

Determination of Input Variables to Estimate Solar Radiation Using Entropy Theory and Principal Component Analysis

BABAK MOHAMMADI¹, ZAHRA AGHASHARIATMADARI^{2*}, ROOZBEH MOAZENZADEH

1. PhD student in Agrometeorology, Department of Irrigation and Reclamation, Faculty of Agricultural Engineering and Technology, University of Tehran, Karaj, Iran.
2. Assistant professor of Department of Irrigation and Reclamation, , Faculty of Agricultural Engineering and Technology, University of Tehran, Karaj, Iran.
3. Assistant Professor, Water Engineering Department, Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran.

(Received: May. 9, 2018- Revised: Nov. 7, 2018- Accepted: Dec. 17, 2018)

ABSTRACT

Solar radiation arriving to the land surface is one of the major variables that is used in projects and hydrological, agricultural, meteorological and climatic models. In this study, the functionality of the principal component analysis (PCA) and the entropy theory (EN) for determination of inputs to multilevel perceptron artificial neural network (MLP), artificial neural network, radial basis function (RBF), support vector machine (SVM) and genetic programming (GEP), was investigated for estimation of solar radiation at two stations (Kerman and Mashhad) during 1984-2005 and 1980-2004 periods, respectively. The average temperature, mean water deficit pressure, minimum temperature, maximum temperature, sunshine, relative humidity, dew point temperature, hourly vapor pressure, horizontal visibility and water content were selected as inputs of pre-processing methods. The obtained results in Kerman station showed that the ENT-MLP model with RMSE=38.36 (Mj /m²) and R² = 0.93 have had the best performance. Also in Mashhad station, PCA-MLP model with RMSE=79.75 (Mj / m²) and R² = 0.77 had the best performance. In general, the both pre-processing principal component analysis and entropy theory were recognized as the proper methods for determination of estimating models input to estimate solar radiation.

Keywords: Data Preprocessing, Artificial Neural Network, Radial Basis Function, Support Vector Machine, genetic programming

تعیین متغیرهای ورودی برای تخمین تابش خورشیدی با استفاده از تئوری آنتروپی و تحلیل مؤلفه اصلی

بابک محمدی^۱، زهرا آقاشریعتمداری^{۲*}، روزبه موذن‌زاده^۳

۱. دانشجوی دکتری هواشناسی کشاورزی، گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی،

دانشگاه تهران، کرج، ایران

۲. استادیار گروه مهندسی آبیاری و آبادانی، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران، کرج، ایران

۳. استادیار گروه مهندسی آب، دانشکده کشاورزی، دانشگاه صنعتی شاهرود، شاهرود، ایران

(تاریخ دریافت: ۱۳۹۷/۲/۱۹ - تاریخ بازنگری: ۱۳۹۷/۸/۱۶ - تاریخ تصویب: ۱۳۹۷/۹/۲۶)

چکیده

تابش خورشیدی رسیده به سطح زمین یکی از متغیرهای اصلی مورد استفاده در پروژه‌ها و مدل‌سازی‌های هیدرولوژی، کشاورزی، هواشناسی و اقلیمی می‌باشد. در این تحقیق قابلیت عملکرد روش تحلیل مؤلفه اصلی (PCA) و تئوری آنتروپی (EN) برای تعیین ورودی مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، ماشین بردار پشتیبان (SVM) و برنامه‌ریزی ژنتیک (GEP) در برآورد تابش خورشیدی در دو ایستگاه همدید کرمان و مشهد به ترتیب در حد فاصل سال‌های ۱۹۸۴ تا ۲۰۰۵ و ۱۹۸۰ تا ۲۰۰۴ مورد بررسی قرار گرفت. متغیرهای میانگین دما، میانگین کمبود فشار بخار آب اشباع، دمای کمینه، دمای بیشینه، ساعت آفتابی، رطوبت نسبی، دمای نقطه شبنم، فشار بخار ساعتی، دید افقی و محتوی بخار آب جو به عنوان ورودی روش‌های پیش‌پردازش انتخاب شدند. با توجه به نتایج به دست آمده در ایستگاه کرمان، مدل ENT-MLP با ریشه میانگین مربعات خطای برابر ۳۸/۳۶ (Mj/m²) و ضریب تبیین $R^2 = 0/93$ بهترین عملکرد را داشته است. همچنین در ایستگاه مشهد مدل PCA-MLP با ریشه میانگین مربعات خطای برابر ۷۹/۷۵ (Mj/m²) و ضریب تبیین $R^2 = 0/77$ بهترین عملکرد را داشته است. به طور کلی هر دو روش پیش‌پردازش تحلیل مؤلفه اصلی (PCA) و تئوری آنتروپی (EN) برای تعیین ورودی مدل‌های تخمین‌گر به منظور تخمین تابش خورشیدی روش مناسبی تشخیص داده شدند.

کلمات کلیدی: پیش‌پردازش داده‌ها، شبکه عصبی مصنوعی، تابع پایه شعاعی (RBF)، ماشین بردار پشتیبان (SVM) و برنامه‌ریزی ژنتیک (GEP)

مقدمه

خورشید در بسیاری از مناطق به دلیل مشکلات اقتصادی و فقدان امکانات به طور محدود انجام می‌شود و این نقیصه حتی در کشورهای در حال توسعه نیز به چشم می‌خورد (Khalili and Rezai Sadr, 1997; Skeiker, 2006). روش معمول ثبت این داده‌ها استفاده از دستگاه پیرانومتر در مکان‌های مورد نیاز است. استفاده از این دستگاه به دلیل هزینه زیاد نصب و نگهداری در همه مناطق امکان‌پذیر نیست در نتیجه این وسیله اندازه‌گیری فقط در مکان‌های محدودی نصب شده است (Aghashariatmadary, 2011).

تحقیقات زیادی به منظور تخمین تابش خورشیدی با استفاده از متغیرهای هواشناسی صورت گرفته و روش‌های زیادی توسعه داده شده است. در نتیجه، محققین بیشتر به دنبال راهی برای تخمین تابش با دقت نسبتاً خوب می‌باشند. این مسئله باعث ترغیب دانشمندان و محققان به سوی ارائه و استفاده از مدل‌های برآورد تابش شده است. مقادیر تابش به روش‌های مختلفی از جمله روابط تجربی و رگرسیونی، استفاده

دانستن مقدار تابش خورشیدی در هر محل برای بسیاری از مسائل کاربردی از جمله تخمین تبخیر - تعرق، طراحی معماری و مدل‌های رشد محصولات کشاورزی اهمیت فراوان دارد (Kamali and Aghashariatmadari, 2018). متغیر تابش خورشیدی به عنوان یکی از متغیرهای مهم در مدل‌های بیلان انرژی و شبیه‌سازی رشد گیاهان اهمیت زیادی دارد. تابش خورشیدی متغیر مهمی است که در مباحث هیدرولوژی، مدیریت منابع آب، مدل‌های بیلان آبی و مدل‌های شبیه‌سازی رشد گیاه کاربرد دارد (Ball et al., 2004). اما علی‌رغم اهمیت اندازه‌گیری این متغیر، وسایل و ابزار مناسب اندازه‌گیری تابش خورشیدی مانند سایر متغیرهای هواشناسی از جمله دما و بارش در همه مناطق موجود نبوده و بنابراین باید به گونه‌ای آن را برآورد کرد. اغلب اندازه‌گیری‌های صورت گرفته در زمینه تابش

(2011) Lazzus *et al.* در پژوهش دیگری نشان دادند شبکه عصبی با ورودی‌های سرعت باد، رطوبت نسبی و دمای هوا و خاک نسبت به روش‌های تجربی توانایی بالاتری در تخمین تابش خورشیدی دارد. همچنین (2011) Azeez با استفاده از شبکه عصبی با ورودی‌های ساعت آفتابی، بیشینه دما و رطوبت نسبی، میانگین ماهانه تابش خورشیدی را با دقت قابل قبولی تخمین زد.

در سال‌های اخیر ماشین بردار پشتیبان (Support Vector Machine) در زمینه‌های مختلفی مورد استفاده قرار گرفته است (2017) Biazar, Mohammadi and Moazenzadeh. به طور مثال (2013) Chen *et al.* تابش خورشیدی را با هفت ترکیب مختلف از دمای بیشینه و کمینه توسط SVM برآورد کردند و با روابط تجربی مقایسه کردند. آن‌ها تکنیک SVM را به عنوان یک گزینه برتر در برابر روابط معمول پیشنهاد دادند. (2014) Long *et al.* با استفاده از شبکه عصبی و SVM و دو روش دیگر مقدار تابش خورشیدی را تخمین زدند. تحقیقات آن‌ها نشان داد دمای کمینه، میانگین و بیشینه، ساعت آفتابی، سرعت باد و مقدار بارندگی متغیرهای با اهمیت رطوبت، فشار در سطح دریا و نقطه شبنم متغیرهای کم اهمیت در تخمین انرژی خورشیدی می‌باشند. در پژوهشی دیگر توسط (2009) Azadeh *et al.* با استفاده از داده‌های سنجش از دور دما و میانگین‌های ماهانه‌ی متغیرهای مختلف هواشناسی در شش ایستگاه همدیدی ایران طی سال‌های ۱۹۹۵ الی ۲۰۰۰، مقادیر تابش خورشیدی ماهانه را به کمک شبکه عصبی (MLP) و معادله انگستروم- پرسکات برآورد کرده و به مقایسه‌ی نتایج آن‌ها با یکدیگر پرداختند و بیان داشتند که شبکه عصبی مصنوعی نسبت به معادله انگستروم- پرسکات دارای دقت بالاتری است.

