} می‌رسد. در مدل M5 از هر گره والد دو شاخه منشعب می‌شود.

ساخت مدل درخت تصمیم‌گیری در دو مرحله انجام می‌شود. در مرحله اول درخت تصمیم‌گیری با انشعاب‌سازی داده‌ها تشکیل می‌شود. معیار انشعاب در مدل M5 بیشینه‌سازی کاهش انحراف معیار داده‌ها در گره فرزند است. زمانی که امکان کاهش انحراف معیار داده‌های گره فرزند میسر نباشد گره والد آن منشعب نمی‌شود و به گره پایانی یا برگ نمی‌رسد. کاهش انحراف معیار با رابطه ۱ برآورد می‌شود (Pal and Deswal, 2009):

$$\text{SDR} = \text{sd}(T) - \sum \frac{|T_i|}{|T|} \text{sd}(T_i) \quad (\text{رابطه ۱})$$

SDR کاهش انحراف معیار در گره فرزند، T مجموعه داده‌های ورودی به گره والد، T_i زیرمجموعه داده‌های ورودی به گره والد، و sd انحراف معیار است.

به علت فرایند انشعاب، انحراف معیار داده‌ها در گره فرزند کمتر از گره والد است. بنابراین همگنی بیشتری دارد. در مدل M5، بعد از آزمودن همه انشعابات ممکن از یک گره، انشعابی که بیشترین کاهش انحراف معیار را تولید کند انتخاب می‌شود.

سازمان ملی هوا و فضای امریکا (ناسا) عهده‌دار پردازش داده‌های سنجنده مودیس است و محصولات متنوعی برای کاربردهای زمینی و اقیانوسی و جوئی تولید و عرضه می‌کند. این محصولات از طریق وبسایت این سازمان قابل دسترس است. در این تحقیق ۳۶۵ تصویر با کد محصول MOD11A1 مربوط به سال ۲۰۰۷، که منطقه خوزستان را پوشش می‌دهند، از سایت فوق اخذ شدند. این محصول دمای سطح زمین را در دو زمان روز و شب، با دقت مکانی ۱ کیلومتر و با مختصات سینوسی، که حاصل تصاویر ماهواره ترا سنجنده مودیس است، ارائه می‌دهد. دمای سطح زمین با استفاده از دو باند حرارتی ۳۱ (محدوده طیفی ۱۰۷۸ تا ۱۱۲۸ میکرومتر) و ۳۲ (محدوده طیفی ۱۱۷۷ تا ۱۲۲۷ میکرومتر) و تصاویر مودیس با به‌کارگیری الگوریتم روزنه مجزا برآورد می‌شود (Vancutsem et al, 2010). الگوریتم روزنه مجزا آثار اتمسفر و گسیلندگی سطح زمین را اصلاح می‌کند.

تصاویر اخذشده، با استفاده از نرم‌افزار MRT^۱، از سیستم سینوسی به سیستم مختصات طول و عرض جغرافیایی تبدیل و به فرمت GeoTIFF، از فرمت‌های استاندارد سیستم اطلاعات جغرافیایی (GIS)، ذخیره شدند. سپس موقعیت ایستگاه‌های هواشناسی استفاده‌شده در این تحقیق با استفاده از نرم‌افزار ArcGIS شناسایی و دمای سطح زمین در دو زمان روز و شب از پیکسل ایستگاه‌ها استخراج شدند. گفتنی است هر تصویر مودیس تعدادی اطلاعات را به علت ابرناکی هوا، غلظت زیاد جو از ذرات معلق، و شکاف بین دو مسیر ماهواره از دست می‌دهد. بنابراین هر ایستگاه در بعضی روزها فاقد داده در تصویر است.

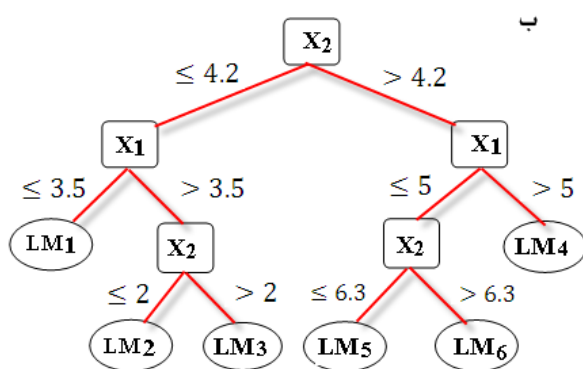
مدل درختی M5

اولین بار مدل درختی M5 را Quinlan (1992) بر اساس روش طبقه‌بندی درختی برای ایجاد رابطه بین متغیرهای مستقل و وابسته ارائه کرد. این مدل، برخلاف مدل درخت تصمیم^۲، که برای داده‌های کیفی^۳ استفاده می‌شود، برای هر دو نوع داده‌های کیفی و کمی^۴ قابل استفاده است (Mitchell, 1997; Quinlan, 1992; Quinlan, 1986). مدل M5 مشابه توابع خطی جداشده^۵ است که ترکیبی از مدل‌های رگرسیون خطی و رگرسیون درختی است و در علوم مختلف کاربرد فراوان دارد (Witten and

