



نشریه پژوهش‌های زعفران (دو فصلنامه)

جلد هشتم، شماره دوم، پاییز و زمستان ۱۳۹۹

شماره صفحه: ۳۵۸-۳۴۳

<http://dx.doi.org/10.22077/jsr.2018.1741.1067>

مقاله پژوهشی:

## کارایی مدل‌های ماشین‌بردار پشتیبان و برنامه‌ریزی بیان ژن در پیش‌بینی عملکرد زعفران

محدثه کاوسی<sup>۱</sup>، عباس خاشعی سیوکی<sup>۲\*</sup>، محمدعلی بهدانی<sup>۳</sup>

۱- دانشجوی دکترای منابع آب، دانشگاه بیرجند، ایران.

۲- استاد گروه مهندسی علوم و مهندسی آب و عضو گروه پژوهشی خشکسالی و تغییر اقلیم دانشگاه بیرجند، ایران.

۳- استاد گروه زراعت و اصلاح نباتات و عضو گروه پژوهشی زعفران، دانشگاه بیرجند، ایران.

\* نویسنده مسئول: [Email: abbaskhashei@birjand.ac.ir](mailto:abbaskhashei@birjand.ac.ir)

تاریخ دریافت: ۹۷/۰۴/۲۱؛ تاریخ پذیرش: ۹۷/۰۶/۰۷

### چکیده

با توجه به حساسیت عملکرد زعفران و تأثیرپذیری آن از پارامترهای اقلیمی و خاصیت غیرخطی توابع عملکرد، در این تحقیق به پیش‌بینی عملکرد این گیاه پرداخته شد. هدف این مطالعه، توانایی مدل شبیه‌سازی ماشین‌بردار پشتیبان (LSSVM) و مدل برنامه‌ریزی بیان ژن (GenXproTools5,0) در پیش‌بینی عملکرد زعفران بر اساس داده‌های هواشناسی (حداقل دما، حداکثر دما، بارش، تبخیر و رطوبت نسبی، عملکرد یکسال قبل) در مقیاس فصلی در بازه زمانی ۲۰۰۶-۱۹۹۲ بود. بهترین مدل بر اساس معیارهای ارزیابی  $RMSE$ ،  $R^2$  و  $MAE$  انتخاب شد. بررسی‌ها نشان داد که در هر دو مدل، در سناریوی  $H$  (میانگین حداقل دما در فصل زمستان، میانگین بارش در فصل پاییز، میانگین بارش در فصل زمستان، میانگین تبخیر در فصل زمستان، عملکرد یکسال قبل) برآورد مطلوب‌تری از عملکرد زعفران حاصل شد. در مدل LSSVM ترکیب‌های با تابع کرنل Linear نتایج دقیق‌تری داشت، اما بین مدل LSSVM و مدل GEP، مدل GEP دارای  $R^2$  بالاتر و  $RMSE$  و  $MAE$  پایین‌تری بود. میزان  $RMSE$  و  $MAE$  در این مدل تحت سناریوی  $H$  در بخش آموزش به ترتیب  $۰/۶۰۶۸۸$ ،  $۰/۴۲۲۶۵$  و  $۰/۴۶۴۳۲$  به دست آمد. در مجموع، مدل GEP نتایج دقیق‌تری را در تخمین عملکرد زعفران نسبت به مدل LSSVM دارا بود.

واژه‌های کلیدی: پیش‌بینی عملکرد، مدل LSSVM، مدل برنامه‌ریزی بیان ژن.

## مقدمه

زعفران (*Crocus sativus L.*) گیاهی چندساله است که اساساً در نواحی خشک و نیمه‌خشک کشور رویش دارد. به عبارتی، یکی از مهم‌ترین محصولات کشاورزی ایران به ویژه در استان‌های خراسان رضوی و جنوبی می‌باشد. پیش‌بینی عملکرد محصولات با استفاده از داده‌ها تأثیر مهمی در مسائل اجتماعی-اقتصادی و تصمیم‌گیری‌های سیاسی در مقیاس منطقه‌ای دارد (Koocheki et al., 2017). همچنین زعفران ارزشمندترین گیاه بومی ایران بوده و قدمت کشت آن در مناطق جنوبی و مرکزی خراسان به بیش از ۷۵۰ سال می‌رسد (Sadeghi, 1993). بررسی روند ۳۰ ساله این گیاه نشان می‌دهد که سطح زیر کشت آن حدود ۱۶ برابر افزایش یافته، در حالی که میزان متوسط عملکرد در واحد سطح در حال حاضر حدود ۶۸ درصد ۳۰ سال قبل است (Behdani, 2004) که نشان‌دهنده تأثیرپذیری عملکرد زعفران از شرایط اقلیمی و مدیریتی کشاورزی است. اقلیم از طریق تغییر در مقادیر  $CO_2$ ، درجه حرارت، بارندگی و تبخیر بر عملکرد محصولات کشاورزی تأثیر می‌گذارد. ارتباط داده‌های اقلیمی با عملکرد و سطح زیر کشت گیاهان زراعی وجود ارتباط دو سویه بین بارندگی و درجه حرارت با عملکرد و سطح زیر کشت را نشان می‌دهد (Adams et al., 2001). در سال‌های اخیر با توجه به کمبود شدید منابع آبی با کیفیت، رشد یا کشت گیاهان و محصولات زراعی در مناطق مختلف جهان و ویژه در مناطق خشک و نیمه‌خشک بر اساس قابلیت دسترسی به منابع آبی طراحی و اجرا می‌شود که اصولاً در بسیاری از موارد کارایی تولید به ازای مترمکعب آب مصرفی محاسبه می‌شود (Koozegaran, 2010). علی‌رغم تحمل‌پذیری بالای گیاه زعفران به شرایط کم‌آبی، تغییرات اقلیمی سال‌های اخیر در مناطق مختلف دنیا رشد و عملکرد گیاهان را به شدت تحت تأثیر قرار داده است.

پیش‌بینی عملکرد و تولید محصولات کشاورزی در شرایط اقلیمی و هواشناسی مختلف بر اساس داده‌های قابل دستیابی به دلیل اهمیتی که در برنامه‌ریزی اقتصادی دارند از دیرباز مورد توجه پژوهشگران بوده است. این مدل‌های کامپیوتری مبتنی بر داده‌های سهل‌الوصول، امکان بررسی راهکارهای مختلف مدیریتی را فراهم