قابلیت یک مدل هیدرولوژیکی باید به صورت ساختاری هم در مورد اساس تئوری‌های هیدرولوژیکی به کار رفته در آن و هم از لحاظ میزان داده مورد نیاز و موجودیت آن‌ها مورد ارزیابی قرار گیرد و مدل‌های هیدرولوژیکی به لحاظ جزئیات ورودی به مدل‌ها متفاوت می‌باشند (2017) Mohammadi. پس اینکه چه تعداد داده و چه تعداد پارامتر ورودی از یک پدیده در اختیار داشته باشیم برای مدل‌سازی یک پدیده‌ی هیدرولوژیکی مانند تابش خورشیدی حائز اهمیت می‌باشد و از آنجایی که داده‌های مربوط به همه متغیرهای مؤثر بر تابش خورشیدی در اکثر ایستگاه‌ها وجود نداشته یا ثبت نشده است، بنابراین باید روی متغیرهای موجود تحلیل مناسبی انجام داد تا بتوان تأثیر نسبی آن‌ها را روی این پدیده بررسی کرد. روش‌های پیش‌پردازش

از فن سنجش از دور و شبکه‌های عصبی و میان‌یابی خطی محاسبه می‌شوند (2007) Sabziparvar and Shetaee.

در سال‌های اخیر توانایی و قابلیت شبیه‌سازی تکنیک‌های هوش مصنوعی در زمینه‌های مختلفی از جمله برآورد پدیده‌های محیطی مانند بررسی متغیرهای هیدرولوژی مورد بررسی قرار گرفته و اثبات شده است. یکی از اولین تکنیک‌های هوش مصنوعی که در زمینه‌های مختلف مورد استفاده قرار گرفته است شبکه عصبی می‌باشد (Mohammadi and Ghorbani, 2016). شبکه‌های عصبی قادرند با ایجاد نگاشت غیرخطی بین متغیرهای ورودی و خروجی، روابط میان آن‌ها را شبیه‌سازی نمایند و همچنین در حل مسائل پیچیده که مدل‌های تجربی یا نیمه تجربی در حل آن کارایی مناسب نداشته باشند، بسیار مفید است (2017) Mohammadi and Moazenzadeh. شبکه‌های عصبی مصنوعی ابزاری قدرتمند و توانمند جهت برآوردها و پیش‌بینی پارامترها در سیستم‌های غیرخطی می‌باشند. با توجه به ماهیت پیچیده و غیرخطی متغیر تابش خورشیدی استفاده از تکنیک‌های قدرتمندی مانند هوش مصنوعی نیز مورد توجه محققین بوده است (Rahimikhoob, 2010).

در پژوهشی، سلطانی و مرید با استفاده از شبکه‌های عصبی مصنوعی و با استفاده از داده‌های دمای حداقل و حداکثر روزانه، تابش کل خورشیدی را به صورت روزانه در ایستگاه‌های سینوپتیک مهرآباد - تهران، شیراز، زاهدان و داده‌های ایستگاه امامه (شمال شرق تهران)، تخمین زده و دقت شبکه‌های عصبی مصنوعی را با روش تجربی هارگریوز و سامانی مقایسه کردند. نتایج پژوهش یاد شده نشان داد که شبکه‌های عصبی مصنوعی در مقایسه با مدل تجربی مذکور از دقت بیشتری برخوردار بوده است (2005) Soltani and Morid. ردی و رانجان در سال ۲۰۰۳ با استفاده از شبکه‌های عصبی و با استفاده از داده‌های هواشناسی موجود (دما، سرعت باد، رطوبت نسبی و بارندگی)، طول و عرض جغرافیایی و ارتفاع از سطح دریا تابش کل خورشیدی را برآورد کردند و سپس با مدل‌های رگرسیونی آنگستروم (1924) Angstrom و هارگریوز و سامانی (1982) Hargreaves and Samani مقایسه نمودند (Redy and Ranjan, 2003). از بین مدل‌های مختلف استفاده شده مدل شبکه عصبی از بهترین دقت برخوردار بود.

(2005) Tymvios *et al.*, با استفاده از شبکه‌های عصبی مصنوعی، تابش کل خورشیدی را برآورد کردند و نشان دادند که مدل شبکه عصبی با ورودی ساعات آفتابی و حداکثر ساعات آفتابی ممکن و دمای حداکثر دارای دقت بالاتری می‌باشد.

مواد و روش‌ها

مشخصات آماری داده‌ها و منطقه مورد مطالعه

در این تحقیق به منظور برآورد مقادیر تابش خورشیدی ماهانه از مدل‌های تخمین‌گر شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، برنامه‌ریزی ژنتیک (GEP) و ماشین بردار پشتیبان (SVM) استفاده شده است. همچنین برای پیش‌پردازش داده‌های ورودی و تعیین متغیرهای مؤثر در تخمین تابش خورشیدی از روش تئوری آنتروپی (ENT) و روش تحلیل مؤلفه اصلی (PCA) استفاده گردید. داده‌های مورد استفاده در این تحقیق متغیرهای اقلیمی اندازه‌گیری شده در ایستگاه‌های همدیدی کرمان و مشهد به ترتیب طی سال‌های ۲۰۰۵-۱۹۸۴ و ۲۰۰۴-۱۹۸۰ در مقیاس ماهانه می‌باشد. مشخصات متغیرهای اقلیمی مذکور به همراه دامنه تغییرات آن‌ها در جدول (۱) آورده شده است. ایستگاه همدیدی کرمان در موقعیت طول جغرافیایی ۵۷ درجه و ۶ دقیقه شرقی و عرض ۳۰ درجه و ۲۹ دقیقه شمالی (ارتفاع ۱۷۵۲ متری از سطح دریا) و ایستگاه همدیدی مشهد در موقعیت طول جغرافیایی ۵۹ درجه و ۳۶ دقیقه شرقی و عرض ۳۸ درجه و ۱۸ دقیقه شمالی (ارتفاع ۱۰۵۰ متری از سطح دریا) واقع شده‌اند.

داده‌ها با شناسایی ورودی‌های مؤثر بر کاهش ابعاد ورودی‌های مدل و از بین بردن روند نایستایی موجود در ورودی‌ها موجب افزایش کارایی مدل برای پیش‌بینی پدیده‌ها در آینده می‌شود (Biazar, 2017). از جمله روش‌های پیش‌پردازش که در این تحقیق به آن‌ها پرداخته شده است روش‌های تئوری آنتروپی و تحلیل مؤلفه اصلی می‌باشند.

با توجه به مطالعات صورت گرفته تاکنون مقایسه‌ای بین مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، برنامه‌ریزی ژنتیک (GEP) و ماشین بردار پشتیبان (SVM) همراه با پیش-پردازش داده‌ها با روش تحلیل مؤلفه اصلی و تئوری آنتروپی بر روی تابش خورشیدی صورت نگرفته است. به این ترتیب با توجه به اهمیت تابش رسیده به سطح زمین، هدف از این تحقیق تخمین تابش خورشیدی در یک اقلیم نیمه‌خشک (مشهد) و اقلیم خشک (کرمان) با استفاده از مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، برنامه‌ریزی ژنتیک (GEP) و ماشین بردار پشتیبان (SVM) همراه با کاهش پیچیدگی مدل-سازی به وسیله پیش‌پردازش داده‌ها با روش‌های تحلیل مؤلفه اصلی و تئوری آنتروپی می‌باشد.

جدول ۱- مشخصات متغیرهای اقلیمی مورد استفاده در ایستگاه‌های کرمان و مشهد

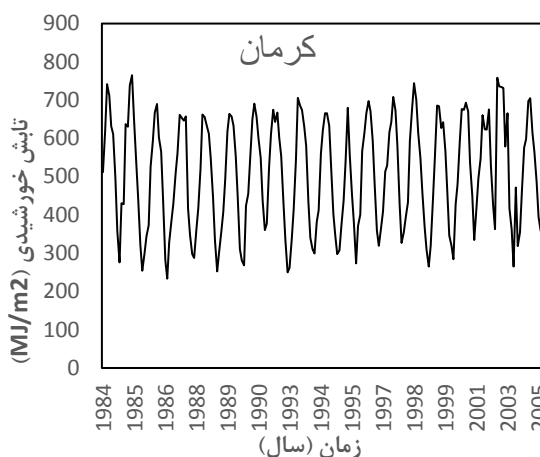
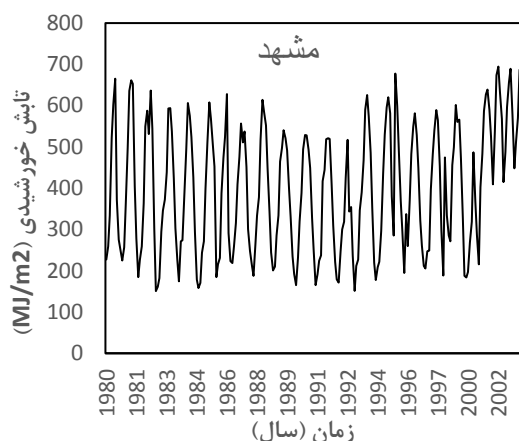
متغیرهای اقلیمی	ساعت آفتابی (hr)	دمای کمینه (°C)	دمای بیشینه (°C)	دمای نقطه شبنم (°C)	فشار بخار ساعتی (Mbar)	میانگین اشباع کمبود فشار بخار (Mbar)	رطوبت نسبی (%)	دید افقی (Number of days) $V < 2KM$	محتوی بخار آب جو	تابش خورشیدی (mj/m ²)	ایستگاه
کرمان	میانگین	۹/۱۲	۷/۳۶	۲۵/۵۱	-۲/۴۸	۵/۵۷	۱۷/۶۸	۳۲/۴۴	۰/۹۶	۴۱۲/۸۱	۵۰۶/۸۲
	کمینه	۴/۶۹	-۶/۹	۹/۷	-۱۳/۳	۲/۴۲	۳/۰۵	۱۳	۰	۲۰۰/۱۱	۲۳۳/۳۹
	بیشینه	۱۲/۳۸	۲۰/۵	۳۸	۵/۸	۹/۵۸	۳۷/۳۷	۷۰	۹	۶۵۹/۷۰	۷۶۵/۰۵
	انحراف معیار	۱/۸۶	۷/۱۱	۸/۱۶	۳/۸۶	۱/۴۷	۹/۹۶	۱۳/۲۸	۱/۴۹	۹۲/۶۱	۱۴۳/۵۴
مشهد	میانگین	۷/۸۴	۸/۱۴	۲۱/۷۵	۳/۷۳	۸/۶۷	۱۱/۵۹	۵۵/۰۲	۴/۴۱	۶۶۶/۲۴	۳۹۸/۲۱
	کمینه	۲/۴	-۷/۵	۲/۵	-۷/۵	۳/۶۱	۱/۱۶	۲۲	۰	۲۹۰/۷۲	۱۵۰/۷۲
	بیشینه	۱۲/۵۳	۲۱/۷	۳۶/۹	۱۴/۱	۱۶/۵۳	۳۱/۳۵	۸۴	۲۱	۱۱۶۲/۳۵	۶۹۴/۶۷
	انحراف معیار	۲/۸۷	۷/۷۵	۹/۵۹	۴/۸۵	۲/۸۶	۸/۶۱	۱۶/۴۸	۵/۴۲	۱۸۶/۵۴	۱۵۱/۵۲

۱۹۸۴-۲۰۰۵ و ۱۹۸۰-۲۰۰۴ در شکل (۱) نشان داده شده است. چگونگی تغییرات مقادیر اندازه‌گیری شده تابش

روند تغییرات مقادیر اندازه‌گیری شده تابش خورشیدی در ایستگاه‌های همدیدی کرمان و مشهد به ترتیب طی سال‌های

در ایستگاه مذکور حتی مقادیر ثبت شده آن به حدود ۷۶۵ مگاژول بر متر مربع نیز می‌رسد و اهمیت برآورد صحیح آن را بیش از پیش نمایان می‌سازد.

