



دانشگاه گیلان

نشریه پژوهش‌های تولید گیاهی

جلد بیست و ششم، شماره سوم، ۱۳۹۸

۱۹۹-۲۱۷

<http://jopp.gau.ac.ir>

DOI: 10.22069/jopp.2019.15568.2397

## بر آورد مکانی محصول کلزای پاییزه بر اساس روش‌های ناپارامتریک

\*حامد ادب<sup>۱</sup>، آزاده عتباتی<sup>۲</sup>، سید مهدی پورباقر کردی<sup>۳</sup>، محمد آرمین<sup>۴</sup> و حسن ذبیحی<sup>۵</sup>

<sup>۱</sup>استادیار گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشگاه حکیم سبزواری، ایران، <sup>۲</sup>استادیار گروه علوم و مهندسی محیط زیست،

دانشگاه حکیم سبزواری، ایران، <sup>۳</sup>استادیار گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشگاه پیام‌نور، واحد همدان، ایران،

<sup>۴</sup>دانشیار گروه زراعت، واحد سبزوار، دانشگاه آزاد اسلامی، سبزوار، ایران،

<sup>۵</sup>گروه علوم اطلاعات زمین، دانشگاه تکنولوژی مالزی، اسکودای، مالزی

تاریخ دریافت: ۹۷/۰۶/۲۱؛ تاریخ پذیرش: ۹۷/۱۰/۲۲

### چکیده

**سابقه و هدف:** قبل از توسعه کشت‌های جدید کلزا در مناطق مختلف ایران، ابتدا نیاز به بررسی پارامترهای مؤثر طبیعی و اقلیمی در عملکرد کلزا است، تا به‌وسیله آن توان بوم‌شناسی مناطق به‌منظور کشت کلزا شناخته شود. مدل‌سازی مکانی در سامانه اطلاعات جغرافیایی از مهم‌ترین راهکارهایی است که می‌تواند با ترکیب روش‌های آماری و داده‌های مکانی، زمینه را برای سنجش عوامل محیطی و تناسب اراضی برای کشت یک محصول خاص فراهم آورد. در این پژوهش رابطه مکانی بین عملکرد محصول کلزای پاییزه و عوامل آب، خاک و هواشناسی طی دوره رشد در مزارع نمونه بررسی شد.

**مواد و روش‌ها:** در این پژوهش با به‌کارگیری دستگاه موقعیت‌یاب جهانی (GPS) از ۲۴ مزرعه کشت کلزای پاییزه نمونه‌برداری شد و عملکرد واقعی آن محاسبه گردید. سپس مقادیر ده عامل محیطی شامل ارتفاع، شیب و جهت شیب، EC و pH آب زیرزمینی، میانگین دما، تابش کل دریافتی مستقیم و پراکنده، تبخیر و تعرق پتانسیل، شاخص عرضه باد و بافت خاک به روش نزدیک‌ترین همسایه برای مزارع انتخابی استخراج گردید. سپس بعد از نرمال‌سازی متغیرها و با در نظر گرفتن دامنه اعداد، نمونه‌ها به دو قسمت آموزش (۶۰ درصد، ۱۴ مزرعه) و آزمون (۴۰ درصد، ۱۰ مزرعه) به‌طور تصادفی تقسیم گردید. سپس از دو روش ناپارامتریک K نزدیک‌ترین همسایه و جنگل تصادفی به‌منظور برآورد توان محیطی عملکرد کلزا استفاده شد و در محیط سامانه اطلاعات جغرافیایی نقشه برآورد عملکرد محصول کلزا تهیه گردید.

**یافته‌ها:** نتایج میانگین درصد خطای مطلق در روش‌های مورد استفاده نشان داد که روش K نزدیک‌ترین همسایه با ۲۶ درصد خطا و جنگل تصادفی با ۱۱ درصد خطا است. نتایج شاخص کارایی نش ساتکلیف برای داده‌های آزمون نشان‌دهنده مقدار ۰/۶۵ برای روش K نزدیک‌ترین همسایه و ۰/۸۲ برای روش جنگل تصادفی است. روش جنگل تصادفی خطای کم‌تری نسبت به روش K نزدیک‌ترین همسایه در برآورد عملکرد کلزا در منطقه مورد مطالعه داشت. یافته‌های این پژوهش بر اساس مدل جنگل تصادفی نشان داد که شاخص عرضه باد و میانگین دما بیش‌ترین تأثیر و عوامل توپوگرافی جهت شیب جغرافیایی و ارتفاع از کم‌ترین تأثیر برخوردار هستند. همچنین EC و pH آب زیرزمینی از عوامل مهم در این مطالعه می‌باشند.