آورده‌اند. در زمینه پیش‌بینی عملکرد گیاهی تحقیقاتی از قبیل پیش‌بینی عملکرد زعفران با استفاده از داده‌های هواشناسی به وسیله شبکه عصبی مصنوعی در استان‌های خراسان رضوی و جنوبی صورت گرفته به این منظور ابتدا با بهره‌گیری از نرم‌افزار *Wingamma* داده‌ها و پارامترهای موجود مورد تجزیه و تحلیل قرار گرفت و بهترین ترکیب‌های ورودی به مدل تعیین گردید. کارایی مدل شبکه عصبی پرسپترون<sup>۱</sup> چند لایه (*MLP*)، برای پیش‌بینی عملکرد محصول مورد ارزیابی قرار گرفت. مدل *MLP* با استفاده از شاخص‌های آماری مقایسه شد. که در مدل *MLP* شبکه عصبی هنگامی که از داده‌های حداکثر دما، بارندگی، تبخیر و تعرق و رطوبت نسبی فصل پاییز و عملکرد سال قبل، به عنوان متغیرهای مستقل در پیش‌بینی عملکرد محصول استفاده شد ( $R^2=0.8832$  و  $RMSE = 0.689 \text{ kg.ha}^{-1}$  و  $MAE = 0.560 \text{ kg.ha}^{-1}$ ). بیشترین کارایی بدست آمد (Nekouei et al., 2014). در تحقیقی دیگر با استفاده از مدل شبکه عصبی مصنوعی به پیش‌بینی عملکرد زعفران در منطقه خراسان جنوبی پرداخته شد. نتایج تحلیل عدم قطعیت مونت کارلو بر مبنای ۱۰۰۰ نمونه‌گیری بدون جایگذاری، بر اساس فاکتورهای  $R^2$ ،  $95\%PPU$  و  $d$ -factor بیانگر پهنای باند اطمینان مطلوب پیش‌بینی‌ها بود و الگوی عمومی و کلی تغییر عملکرد زعفران را به خوبی پیش‌بینی نمود. متوسط ضریب  $R^2$  مدل در مرحله آموزش و آزمایش بر اساس ۱۰۰۰ شبیه‌سازی مونت کارلو به ترتیب ۰/۹۲ و ۰/۵۸ بود که برای مدل‌های عملکرد گیاهی مبتنی بر داده‌های اقلیمی دارای معنی‌داری در سطح احتمال یک درصد است (Riahi Modavar et al., 2017). در تحقیقی دیگر نوروزی و همکاران (Norouzi et al., 2010) با پیش‌بینی مؤلفه‌های عملکرد گندم دیم (*Triticum aestivum L.*) با استفاده از روش شبکه عصبی مصنوعی پرداختند. تحلیل حساسیت مدل نشان‌دهنده بیشترین حساسیت عملکرد محصول به عامل بارندگی سپس دما و در نهایت، رطوبت می‌باشد (Akbarpour et al., 2013). در تحقیقی دیگر، نتایج ارزیابی مدل شبکه عصبی مصنوعی در پیش‌بینی عملکرد زعفران با استفاده از شاخص خشکسالی استاندارد در

همبستگی بالای ۰/۹۵ در این تحقیق نتایج خوبی را ارائه داد (Alkroosh et al., 2013). در تحقیقی دیگر، برنامه‌ریزی بیان ژن برای پیش‌بینی عمق آبشویی پایین دست سیل‌ها بکار برده شد به بیان دیگر این روش (GEP) برای استخراج یک مدل جدید در پیش‌بینی عمق آبشویی مورد استفاده قرار گرفت که بر اساس نتایج، معادله مطرح شده، یک پایه‌ریزی مفید برای تخمین عمق آبشویی در رودخانه‌های کوهستانی مختلف با شیب بستر متفاوت است در نتیجه این تحقیق مقایسه‌ای نشان می‌دهد که مدل GEP از دیگر مدل‌های موجود بهتر است (Azamathulla et al., 2012). ابارشی و همکاران (Abarashi et al., 2014) سه روش بیان ژن، شبکه عصبی مصنوعی، سامانه استنتاج فازی-عصبی را در پیش‌بینی تراز سطح ایستابی در حوضه زرين گل در استان گلستان مقایسه کردند که در آن با توجه به نتایج روش بیان ژن نسبت به بقیه روش‌ها دارای ارجحیت بود. کیسی و همکاران، برای مدل‌سازی بارش رواناب مدل‌های GEP و ANN و همچنین ANFIS را باهم مقایسه کردند و نشان داد که GEP برای مدل‌سازی بارش رواناب مناسب است (Kisi et al., 2013). موزامیل و همکاران (Muzzammil et al., 2015) با بررسی کاربرد برنامه‌ریزی بیان ژن در آنالیز فرکانس جریان و مقایسه آن با شبکه عصبی مصنوعی و توزیع گامبل دریافتند که برنامه‌ریزی بیان ژن نسبت به شبکه عصبی مصنوعی و توزیع گامبل مدلی برتر می‌باشد.

بررسی تحقیقات نشان می‌دهد که فاکتورهای اقلیمی از عوامل تأثیرگذار بر عملکرد زعفران محسوب می‌شود که به دلیل پیچیدگی تأثیرگذاری آنها بر فرآیند رشد گیاهی پیش‌بینی اثرات آن دشوار است. همچنین محصول زعفران به عنوان با ارزش‌ترین محصول کشاورزی و دارویی جهان، جایگاه ویژه‌ای در بین محصولات صنعتی و صادراتی ایران در تولید و اشتغال دارد. به همین دلیل هدف از این تحقیق پیش‌بینی عملکرد زعفران با استفاده از مدل شبیه‌ساز ماشین‌بردار پشتیبان (LSSVM) و برنامه‌ریزی بیان ژن (GEP) و انتخاب بهترین مدل در این برآورد می‌باشد. بررسی‌ها نشان داد که تاکنون تحقیقی در خصوص برآورد عملکرد زعفران با این مدل‌ها

استان‌های خراسان رضوی و جنوبی موید کارایی مدل چند لایه پرسپترون (MLP) شبکه عصبی، برای پیش‌بینی عملکرد محصول می‌باشد. در این مدل از داده‌های ماهانه SPI و میانگین ماهانه دما به عنوان متغیرهای مستقل در پیش‌بینی عملکرد زعفران استفاده شد (Aghamohammadi et al., 2017).

در زمینه شبیه‌سازی با استفاده از مدل LSSVM می‌توان به مواردی از قبیل ارزیابی مدل حداقل مربعات ماشین-بردار پشتیبان در پیش‌بینی سطح ایستابی اشاره کرد. نتایج نشان‌دهنده عملکرد دقیق مدل (LSSVM-RBF) با پارامترهای مربوط به تابع  $RBF$  ( $\sigma=4.99$ ) و شاخص‌های عملکرد  $RMSE=0.3401$  و  $R^2=0.999$  نسبت به مدل ANN می‌باشد و نیز بیانگر پیش‌پردازش بهینه آزمون گاما بود (Rezai et al., 2015). رضائی و همکاران (Rezai et al., 2014) شبکه پایش سطح آب‌های زیرزمینی را با استفاده از مدل حداقل مربعات ماشین‌بردار پشتیبان طراحی کردند. نتایج حاصل از این تحقیق نشان داد که ترکیب برتر مدل LSSVM در برگزیده شاخص‌های عملکرد  $MAE=0.3405$  و  $R^2=0.9992$  است. سپس، با استفاده از تابع تقریب بهینه، ۴۲ چاه مشاهداتی به منظور پایش مکانی مناسب در منطقه دشت رامهرمز مشخص شد. نتایج تحقیقی دیگر روی پیش‌بینی برآورد تبخیر با استفاده از مدل LSSVM و مقایسه آن با مدل‌های تجربی نشان داد که عملکرد بالای مدل  $LSSVM-poly$  (چندجمله‌ای)، در ترکیب با متغیرهای دمای کمینه، میانگین دما، سرعت باد و ساعات آفتابی با شاخص‌های عملکرد  $R^2=0.8915$  و  $RMSE=0.995$  نسبت به توابع کرنل و دیگر روش‌های تجربی می‌باشد (Kavusi et al., 2017). تحقیقاتی که در زمینه برنامه‌ریزی بیان ژن (GEP) صورت گرفته عبارتند از تحقیقی که در زمینه عملکرد روش بیان ژن در روندیابی سیلاب رودخانه زنگمار در مقایسه با روش موج دینامیکی انجام گرفت که به طور میانگین روش بیان ژن از نتایج بهتری نسبت به روش موج دینامیکی برخوردار بود (Ghobadian et al., 2013). در تحقیقی دیگر ظرفیت جانبی شمع‌ها در خاک رس را با استفاده از مدل GEP پیش‌بینی شد که این مدل با ضریب