خورشیدی که به صورت مستقیم تابع برخی از متغیرهای اقلیمی و به صورت غیرمستقیم تابعی از زمان است، به همراه مقادیر متوسط ثبت شده این مؤلفه در جدول (۱) بیانگر اهمیت آن در وضعیت بیلان آبی منطقه مورد مطالعه است به نحوی که



شکل ۱- سری زمانی مقادیر تابش ماهانه (ایستگاه‌های همدیدی کرمان و مشهد)

اطلاعات جدید، در واقع عدم قطعیت‌ها کاهش یافته و ارزش اطلاعات جدید برابر با مقداری است که از عدم قطعیت کاسته شده است. در نتیجه عدم قطعیت و اطلاعات پارامترهایی وابسته به هم هستند. به عبارت دیگر، از تئوری آنتروپی می‌توان به‌عنوان شاخصی برای کمی کردن میزان عدم آگاهی و دانش نسبت به مشخصات یک سامانه، استفاده نمود (Xu et al., 2015).

برای به دست آوردن وزن آنتروپی ω_j از معادله زیر استفاده می‌شود (Harmancioglu, 1984):

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^n 1 - e_j} \quad (\text{رابطه ۱})$$

متغیر e_j مقدار آنتروپی انتقال اطلاعات را بیان می‌کند. در واقع آنتروپی انتقال اطلاعات نوعی از آنتروپی است که برابر با تفاضل مجموع آنتروپی مرزی و آنتروپی‌های مشترک می‌باشد. هرچقدر مقدار آنتروپی کمتر باشد، تأثیرز بیشتر خواهد بود. آنتروپی انتقال اطلاعات بین دو متغیر i و j به صورت زیر تعریف می‌شود (Harmancioglu, 1984):

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m P_{ij} \ln P_{ij} \quad (\text{رابطه ۲})$$

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان مجموعه‌ای از روش‌های یادگیری نظارت شده^۲ است که برای طبقه‌بندی^۱ و رگرسیون استفاده می‌شود.

تحلیل مؤلفه‌های اصلی (PCA)

در روش تحلیل مؤلفه‌های اصلی، فرض بر این است که P متغیر اصلی همبسته به مؤلفه غیرهمبسته یا متعامد تبدیل می‌شود. با اعمال PCA متغیرهای ورودی اصلی به متغیرهای جدید که بدون همبستگی می‌باشند، تبدیل می‌شوند. مؤلفه‌های ایجاد شده ترکیب خطی از متغیرهای اصلی می‌باشد (Mohammadi and Emamgholizadeh, 2017). تحلیل مؤلفه‌های اصلی متغیرهای ورودی را به مؤلفه‌های اصلی که ترکیب مستقل و خطی از متغیرهای ورودی هستند تغییر می‌دهد. در این روش، اطلاعات متغیرهای ورودی با حداقل تلفات در مؤلفه‌های اصلی ارائه خواهد شد (Helena et al., 2000). در این تحقیق از مجموعه مؤلفه‌های اصلی اولی که حداقل هشتاد درصد واریانس کل داده‌ها را دارا هستند، به‌عنوان ورودی مدل‌های تخمین‌گر استفاده شده است.

تئوری آنتروپی شانون (ENT)

شانون در سال ۱۹۴۸ نشان داد که وقایع با احتمال وقوع زیاد، اطلاعات واضح و مشهودی در اختیار می‌گذارند و برعکس هر چقدر احتمال وقوع یک رخداد کمتر باشد، اطلاعات حاصل از آن جدیدتر و برای محققین مفیدتر است. با به دست آوردن

۱. Principal Component Analysis

۲. Shannon Entropy Theory

در معادله (۵)، $\bar{\alpha}_i$ میانگین ضرایب لاگرانژ است. محاسبه $\phi(x)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل رگرسیون بردار پشتیبان انتخاب یک تابع کرنل است. با توجه به اینکه در ماشین بردار پشتیبان معمولاً از تابع کرنل پایه شعاعی استفاده می‌گردد، در این تحقیق نیز از این تابع استفاده گردیده است. فرمول کرنل پایه شعاعی در زیر آورده شده است (Mohammadi and moazenzadeh, 2017).

$$k(x, x_i) = \exp\left(-\|x - x_i\|^2 / \sigma^2\right) \quad (\text{رابطه ۶})$$

لازم به ذکر است فرآیند محاسبات ماشین بردار پشتیبان، بر اساس کد نویسی در محیط متلب انجام و پارامترهای تابع کرنل از طریق سعی و خطا بهینه شدند.

شبکه عصبی پرسپترون چندلایه (MLP)

این دسته از شبکه‌های عصبی از اوایل دهه ۱۹۹۰ تاکنون در زمینه هیدرولوژی نیز به‌طور موفقیت‌آمیز به کار گرفته شده‌اند (Dawson et al., 2006). معماری معمول این شبکه متشکل از سه لایه ورودی، پنهانی و خروجی است. تعداد نرون‌های موجود در لایه ورودی و خروجی بستگی به نوع مسئله دارد ولی تعداد گره‌های لایه پنهان با سعی و خطا به دست می‌آید (Mohammadi and Moazenzadeh, 2017). در تحقیقی توسط جهانی و محمدی استفاده از شبکه‌های عصبی چندلایه در برآورد تابش خورشیدی توصیه شده است، همچنین برای کسب اطلاعات بیشتر در زمینه کاربرد شبکه‌های عصبی مصنوعی در برآورد تابش می‌توان به تحقیق نام برده رجوع کرد (Jahani and Mohammadi, 2018).

شبکه عصبی مصنوعی تابع پایه شعاعی (RBF):

شبکه RBF یک شبکه سه لایه می‌باشد که لایه ورودی یک بافر است. لایه دوم یا لایه مخفی، نگاشتی غیرخطی از فضای ورودی به فضایی با ابعاد بزرگ‌تر است که امکان جداسازی خطی الگوها را میسر می‌سازد. لایه خروجی، مجموع وزنی ساده با خروجی خطی است. اگر هدف از RBF تقریب تابع باشد، این خروجی مناسب است. اما اگر هدف دسته‌بندی الگوها باشد، باید از تابع سیگموئید استفاده نمود. ویژگی منحصر به فرد شبکه RBF فرایند انجام شده در لایه مخفی است (Shamshirband et al., 2015). اگر الگوها در فضای ورودی، خوشه‌ها را تشکیل دهند و اگر مرکز این خوشه‌ها شناخته شده باشد، فاصله تا مرکز خوشه

ماشین بردار پشتیبان توسط Vapnik در سال ۱۹۹۸ بر پایه‌ی تئوری یادگیری آماری معرفی شد. این روش مبتنی بر طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این‌رو روشی مناسب برای مسائل پیش‌بینی به شمار می‌رود (Pai & Hong, 2007). ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌شود (Mohammadi and Emamgholizadeh, 2017). شکل (۳) ساختار ماشین بردار پشتیبان را نشان می‌دهد.

در مدل ماشین بردار پشتیبان تابعی مرتبط با متغیر وابسته y که خود تابعی از چند متغیر مستقل x است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به‌علاوه مقداری اغتشاش (خطای مجاز ε) مشخص شود (Yoon et al., 2001).

$$f(x) = W^T \cdot \phi(x) + b$$

$$y = f(x) + \text{noise}$$

چنانچه W بردار ضرایب، b ثابت مشخصه‌های تابع رگرسیونی و $\phi(x)$ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با واسنجی مدل ماشین بردار پشتیبان توسط مجموعه‌ای از نمونه‌ها (مجموعه واسنجی) محقق می‌شود. برای محاسبه W و b لازم است تابع خطا (معادله ۳) در مدل ماشین بردار پشتیبان رگرسیونی با در نظر گرفتن شرایط مندرج (قیود) در معادله (۴) بهینه شود.

$$W^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \varepsilon_i^* \quad (\text{رابطه ۳})$$

$$(\text{رابطه ۴})$$

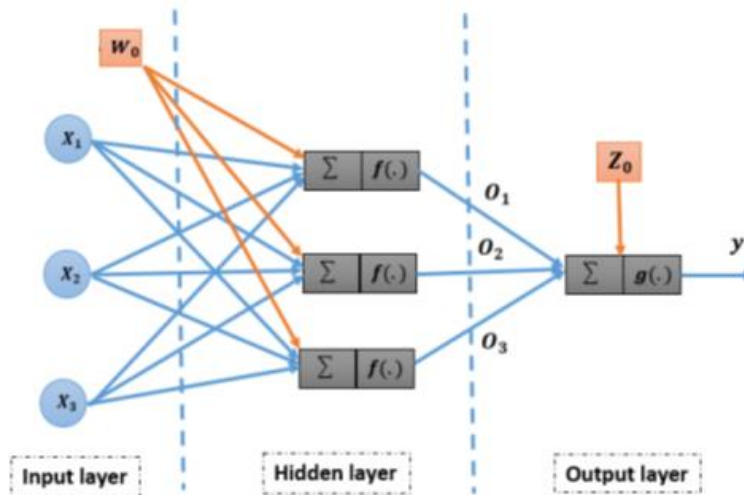
$$\varepsilon_i, \varepsilon_i^* \geq 0, i = 1, 2, \dots, N$$

$$\frac{1}{2} W^T w + C \sum_{i=1}^N \varepsilon_i + C \sum_{i=1}^N \varepsilon_i^* y_i - W^T \cdot \phi(x_i) - b \leq \varepsilon + \varepsilon_i$$