\* مسئول مکاتبه: [h.adab@hsu.ac.ir](mailto:h.adab@hsu.ac.ir)

نتیجه‌گیری: نتایج نشان داد که نواحی مناسب کشت گیاه کلزای پاییزه در مناطق شمالی و شمال‌غربی منطقه سبزوار واقع شده است. مناطق مرکزی با عملکرد پایین مشخص شدند که عمدتاً به دلیل وجود سازندهای گچی، نمکی در این مناطق و همچنین وجود زهکش کال شور و وجود نمکزارها است. از این رو توصیه می‌شود که برای توسعه کشت این گیاه مناطق شمالی و شمال‌غربی منطقه سبزوار در اولویت کشت قرار گیرد.

واژه‌های کلیدی: روش جنگل تصادفی، روش K- نزدیک‌ترین همسایه، محصول کلزای پاییزه، سبزوار

### مقدمه

که تهیه آن در مناطق فاقد تجهیزات امکان‌پذیر نیست (۱).

یکی دیگر از روش‌ها استفاده از روابط میان عملکرد محصول و متغیرهای بیوفیزیکی اندازه‌گیری شده توسط دورسنجی ماهواره‌ای است (۷۱)، اما داده سنسج‌ازدور به‌تنهایی قادر به تولید برآورد دقیق نیست (۷۱) مطالعاتی در زمینه استفاده از عوامل محیطی در برآورد عملکرد نیز انجام شده است (۷، ۳۲ و ۳۳). استفاده از مدل‌های تجربی آماری به دلیل این که به‌صورت محلی کالیبره می‌شوند، توصیه شده است. زیرا این مدل‌ها می‌توانند مبنای دقیقی برای برآورد عملکرد محصول در مقیاس‌های خرد و محلی باشند (۳۸). با ترکیب مدل‌های تجربی آماری با روش‌های مکانی می‌توان دامنه مطلوب عوامل محیطی برای کشت انواع محصولات در هر منطقه را مشخص نمود و سپس اساس آن مناطق مستعد کشت با بازدهی بالا تعیین می‌شود. از این رو ایجاد روابط تجربی بین عملکرد واقعی محصول و عوامل اندازه‌گیری شده در یک تاریخ یا در طول فصل رشد، ساده‌ترین روش برای برآورد میزان محصول است. با این حال، همان‌طور که در هر رویکرد صرفاً تجربی، برون‌یابی معادلات به مکان‌های و یا سال‌های جدید می‌تواند مشکل‌ساز باشد، به همین دلیل بسیاری از تلاش‌ها به سمت ارائه فن‌های کلی‌تر پیش می‌رود (۴۵).

تا زمانی که ارتباطات بین عوامل محیطی و فرآیندهای فیزیولوژیکی گیاهان زراعی در مناطق مختلف شناخته نشده باشد، تولید محصولی با عملکرد مناسب در یک شرایط محلی موضوعی کاملاً مبتنی بر کاشت و آزمایش خواهد بود. با توجه به این امر، با مدل‌سازی عملکرد محصولات آثار سوء عوامل محیطی و اقلیمی در محیط شناخته می‌شود و ثبات تولید در نظام کشاورزی پایدار می‌شود (۶۵). بنابراین در اختیار داشتن اطلاعات کمی از عوامل و تعاملات مؤثر بر عملکرد برای مدیریت محصول برای یک مکان خاص ضروری است. روش‌های مرسوم<sup>۱</sup> و روش‌های مکانی و دورسنجی ماهواره‌ای دو روش برای برآورد عملکرد محصولات زراعی هستند. روش‌های مرسوم برآورد عملکرد بر اساس حقایق زمینی و بر اساس مدل‌های تجربی آماری و مدل رشد محصول تهیه می‌شوند. مدل‌های رشد محصول امکان برآورد عملکرد به‌عنوان تابعی از فرآیندهای فیزیولوژیکی پیچیده مختلف با محیط را فراهم می‌آورند؛ اما این مدل‌ها اغلب پیچیده و نیاز به داده‌های متنوع و اندازه‌گیری شده در محل مزارع هستند (۵۲). برآورد عملکرد محصول با استفاده از مدل‌های شار انرژی دارای محدودیت ویژه‌ای هستند، زیرا اجرای آن نیازمند داده‌های شار انرژی اندازه‌گیری شده در محل است

1- Conventional

بررسی خصوصیات طبیعی (شامل آب، خاک و هوا) سرزمین و انطباق آن با نیازهای هر محصول توسط روش های متعددی انجام می شود، تا تناسب یک سرزمین با محصول موردنظر به دست آید.