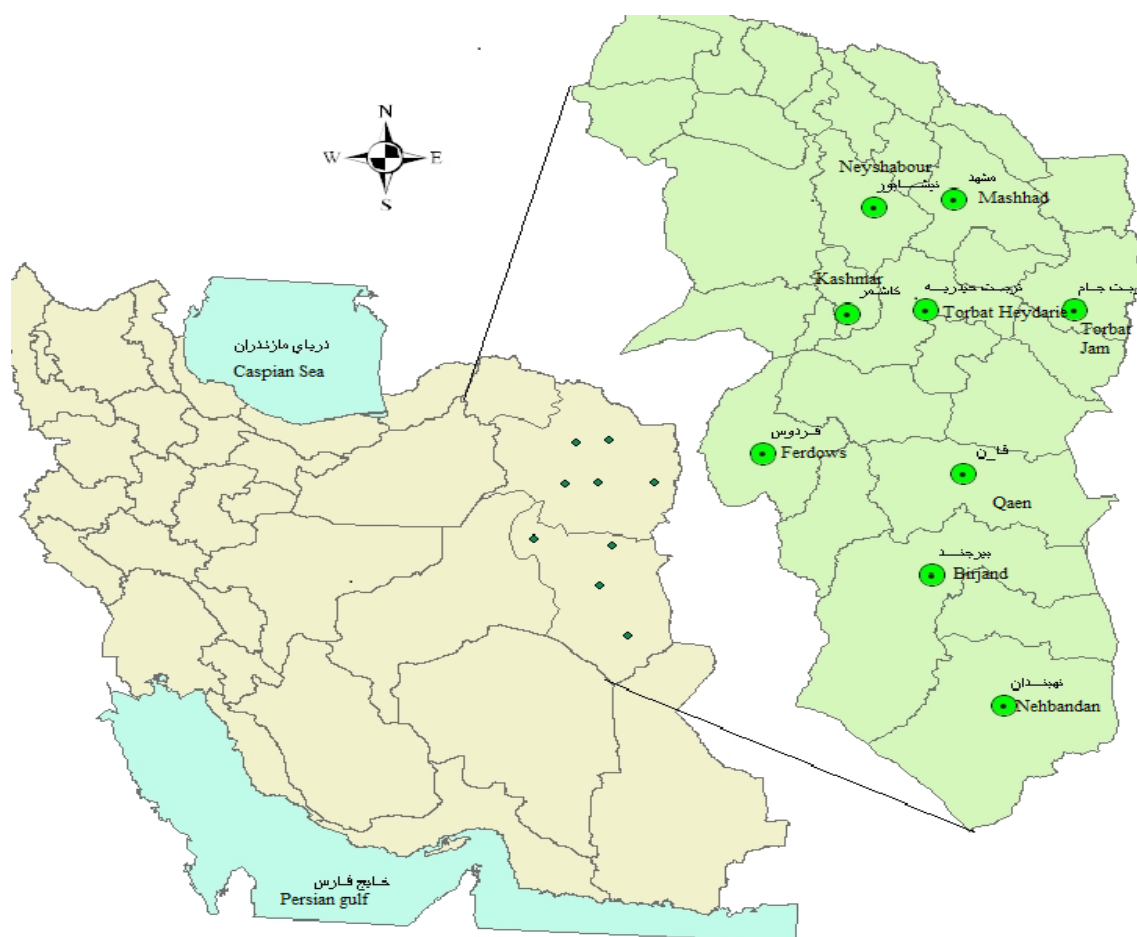
زعفران در ایران انجام شد. شکل ۱ موقعیت شهرهای مختلف مورد مطالعه در دو استان را نشان می‌دهد. داده‌های ۱۶ ساله در این مطالعه از ایستگاه‌های هواشناسی و سازمان جهاد کشاورزی استان خراسان اخذ گردید. داده‌های هواشناسی مورد استفاده در این تحقیق شامل حداقل دمای روزانه، حداکثر دمای روزانه، رطوبت نسبی روزانه، بارندگی، تبخیر و عملکرد یکسال قبل بودند که به عنوان ورودی مدل و اطلاعات عملکرد محصول نیز که از سازمان‌های جهاد کشاورزی و آمارنامه‌های وزارت جهاد کشاورزی تهیه شد و به عنوان خروجی مدل مورد استفاده قرار گرفتند (جدول ۱).

صورت نگرفته است. همچنین مدل بیان ژن قابلیت‌های بالایی از قبیل دادن معادله نهایی غیرخطی بین متغیرهای مستقل و وابسته با ضریب تبیین بالا را داراست. در صورتی که مدل *LSSVM* این قابلیت را ندارد.

## مواد و روش‌ها

### معرفی منطقه مورد مطالعه

این تحقیق بر اساس داده‌های اقلیمی مربوط به استان‌های خراسان رضوی و جنوبی به عنوان مهمترین نواحی کشت



شکل ۱. موقعیت شهرهای مختلف منطقه مورد مطالعه (Nekouei et al., 2014)

Fig. 1. Location of different cities of the study area (Nekouei et al., 2014)

$$\min w, e, b, j(w, e) = \frac{1}{2}w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (2)$$

با محدودیت،  
(۳)

$y_i = w^T \sigma(x_i) + b + x_i$   
که در آن‌ها،  $Y$ : پارامتر تنظیم‌کننده بخش خطا و  $e_i$ : خطای داده‌های آموزشی می‌باشد.  
در بخش مدل‌سازی ۷۰ و ۳۰ درصد داده‌ها به ترتیب به داده‌های آموزشی و داده‌های بخش آزمون تعلق گرفت. در بخش اعتبارسنجی از تکنیک اعتبار سنجی تقاطعی با روش بهینه‌سازی سیمپلکس استفاده گردید. برای تعیین پارامتر تنظیم‌کننده  $\sigma$  و تعیین  $\sigma < 0$  ابتدا باید بهترین تابع کرنل به وسیله سعی و خطا تعیین شود. در این تحقیق از سه تابع کرنل مختلف استفاده شده است. جدول ۲ نام و فرمول توابع کرنل را نشان می‌دهد.

### مدل حداقل مربعات ماشین بردار پشتیبان

اساس روش ماشین بردار پشتیبان بر پایه تئوری یادگیری آماری ارائه گردید که از حداقل‌سازی ریسک پیروی می‌کند. سانچز و همکاران (Sanchez et al., 2011) روش حداقل مربعات ماشین بردار پشتیبان را ارائه نمودند. این مدل برخلاف مدل SVM به جای مسأله برنامه‌نویسی درجه دوم در حل از معادلات خطی استفاده می‌نماید. بنابراین، دقت محاسباتی بالاتری نسبت به ماشین بردار پشتیبان کلاسیک دارد (Seifi et al., 2012). مدل رگرسیون LSSVM از معادله زیر پیروی می‌کند:

$$y(x_i) = w^T * Q(x_i) + b \quad (1)$$

که در آن،  $w$  و  $b$ : به ترتیب مقادیر وزن‌ها و اریبی تابع در معادله ۲ تعیین گردید:

جدول ۱. میانگین پارامترهای هواشناسی ایستگاه‌های سینوپتیک استان‌های خراسان رضوی و جنوبی طی سال‌های ۱۹۹۲-۲۰۰۶

Tables 1. Meteorological parameters of the synoptic stations of Razavi Khorasan and South Khorasan provinces during 1992-2006