در معادلات بالا C عددی صحیح و مثبت است که عامل تعیین جرمه در هنگام رخ دادن خطای واسنجی مدل است، $\phi(x)$ تابع کرنل، N تعداد نمونه‌ها و دو مشخصه ε_i و ε_i^* متغیرهای کمبود هستند (Mohammadi and Emamgholizadeh, 2017). در نهایت، تابع ماشین بردار پشتیبان را می‌توان به فرم زیر باز نویسی کرد:

$$f(x) = \sum_{i=1}^N \bar{\alpha}_i \cdot \phi(x_i)^T \cdot \phi(x) + b \quad (\text{رابطه ۵})$$

شناخته شده که معمولترین تابع آن، تابع گوسین است. این شبکه در مقایسه با شبکه عصبی MLP به نرون بیشتری نیاز دارد؛ اما طراحی آن سریعتر از شبکه MLP بوده و سریعتر آموزش می‌بیند. با تعداد نرون کافی می‌توان هر تابع منطقی را به‌طور تقریبی نشان داد (Shamshirband et al., 2015).



شکل ۲. شمای کلی شبکه‌های عصبی مصنوعی پرسپترون چندلایه

t_j مقدار مورد انتظار می‌باشد.

معیارهای ارزیابی

با استفاده از آماره‌های مختلفی می‌توان عملکرد مدل‌ها را مورد ارزیابی و مقایسه قرار داد. یکی از این آماره‌ها استفاده از معیارهای ارزیابی می‌باشد. از جمله معیارهای ارزیابی پرکاربرد در علوم مهندسی آب، ریشه میانگین مربعات خطا ($RMSE^2$)، میانگین قدر مطلق خطای نسبی ($MAPE^3$) و ضریب تبیین (R^2) می‌باشند که به ترتیب، در روابط ۸، ۹ و ۱۰ آورده شده‌اند. دقیق‌ترین مدل با توجه به این معیارها، مدلی خواهد بود که مقدار این سه معیار به ترتیب نزدیک به صفر، صفر و یک باشد (Mohammadi, 2017).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (\text{رابطه ۸})$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - O_i}{O_i} \right| \times 100 \quad (\text{رابطه ۹})$$

قابل محاسبه است. سنجش مسافت به صورت غیرخطی انجام می‌شود. بنابراین، اگر الگو در ناحیه نزدیک مرکز خوشه باشد، مقدار آن نزدیک به یک خواهد بود. اطراف این ناحیه، مقدار مذکور به شدت کاهش می‌یابد. این ناحیه اطراف مرکز خوشه، متقارن است، به گونه‌ای که تابع غیرخطی به شعاعی - پایه

برنامه‌ریزی ژنتیک (GEP)

روش برنامه‌ریزی بیان ژن، در سال ۱۹۹۹ توسط فریرا ارائه شد (Ferreira, 2001). این روش ترکیبی از روش‌های برنامه‌ریزی ژنتیک بوده که در آن، کروموزوم‌های خطی و ساده با طول ثابت؛ مشابه با آنچه که در الگوریتم ژنتیک استفاده می‌شود؛ و ساختارهای شاخه‌ای با اندازه‌ها و اشکال متفاوت؛ مشابه با درختان تجزیه در برنامه‌ریزی ژنتیک؛ ترکیب می‌شوند. از آنجایی که در این روش تمام ساختارهای شاخه‌ای با اندازه و اشکال متفاوت، در کروموزوم‌های خطی با طول ثابت کدگذاری می‌شوند، سبب گردیده که در این روش فنوتیپ و ژنوتیپ از هم جدا شوند و سیستم بتواند از تمام مزایای تکاملی، استفاده کند (Ferreira, 2001). در این روش پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و مجموعه‌ای از ترمینال‌ها، مدل‌سازی می‌شوند. در این تحقیق برای به کارگیری روش برنامه‌ریزی بیان ژن از نرم‌افزار GenXproTools استفاده گردید. تابع برازش GEP به شکل زیر تعریف می‌شود:

$$f_t = \sum_{j=1}^n (R - |P_{ij} - T_j|) \quad (\text{رابطه ۷})$$

که در آن R محدوده انتخاب، P_{ij} مقدار پیش‌بینی شده و

۲. Root Mean Square Error

۳. Mean absolute percentage error

۴. Coefficient of determination

۱. Generation Expansion Planning.

پس از تعیین مؤلفه‌های اصلی تابش خورشیدی ایستگاه‌های همدیدی کرمان و مشهد، درصد واریانس تجمعی مؤلفه‌های اصلی محاسبه شد. سپس در هر ایستگاه تعداد تجمعی مؤلفه‌های اصلی‌ای که هشتاد درصد واریانس کل داده‌ها را دارا باشند، به عنوان ورودی منتخب روش تحلیل مؤلفه اصلی انتخاب شده‌اند. با توجه به شکل (۳) برای ایستگاه کرمان مقدار اولین مؤلفه اصلی ۶۳/۹۷ درصد از کل واریانس داده‌ها و مؤلفه دوم ۲۶/۱۹ درصد از کل واریانس داده‌ها را شامل می‌گردد. این دو مؤلفه مجموعاً ۹۰/۱۶ درصد از پراکندگی داده‌های اصلی را شامل می‌شوند که این دو مؤلفه به عنوان ورودی مدل‌های تخمین‌گر برای تخمین تابش در ایستگاه کرمان در نظر گرفته شدند. برای ایستگاه مشهد مؤلفه اول ۸۱ درصد واریانس کل داده‌ها را شامل شده است که به عنوان ورودی مدل‌های تخمین‌گر برای تخمین تابش در ایستگاه مشهد در نظر گرفته شده است. همچنین با توجه به شکل (۳) مؤلفه‌های ۳، ۴، ۵، ۶، ۷، ۸ و ۹ برای ایستگاه‌های کرمان و مشهد جزو ورودی‌های تقریباً بی‌تأثیر می‌باشند و حتی ورود این مؤلفه‌ها به مدل امکان افزایش خطا را به همراه خواهد داشت. بنابراین با توجه به شکل (۳) در این تحقیق جهت تخمین تابش خورشیدی ایستگاه‌های همدیدی کرمان و مشهد با استفاده از روش PCA، به ترتیب از دو و یک مؤلفه اصلی اول استفاده شد.

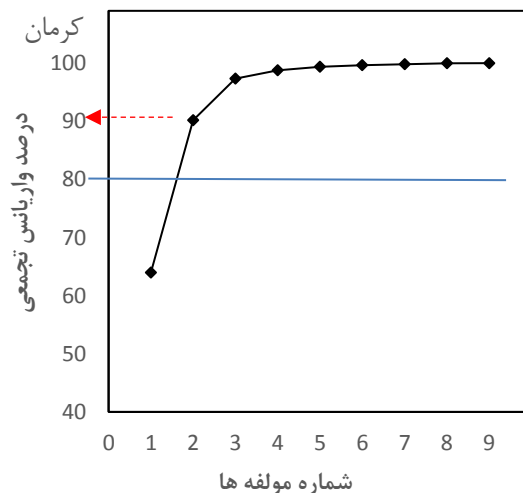
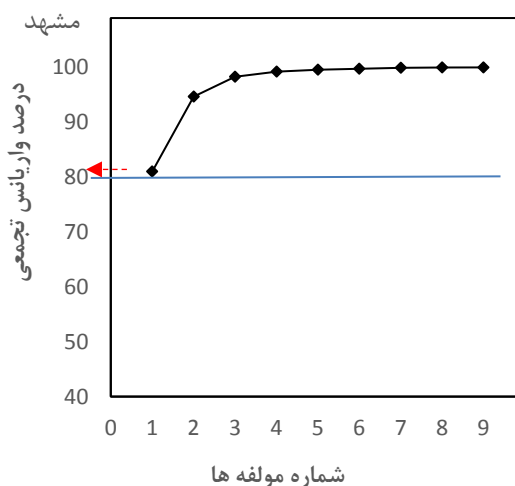
$$R^2 = \left(\frac{\sum (O_i - \bar{O}_i) \cdot (P_i - \bar{P}_i)}{\sqrt{\sum (O_i - \bar{O}_i)^2 \cdot \sum (P_i - \bar{P}_i)^2}} \right)^2 \quad (\text{رابطه ۱۰})$$

در روابط بالا O_i مقادیر مشاهداتی تابش خورشیدی، P_i مقادیر پیش‌بینی شده تابش خورشیدی، \bar{O}_i میانگین مقادیر مشاهداتی تابش خورشیدی، \bar{P}_i میانگین مقادیر پیش‌بینی شده تابش خورشیدی و n تعداد داده‌ها می‌باشد. همچنین برای مدل‌سازی و پیاده‌سازی روش‌های ذکر شده از نرم‌افزار متلب (نسخه ۲۰۱۶b) استفاده شده است.

نتایج و بحث

پس از تعیین ساختارهای ورودی هر یک از مدل‌ها و تعیین متغیرهای ورودی مؤثر با روش‌های پیش‌پردازش تحلیل مؤلفه اصلی و تئوری آنتروپی، هر یک از مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، برنامه‌ریزی ژنتیک (GEP) و ماشین بردار پشتیبان (SVM) به ازای داده‌های آموزش مورد واسنجی قرار گرفته و سپس عملکرد مدل‌های آموزش دیده به ازای داده‌های بخش صحت‌سنجی ارزیابی گردیده است. در ادامه خلاصه نتایج مربوط به هر یک از مدل‌ها در تخمین تابش خورشیدی ارائه شده است.