6. Leaf
7. Root
8. Branches
9. Nodes
10. Leaves
11. Parent node
12. Child node

1. MODIS Reprojection Too
2. Decision tree
3. Categorical
4. Quantitative
5. Piece-wis linear functions

استفاده شد. در این الگوریتم ابتدا به درخت اجازه داده می شود به اندازه کافی رشد کند. سپس گره هایی که باعث افزایش دقت دسته بندی نمی شوند هرس می شوند. شکل ۲ الف انشعابات داده های ورودی X_1 و X_2 (متغیرهای مستقل) را، که به ۶ مدل رگرسیون خطی در برگ ها (برچسب های LM1 تا LM2) منتهی می شود، نشان می دهد. فرم کلی این مدل ها به صورت $y = a_0 + a_1 X_1 + a_2 X_2$ است که در آن a_0 ، a_1 و a_2 ضرایب رگرسیون خطی اند. شکل ۲ ب ارتباط انشعابات را به صورت دیاگرام درختی نشان می دهد.



شکل ۲. شمای کلی شبکه درختی M5

پارامترهای d_r ، ω_s و δ با رابطه های ۳ و ۴ و ۵ تعیین شدند (allen et al, 1998):

$$d_r = 1 + 0.033 \cos(2\pi/365 \times J) \quad \text{(رابطه ۳)}$$

$$\omega_s = \arcsin[-\tan(\phi) \tan(\delta)] \quad \text{(رابطه ۴)}$$

$$\delta = 0.409 \sin\{(2\pi/365 \times J) - 1.39\} \quad \text{(رابطه ۵)}$$

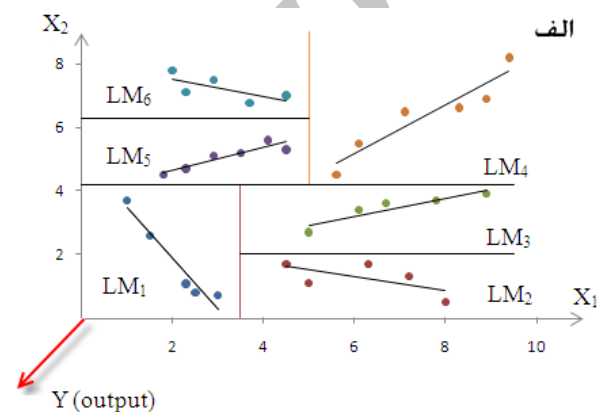
J روزشمار سال است.

برای ساخت مدل درختی M5 داده های ۱۷ ایستگاه (ایستگاه های شماره ۱ تا ۱۷ در جدول ۱)، که در کل استان خوزستان پراکنده اند، به منزله داده های آموزش و داده های سایر ایستگاه ها (ایستگاه های شماره ۱۸ تا ۲۹ در جدول ۱) به منزله داده های آزمون استفاده شدند.

مدل شبکه عصبی مصنوعی

در این تحقیق، شبکه عصبی چندلایه، شامل یک لایه ورودی و یک لایه میانی و یک لایه خروجی، برای تخمین متوسط دمای روزانه هوا، با استفاده از تصاویر ماهواره ای، استفاده شد. در هر لایه یک یا چند عنصر پردازشگر (نرون) وجود دارد که به همه نرون های لایه بعدی با اتصالات وزن دار مربوط می شود. پردازش هر نرون در دو مرحله انجام می شود. در مرحله نخست به هر

این گونه انشعاب سازی اغلب به ایجاد درختی بزرگ منجر می شود و امکان دارد برازش بیش از حد روی داده های آموزشی رخ دهد (Quinlan, 1986). برازش بیش از حد باعث کاهش کلیت بخشی مدل می شود؛ طوری که مدل فقط برای داده هایی اعتبار خواهد داشت که برای ساخت آن استفاده شده و برای داده های جدید از دقت لازم برخوردار نیست. بنابراین مرحله دوم طراحی مدل درختی کوچک کردن درخت بیش از حد بزرگ شده از طریق هرس شاخه ها و جایگزین شدن با توابع رگرسیون خطی است. الگوریتم های مختلفی برای هرس کردن درخت ارائه شده است. در این تحقیق از الگوریتم کوئینلان (Quinlan, 1999)



در این بررسی نرم افزار Weka، که در دانشگاه Waikato کشور نیوزلند توسعه یافته و نام آن از عبارت Waikato Environment for knowledge Analysis گرفته شده، به کار رفته است.

در این تحقیق، دمای سطح زمین در روز (T_d) و شب (T_n) برحسب درجه سانتی گراد و تابش بیرون زمینی برحسب مگاژول بر متر مربع در روز به منزله داده های ورودی مدل درختی M5، برای تخمین متوسط روزانه دمای هوا، انتخاب شدند. تابش بیرون زمینی نشان دهنده بازتاب تغییرات روزانه دمای هواست. این پارامتر تابعی از عرض جغرافیایی و روزشمار سال است و با رابطه ۲ محاسبه شد (allen et al, 1998):

(رابطه ۲)

$$R_a = \frac{24(60)}{\pi} G_{sc} d_r [\omega_s \sin(\phi) \sin(\delta) + \cos(\phi) \cos(\delta) \sin(\omega_s)]$$

R_a تابش بیرون زمینی (بر حسب مگاژول بر متر مربع در روز)، G_{sc} ثابت خورشیدی برابر 0.10820 مگاژول بر متر مربع در روز، d_r فاصله نسبی معکوس زمین و خورشید، ω_s زاویه ساعتی غروب خورشید (رادیان)، ϕ عرض جغرافیایی (رادیان)، و δ زاویه میل خورشید (رادیان) است.