مجموع میانگین حداکثر دما (بهار)	مجموع میانگین حداکثر دما (تابستان)	مجموع میانگین حداکثر دما (پاییز)	مجموع میانگین حداکثر دما (زمستان)	مجموع میانگین حداکثر دما (بهار)	مجموع میانگین حداکثر دما (تابستان)	مجموع میانگین حداکثر دما (پاییز)	مجموع میانگین حداکثر دما (زمستان)
Total mean maximum temperature (spring)	Total mean maximum temperature (summer)	Total mean maximum temperature (autumn)	Total mean maximum temperature (winter)	Total mean maximum temperature (spring)	Total mean maximum temperature (summer)	Total mean maximum temperature (autumn)	Total mean maximum temperature (winter)
1054.55	1537.11	445.15	40.93	2175.82	2739.9	1558.43	932.56
مجموع میانگین بارندگی (بهار)	مجموع میانگین بارندگی (تابستان)	مجموع میانگین بارندگی (پاییز)	مجموع میانگین بارندگی (زمستان)	مجموع میانگین بارندگی (بهار)	مجموع میانگین بارندگی (تابستان)	مجموع میانگین بارندگی (پاییز)	مجموع میانگین بارندگی (زمستان)
Mean precipitation (spring)	Mean annual precipitation (summer)	Mean annual precipitation (autumn)	Mean annual precipitation (winter)	Mean annual precipitation (autumn)	Mean annual precipitation (winter)	Total mean evaporation (spring)	Total mean evaporation (summer)
46.85	3.14	31.37	93.72	31.37	93.72	428.84	571.48
مجموع میانگین تبخیر (پاییز)	مجموع میانگین تبخیر (زمستان)	مجموع میانگین تبخیر (بهار)	مجموع میانگین تبخیر (تابستان)	مجموع میانگین تبخیر (بهار)	مجموع میانگین تبخیر (تابستان)	مجموع میانگین تبخیر (پاییز)	مجموع میانگین تبخیر (زمستان)
Total mean evaporation (autumn)	Total mean relative humidity (summer)	Total mean relative humidity (spring)	Total mean relative humidity (winter)	Total mean relative humidity (spring)	Total mean relative humidity (summer)	Total mean relative humidity (autumn)	Total mean relative humidity (winter)
220.73	156.85	3135.27	2136.03	3135.27	156.85	3437.10	4443.19

جدول ۲. نام و فرمول توابع کرنل

Table 2. The name and formula of the Kernel functions

فرمول توابع Formula functions	نام توابع Functions name
$K(x_i, x_j) = x_i^t x_j$	خطی Linear
$K(x_i, x_j) = (x_i^t, x_j + \tau)^d$	درجه دو Polynomial
$K(x_i, x_j) = \exp(- x_i - x_j ^2 / 2\sigma^2)$	پایه شعاعی Radial basis

بردارهای پشتیبان میزان عملکرد زعفران براساس مقادیر مختلف خطا تولید گردید. در این تحقیق پارامترهای بهینه تأثیرگذار بر عملکرد طبق جدول ۳ از طریق آزمون گاما به دست آمد (Nekouei et al., 2014).

در مرحله طراحی با استفاده از مدل LSSVM، میزان عملکرد با توجه به پارامترهای بهینه تقریب زده شد. سرانجام بنابر روند بهینه‌سازی درجه دوم، بردارهای پشتیبان استخراج شدند. در نتیجه بر اساس این

جدول ۳. نتایج آزمون گاما برای تعیین مدل‌های بهینه جهت پیش‌بینی عملکرد (Nekouei et al., 2014)

Table 3. Gamma test results for determining optimal models for predicting yield (Nekouei et al., 2014)

پارامترها Parameters	آماره گاما Gamma parameter	شماره ترکیب Number of combination
$ET_{SP}, ET_{su}, ET_{au}, ET_{wi}, RH_{sp}, RH_{su}, RH_{au}, RH_{wi}, y$	0.008764	A
$T_{min\ wi}, T_{max\ wi}, P_{sp}, ET_{sp}, RH_{sp}, y$	0.00967	B
$T_{max\ sp}, T_{max\ su}, T_{max\ au}, T_{max\ wi}, y$	0.010949	C
$P_{au}, P_{wi}, RH_{au}, RH_{wi}, T_{min\ au}, T_{max\ au}, y$	0.011454	D
$T_{min\ au}, T_{max\ au}, P_{au}, P_{wi}, RH_{au}, RH_{wi}, y$	0.012785	E
$P_{sp}, P_{su}, T_{min\ au}, T_{max\ au}, y$	0.013507	F
$Y, ET_{au}, RH_{au}, P_{au}, T_{max\ au},$	0.013751	G
$T_{min\ wi}, P_{wi}, P_{au}, ET_{wi}, y$	0.01449	H
$ET_{SP}, ET_{su}, ET_{au}, ET_{wi}, P_{sp}, P_{su}, P_{au}, P_{wi}, y$	0.015436	I
$RH_{au}, ET_{au}, P_{au}, T_{min\ au}$	0.018415	J

بین روش‌های الگوریتم گردش می‌باشد که به دلیل دارا بودن دقت کافی، به عنوان مرسوم‌ترین شیوه بوده و از کاربرد بیشتری برخوردار است. زمینه اصلی برنامه‌ریزی بیان ژن، همان الگوریتم ژنتیک است، با این تفاوت که در این روش از شاخه‌های مجزا به جای نوارهای بیت استفاده شده است. هر شاخه نیز از یک مجموعه پایانه‌ها (متغیرهای مسأله) و مجموعه توابع (عملگرهای اصلی) تشکیل می‌شود (Borelli et al., 2006). فرآیند گام به گام برنامه‌ریزی بیان ژن به صورت مراحل زیر است:

### برنامه‌ریزی بیان ژن

این روش جزو روش‌های الگوریتم گردش محسوب می‌شود که مبنای تمامی آنها بر اساس نظریه تکامل داروین استوار است. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای اندازه‌گیری و مقایسه روش‌های مختلف حل در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، روش حل مناسب را ارائه می‌نمایند. برنامه‌ریزی بیان ژن جدیدترین شیوه از

انتخاب تابع برازش می‌تواند بر اساس روابط متعددی صورت گیرد و اگر این تابع به درستی انتخاب شود مدل از برازش بهتری برخوردار است که در این پژوهش، بعد از سعی و خطا کردن و انتخاب توابع مختلف تابع متوسط قدر مطلق خطا ( $MAE$ ) به عنوان بهترین تابع برازش انتخاب گردید. در جدول ۴ مقادیر مطلوب پارامترهای مورد استفاده در نرم‌افزار  $GEP$  که با آزمایش و سعی و خطا، مقدار مطلوب هر کدام حاصل گردید، نشان داده شده است.

#### ارزیابی عملکرد مدل

در پایان جهت ارزیابی و آزمون اعتبار مدل  $LSSVM$  و برنامه‌ریزی بیان ژن، به ارزیابی عملکرد آنها پرداخته شد. برای ارزیابی عملکرد این مدل‌ها از آماره‌های ریشه متوسط مربعات خطا ( $RMSE$ )، متوسط قدر مطلق خطا ( $MAE$ ) و ضریب تبیین ( $R^2$ ) مطابق معادله ۴ تا ۶ استفاده گردید.

(۴)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{N}}$$

(۵)

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{N}$$

(۶)

$$R2 = \frac{(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)(\sum_{i=1}^n (y_i - \bar{y})^2)}$$

#### نتایج و بحث

هدف اصلی این تحقیق، انتخاب بهترین مدل برای پیش‌بینی عملکرد زعفران در گستره استان‌های خراسان رضوی و جنوبی با استفاده از مدل  $LSSVM$  و برنامه‌ریزی بیان ژن می‌باشد. در این تحقیق، ۱۰ سناریو بر اساس بهترین ترکیب‌های ورودی به مدل که از طریق آزمون گاما به دست آمده بود، در نظر گرفته شد ( $Nekouei et al., 2014$ ). سپس مدل  $LSSVM$  و برنامه‌ریزی بیان ژن برای این ۱۰ ترکیب اجرا شد که این نتایج در جدول ۵ آمده است.

۱- یک جمعیت اولیه از توابع مرکب نشان‌دهنده مدل‌های پیش‌بینی، به صورت تصادفی در نظر گرفته می‌شود.

۲- هر یک از افراد جمعیت مذکور با استفاده از توابع برازش، مورد ارزیابی قرار می‌گیرد.

۳- در هر تولید، مراحل زیر برای انتخاب یک جمعیت جدید دنبال می‌شود:

الف- یکی از عملگرهای جهش، عبور و کپی انتخاب می‌شود.

ب- تعداد مناسبی از افراد جمعیت حاضر انتخاب می‌شود.