تعیین تعداد مؤلفه اصلی ورودی



شکل ۳. درصد واریانس تجمعی مؤلفه‌های اصلی تابش خورشیدی در ایستگاه‌های کرمان و مشهد

۵۳/۰۶ (MJ/m^2)، میانگین قدر مطلق خطای نسبی ۹/۱۸ و ضریب تبیین ۰/۸۶۵ نسبت به سایر مدل‌ها در ایستگاه کرمان عملکرد بهتری در تخمین تابش خورشیدی داشته است. مدل PCA-SVM نسبت به مدل‌های PCA-RBF و PCA-GEP با مقدار

مدل‌سازی همراه با پیش‌پردازش توسط تحلیل مؤلفه اصلی نتایج جدول (۲) گویای مدل‌سازی با استفاده از داده‌های حاصل از پیش‌پردازش PCA می‌باشد. با توجه به نتایج قسمت آموزش جدول (۲) مدل PCA-MLP با مجذور میانگین مربعات خطای

ایستگاه مشهد عملکرد بهتری در تخمین تابش خورشیدی داشته است. دومین مدل برتر، پس از مدل PCA-SVM، با اختلاف خطایی خیلی کم نسبت به مدل PCA-RBF با مقدار مجذور میانگین مربعات خطای کمتر در تخمین تابش در ایستگاه مشهد مناسبتر تشخیص داده شده است. همچنین مدل PCA-GEP با بیشترین خطا به عنوان ضعیفترین مدل در بین مدل‌های مذکور در ایستگاه مشهد بوده است.

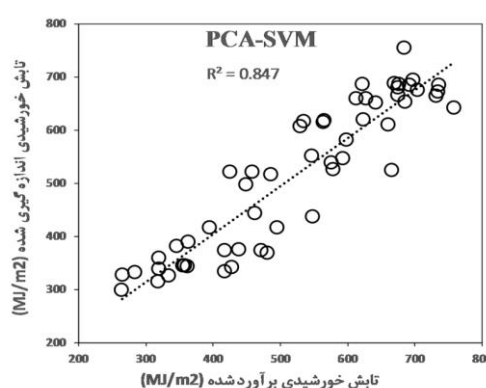
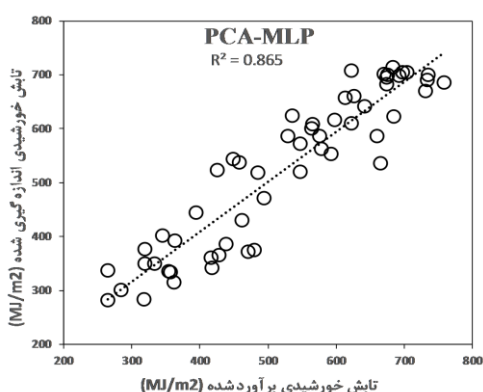
مجذور میانگین مربعات خطای کمتر در تخمین تابش در ایستگاه کرمان مناسبتر تشخیص داده شده است. همچنین مدل PCA-GEP با بیشترین خطا به عنوان ضعیفترین مدل در بین مدل‌های مذکور در ایستگاه کرمان بوده است. در ایستگاه مشهد مدل PCA-MLP با ریشه میانگین مربعات خطای $75/79$ (MJ/m^2)، میانگین درصد خطای مطلق $14/23$ و ضریب تبیین $0/772$ نسبت به سایر مدل‌ها در

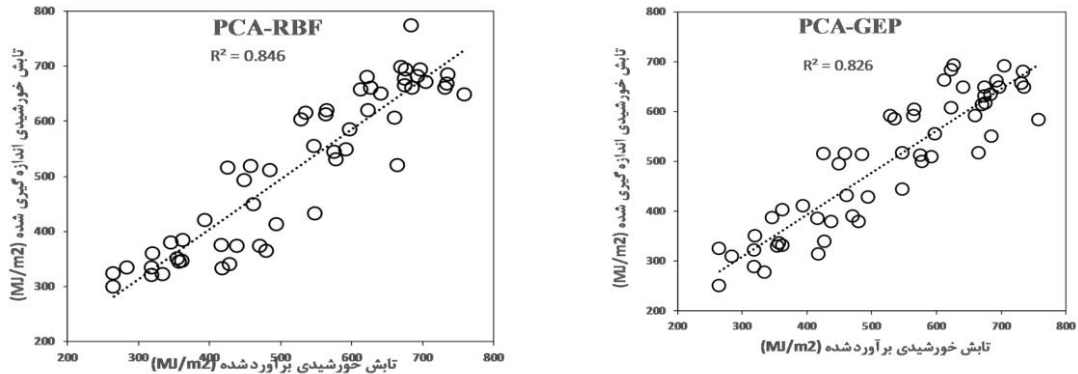
جدول ۲- نتایج مدل‌سازی با پیش‌پردازش تحلیل مؤلفه اصلی برای دو ایستگاه همدیدی کرمان و مشهد

ایستگاه	مدل	آموزش			آزمون		
		RMSE (MJ/m^2)	MAPE	R ²	RMSE (MJ/m^2)	MAPE	R ²
کرمان	PCA-MLP	۴۶/۱۹	۷/۸۸	۰/۸۹۸	۵۳/۰۶	۹/۱۸	۰/۸۶۵
	PCA-RBF	۴۸/۵	۸/۵۶	۰/۸۸۵	۵۷	۹/۱۵	۰/۸۴۷
	PCA-SVM	۴۸/۸۶	۸/۷۷	۰/۸۸۳	۵۶/۶۸	۹/۲۲	۰/۸۴۸
مشهد	PCA-GEP	۵۷/۱۲	۱۰/۰۹	۰/۸۴۷	۶۵/۱۱	۱۰/۶۶	۰/۸۲۶
	PCA-MLP	۴۷/۹۲	۱۱/۶۴	۰/۸۹۴	۷۵/۷۹	۱۴/۲۳	۰/۷۷۲
	PCA-RBF	۴۸/۴۶	۱۱/۸	۰/۸۹۱	۷۷/۹۱	۱۴/۳۶	۰/۷۶
	PCA-SVM	۴۸/۴۷	۱۱/۸	۰/۸۹۱	۷۷/۶۹	۱۴/۳۵	۰/۷۶۱
	PCA-GEP	۱۳۴/۶۸	۲۷/۲۵	۰/۸۸۱	۱۷۸/۱۹	۳۰/۸۳	۰/۷۴۷

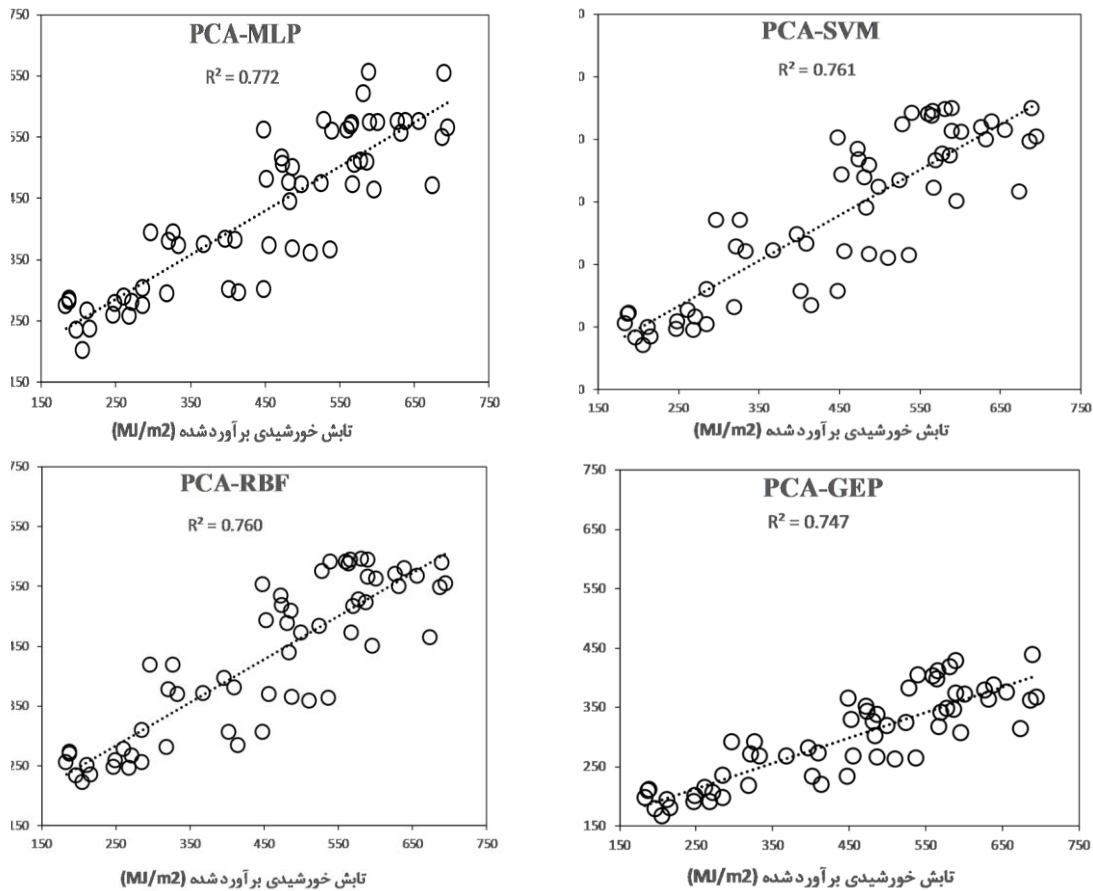
به ترتیب بهترین و ضعیفترین ضریب تبیین را در بین مقادیر مشاهداتی و مقادیر تخمین زده شده در ایستگاه کرمان دارا می‌باشند. همچنین مدل‌های PCA-MLP با همبستگی $0/772$ و مدل PCA-GEP با ضریب تبیین $0/747$ به ترتیب بهترین و ضعیفترین همبستگی را در ایستگاه مشهد دارند.