داده‌های ارزیابی کاهش می‌یابد؛ ولی زمانی که آموزش بیش از حد اتفاق می‌افتد این خطا برای داده‌های ارزیابی افزایش می‌یابد. با شروع افزایش خطا آموزش داده‌ها متوقف و وزن‌های شبکه در شرایط حداقل خطا برای داده‌های ارزیابی تعیین می‌شوند. به عبارت دیگر، آموزش شبکه با استفاده از داده‌های آموزش و ارزیابی صورت می‌گیرد. بعد از آموزش، شبکه با داده‌هایی که در آموزش و ارزیابی از آن‌ها استفاده نشده آزمایش و عملکرد آن با استفاده از شاخص‌های آماری بررسی می‌شود. در این تحقیق، همان داده‌های آموزش شبکه درختی M5 برای آموزش و ارزیابی اوزان شبکه عصبی استفاده شد؛ طوری که به صورت تصادفی ۷۰ درصد این داده‌ها برای آموزش و ۳۰ درصد بقیه برای ارزیابی استفاده شدند. داده‌های آزمون شبکه عصبی همان داده‌های آموزش شبکه درختی M5 هستند.

کیفیت اجرای مدل شبکه عصبی به تعداد نرون‌های لایه میانی بستگی دارد. ولی تعداد زیاد آن‌ها سبب برآزش بیش از حد داده‌های آموزش می‌شود و بر کیفیت نتایج مدل تأثیر می‌گذارد. بنابراین انتخاب تعداد نرون‌های میانی بسیار مهم است. با توجه به اینکه روش مشخصی برای تعیین بهینه نرون‌های میانی وجود ندارد، با استفاده از روش سعی و خطا تعداد آن‌ها تعیین می‌شود. در این تحقیق مدل شبکه عصبی ۱۰ بار با تعداد نرون ۱ تا ۱۰ در لایه میانی آموزش یافت و سپس با استفاده از داده‌های اختصاص یافته برای آزمون نتایج مدل‌ها با داده‌های واقعی مقایسه و مناسب‌ترین تعداد نرون انتخاب شد.

از آنجا که دامنه تغییرات داده‌های ورودی و خروجی اختلاف زیادی با هم دارند، وارد کردن آن‌ها به شبکه عصبی باعث کاهش سرعت و دقت مدل می‌شود. نرمال‌سازی داده‌ها، قبل از آموزش شبکه، طوری که همه داده‌ها در بازه‌ای مناسب قرار گیرند، به آموزش شبکه کمک زیادی می‌کند. در این مقاله با استفاده از رابطه ۷ همه داده‌های ورودی و خروجی در بازه ۰ تا ۱ نرمال شدند و سپس بعد از شبیه‌سازی مدل به مقادیر اصلی برگشتند:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (\text{رابطه ۷})$$

X_{norm} مقدار نرمال شده، X مقدار اصلی، و X_{min} و X_{max}

به ترتیب بیشترین و کمترین مقادیر اصلی‌اند.

یک از ورودی‌های نرون وزنی داده می‌شود که با مقداری بایاس^۱ جمع جبری می‌شوند. حاصل جمع در مرحله دوم وارد فیلتری غیر خطی می‌شود که معمولاً تابع انتقال^۲ یا تابع محرک^۳ نامیده می‌شود. تابع محرک مثل تابع فشرده عمل می‌کند؛ طوری که خروجی نرون در یک شبکه عصبی بین دو مقدار معین (معمولاً بین ۰ و ۱ یا بین -۱ و ۱) قرار گیرد. شکل ۳ فرایند ریاضی مراحل فوق را به صورت شماتیک نشان می‌دهد. در این تحقیق، تابع محرک لوگ سیگموئید برای هر دو لایه میانی و خروجی استفاده شد. این تابع از توابع محرک رایج است که پیوسته است و بین دو مقدار ۰ و ۱ تغییر می‌کند. تابع لوگ سیگموئید با رابطه ۶ تعریف می‌شود:

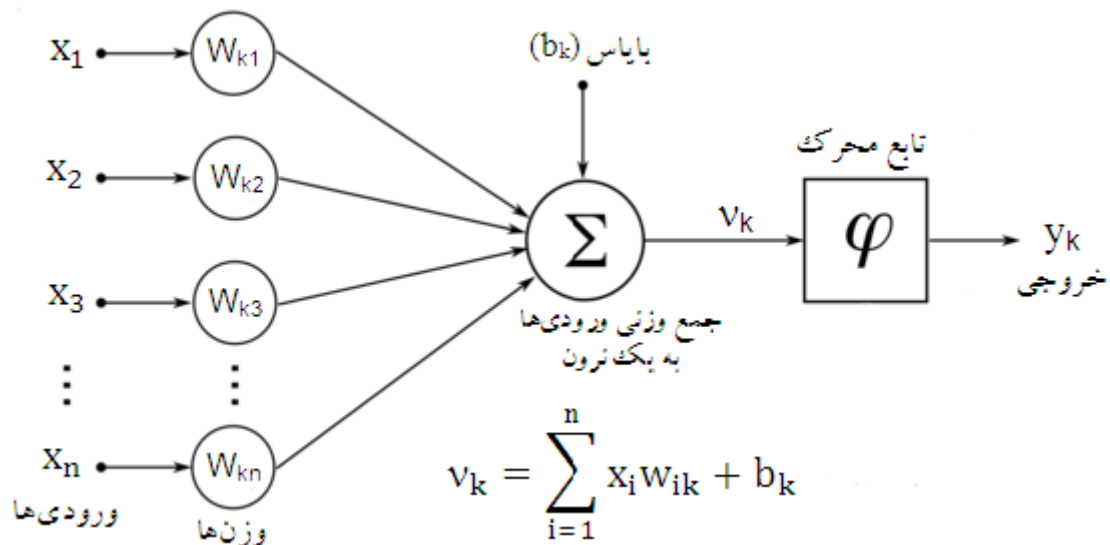
$$y_k = \frac{1}{1 + \exp(-k)} \quad (\text{رابطه ۶})$$