مطالعات انجام شده جهت برآورد عملکرد محصولات مانند کلزا به دو قسمت روش های شبیه سازی شده مانند مدل CropSyst جهت شبیه سازی رشد گیاه کلزا (۳۰) و مدل CRPSM (۳۱) و روش های آماری و داده کاوی مانند روش رگرسیون خطی چندمتغیره (۲) و شبکه عصبی قابل طبقه بندی است (۳). الگوریتم رگرسیونی جنگل تصادفی (RF) و الگوریتم نزدیک ترین همسایه (KNN) یکی از روش های ناپارامتری یادگیری ماشین است که اخیراً در علومی مانند هیدرولوژی کاربرد بسیاری پیدا کرده است. این روش ها قادر به تولید انواع وابستگی خطی و غیرخطی بین متغیر وابسته و چندین متغیر وابسته است. از این رو هنگامی که ارتباط بین داده های ورودی و خروجی مشخص نباشد، چنین روشی در برآورد متغیر موردنظر از سودمندی بیش تری برخوردار است.

استان خراسان رضوی یکی از استان هایی است که به دلیل شرایط مناسب آب و هوایی، از توان لازم برای کشت و تولید کلزا دارا است و به اعتقاد کارشناسان شهرهای شمالی و مرکزی استان خراسان رضوی از قابلیت خوبی به منظور کشت کلزا برخوردار هستند (۵۱). این مطالعه باهدف ارزیابی توان محیط (آب و هوا، خاک و سیمای توپوگرافی) برای کشت کلزای پاییزه از طریق برآورد عملکرد آن با استفاده از فن تحلیل مکانی در شهرستان های سبزوار، داورزن، خوشاب، جوین و جغتای (که در این پژوهش منطقه سبزوار اطلاق می شود) انجام شده است. در این پژوهش رابطه مکانی بین عملکرد محصول کلزا و عوامل آب، خاک و

در حال حاضر، ایران یکی از بزرگ ترین واردکنندگان روغن نباتی به شمار می رود و عمده منابع داخلی تولید روغن نباتی شامل پنبه دانه، سویا، آفتابگردان و گل رنگ است، اما در دهه اخیر کشت کلزا با نام علمی *Brassica napus* مورد توجه قرار گرفته است. کلزا گیاهی روغنی و زراعی است و متعلق به خانواده شب بو است. این خانواده از ۳۵۰ جنس و بیش از ۳۰۰۰ گونه شامل می شود (۴۰). از میان گونه های جنس *Brassica* کاشت کلزا و رقم اصلاح شده آن که امروزه کانولا خوانده می شود، سازگاری گسترده به شرایط محیطی دارد و می تواند نقش مهمی در تولید محصولات دانه های روغنی در ایران داشته باشد (۶۴). در این راستا طرح ۱۰ ساله تأمین منابع روغن نباتی کشور در وزارت جهاد کشاورزی طی ۱۳۸۴-۹۳ به اجرا گذاشته شده است. هدف این طرح افزایش تولید دانه های روغنی در کشور و افزایش درصد خوداتکایی به ۷۰ درصد به تولید داخلی این محصول بود (۵۷) اما یکی از دلایل عدم موفقیت کامل طرح فوق می تواند عدم توجه به توانمندی های محیطی و شناسایی نواحی مستعد کشت این محصول به صورت فضایی و در مقیاس راهبردی باشد.