ج- از عملگر انتخابی برای تولید فرزند استفاده می‌شود.

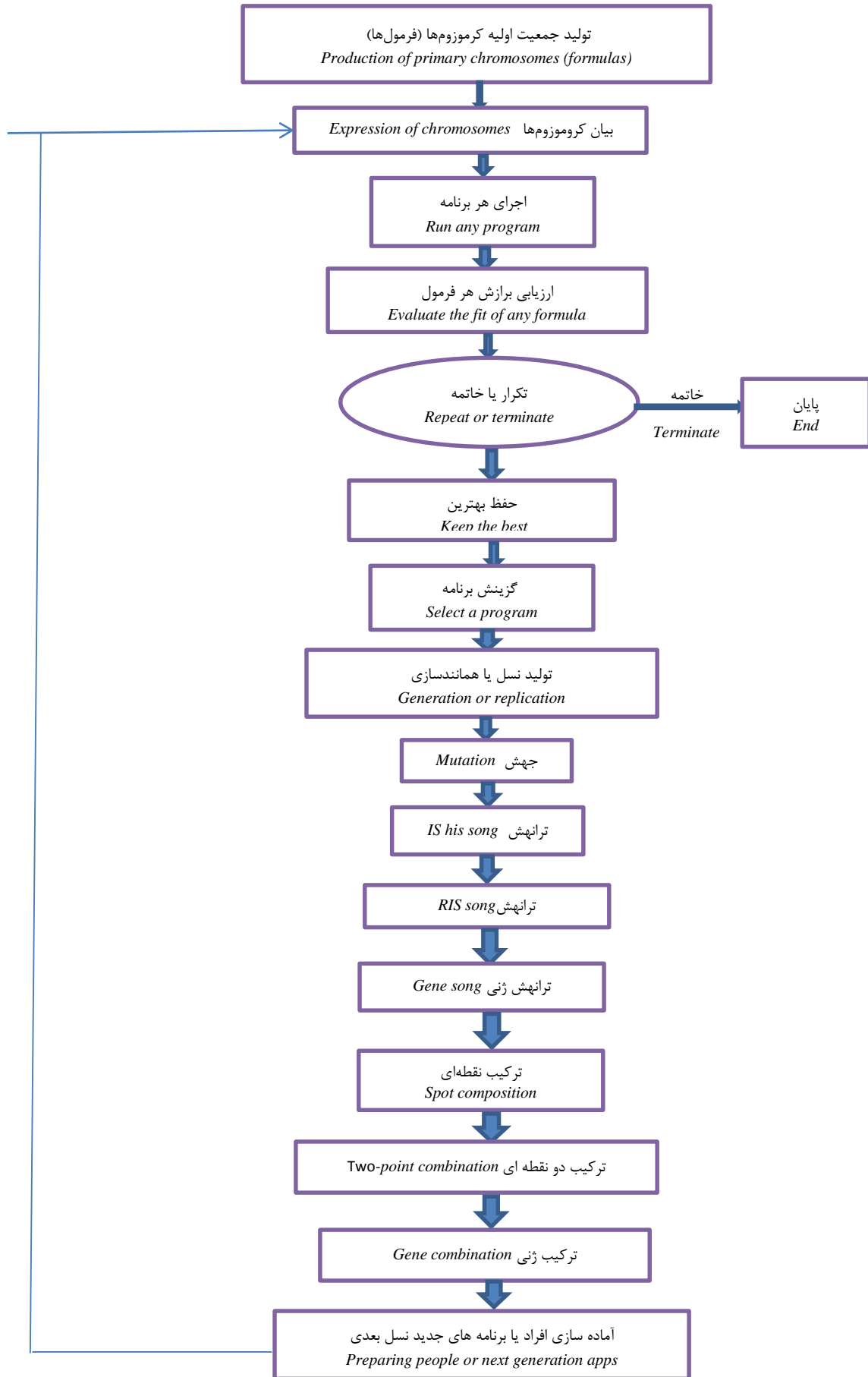
د- فرزند یاد شده در یک جمعیت جدید وارد می‌شود.

ه- مدل مورد نظر با استفاده از برازش‌های مختلف مورد ارزیابی واقع می‌شود.

۴- گام سوم تا نیل به حداکثر تعداد تولید، تکرار خواهد شد. در این روش در ابتدای فرآیند، هیچگونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مؤلفه‌های آن می‌باشد.

اولین مرحله در تخمین عملکرد به کمک این نرم‌افزار، تولید جمعیت اولیه از راه‌حل‌هاست که به کمک توابع و ترمینال‌ها صورت می‌گیرد. سپس کروموزم‌ها به صورت بیان درختی نشان داده می‌شوند.

در مرحله بعد باید کارایی یا سازگاری هر عضو جمعیت کروموزوم‌ها را به کمک تابع برازش ارزیابی کرد که حالت آرمانی این تابع، ۱۰۰۰ می‌باشد. بدین ترتیب، به کمک تابع برازش و حفظ جواب‌های برتر و حذف جواب‌های ضعیف به جواب مطلوب نزدیک می‌شویم. از این رو رابطه بین متغیرهای وابسته و مستقل ثابت نیست و پیوسته تغییر می‌کند. به منظور مقایسه نتایج حاصل از الگوریتم بیان ژن در پیش‌بینی تابش خورشیدی، ۱۰ سناریو با توجه به پارامترهای تأثیرگذار بر تابش خورشیدی مانند بارش، رطوبت نسبی، حداکثر دما، حداقل دما و تبخیر تعریف شد. در جدول ۲، پارامترهای ورودی در هر سناریو ارائه گردید. برای بهره‌گیری از الگوریتم بیان ژن از نرم‌افزار  $GeneXproTools 5.0$  استفاده شده است. فلوجارت مراحل الگوریتم بیان ژن در شکل ۲ نمایش داده شده است.



شکل ۲. فلوجارت الگوریتم بیان ژن

Fig. 2. Flowchart of gene expression algorithm



جدول ۴. پارامترهای مورد استفاده در GEP

Table 4. Parameters used in GEP

پارامتر Parameter	مقدار Value
اندازه سر <i>Head size</i>	10
اندازه دنباله <i>Tail size</i>	9
تعداد کروموزوم‌ها <i>Number of chromosomes</i>	30
تعداد ژن‌ها در هر کروموزوم <i>Number of genes</i>	3
نرخ جهش <i>Mutation rate</i>	0.05
نرخ وارون‌سازی <i>Inversion rate</i>	0.05
نرخ ترکیب تک نقطه‌ای <i>One-point recombination rate</i>	0.3
نرخ ترکیب دو نقطه‌ای <i>Two-point recombination rate</i>	0.2
نرخ ترکیب ژن <i>Gene recombination rate</i>	0.2
نرخ ترانزیشن درجه متوالی <i>IS transposition rate</i>	0.1
نرخ ترانزیشن ریشه درجه متوالی <i>RIS transposition rate</i>	0.1
نرخ ترانزیشن ژن <i>Gene transposition rate</i>	0.1
معیار خطای تابع <i>Fitness function error type</i>	MAE
تابع پیوند <i>Linking function</i>	تابع جمع "+" Sum function "+"
مجموعه توابع <i>Function set</i>	+, *, /, -, ln, sin, cos, arctan, x <sup>2</sup> , √
تعداد اجرا <i>Number of runs</i>	20
تعداد جمعیت یا نسل <i>Number of generation</i>	3000
نرخ عملگر جهش در ثوابت تصادفی <i>R.C. mutation rate</i>	0.1
نرخ عملگر ترانزیشن در دامنه DC <i>DC specific IS transposition rate</i>	0.05