V_k جمع وزنی ورودی‌ها به K امین نرون در لایه بعدی و Y_k خروجی از نرون است.

وزن‌های ارتباط‌دهنده نرون‌های شبکه و بایاس با آموزش تعیین می‌شوند؛ به صورتی که تابع خطا بین خروجی‌های مدل با مقادیر واقعی در مجموعه‌ای که برای آموزش اختصاص یافته به حداقل برسد. در این تحقیق، از الگوریتم پس‌انتشار خطا برای آموزش شبکه عصبی استفاده شد. کیفیت شبکه عصبی به کلیت بخشی آن برای دامنه وسیعی از داده‌های ورودی و خروجی بستگی دارد که با اعمال نظارت بر آموزش به دست می‌آید. نظارت بر آموزش باعث می‌شود خروجی‌های دقیقی برای ورودی‌هایی که در شبکه دیده نشده برآورد گردد (Atkinson and Tatnall, 1997). عدم کلیت بخشی به علت آموزش بیش از حد پیش می‌آید که در این شرایط شبکه فقط روی داده‌های آموزش خوب کار می‌کند و برای سایر داده‌ها مناسب نیست. توقف به موقع آموزش قبل از همگراشدن شبکه از روش‌های رایج دورزدن آموزش بیش از حد است (Prechelt, 1998). در این تحقیق از تابع لونبرگ مارکوارت به دلیل سرعت بالاتر و نتایج دقیق‌تر نسبت به سایر روش‌های آموزشی (Hagan and Menhaj, 1994) برای تصمیم به توقف مرحله آموزش شبکه استفاده شد. در این تابع، از یک سری داده به منزله داده‌های ارزیابی استفاده شد. پس از هر بار تکرار فرایند یادگیری، شبکه با اوزان جدید برای داده‌های ارزیابی اجرا می‌شود. به طور معمول در مراحل اولیه آموزش خطای برآورد خروجی مدل برای

1. Bias
2. Transfer function
3. Activation function

4. Overfitting



شکل ۳. روند محاسبات در یک نرون شبکه عصبی

یافته‌ها و بحث

ماتریس همبستگی برای تعیین اهمیت نسبی هر یک از متغیرهای ورودی (T_d, T_n, R_a) بر متغیر خروجی (T_a), برای داده‌های آموزش و صحت‌سنجی، در جدول ۲ می‌آید. همان‌طور که ملاحظه می‌شود، دمای سطح زمین در شب (T_n) از بین داده‌های ورودی، در هر دو داده آموزش و صحت‌سنجی، بیشترین نقش و همبستگی را دارد. پس از آن، دمای سطح زمین در روز (T_d) نسبت به تابش بیرون زمینی (R_a) نقش بیشتری دارد. همبستگی بالای داده‌های ورودی و متوسط روزانه دمای هوا (بیش از ۰٫۸۴ برای داده‌های آموزش و بیش از ۰٫۸۸ برای داده‌های صحت‌سنجی) نشان‌دهنده این است که هر مدل با داده‌های ورودی اشاره‌شده بالا دمای متوسط روزانه هوا را با دقت مطلوبی برآورد می‌کند.

شاخص‌های آماری عملکرد شبکه‌های عصبی

میزان دقت و چگونگی عملکرد مدل‌های شبکه درختی M5 و شبکه‌های عصبی طراحی‌شده در این تحقیق با استفاده از شاخص‌های آماری و ترسیم نمودار تعیین شد. آماره‌های ضریب تعیین (R^2), ریشه میانگین مربعات خطا (RMSE), و میانگین انحراف خطا (MBE) شاخص‌های به‌کاررفته در این تحقیق‌اند:

$$R^2 = \frac{\left[\sum (P_i - \bar{P})(O_i - \bar{O}) \right]^2}{\sum (P_i - \bar{P})^2 \sum (O_i - \bar{O})^2} \quad (\text{رابطه ۸})$$

$$RMSE = \left[N^{-1} \sum_{i=1}^N (O_i - P_i)^2 \right]^{0.5} \quad (\text{رابطه ۹})$$

$$MBE = N^{-1} \sum_{i=1}^N (O_i - P_i) \quad (\text{رابطه ۱۰})$$

P_i مقادیر برآوردشده مدل، \bar{P} متوسط مقادیر برآوردشده مدل، O_i مقادیر واقعی، \bar{O} متوسط مقادیر واقعی، و N تعداد مشاهدات است.