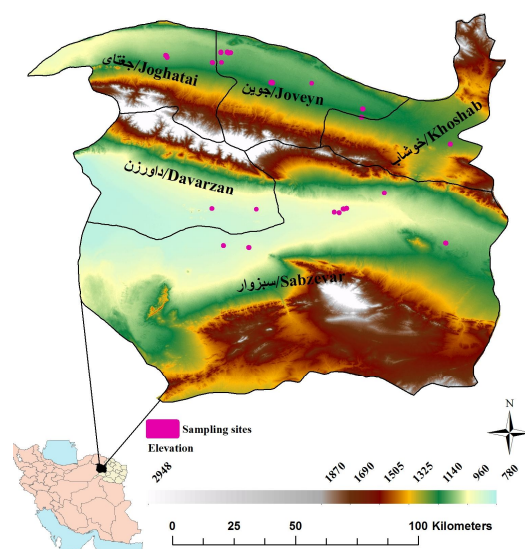
قبل از گسترش کشت های جدیدی مانند کلزا در مناطق مختلف ایران، ابتدا نیاز به بررسی پارامترهای مؤثر طبیعی و اقلیمی در عملکرد کلزا است تا به وسیله آن توان بوم شناختی مناطق برای کشت کلزا شناخته شود. اگرچه کشاورزان می توانند برخی عوامل مانند خاک، بذر و فنون کشاورزی کنترل کنند، اما آب و هوا و عوامل طبیعی به دلیل عدم امکان کنترل آن در مزرعه، یک عامل مهم و مستقل در بخش کشاورزی است که بر روی عملکرد به لحاظ کیفیت و کمیت مؤثر است (۶۲). از این رو

ارتفاعات اصلی منطقه را تشکیل می‌دهد (ارتفاعات سفید رنگ در شکل ۱). شرایط آب‌وهوایی خشک و نیمه‌خشک که منجر به محدودیت پوشش گیاهی می‌شود، فرآیند پدوژنز خاک را در منطقه محدود می‌سازد. در منطقه سبزوار ۱۰ نوع پوشش اراضی مشاهده می‌گردد، که بیش‌ترین درصد آن را لندفرم‌های کوه و تپه تشکیل می‌دهد (۳). اقلیم منطقه مطابق با طبقه‌بندی اقلیمی کوپن دارای آب‌وهوای خشک نیمه بیابانی است (۳). میزان بارش ماهانه ایستگاه هواشناسی سبزوار ۱۹۰/۵ میلی‌متر و متوسط دمای هوا ۱۷/۴ درجه سلسیوس است. میانگین تبخیر و تعرق پتانسیل برای منطقه سبزوار ۱۱۶۴/۸ میلی‌متر است (۴۳).

هواشناسی طی دوره رشد در مزارع نمونه مدل‌سازی شد. همچنین پژوهش حاضر ضمن بررسی کاربرد رگرسیون جنگل تصادفی و الگوریتم نزدیک‌ترین همسایه در برآورد عملکرد کلزا با استفاده از انواع داده‌های محیطی، برنامه‌ریزی فضایی سرزمین جهت کشت کلزا مورد تاکید قرار گرفت.

### مواد و روش‌ها

منطقه مورد مطالعه در این پژوهش منطقه سبزوار است و با ارتفاع متوسط ۹۷۷ متر از سطح دریا و مساحت ۲۰۵۰۲ کیلومترمربع در غرب استان خراسان رضوی واقع شده است (۴۵) (شکل ۱). دو رشته‌کوه جغتای در شمال و کوه میش و ششتمد در جنوب که با جهت غربی- شرقی امتداد داشته،



شکل ۱- موقعیت منطقه مورد مطالعه و موقعیت مزارع انتخابی کلزا.

Fig. 1. The geographical locatin of study area and the selected rapeseed farms.

نیمه‌خشک سرد و خشک بیابانی سرد قرار گرفتند. سپس ۲۴ مزرعه به‌عنوان نمونه برای مدل‌سازی برآورد عملکرد محصول انتخاب گردید که به تعداد مساوی ۱۲ مزرعه در هر تیپ اقلیمی انتخاب شد.

نمونه‌گیری مزارع: در این مطالعه تعداد مناسبی از مزارع کشت کلزا در سطح منطقه انتخاب گردید. انتخاب مزارع بر اساس نقشه اقلیمی روش دمارتن انجام شد، به‌گونه‌ای که مزارع کلزا در ۲ تیپ اقلیمی

عملکرد محصول یک پاسخ تجمع یافته به محیط طی دوره رشد از کاشت تا برداشت است. با توجه به مطالعات انجام شده در مورد موضوع مورد پژوهش در این مطالعه، ۱۰ عامل محیطی شامل سیمای توپوگرافی (ارتفاع، شیب و جهت شیب)، EC و pH منابع آب زیرزمینی، میانگین دما طی فصل رشد، تابش کل دریافتی مستقیم و پراکنده طی فصل رشد، تبخیر و تعرق پتانسیل طی فصل رشد، شاخص عرضه باد و بافت خاک انتخاب گردید (۲، ۳۱، ۴۱ و ۵۱).