شکل‌های (۴) و (۵) نمودار پراکنش نتایج بخش آموزش مدل‌سازی مدل‌های مورد استفاده را همراه با پیش‌پردازش روش تحلیل مؤلفه اصلی در ایستگاه‌های کرمان و مشهد نشان می‌دهد. همان‌طور که در شکل (۴) مشاهده می‌شود مدل PCA-MLP با ضریب تبیین $0/865$ و مدل PCA-GEP با همبستگی





شکل ۴. نمودار پراکنش نقاط تخمین زده شده و مشاهداتی با استفاده از پیش پردازش تحلیل مؤلفه اصلی مربوط به بخش آزمون در ایستگاه کرمان



شکل ۵. نمودار پراکنش نقاط تخمین زده شده و مشاهداتی با استفاده از پیش پردازش تحلیل مؤلفه اصلی مربوط به بخش آزمون در ایستگاه مشهد

متغیرهای بعدی به ترتیب میانگین کمبود فشار بخار آب اشباع، دمای کمینه، دمای بیشینه، ساعت آفتابی و رطوبت نسبی به ترتیب بیشترین تأثیر را در ارتباط با تابش خورشیدی دارند. همچنین متغیر محتوی بخار آب جو کم‌ترین تأثیر را در ارتباط با تابش خورشیدی در ایستگاه کرمان دارد. در ایستگاه مشهد مؤثرترین متغیر و کم تأثیرترین متغیر در ارتباط با تابش خورشیدی به ترتیب متغیر دمای میانگین و محتوی بخار آب جو بودند. برای تعیین ترکیب بهینه با استفاده از تئوری آنتروپی نمودار مقادیر احتمال آنتروپی در برابر متغیرهای

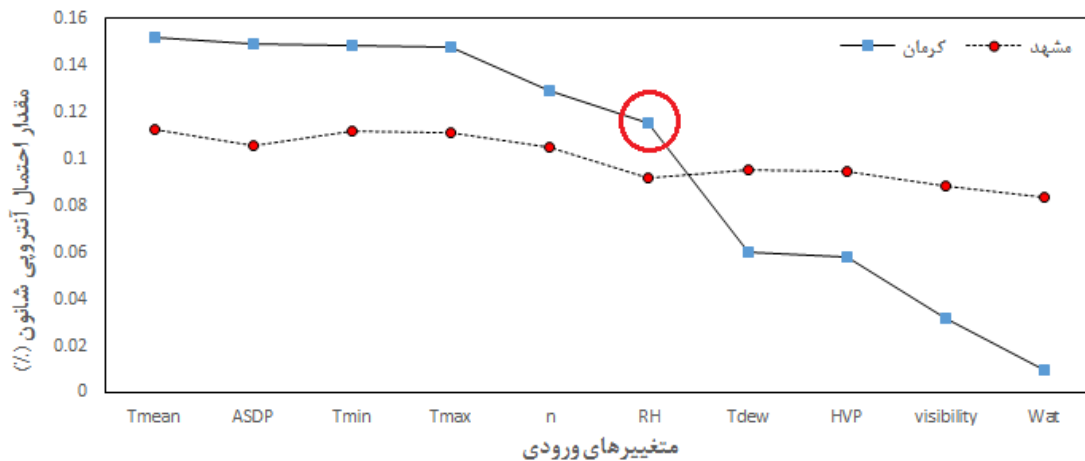
تعیین وزن‌های ورودی‌ها توسط آنتروپی شانون

در این تحقیق تئوری آنتروپی شانون برای تک‌تک متغیرهای ورودی محاسبه شد که نتایج مربوط به آن در شکل‌های (۶) و (۷) آورده شده است. براساس تئوری آنتروپی متغیری که دارای بیشترین احتمال وقوع باشد، آن متغیر ارتباط معنی‌داری با تابش خورشیدی دارد.

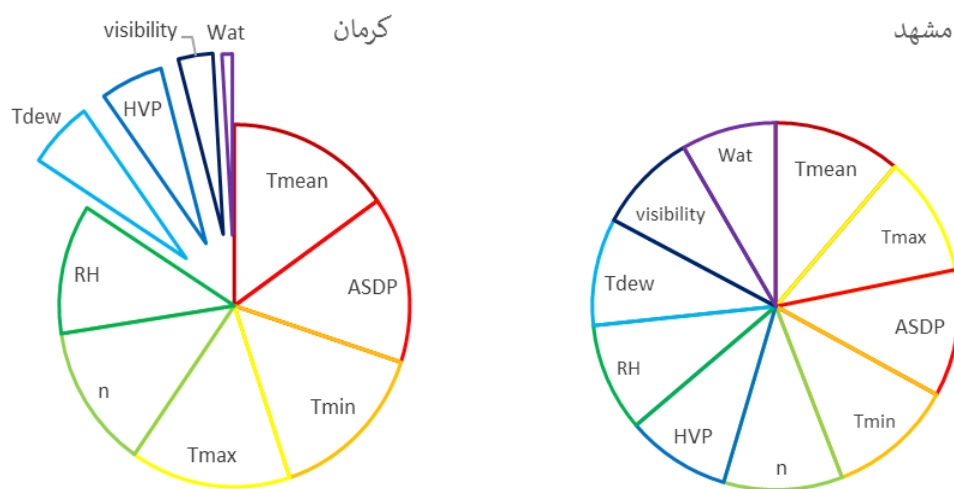
با توجه به شکل (۶) در ایستگاه کرمان متغیر دمای میانگین با بیش‌ترین مقدار احتمال آنتروپی، مؤثرترین متغیر برای تخمین تابش خورشیدی شناسایی شد. در ایستگاه کرمان

مؤثر در تخمین تابش خورشیدی شناسایی شده‌اند. شکل (۷) نیز نتایج فوق را تأیید می‌کند. در این شکل، در ایستگاه مشهد همه متغیرها تقریباً احتمال نزدیک به یکدیگر و سهم یکسانی در درصد کل احتمال رخداد همه متغیرها دارند، پس همه متغیرها به عنوان ورودی مدل‌های تخمین‌گر انتخاب شدند. در ایستگاه کرمان چهار متغیر دمای نقطه شبنم، فشار بخار ساعتی، دید افقی و محتوی بخار آب جو درصد کمتری از مقدار احتمال آن‌تروپی کل را شامل می‌شدند، پس این متغیرها به عنوان ورودی مدل‌ها در نظر گرفته نشدند و متغیرهای میانگین دمای میانگین کمبود فشار بخار آب اشباع، دمای کمینه، ساعت آفتابی و رطوبت نسبی به عنوان ترکیب ورودی به مدل در ایستگاه کرمان انتخاب شدند.

ورودی رسم شد. نقطه‌ای که نمودار دچار تغییر شیب شدیدی شده است به عنوان نقطه عطف انتخاب شد. همان‌طور که مشاهده می‌شود، در ایستگاه کرمان مقادیر آن‌تروپی قبل از نقطه عطف (دایره قرمز در شکل ۶) بسیار نزدیک به هم بوده‌اند. مقادیر بعد از این نقطه در مقایسه با مقادیر قبلی دچار کاهش شدیدی شده‌اند. پس با توجه به اینکه متغیرهایی که مقدار آن‌تروپی آن‌ها بیشتر است در تخمین تبخیر مؤثرتر هستند، متغیرهای میانگین دمای، میانگین کمبود فشار بخار آب اشباع، دمای کمینه، دمای بیشینه، ساعت آفتابی و رطوبت نسبی به عنوان ترکیب ورودی به مدل در ایستگاه کرمان انتخاب شدند. در ایستگاه مشهد مقادیر احتمال آن‌تروپی تفاوت چشم‌گیری با یکدیگر ندارند، در نتیجه تمام متغیرها به عنوان متغیر



شکل ۶. تعیین نقطه عطف و انتخاب ترکیب مؤثر در تخمین تابش خورشیدی



شکل ۷. متغیرهای مؤثر انتخاب شده توسط روش آن‌تروپی در نمودار دایره‌ای

نتایج مدل‌سازی با استفاده از داده‌های حاصل از پیش‌پردازش

مدل‌سازی همراه با پیش‌پردازش توسط تئوری آن‌تروپی

ضعیف‌ترین مدل در بین مدل‌های مذکور در ایستگاه مشهد بوده است.

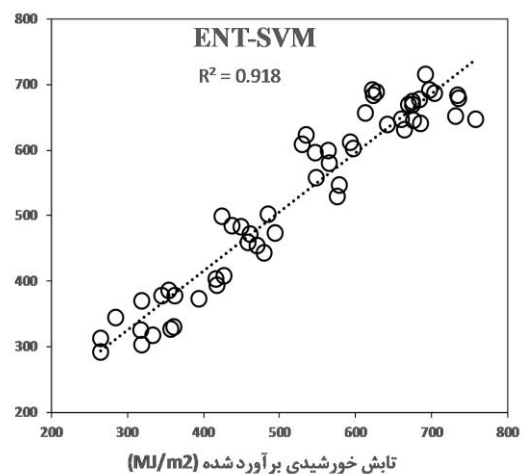
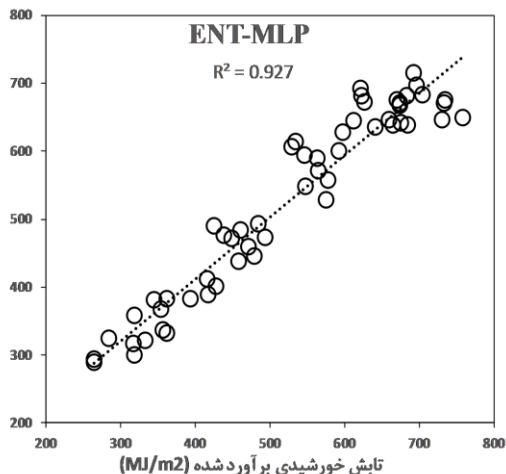
شکل‌های (۸) و (۹) نمودار پراکنش نتایج بخش آزمون مدل‌سازی مدل‌های مورد استفاده را همراه با پیش‌پردازش روش تئوری آنتروپی در ایستگاه‌های کرمان و مشهد نشان می‌دهد. همان‌طور که مشاهده می‌شود مدل ENT-MLP و ENT-RBF با ضریب تبیین ۰/۹۲۷۵ بیش‌ترین ضریب تبیین را در بین مدل‌های مذکور در ایستگاه کرمان دارند. مدل PCA-GEP با ضریب تبیین ۰/۷۹۸ ضعیف‌ترین ضریب تبیین را در بین مدل‌های مذکور در ایستگاه کرمان دارد. همچنین مدل‌های ENT-SVM با ضریب تبیین ۰/۷۹۰ و مدل ENT-GEP با همبستگی ۰/۷۳۴ به ترتیب بهترین و ضعیف‌ترین مدل منتخب در ایستگاه مشهد می‌باشند.