جدول ۲. ماتریس همبستگی بین متغیرهای ورودی و متوسط روزانه دمای هوا برای هر دو داده‌های آموزش و صحت‌سنجی

داده‌های صحت‌سنجی				داده‌های آموزش				متغیر
T_a	T_d	T_n	R_a	T_a	T_d	T_n	R_a	
۰٫۸۸	۰٫۸۳	۰٫۸۸	۱	۰٫۸۴	۰٫۷۴	۰٫۸۲	۱	R_a
۰٫۹۶	۰٫۹۳	۱		۰٫۹۶	۰٫۹۱	۱		T_n
۰٫۹۳	۱			۰٫۹۰	۱			T_d
۱				۱				T_a

متوسط دمای روزانه هوا استفاده شد. مدل ساخته‌شده در شکل ۴ می‌آید. همان‌طور که ملاحظه می‌شود، مدل بر اساس یک

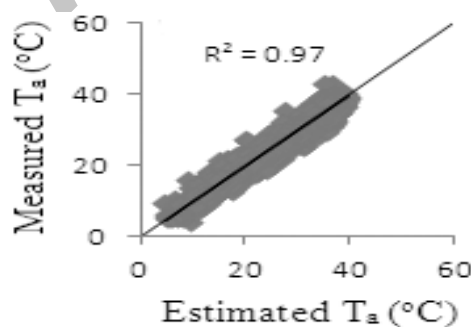
نرم‌افزار Weka برای ساخت مدل شبکه درختی M5 به منظور تبدیل داده‌های دمای سطح زمین و تابش فرازمینی به

MAE، و R^2 برای کل داده‌های آزمون محاسبه شد. شکل ۶ آثار تغییرات تعداد نرون در لایه میانی شبکه عصبی را بر شاخص‌های فوق نشان می‌دهد. همان‌طور که ملاحظه می‌شود، تعداد نرون در لایه میانی اثری معنادار بر دقت مدل دارد. بهترین مقادیر شاخص‌های MAE، RMSE، و R^2 (به ترتیب ۱٫۹۴ درجه سانتی‌گراد، -۰٫۱۷ درجه سانتی‌گراد و ۰٫۹۶) با چهار نرون در لایه میانی به دست آمده است.

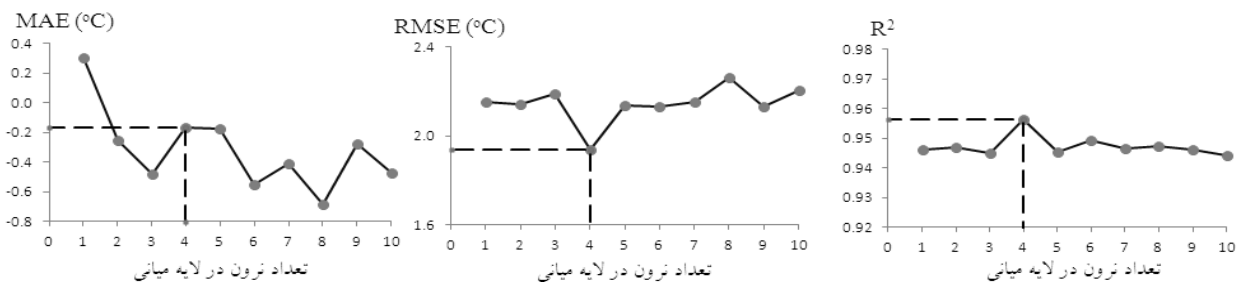
سری روابط شرطی به ۱۶ برگ منتهی می‌شود. شکل ۵ پراکنش مقادیر اندازه‌گیری شده و برآورد شده برای کل داده‌های آموزش را نشان می‌دهد. همان‌طور که مشاهده می‌شود، تطابق خوب (ضریب تعیین برابر ۰٫۹۷) است و پراکندگی کم بین مقادیر اندازه‌گیری شده و برآورد شده وجود دارد. در این تحقیق، برای تعیین تعداد بهینه نرون در لایه میانی، ۱۰ مدل شبکه عصبی با ۱ تا ۱۰ نرون در لایه میانی آموزش داده شد و پس از هر آموزش شاخص‌های RMSE،