عوامل مورد استفاده در این پژوهش به منظور برآورد عملکرد کلزا در جدول ۱ مشخص است.

مزارع انتخاب شده با توجه به سطح سبز به ۵ طبقه تقسیم شد و هر طبقه با توجه به فراوانی مزارع در آن طبقه درصدی از انتخاب را به خود اختصاص داد. هرچه تعداد فراوانی مزارع در هر طبقه بیش تر باشد، به همان نسبت تعداد مزارع انتخابی هم افزایش می یابد. در این پژوهش، برداشت میدانی سطح سبز مزارع انتخابی کلزا در سال زارعی ۸۴-۱۳۸۳ به صورت پلیگونی با استفاده از دستگاه موقعیت یاب ماهواره ای GPS Garmin مدل GPSmap 76S (صحت ۳ متر) با انجام شد. در نهایت عملکرد واقعی مزارع ثبت شده توسط کارشناسان زراعت مدیریت جهاد کشاورزی شهرستان سبزوار به دست آمد.

جدول ۱- متغیرهای جغرافیایی مورد استفاده در این پژوهش به منظور برآورد عملکرد کلزا.

Table 1. The geographical variables used for estimation of rapeseed yield.

منبع/ روش تهیه Datasource/ method	متغیر Variable	منبع/ روش تهیه Datasource/ method	متغیر Variable
ایستگاه های هواشناسی قوچان، کاشمر، سبزوار، مزینان و کشت و صنعت جوبین، روش گرادیان ارتفاع Weather stations; Quchan, Kashmar, Sabzevar, Mazinan, and Jovain industrial and agricultural Co Gradient method	میانگین دمای هوا Mean air temperature	ASTER (۹۳) نسخه ۲ ASTER v2	مدل رقومی جهانی ارتفاع Global Digital Elevation Map
مدل ارائه شده توسط (۱۲) The model provided by (12)	تابش پتانسیل ورودی خورشیدی Potential Incoming Solar Radiation	چندجمله ای مرتبه دوم (۱۰۲) Second-order polynomial	شیب توپوگرافی Slope
مدل ارائه شده توسط (۴۳) The model provided by (43)	تبخیر و تعرق پتانسیل evapotranspiration	چندجمله ای مرتبه دوم (۱۰۲) Second-order polynomial	جهت شیب توپوگرافی Aspect
مدل ارائه شده توسط (۱۱) The model provided by (11)	شاخص عرضه باد Wind Exposition Index	شرکت مدیریت منابع آب ایران، روش معکوس فاصله (۱۴ و ۹۴) Iran Water Resources Management Co Inverse Distance Weight	EC و pH منابع آب زیرزمینی EC and pH of groundwater
سازمان جنگل ها و مراتع کشور Forests, Range and Watershed Management Organization	بافت خاک Soil texture		

آماده‌سازی داده‌های ورودی: ابتدا تمام متغیرهای مستقل ورودی به الگوریتم‌ها استاندارد شد، زیرا تفاوت در واحد و دامنه تغییرات متغیرهای مستقل مورد استفاده موجب ایجاد انحراف در مدل می‌شود. روش‌های یکسان‌سازی متعددی به منظور استاندارد کردن داده‌ها وجود دارد. در این مطالعه تمام ورودی‌های مدل با استفاده از روش پارتو تبدیل به متغیرهایی با توزیع میانگین صفر شد. استفاده از روش پارتو به دلیل انجام کاهش اهمیت نسبی مقادیر بزرگ و از طرفی حفظ نسبی ساختار داده‌ها، توصیه می‌شود (۷۷). سپس استخراج مقادیر متغیرهای مستقل به روش نزدیک‌ترین همسایه به جدول اطلاعات توصیفی نمونه‌ها انجام شد. روش‌های متعددی به منظور تقسیم داده‌ها برای پیش‌بینی وجود دارد، در این پژوهش داده‌ها با توجه به دامنه اعداد به دو قسمت آموزش (۶۰ درصد، ۱۴ مزرعه) و آزمون (۴۰ درصد، ۱۰ مزرعه) به صورت تصادفی تقسیم گردید. به طور معمول صحت الگوریتم به وسیله داده‌های آزمون که برای آموزش مورد استفاده قرار نمی‌گیرد، ارزیابی می‌شود.