## جدول ۵. نتایج برنامه GEP برای سناریوهای مختلف

Table 5. GEP Program results for different scenarios

مدل Model	نام ترکیب Compound name	R <sup>2</sup>	RMSE (kg/ha)	MAE (kg/ha)
GEP	A	0.44018	0.87754	0.53309
GEP	B	0.44258	0.67481	0.52644
GEP	C	0.49457	0.51229	0.50939
GEP	D	0.44137	0.69779	0.53079
GEP	E	0.45731	0.65511	0.51999
GEP	F	0.46897	0.54337	0.51821
GEP	G	0.44795	0.63120	0.51606
GEP	H	0.60688	0.43265	0.46432
GEP	I	0.56964	0.50417	0.49221
GEP	J	0.47597	0.55456	0.51030
LSSVM-LIN	A	0.29185	1.19210	0.87469
LSSVM-RBF	A	0.20155	1.40152	1.0230
LSSVM-POLY	A	0.11711	1.36783	1.00692
LSSVM-LIN	B	0.15522	1.22982	0.86508
LSSVM-RBF	B	0.05217	1.25955	0.94834
LSSVM-POLY	B	0.12679	1.45897	1.04836
LSSVM-LIN	C	0.21437	1.774956	1.45425
LSSVM-RBF	C	0.0455699	1.85943	1.53190
LSSVM-POLY	C	0.04817	1.846447	1.52212
LSSVM-LIN	D	0.151352	1.44637	1.10862
LSSVM-RBF	D	0.219426	1.49516	1.15699
LSSVM-POLY	D	0.221307	1.47379	1.11513
LSSVM-LIN	E	0.24952	1.34142	1.08399
LSSVM-RBF	E	0.22834	1.49958	1.14749
LSSVM-POLY	E	0.21810	1.48194	1.11893
LSSVM-LIN	F	0.07972	1.62685	1.21100
LSSVM-RBF	F	0.07233	1.58793	1.17837
LSSVM-POLY	F	0.00937	2.79159	2.06644
LSSVM-LIN	G	0.13571	1.60762	0.75787
LSSVM-RBF	G	0.25603	1.33543	0.95370
LSSVM-POLY	G	0.21070	1.19288	0.76327
LSSVM-LIN	H	0.45686	0.58784	0.86743
LSSVM-RBF	H	0.05996	1.01827	1.05865
LSSVM-POLY	H	0.32336	1.17583	1.01972
LSSVM-LIN	I	0.02154	1.26208	0.90128
LSSVM-RBF	I	0.08681	1.95901	1.49687
LSSVM-POLY	I	0.03592	1.44612	1.05292
LSSVM-LIN	J	0.12970	1.93232	1.49387
LSSVM-RBF	J	0.07399	1.35979	1.01221
LSSVM-POLY	J	0.19696	1.42117	1.08436

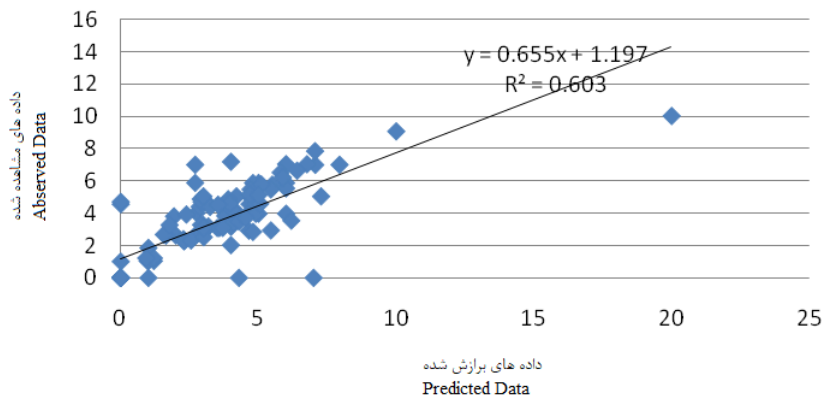
## انتخاب بهترین مدل

عملکرد مدل بیان ژن و مدل LSSVM از طریق ارزیابی آماره‌های ضریب همبستگی ( $R^2$ )، ریشه متوسط خطای مربعات (RMSE) و متوسط قدر مطلق خطا (MAE) که در بهترین حالت (حالت آرمانی مدل) مقادیر  $R^2$ ، RMSE و MAE به ترتیب برابر ۱، صفر و صفر می‌باشند ارزیابی شد. از مقایسه نتایج آماره‌های ارزیابی مدل‌ها در سناریوهای مختلف در جدول ۵ مشاهده می‌شود که در تمام سناریوها، سناریوی H به دلیل در نظر گرفتن پارامترهای میانگین حداقل دما در فصل زمستان، میانگین بارندگی در فصل زمستان، میانگین بارندگی در فصل پائیز، میانگین تبخیر در فصل زمستان پیش‌بینی

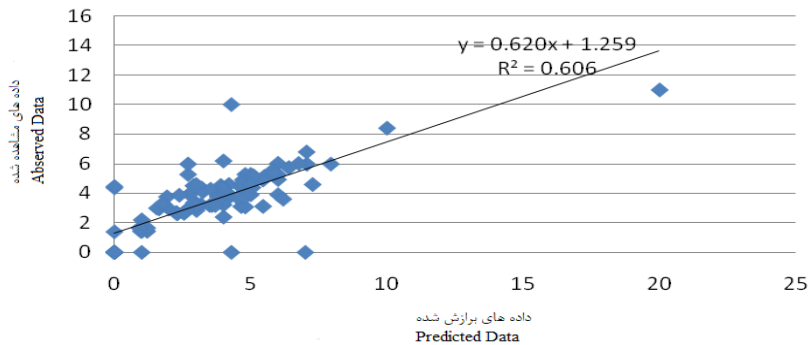
مطلوب‌تری را ارائه کرده است. در این سناریو مقادیر  $R^2$ ، RMSE، MAE به ترتیب ۰/۶۰۶۸۸، ۰/۴۳۲۶۵ و ۰/۴۶۴۳۲ کیلوگرم در هکتار به دست آمد. همچنین سناریوی A دارای ضعیف‌ترین نتیجه می‌باشد (به دلیل در نظر نگرفتن پارامترهای حداقل دما، بارندگی و تبخیر). از این‌رو، می‌توان این چنین برداشت کرد که سناریوی H با توجه به بالا بودن ضریب همبستگی، بهترین برآزش و کم بودن مقادیر خطا، مدل مطلوب می‌باشد. همچنین در مدل LSSVM تابع Linear در بین توابع دارای بهترین نتایج و ترکیب H با خطای کمتر و ضریب همبستگی بیشتر نتیجه بهتری را دارا بود. اما بین مدل GEP و مدل LSSVM، برنامه‌ریزی بیان ژن دارای نتایج مطلوبتر و

بهتری در پیش‌بینی عملکرد زعفران بود و از دقت بالاتری نسبت به مدل *LSSVM* برخوردار بود. در مقایسه با نتایج تحقیقات دیگر سیدو و آیتک ( *Seydou & Aytac, 2012* ) با بررسی توانایی عملکرد برنامه‌ریزی بیان ژن برای مدل‌سازی تبخیر پتانسیل با استفاده از داده‌های ده ساله کشور ساحلی، بورکینافاسو نشان دادند که روش *GEP* توانایی خوبی برای ارائه مدل بر اساس داده‌های منطقه‌ای را دارد. در تحقیقی دیگر، روی پیش‌بینی تبخیر تعرق مرجع روزانه با استفاده از اطلاعات آب و هوا و برنامه‌نویسی بیان ژن در چین نشان داده شد که بیان ژن توانایی خوبی برای پیش‌بینی تبخیر تعرق روزانه با کمترین خطا (  $MAE=0.902$  ,  $RMSE=1.207$  ) و بیشترین همبستگی (  $R=0.77$  ) داشته است ( *Nekouei et al., Traore 2017* ). نکویی و همکاران ( *Nekouei et al.,* )

به پیش‌بینی عملکرد زعفران با استفاده از داده-های هواشناسی به وسیله شبکه عصبی مصنوعی نشان دادند که در مدل MLP شبکه عصبی هنگامی که از داده‌های حداکثر دما، بارندگی، تبخیر و تعرق و رطوبت نسبی فصل پاییز و عملکرد سال قبل، به عنوان متغیرهای مستقل در پیش‌بینی عملکرد محصول استفاده شد (  $MAE=0560 kg.ha^{-1}$  و  $R^2=0.8832$  ) و بیشترین کارایی بدست آمد. شکل ۳ و ۴، عملکرد مقادیر اندازه‌گیری شده و پیش‌بینی شده را نشان می‌دهد که پیش‌بینی عملکرد زعفران برای مدل مطلوب *GEP* (سناریوی *H*) در مطالعه حاضر است. از این شکل‌ها می‌توان دید که مقادیر پیش‌بینی *GEP* بسیار نزدیک به مقادیر اندازه‌گیری عملکرد زعفران برای تمام مجموعه داده‌هاست.