TN <= 20.31 :	Mnum: 1	Ta = 0.0032 * DOY+ 0.1653 * RA + 0.2564 * TDAY + 0.5591 * TN - 0.5647
TN <= 11.89 : LM1	LMnum: 2	Ta = 0.0305 * DOY + 0.0912 * RA - 0.0187 * TDAY + 0.0961 * TN + 11.7644
TN > 11.89 :	LMnum: 3	Ta = 0.023 * DOY - 0.0164 * RA - 0.015 * TDAY + 0.1476 * TN + 15.1653
TN <= 15.12 :	LMnum: 4	Ta = 0.0852 * DOY - 0.0481 * RA - 0.015 * TDAY + 0.3551 * TN + 8.7702
DOY <= 83 :	LMnum: 5	Ta = 0.0292 * DOY + 0.0912 * RA - 0.1327 * TDAY + 0.0961 * TN + 15.9714
DOY <= 52.5 : LM2	LMnum: 6	Ta = 0.0389 * DOY + 0.6648 * RA + 0.1944 * TDAY + 0.1496 * TN - 13.6162
DOY > 52.5 :	LMnum: 7	Ta = 0.266 * DOY + 0.6652 * RA + 0.0796 * TDAY + 0.1964 * TN - 82.0644
RA <= 29.977 :	LMnum: 8	Ta = 0.052 * DOY + 0.0641 * RA + 0.336 * TDAY + 0.2977 * TN - 12.3179
DOY <= 60 : LM3	LMnum: 9	Ta = -0.0535 * DOY - 0.2952 * RA + 0.1349 * TDAY + 0.2145 * TN + 38.9534
DOY > 60 : LM4	Mnum: 10	Ta = -0.107 * DOY - 1.2224 * RA + 0.1259 * TDAY + 0.5236 * TN + 70.5907
RA > 29.977 : LM5	LMnum: 11	Ta = 0.0474 * DOY + 0.7057 * RA + 0.0698 * TDAY + 0.2891 * TN - 13.7468
DOY <= 317 : LM7	LMnum: 12	Ta = -0.117 * DOY - 0.3645 * RA + 0.0007 * TDAY + 0.2082 * TN + 66.5253
DOY > 317 : LM8	LMnum: 13	Ta = -0.1218 * DOY - 0.3814 * RA + 0.0007 * TDAY + 0.2118 * TN + 68.8956
TN > 15.12 :	LMnum: 14	Ta = -0.0081 * DOY - 0.0014 * RA + 0.0007 * TDAY + 0.4018 * TN + 22.5121
RA <= 23.939 :	LMnum: 15	Ta = 0.0467 * DOY + 0.7857 * RA + 0.0007 * TDAY + 0.2294 * TN - 11.413
DOY <= 305.5 : LM9	LMnum: 16	Ta = 0.0482 * DOY + 0.6371 * RA + 0.0007 * TDAY + 0.3831 * TN - 10.1171
DOY > 305.5 : LM10		
RA > 23.939 : LM11		
TN > 20.31 :		
TN <= 25.79 :		
RA <= 31.104 :		
RA <= 26.314 :		
RA <= 25.275 : LM12		
RA > 25.275 : LM13		
RA > 26.314 : LM14		
RA > 31.104 : LM15		
TN > 25.79 : LM16		

شکل ۴. مدل شبکه درختی M5 تدوین شده برای برآورد دمای هوا



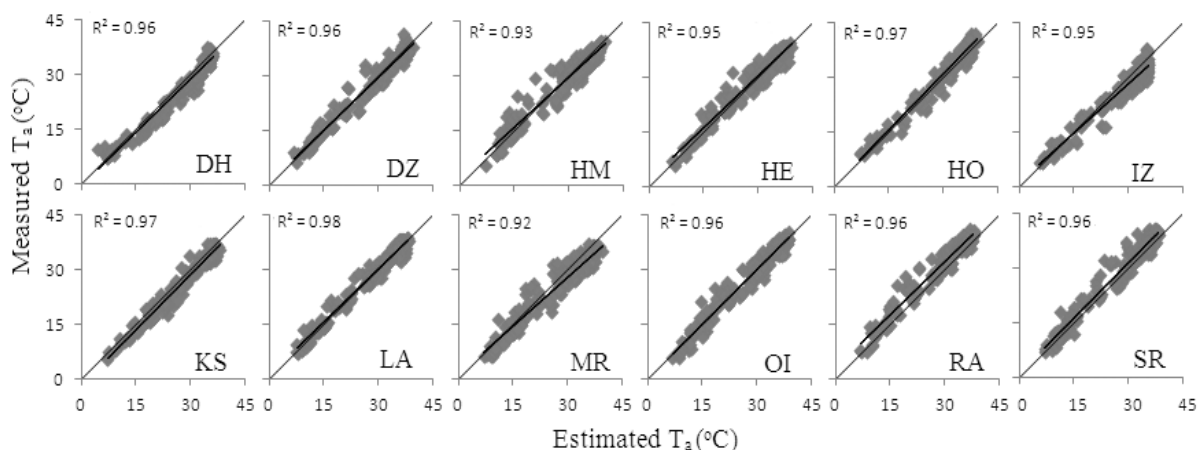
شکل ۵. پراکنش مقادیر برآورد شده با استفاده از مدل شبکه درختی و اندازه‌گیری شده دمای هوا برای داده‌های آموزش