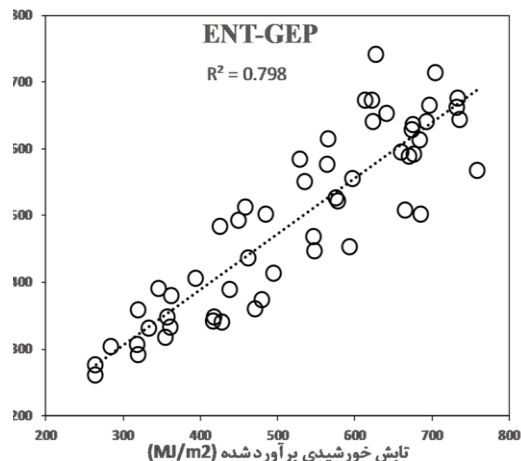
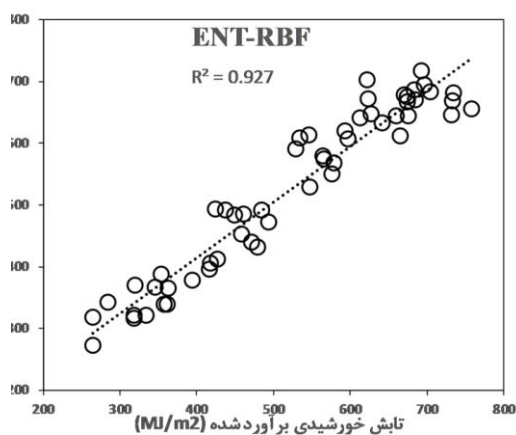
تئوری آنتروپی در جدول (۳) آورده شده است. با توجه به نتایج قسمت آموزش جدول (۳) مدل ENT-MLP با مجذور میانگین مربعات خطای ۳۸/۳۶ (MJ/m²)، میانگین درصد خطای مطلق ۵/۹۴ و ضریب تبیین ۰/۹۲۷ نسبت به سایر مدل‌ها در ایستگاه کرمان عملکرد بهتری در تخمین تابش خورشیدی داشته است. همچنین مدل ENT-GEP با بیش‌ترین خطا به عنوان ضعیف‌ترین مدل در بین مدل‌های مذکور در ایستگاه کرمان بوده است.

در ایستگاه مشهد مدل ENT-SVM با مجذور میانگین مربعات خطای ۷۴/۱۹ (MJ/m²)، میانگین درصد خطای مطلق ۱۴/۱ و ضریب تبیین ۰/۷۹۱ نسبت به سایر مدل‌ها در ایستگاه مشهد عملکرد بهتری در تخمین تابش خورشیدی داشته است. همچنین مدل PCA-GEP با بیش‌ترین خطا به عنوان

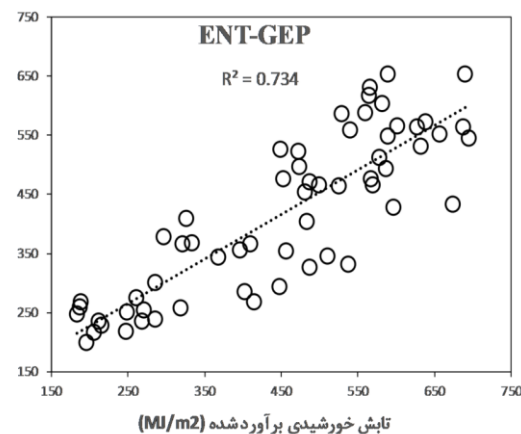
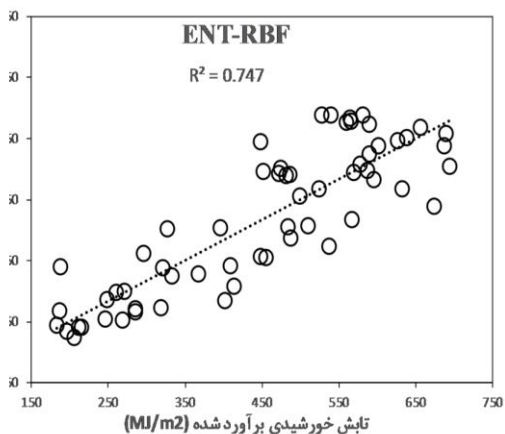
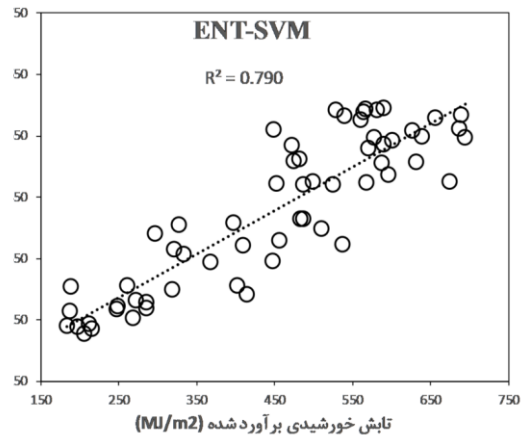
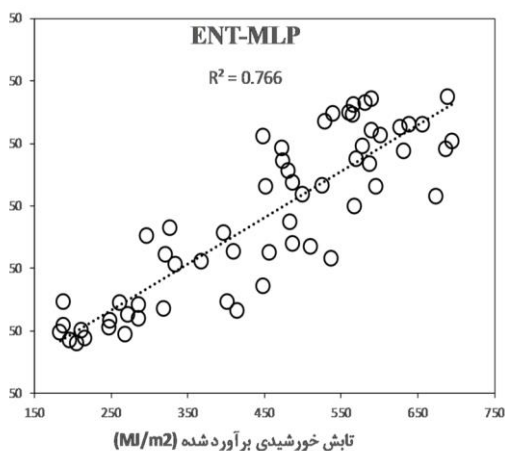
جدول ۳. نتایج مدل‌سازی با پیش‌پردازش تئوری آنتروپی برای دو ایستگاه همدیدی کرمان و مشهد

ایستگاه مدل	آموزش			آزمون			متغیرها و ساختار مدل
	R ²	MAPE	RMSE (MJ/m ²)	R ²	MAPE	RMSE (MJ/m ²)	
ENT-MLP	۰/۹۳۹	۵/۸۵	۳۵/۴۵	۰/۹۲۷	۵/۹۴	۳۸/۳۶	۶-۴-۱
ENT-RBF	۰/۹۴۵	۵/۶۸	۳۳/۴۲	۰/۹۲۷	۵/۹۷	۳۸/۵۸	۲۰, spread=۵ neuron=
ENT-SVM	۰/۹۳۷	۶/۰۶	۳۶/۰۲	۰/۹۱۸	۶/۶۳	۴۰/۸۴	۶/۵۹, σ=۱۲/۳۷ γ=
ENT-GEP	۰/۸۳۴	۹/۹۶	۵۹/۹۲	۰/۷۹۸	۱۰/۵	۷۱/۶۲	GN=3000
ENT-MLP	۰/۸۹۹	۱۱/۵۲	۴۶/۷۵	۰/۷۶۷	۱۴/۳۵	۷۵/۸۷	۱۰-۷-۱
ENT-RBF	۰/۹۲۵	۹/۷۶	۴۰/۳۴	۰/۷۴۷	۱۵/۳۸	۸۴/۳۲	۲۰, spread=۶ neuron=
ENT-SVM	۰/۹۱۸	۱۰/۲۴	۴۲/۰۷	۰/۷۹۱	۱۴/۱	۷۴/۱۹	۹/۷۴, σ=۱۸/۴۴ γ=
ENT-GEP	۰/۸۶۲	۱۲/۸	۵۶/۶۱	۰/۷۳۴	۱۵/۲۹	۸۵/۷۲	GN=3000





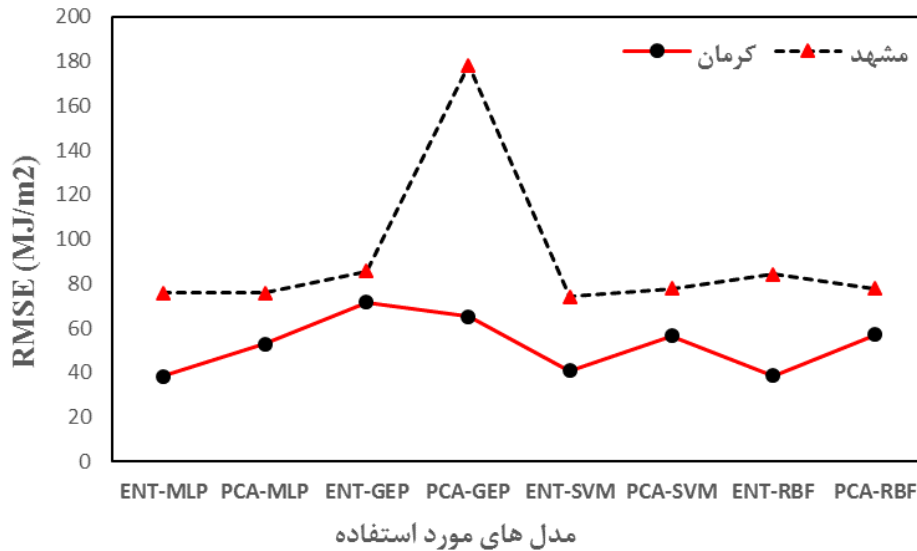
شکل ۸- نمودار پراکنش نقاط تخمین زده شده و مشاهداتی با استفاده از پیش پردازش تئوری آنتروپی مربوط به بخش آزمون در ایستگاه کرمان



شکل ۹- نمودار پراکنش نقاط تخمین زده شده و مشاهداتی با استفاده از پیش پردازش تئوری آنتروپی مربوط به بخش آزمون در ایستگاه مشهد

در جایگاه دوم قرار می‌گیرد. مدل برنامه‌ریزی ژنتیک هم با هر دو روش پیش‌پردازش به کار رفته و در هر دو ایستگاه عملکرد ضعیفی نسبت به سایر مدل‌ها داشته است، به‌خصوص مدل PCA-GEP در ایستگاه کرمان که ضعیف‌ترین عملکرد را در بین کل مدل‌های مذکور داشته است.