شکل ۳. نتایج برنامه برای مدل مطلوب در بخش آموزش  
**Fig. 3. Program results for the optimal model in the learning section**



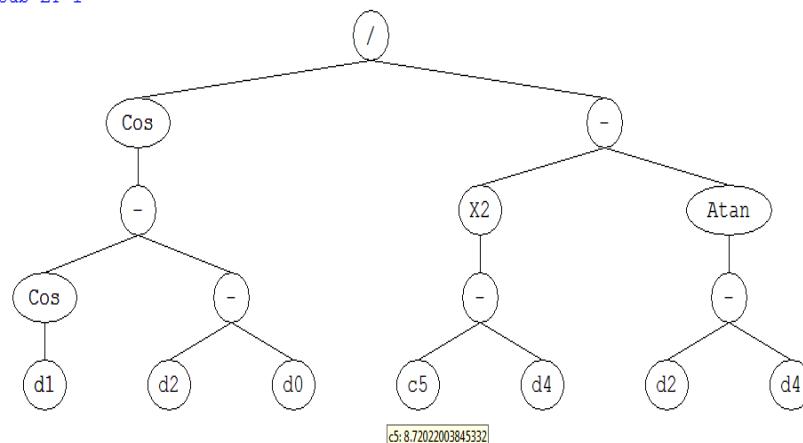
شکل ۴. نتایج برنامه برای مدل مطلوب در بخش آزمایش  
**Fig. 4. Program results for the optimal model in the testing section**

ساختار درختی مدل مطلوب

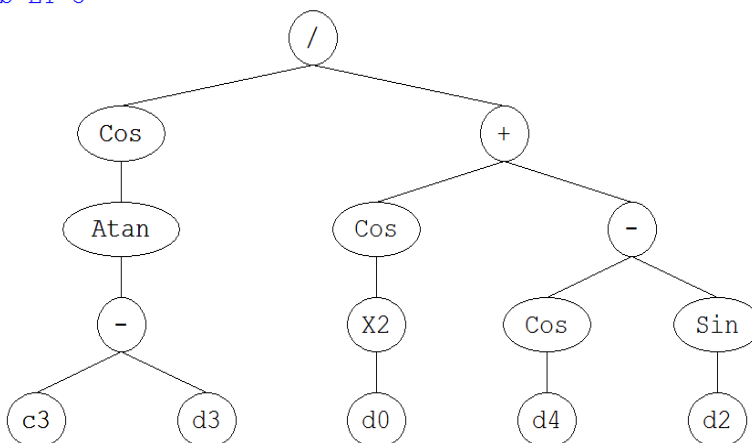
حاصل شده است. از آنجا که سه ژن در اینجا تشکیل شده، هریک از ژن‌ها یک زیردرخت و معادله مربوط به خود را دارد که در نهایت، با توجه به تابع پیوند معادله نهایی حاصل می‌شود.

شکل ۵ ساختار مدل خروجی مطلوب را به صورت درختی نشان می‌دهد و همچنین معادله حاصل از این ساختار

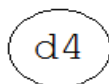
Sub-ET 1



Sub-ET 3



Sub-ET 2



شکل ۵. ساختار درختی مدل خروجی

Fig. 5. The tree structure of the output model

$$y = (\cos((\cos(d[1])-(d[2]-d[0])))/(\text{pow}((G_1C_5-d[4]),2)-\text{atan}((d[2]-d[4]))));$$

(۸)

$$y = d[4];$$

(۹)

$$y = (\cos(\text{atan}((G_3C_3-d[3])))/(\cos(\text{pow}(d[0],2))+(\cos(d[4])-\sin(d[2]))));$$

$$G_1C_5 = 8.72022003845332;$$

$$G_3C_3 = -6.58319955054476;$$

$$\text{Result} = y;$$

شکل ۵ زیردرخت مربوط به ژن اول (معادله ۷)، دوم (معادله ۸)، سوم (معادله ۹) را نشان می‌دهد که در این زیردرخت ثوابت عددی تصادفی  $C_5$  و  $C_3$  و ورودی‌های، اول ( $d_0$ )، دوم ( $d_1$ )، سوم ( $d_2$ )، چهارم ( $d_3$ ) و پنجم ( $d_4$ ) که در سناریوی  $H$  به ترتیب میانگین حداقل دما در فصل زمستان، میانگین بارش در فصل پاییز، میانگین بارش در فصل زمستان، میانگین تبخیر در فصل زمستان، عملکرد یکسال قبل هستند، ایجاد شده که در زیر معادلات مربوط به هر زیر درخت نوشته شده است.

(۷)

## نتیجه‌گیری

دارد. در مدل *LSSVM* ترکیب‌های با تابع کرنل *Liner* دارای نتایج بهتری نسبت به ترکیب‌های با تابع کرنل *RBF* و *poly* می‌باشد. همچنین در این مدل همانند مدل *GEP*، سناریوی *H* با در نظر گرفتن میانگین حداقل دما در فصل زمستان، میانگین بارندگی در فصل پاییز و زمستان و میانگین تبخیر در فصل زمستان نتایج بهتری نسبت به بقیه سناریوها دارد. اما بین مدل *LSSVM* و مدل *GEP*، مدل *GEP* از نتایج بهتری برخوردار بود. همچنین در این مدل معیارهای ارزیابی تحت سناریوی *H* با  $RMSE=0.43265$ ،  $R^2=0.60688$  و  $MAE=0.46432$  در بخش آموزش از سایر سناریوها بهتر می‌باشد. در مجموع، روش *GEP* دقیق‌ترین نتایج را در تخمین عملکرد زعفران دارد.

در این مطالعه، توانایی مدل *LSSVM* برنامه‌ریزی بیان ژن برای پیش‌بینی مقدار عملکرد زعفران در استان‌های خراسان رضوی و جنوبی مورد ارزیابی قرار گرفت. با اعمال مدل *LSSVM* و برنامه‌ریزی بیان ژن بر روی پارامترهای تأثیرگذار در نهایت بعد از تعریف ۱۰ سناریو مقدار عملکرد زعفران با بهترین ترکیب (حداقل دما در فصل زمستان، بارندگی در فصل زمستان، بارندگی در فصل پاییز و تبخیر در فصل زمستان) به دست آمد. بهترین مدل بر اساس معیارهای ارزیابی *RMSE*، *MAE* و  $R^2$  انتخاب شد. از نتایج تجزیه و تحلیل، مشخص شد که مقادیر پیش‌بینی شده سازگاری خوبی با مقادیر اندازه‌گیری شده در مدل برنامه‌ریزی بیان ژن (*GEP*)