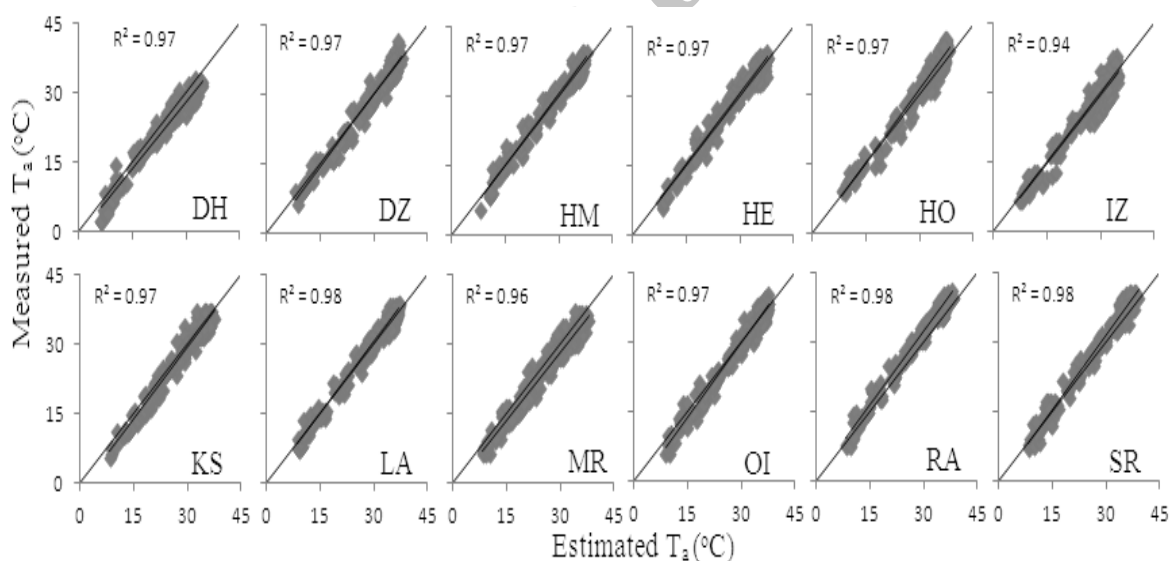
شکل ۶. تغییرات تعداد نرون لایه میانی بر دقت نتایج

تطابق خوب بین داده‌های اندازه‌گیری‌شده و برآوردشده است. ولی توزیع داده‌ها در مدل شبکه عصبی قدری بهتر از مدل شبکه درختی است. دامنه ضریب تعیین در مدل درختی (شکل ۶) بین ۰/۹۲ تا ۰/۹۸ (شکل ۷) و در مدل شبکه عصبی بین ۰/۹۴ تا ۰/۹۸ (شکل ۸) تغییر می‌کند.

مدل‌های تدوین‌شده در این تحقیق برای ایستگاه‌های آزمون اجرا شد. پراکنش مقادیر اندازه‌گیری‌شده دمای هوا و آن‌هایی که با استفاده از مدل شبکه درختی و شبکه عصبی برآورد شدند برای هر ایستگاه در شکل‌های ۷ و ۸ می‌آید. همان‌طور که مشاهده می‌شود، در هر دو مدل همه داده‌ها با فاصله کمی از بهترین خط انطباق توزیع شده‌اند که نشان‌دهنده



شکل ۷. پراکنش مقادیر برآوردشده با استفاده از مدل شبکه درختی و اندازه‌گیری‌شده دمای هوا برای ایستگاه‌های آزمون



شکل ۸. پراکنش مقادیر برآوردشده با استفاده از مدل شبکه عصبی و اندازه‌گیری‌شده دمای هوا برای ایستگاه‌های آزمون

اندازه‌گیری‌شده است. دقت نتایج مدل شبکه عصبی بهتر از شبکه درختی M5 برآورد می‌شود. در این مدل دامنه تغییرات و متوسط شاخص RMSE کمتر از مدل شبکه درختی M5 برآورد شده است؛ طوری که RMSE بین ۱/۵ تا ۲/۵ درجه سانتی‌گراد و به طور متوسط ۱/۹ درجه سانتی‌گراد و MBE بین ۲/۲- تا ۱/۶ درجه سانتی‌گراد و به طور متوسط ۰/۲- درجه سانتی‌گراد است.

جدول ۳ خلاصه نتایج آماری هر دو مدل شبکه درختی و شبکه عصبی را نشان می‌دهد. دامنه تغییرات شاخص RMSE برای مدل درختی M5 بین ۱/۵ تا ۳/۱ درجه سانتی‌گراد و به طور متوسط ۲/۳ درجه سانتی‌گراد است. برای همین مدل، دامنه تغییرات شاخص MBE بین ۲/۲- تا ۱/۷ درجه سانتی‌گراد و به طور متوسط ۰ درجه سانتی‌گراد است. عدد منفی MBE نشان‌دهنده این است که مقادیر برآوردشده کمتر از مقادیر

جدول ۳. خلاصه نتایج آماری مدل‌های شبکه درختی M5 و شبکه عصبی برای ایستگاه‌های آزمون

کد ایستگاه	مدل درختی M5			مدل شبکه عصبی		
	R ²	MBE (°C)	RMSE (°C)	R ²	MBE (°C)	RMSE (°C)
DH	۰٫۹۶	۰٫۹	۱٫۹	۰٫۹۷	۱٫۶	۲٫۲
DZ	۰٫۹۶	۰٫۱	۱٫۸	۰٫۹۷	۰٫۱	۱٫۵
HM	۰٫۹۳	۰٫۲	۲٫۵	۰٫۹۷	-۰٫۵	۱٫۶
HE	۰٫۹۵	-۰٫۵	۲٫۰	۰٫۹۷	-۰٫۷	۱٫۷
HO	۰٫۹۷	-۰٫۹	۱٫۹	۰٫۹۷	-۱٫۲	۲٫۱
IZ	۰٫۹۵	۱٫۳	۲٫۴	۰٫۹۴	۰٫۳	۱٫۹
KS	۰٫۹۷	۱٫۱	۱٫۸	۰٫۹۷	۰٫۷	۱٫۷
LA	۰٫۹۸	-۰٫۲	۱٫۵	۰٫۹۸	-۰٫۴	۱٫۴
MR	۰٫۹۲	۱٫۷	۳٫۱	۰٫۹۶	۱٫۵	۲٫۳
OI	۰٫۹۶	-۰٫۲	۱٫۸	۰٫۹۷	۰٫۰	۱٫۷
RA	۰٫۹۶	-۲٫۲	۲٫۸	۰٫۹۸	-۲٫۲	۲٫۵
SR	۰٫۹۵	-۱٫۵	۲٫۵	۰٫۹۸	-۱٫۵	۲٫۲
متوسط	۰٫۹۶	۰٫۰	۲٫۳	۰٫۹۷	-۰٫۲	۱٫۹