با توجه به نتایج حاصل و شکل (۱۰)، با مقایسه در نحوه عملکرد مدل‌ها، مدل‌های PCA-MLP و ENT-MLP را می‌توان به عنوان مدل‌های منتخب برای هر دو ایستگاه انتخاب کرد. پس از مدل شبکه عصبی مصنوعی، مدل ماشین بردار پشتیبان با هر دو پیش‌پردازش PCA و ENT عملکرد مناسبی داشته و



شکل ۱۰- مقایسه نتایج بخش آزمون مدل سازی مدل های مذکور با شاخص ریشه میانگین مربعات خطا

گرفت در ایستگاه کرمان مدل ENT-MLP با مجذور میانگین مربعات خطای $38/36$ (MJ/m^2)، میانگین درصد خطای مطلق $5/94$ و همبستگی $0/927$ نسبت به سایر مدل ها عملکرد بهتری در تخمین تابش خورشیدی داشته است. همچنین در ایستگاه مشهد مدل ENT-SVM با مجذور میانگین مربعات خطای $19/47$ (MJ/m^2)، میانگین درصد خطای مطلق $14/1$ و همبستگی $0/791$ نسبت به سایر مدل ها عملکرد بهتری در تخمین تابش خورشیدی داشته است. در کل پیش پردازش آنتروپی نشان داد که این روش ورودی های بهتری را برای مدل های تخمین گر می تواند تعیین کند. این نتیجه گیری که روش تئوری آنتروپی باعث بهبود نتایج مدل سازی می شود با نتایج سایر پژوهشگران همخوانی دارند، از جمله پژوهش (Remesan et al. (2008 و Biazar (2017). همچنین می توان نتیجه گرفت که روش تحلیل مؤلفه اصلی باعث بهبود نتایج مدل سازی می شود. این نتایج با نتایج سایر پژوهشگران همخوانی دارد، از جمله پژوهش Noori et al. (2010) و Mohammadi and Emamgholizadeh (2017) که همگی اذعان داشتند روش تحلیل مؤلفه اصلی باعث بهبود نتایج مدل سازی می شود.

نتیجه گیری

با توجه به اهمیت تابش خورشیدی در برنامه ریزی و مدیریت منابع آب و لزوم استفاده از روش های نوین در برآورد تابش خورشیدی، در این پژوهش از روش های تحلیل مؤلفه اصلی و تئوری آنتروپی برای پیش پردازش داده ها استفاده شد و ورودی های مدل های تخمین گر توسط دو روش مذکور شناسایی شدند. مدل سازی با مدل های شبکه عصبی مصنوعی پرسپترون چندلایه (MLP)، شبکه عصبی مصنوعی تابع پایه شعاعی (RBF)، برنامه ریزی ژنتیک (GEP) و ماشین بردار پشتیبان (SVM) انجام گرفت. نتایج تئوری آنتروپی نشان داد در ایستگاه کرمان متغیرهای میانگین دمای، میانگین کمبود فشار بخار آب اشباع، دمای کمینه، دمای بیشینه، ساعت آفتابی و رطوبت نسبی متغیرهای مؤثر در برآورد تابش بودند و همچنین در ایستگاه مشهد متغیرهای میانگین دما، میانگین کمبود فشار بخار اشباع، دمای کمینه، دمای بیشینه، ساعت آفتابی، رطوبت نسبی، دمای نقطه شبنم، فشار بخار ساعتی، دید افقی و محتوی بخار آب جو متغیرهای مؤثر در برآورد تابش بودند. در نهایت با توجه به موارد گفته شده و نتایج حاصله می توان چنین نتیجه

REFERENCES

- Aghashariatmadary, Z. (2011). Evaluation of model for estimating total solar radiation at horizontal surfaces based on meteorological data, with emphasis on the performance of the angstrom model over iran. Ph. D. dissertation, University of Tehran. College of Agriculture and Natural Resources. (In Farsi).
- Angstrom, A. 1924. Solar and terrestrial radiation. Quart. J. Roy. Met., 50: 121-125.
- Ball RA, Purcell LC, Carey SK (2004) Evaluation of solar radiation prediction models in North America. *Agronomy Journal* 96:391-397
- Azadeh, A., Maghsoudi, A. and Sohrabkhani, S. 2009. An integrated artificial neural networks approach for predicting global radiation. *Energy Conversion and Management*, 50: 1497-1505.
- Azeez MAA (2011) Artificial neural network estimation of global solar radiation using meteorological parameters in Gusau, Nigeria. *Artificial Applied Science Research* 3(2):586-95

- Biazar, S.M. (2017) Input variables determination using Gamma test and Entropy theory for daily evaporation prediction. Thesis is approved for the degree of Master of Science in Water Resources. Department of Water Engineering, Faculty of Agriculture, University of Tabriz. September 2017. (In Farsi)
- Chen JL, Li GS, Wu SJ (2013) Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy Conversion and Management* 75:311–318
- Dawson, C.W., Abrahart, R.J., Shamseldin, A.Y. and R.L. Wibly. 2006, Flood estimation at ungauged sites using artificial neural networks. *Journal of Hydrology*. 319 (1-4): 391-409.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. *Complex Systems*, Vol.13(2): 87–129.
- Hargreaves, G.H., and Samani, Z.A. 1982. Estimating potential evapotranspiration. *Journal of Irrigation and Drainage Engineering ASCE*, 108: 223-230.
- Harmancioglu, N. B. 1984. Entropy concept as used in determination of optimum sampling intervals. Proc. of Hydrosoft 84, International Conf. on Hydraulic Engineering Software, September 10-14, 1984. Portoroz, Yugoslavia, pp. 6-99 and 6-110.
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and L. Fernandez. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research* 34(3): 807–816.
- Jahani, B., and Mohammadi, B. 2018. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. *Theor Appl Climatol* (2018). <https://doi.org/10.1007/s00704-018-2666-3>
- Kamali, S., Aghasariatmadari, Z. (2018). Evaluation of the effects of atmospheric pollutants on the performance of Angstrom- Prescott equation in estimating solar radiation (Case Study: Karaj). *Iranian Journal of Soil and Water Research*, 48(5), 1053-1061. doi: 10.22059/ ijswr. 2018. 233499. 667682
- Khalili and Rezai Sadr. (1997). Estimation of solar radiation in Iran, based on climate data. *Journal of Geographical Research* 84: 15-35. (In Farsi).
- Lazzus JA, Ponce AAP, Marin J (2011) Estimation of global solar radiation over the city of La Serena (Chile) using a neural network. *Applied Solar Energy* 47(1):66–73
- Long H, Zhang Z, Su Y (2014) Analysis of daily solar power prediction with data-driven approaches. *Applied Energy* 126:29–37
- Mohammadi, B. (2017) Daily Evaporation prediction based on a hybridization of Artificial Neural Network and firefly Optimization Algorithm. Thesis is approved for the degree of Master of Science in Water Resources. Department of Water Engineering, Faculty of Agriculture, University of Tabriz. July 2017. (In Farsi)
- Mohammadi, B. Emamgholizadeh, S. (2017) Using principal component analysis to inputs the effective rainfall estimates based on entries to help support vector machine and artificial neural network. *Journal of Rainwater Catchment Systems*. 2017; 4 (4) ,67-75. (In Farsi)
- Mohammadi, B. Ghorbani, M A. (2016) Gamma test application in input preprocessing for time series modeling of rainfall. *Journal of Rainwater Catchment Systems*. 2016; 4 (3) ,61-72. (In Farsi)
- Mohammadi B, Moazenzadeh R. (2017) Uncertainty analysis of artificial neural network models and support vector machine in rainfall estimation. *Journal of Rainwater Catchment Systems*. 2017; 5 (1) :43-50. (In Farsi)
- Noori, R., Karbassi, A. and M. Sabahi. 2010. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste prediction. *Journal of Environmental Management*. 91(3): 767-771. 20. Singh, VP. and CY. Xu. 1997. Evaluation and generalization of 13 mass transfer equations for determining free water evaporation. *Hydrological Process*. 11:311–324
- Pai, P.F. and Hong, W.C., 2007. A recurrent support vector regression model in rainfall forecasting. *Hydrological Processes: An International Journal*, 21(6), pp.819-827
- Rahimikhoob A. 2010. Estimating global solar radiation using artificial neural network and air temperature data in a semi-arid environment. *Renew. Energy*. 35, 2131-2135.
- Remesan, R. Shamim, M.A. and Han, D. (2008). Model data selection using gamma test for daily solar radiation estimation. *Hydrological Processes*, 22, 4301-4309.
- Redy, K.S., and Ranjan, M. 2003. Solar resource estimation using artificial neural networks and comparison with other correlation models. *J. Energy Conversion and Management*, 44: 2519-2530.
- Sabziparvar A.A., and Shetaee H. 2007. Estimation of global solar radiation in arid and semi-arid climates of East and West Iran, *Energy* 32: 649–655.
- Shamshirband S, Mohammadi K, Chen HL, Narayana Samy G, Petković D, Ma C.(2015) Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran. *J Atmos Solar-Terrestrial Phys*. 2015;134:109-117. doi:10.1016/j.jastp.2015.09.014
- Soltani, S., and Morid, M. (2005). Comparative estimation of global solar radiation using Hargeavessamani and artificial neural network methodologies, *Journal of Science of Agriculture*, 15: 69-77. (In Farsi)
- Skeiker K. 2006. Correlation of global solar radiation with common geographical and meteorological parameters for Damascus province, Syria, *Energy Conversion and Management, Mgmt*, 47: 331-345.
- Tymvios, F. S., Jacovides, C. P., Michaelides, S. C.,

and Scouteli, C. 2005. Comparative study of Angstrom and artificial neural network methodologies in estimating global solar radiation. *J. Solar Energy*, 78: 752-762.

Xu, H. Xu, C. Y, Sæltun, N. R. Xu, Y. Zhou, B and

Chen, H. 2015. Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling – A case study of humid area in southern China. *Journal of Hydrology*. 525, 138-151.