**الگوریتم k نزدیک‌ترین همسایگی:** روش k نزدیک‌ترین همسایگی به عنوان یک الگوریتم یادگیری ماشین یکی از ساده‌ترین الگوریتم‌های آموزش برای پیش‌بینی است. در این روش مشابهت و عدم مشابهت بین داده‌ها بر اساس محاسبه تابع فاصله مانند فاصله اقلیدسی بین داده‌ها تعیین می‌شود. به این صورت که داده‌های مشابه در یک فضای ورودی  $n$  بعدی در نزدیکی یکدیگر قرار می‌گیرند. در این پژوهش یک بانک داده مرجع (حاوی پیکسل‌های معلوم عملکرد کلزا که همان ۱۴ مزرعه انتخابی است) برای جستجو و یافتن نزدیک‌ترین (مشابه‌ترین) نمونه به نمونه هدف،

مورد استفاده قرار گرفت. منظور از داده هدف، نمونه یا پیکسل‌های مجهول عملکرد کلزا است که تمام متغیرهای توپوگرافی و اقلیمی آن به جز متغیر مورد برآورد (عملکرد کلزا) مشخص است. در این حالت الگوریتم از میان  $k$  تعداد از مشابه‌ترین پیکسل‌های عملکرد موجود در بانک داده مرجع به نمونه داده هدف، به برآورد مقدار عملکرد کلزا در پیکسل‌های هدف (مجهول) می‌پردازد. در فرآیند دسته‌بندی داده‌ها، تعداد همسایگان ( $k$ ) نقش مهمی دارد. همسایه به داده‌هایی در مجموعه آموزشی گفته می‌شود که در کنار یکدیگر قرار می‌گیرند.

تعیین تعداد مناسب تعداد همسایگان (مقدار  $K$ ) که نزدیک‌ترین فاصله را با پیکسل‌های هدف دارند، به عنوان یکی از متغیرهای مهم در این فرآیند است. یکی از معمول‌ترین روش‌ها استفاده از الگوریتم تپه نوردی که به صورت سعی و خطا تعداد صحیح  $k$  انتخاب می‌شود. در این روش میزان  $k$  را افزایش داده تا به بهترین انتخاب برسند (۱۴). در این مطالعه، برای تعیین تعداد بهینه  $k$  در تخمین نمونه هدف، مقادیر بین  $k=1$  تا  $k=12$  در مجموعه آموزشی انتخاب شد و معیار جذر میانگین مربعات خطا برای مقادیر  $k=1$  تا  $k=12$  محاسبه شد. کم‌ترین جذر میانگین مربعات خطا به عنوان بهترین تعداد  $k$  در این مطالعه استفاده شد. از مهم‌ترین قسمت‌های الگوریتم  $k$  نزدیک‌ترین همسایگی، تعیین فاصله بین پیکسل‌های هدف (مجهول) با هر یک از پیکسل‌های (معلوم) موجود در بانک داده‌ها است (۱۴). در این مطالعه از فاصله اقلیدسی که یکی از روابط فاصله کلاسیک می‌باشد برای اندازه‌گیری فاصله بین پیکسل‌های مجهول (هدف) تا هر یک از نمونه پیکسل‌های بانک مرجع استفاده شد. محاسبه فاصله اقلیدسی به صورتی انجام شد که نمونه‌ها (مزارع به عنوان

داده‌های آموزشی) و پیش‌بینی است. در پژوهشی، تجزیه و تحلیل ۲۹ مجموعه داده نشان داد که تفاوت معنی‌داری زیادی بین استفاده از ۱۲۸ درخت با ۲۵۶، ۵۱۲، ۱۰۲۴، ۲۰۴۸ و ۴۰۹۶ وجود ندارد. گاهی اوقات، تعداد زیادی از درختان تنها هزینه محاسباتی آن را بدون این‌که موجب افزایش معنی‌داری در عملکرد مدل داشته باشد، به دنبال دارد (۶۰).