## منابع

- Abarashi, M., Moftah Holghi, H., Sani Khani, A., and Dehghani, A., 2014. Comparison of performance of three intelligent methods in order to predict static level fluctuations (Case study: Zarin Gol plain). *Water and Soil Conservation Research*. 21(1), 180-163.
- Adams, R.M., Aka, C., Mc Carl, B.A., and Schimmelpfennig, D., 2001. Climate variability and climate change: Implications for agriculture. *Advances in the Economics of Environmental Resources*. 3, 115-148.
- Aghamohammadi, S., Khashei-Siuki, A., Shahidi, A., and Hashemi, R., 2017. Evaluation of artificial neural network model in predicting saffron yield using standardized drought index in South Khorasan and Khorasan Razavi provinces. *The First National Conference on New Opportunities for Agricultural Production and Employment in the East of the Country (in Line with the Objectives of the Resistance Economy)*. [in Persian].
- Akbarpour, A., Khorashadizadeh, O., Shahidi, A., and Ghochanian, E., 2013. Performance evaluation of artificial neural network models in estimate production of yield saffron based on climate parameters. *J. Saffron Res.* 1(1), 27-35. [in Persian with English Summary].
- Akrami, M.R., Malakouti, M.J., and Keshavarz, P., 2014. Study of flower and stigma yield of saffron as affected by potassium and zinc fertilizers in Khorasan Razavi Province. *J. Saffron Res.* 2(1), 85-96. [in Persian with English Summary].
- Alkroosh, I., and Nikraz, H., 2013. Evaluation of pile lateral capacity in clay applying evolutionary approach. *International Journal of GEOMATE*, 14(1 SERL 7), 462-466.
- Azamathulla, H.M., 2012. Gene expression programming for prediction of scour depth downstream of sills. *J. Hydrol.* 460, 156-159.
- Behdani, M.A., Koocheki, A.R., Nassiri Mahallati, M., and Rezvani Moghaddam, P., 2005. Evaluating the relationships between revenue and consumption of nutrients in *Crocus sativus*. *Iran. J. Field Crops Res.* 3(1), 1-14. [in Persian with English Summary].
- Borelli, A., De Falco, I., Della, C.A., Nicodemi, M., and Trautteur, G., 2006. Performance of genetic programming to extract the trend in noisy data series. *Physica A*. 370, 104-108.

- Ghobadian, M., Ghorbani, A., and Khalaj, M., 2013. Investigating the gene expression function in Zangmar. *River Flood Journals Compared to Dynamic Wave. Water and Soil Journal (Agricultural Science & Technology)* 27(3), 592-602. [in Persian with English Summary].
- Kavusi, M., Khashei-Siuki, A., Pourreza Bildeni, M., and Najafi Mod, M., 2017. Evaluation of the least squares of the vector carriers model in estimating evaporation and comparing it with experimental models. *Journal of Environmental Water Engineering*. 3, 235-247.
- Kisi, O., Shiri, J., and Tombul, M., 2013. Modeling rainfall-runoff process using soft computing techniques. *Computers and Geosciences* 51, 108-117.
- Koocheki, A., Karbasi, A.R., and Seyyedi, S.M., 2017. Some reasons for saffron yield loss over the last 30 years period. *Saffron Agron. & Technol.* 5(2), 107-122. [in Persian with English Summary].
- Koozehgaran, S., Mousavi Baygi, M., Sanaeinejad, S.H., and Behdani, M.A., 2011. Study of the minimum, average and maximum temperature in South Khorasan to identify relevant areas for saffron cultivation using GIS. *Journal of Water and Soil*. 25(4), 892-904. [in Persian with English Summary].
- Muzzammil, M., Alam, J., and Danish, M., 2015. Application of gene expression programming in flood frequency analysis. *Journal of Indian water Resources Society*, 35(2), 1-6.75.
- Nekouei, N., Behdani, M.A., and Khashei-Siuki, A., 2014. Predicting saffron yield from meteorological data Using expert system, Razavi and South Khorasan Province, Iran. *J. Saffron Res.* 2(1), 1-19. [in Persian with English Summary].
- Norouzi, M., Ayoubi, S., Jalalian, A., Khademi, Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., and Srinivasan, R., 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *J. Hydrol.* 333(2), 413-430. [in Persian with English Summary].
- Rezai, A., Khashei-Siuki, A., and Shahidi, A., 2015. Ground water level monitoring network design using the least squares support vector machine (LS-SVM). *Iran Soil and Water Research*. 45(4), 389-396. [in Persian with English Summary].
- Rezai, A., Shahidi, A., Khashei-Siuki, A., and Riahi Madvar, H., 2014. Performance evaluation least squares support vector machine model to predict the water table. *Journal of Irrigation and Drainage*. 7(4), 510-520. [in Persian with English Summary].
- Riahi Modavar, H., Khashei-Siuki, A., and Seifi, A., 2017. Accuracy and uncertainty analysis of artificial neural network in predicting saffron yield in the south Khorasan province based on meteorological data. *Saffron Agron. & Technol.* 5(3), 255-271. [in Persian with English Summary].
- Sadeghi, B., 1993. Effect of corm weight on saffron flower collection. *Publication of Scientific Research-Technology Research Center of Khorasan, Iran*. [in Persian].
- Sanchez, A.S., Nieto, P.J.G., Fernandez, P.R., Diaz, J.J.D., and Iglesias-Rodr, F.J., 2011. Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling*. 54, 1453-1466.
- Seifi, A., Myrlofy, S., and Riahi, H., 2012. Assessment and monitoring of weather station network using principal components analysis and factor analysis Case study: Kerman. *Journal of Irrigation and Drainage*, 5(1), 3-0-42. [in Persian with English Summary].
- Seydou, T., and Aytac, G., 2012. Regional-Specific numerical models of evapotranspiration using Gene-Expression Programming interface in Sahel. *Water Resources Management*. 26, 4367-4380.

*Traore, S., Luo, Y., and Fipps, G., 2017.  
Gene-Expression Programming for  
Short-Term Forecasting of daily*

*reference evapotranspiration using  
public weather forecast information.  
J. Water Resources Management.*



Original Article:

## Efficiency of support vector machine models and gene expression programming in predicting saffron yield

Mohadeseh Kavousi<sup>1</sup>, Abbas Khashei Siuki<sup>2\*</sup>, Mohammad Ali Behdani<sup>3</sup>

*1- PhD student of Water Resources, University of Birjand, Iran.*

*2- Professor, Department of Water Engineering, Research Group of Drought and Climate Change, University of Birjand, Iran.*

*3- Professor, Saffron Research Group, Faculty of Agriculture, University of Birjand, Birjand, Iran, Birjand University, Iran.*

*\* Corresponding author Email: [abbaskhashei@birjand.ac.ir](mailto:abbaskhashei@birjand.ac.ir)*

*Received 12 July 2018; Accepted 29 August 2018*

### Abstract

Due to the sensitivity of saffron yield and its effect affected as climatic parameters and nonlinear properties of plant functions, in this research, saffron yield prediction was considered. The purpose of this study was to evaluate the ability of the support vector machine simulation model (LSSVM) and genome genotype programming model (GenXproTools5.0) to predict saffron yield based on meteorological data (such as minimum temperature, maximum temperature, precipitation, evaporation and relative humidity, yield one year ago) on a seasonal scale between 1992 and 2006. The best model was selected based on  $R^2$ , RMSE and MAE assessment criteria. The results showed that in both scenarios, in the H scenario (including mean winter temperature, mean precipitation in autumn, winter rainfall mean, winter evaporation mean, yield one year ago), better results were obtained from saffron yield. In the LSSVM model, combinations with the Liner kernel function had more accurate results. But between LSSVM model and GEP model, GEP model had higher  $R^2$  and lower RMSE and MAE.  $R^2$ , RMSE and MAE ratios in this model under the H-scenario in education section were calculated with 0.60688, 0.43265 and 0.46432, respectively. In general, the GEP model had more accurate results in saffron yield estimates than the LSSVM model.

**Keywords:** Functional prediction, LSSVM model, Gene expression programming model.