نتیجه‌گیری

در این تحقیق دو روش یادگیری ماشینی، شامل مدل شبکه عصبی و مدل درختی M5، برای تبدیل دمای سطح زمین حاصل از تصاویر سنجنده مودیس به متوسط دمای روزانه هوا مقایسه شدند. نتایج نشان داد ضریب تعیین هر دو مدل بالاست (بیش از ۰٫۹۶). این وضعیت نشان‌دهنده آن است که مدل‌های تدوین شده بیش از ۹۶ درصد تغییرات دمای هوا را با استفاده از دمای سطح زمین توجیه می‌کنند. دقت مدل شبکه عصبی و

مدل شبکه درختی M5 به ترتیب حدود ۱٫۹ و ۲٫۳ درجه سانتی‌گراد تعیین شد. نتایج این تحقیق به نتایج Jang *et al* (2004) نزدیک است. آن‌ها برای تبدیل داده‌های پنج ماهواره نوا به دمای هوا از مدل شبکه عصبی استفاده کردند و دمای هوا با دقت ۲ درجه سانتی‌گراد تعیین شد. پیشنهاد می‌شود در تحقیقات آتی متغیرهای ورودی استفاده شده در این تحقیق مورد آنالیز حساسیت قرار گیرند.

REFERENCES

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). Crop evapotranspiration. Guidelines for computing crop water requirements. *Irrigation and Drainage*, Paper No. 56. FAO, Rome.
- Aher, P. D., Adinarayana, J., and Gorantiwar, S. D. (2011). Remote Sensing and Artificial Neural Network in Spatial Assessment of Air Temperature in a Semi-arid Watershed. *International Journal of Earth Sciences and Engineering*, 4(6), 351-354.
- Atkinson, P. M. and Tatnall, A. R. L. (1997). Introduction neural networks in remote sensing. *International Journal of Remote Sensing*, 18, 699-709.
- Bhattacharya, B. and Solomatine, D. P. (2005). Neural networks and M5 model trees in modeling water level-discharge relationship. *Neurocomputing*, 63, 381-396.
- Cresswell, M. P., Morse, A. P., Thomson, M. C., Connor, S. J. (1999). Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. *International Journal of Remote Sensing*, 20(6), 1125-1132.
- Emamifar, S., Rahimikhoob, A., and Noroozi, A. A. (2013). Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. *International Journal of Climatology*. 33(15), 3174-3181.
- Hagan, M. T. and Menhaj, M. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5, 989-993.
- Jang, J. D., Viau, A. A., and Anctil, F. (2004). Neural Network Estimation of Air Temperatures from AVHRR data. *International Journal of Remote Sensing*, 25(21), 4541-4554.
- Mitchell, T. M. (1997). Machine learning. The McGraw-Hill Comp. Press.

- Pal, M. and Deswal, S. (2009). M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes*, 23, 1437-1443.
- Parviz, L., Kholghi, M., and Valizadeh, K. (2011). Estimation of Air Temperature Using Temperature-Vegetation Index (TVX) Method. *Journal of Science and Technology of Agriculture and Natural Resources, Water and Soil Science*, 15 (56) , 21-34. (In Farsi)
- Prechelt, L. (1998). Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks*, 11, 761-767.
- Prihodko, L. and Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, 60(3), 335-346.
- Quinlan, J. R. (1992). Learning with continuous classes. In Proceedings of the *Fifth Australian Joint Conference on Artificial Intelligence*, Hobart, Australia, 16-18 November, World Scientific, Singapore, 343-348.
- Quinlan, J. R. (1986). Introduction of decision trees. *Machine learning*, 1, 81-106
- Rahimikhoob, A., Behbahani, M. R., and Nazarifar, M. H. (2008). Estimating Maximum Air Temperature in Khoozestan Province Using NOAA Satellite Images Data and Artificial Neural Network. *Journal of Science and Technology of Agriculture and Natural Resources, Water and Soil Science*, 11(42), 357-364. (In Farsi)
- Solomatine, D. P. and Xue, Y. (2004). M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huai River in China. *Journal of Hydrologic Engineering*, 9(6), 491-501.
- Solomatine, D. P. and Dulal, K. N. (2003). Model trees as an alternative to neural networks in rainfall-runoff modelling. *Hydrological Sciences Journal*, 48(3), 399-411.
- Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114, 449-465.
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Technique*. Morgan Kaufmann Publishers, San Francisco.
- Yao, Y. and Zhang, B. (2012). MODIS-based air temperature estimation in the southeastern Tibetan Plateau and neighboring areas. *Journal of Geographical Sciences*, 22(1), 152-166.
- Yan, H., Zhang, J., Hou, Y., and He, Y. (2012). Estimation of air temperature from MODIS data in east China. *International Journal of Remote Sensing*, 30(23), 6261-6275.

Archive