به منظور آموزش مدل، از تعدادی یادگیرنده ضعیف استفاده شد. این روش کنترل‌کننده تعداد مدل‌های رگرسیون ضعیف است. در این روش فرآیند گروه‌بندی هنگامی که تعداد مدل‌های رگرسیون ایجاد شده به مقدار تعیین شده به یادگیرنده ضعیف برسد، متوقف خواهد شد. سپس الگوریتم، میانگین وزنی تمام یادگیرنده ضعیف را محاسبه و آن مقدار را به مدل اختصاص می‌دهد. این یادگیرنده‌ها به‌خود خود نمی‌توانند عاملی برای دسته‌بندی مطلوب یک پدیده باشد، اما ترکیب همه آن‌ها سبب می‌شود یک دسته‌بند قوی ایجاد شود. در این مطالعه، برای تعیین تعداد بهینه یادگیرنده ضعیف از تغییرات میانگین مربعات خطا در مقابل تعداد یادگیرنده ضعیف در نمونه آموزشی استفاده شد. همچنین در این مطالعه از روش نمونه بوت استرپ از داده‌های ورودی به منظور ساخت درخت‌های تصمیم به‌طور تصادفی استفاده شد که در آن تعدادی از متغیرها به‌طور تصادفی در ساخت هر درخت شرکت می‌کنند. با انجام نمونه‌گیری با تکرارهای متعدد، تعدادی مجموعه داده OOB<sup>۱</sup> از مجموعه آموزشی تولید می‌شود، که به وسیله آن می‌توان برای هر مجموعه داده آموزشی یک رشد درخت تصمیم متصور شد (۴۴).

پیکسل‌های معلوم) بر حسب مقدار فاصله محاسبه شده به ترتیب صعودی از کم‌ترین فاصله (حداکثر تشابه) تا بیش‌ترین فاصله (حداقل تشابه) از نمونه موردنظر دسته‌بندی و ارزش‌گذاری کردند. سپس هنگامی که داده جدیدی که به الگوریتم وارد گردد، در ابتدا فاصله آن با دیگر داده‌ها محاسبه و سپس بر اساس یک سیستم رأی‌گیری ساده رأی اکثریت به گروهی که در نزدیک‌ترین فاصله قرار داشته باشد، منتصب و پیش‌بینی می‌گردد.

**الگوریتم درخت تصمیم:** در این مطالعه از الگوریتم درختان تصادفی با نام مستعار (جنگل تصادفی) که یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است، استفاده شد. این روش پیش‌بینی‌کننده بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوطه استوار است (۱۳). جنگل‌های تصادفی به‌طور کلی بهبود و عملکرد قابل‌توجهی به وجود نويز در داده‌های ورودی نسبت به طبقه‌بندی درخت منفرد مانند روش‌های CART و C4.5 دارند (۸۵). این روش یکی از روش‌های غیرپارامتری است و بر خلاف مدل‌های کلاسیک مانند رگرسیون خطی ساده که تنها بر یک مدل تکیه دارند، با استفاده از صدها و هزاران درخت از اطلاعات بیش‌تری در داده‌ها استفاده می‌شود تا بتوان استنباط بهتری از متغیرها برای پیش‌بینی داشت. این روش با استفاده از مجموعه داده آموزش، درخت‌های متعدد رگرسیونی ایجاد می‌کند و پس از آن اهمیت متغیرها در پیش‌بینی پاسخ اندازه‌گیری می‌شود. در این مطالعه، برای ساخت مدل جنگل تصادفی به منظور پیش‌بینی عملکرد کلزا، از مراحل رشد درخت ارائه شده توسط (۸۰) استفاده شد، که شامل سه مرحله: ورودی (داده‌های آموزشی، تعداد درختان)، ساخت درخت (نمونه‌گیری بوت استرپ از

1- Out Of Bag

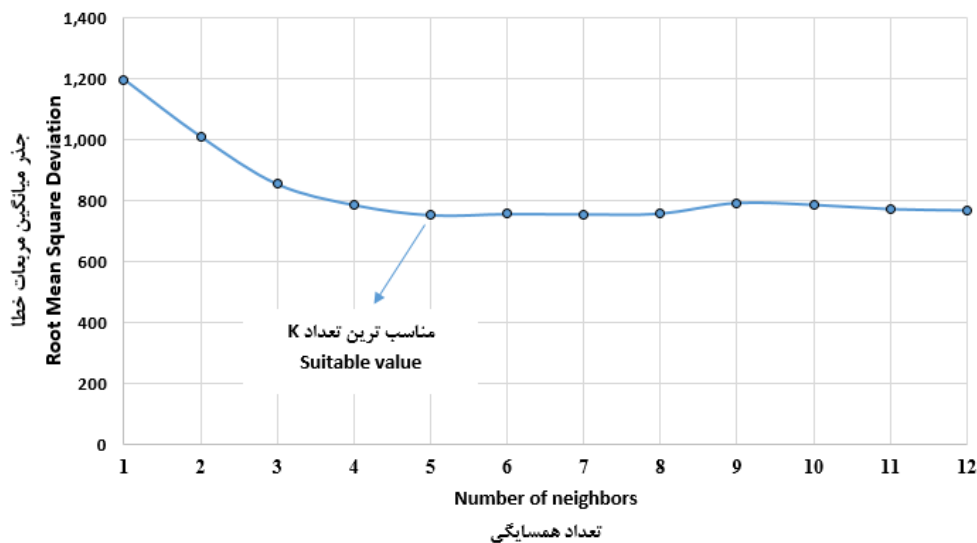
ارزیابی مدل‌های مستخرج از الگوریتم‌ها: در این مطالعه برای آزمون و ارزیابی برآورد عملکرد کلزا با دو روش K نزدیک‌ترین همسایه و جنگل تصادفی از معیارهای ریشه دوم میانگین مربعات خطا (RMSE)، میانگین قدرمطلق خطا (MAE)، میانگین درصد خطای مطلق (MAPE)، ضریب تبیین ( $R^2$ ) و شاخص کارایی نش ساتکلیف (NSE) استفاده گردید. شاخص کارایی نش ساتکلیف، اختلاف نسبی مقادیر مشاهده‌ای و شبیه‌سازی شده را نشان می‌دهد و مقدار این فاکتور بین یک تا منفی بی‌نهایت تغییر می‌کند. مقدار بیش‌تر از ۰/۵ این فاکتور نشان‌دهنده شبیه‌سازی خوب توسط مدل است.

### نتایج و بحث

بر اساس توضیحات داده‌شده در بخش مواد و روش‌ها، الگوریتم k نزدیک‌ترین همسایگی و الگوریتم جنگل تصادفی برای برآورد عملکرد کلزا در منطقه سبزوار استفاده شد. شکل ۲ نشان‌دهنده اثر تعداد k تا ۱۴ در میزان صحت برآورد عملکرد کلزا در الگوریتم k نزدیک‌ترین همسایگی است. بر اساس این نمودار می‌توان تعداد k بهینه را بر اساس مقدار معیار جذر میانگین مربعات خطا به‌دست آورد. همان‌طور که در شکل ۲ مشاهده می‌شود، بیش‌ترین میزان جذر میانگین مربعات خطا با ۹۵۶/۷ کیلوگرم در هکتار در کم‌ترین تعداد همسایگی (k) مساوی با ۱ است و با افزایش تعداد k از میزان خطا کاسته می‌شود. روند کاهش میزان خطا به‌طور نسبی تا  $k=4$  با ۴۹۷/۹ کیلوگرم در هکتار ادامه داشته است، ولی بعد از آن میزان خطا به‌طور اندکی به ۵۶۱/۶ افزایش یافته است. بنابراین بهترین تعداد k تا ۱۴ در این مطالعه،  $k=4$  است که انتظار افزایش صحت برآورد نسبت به سایر تعداد k در الگوریتم k نزدیک‌ترین همسایگی می‌رود.

بوت استرپ با ایجاد نمونه‌های فراوان فارغ از بسیاری از فرضیات (مانند پذیرش پیش‌فرض‌های مربوط به نرمال بودن توزیع و نمونه بزرگ) است و شرایط نمونه را به شرایط جامعه نزدیک می‌کند، از این‌رو نتایج از اعتبار مناسبی برخوردار است (۱۱). در مطالعات محیطی (مانند پژوهش حاضر) دستیابی به حجم نمونه‌های بزرگ برای رسیدن به نتیجه مطلوب به‌عنوان یک مشکل اساسی محسوب می‌شود (به‌گونه‌ای که در صورت فقدان نمونه کافی برای مدل‌سازی، شرایط مدل مناسب نخواهد بود)، اما با روش بوت استرپ استفاده شده در این مطالعه که در سال ۱۹۷۹ توسط افرون ارائه شد، مشکل نبود نمونه‌های بزرگ حل گردد (۲۱). تعداد درخت در جنگل ارتباط با قدرت پیشگویی آن دارد؛ اما همیشه با افزایش تعداد درخت، صحت پیش‌بینی مدل افزایش نمی‌یابد، چرا که پارامترهای دیگر مانند تعداد متغیر انتخاب‌شده در هر گره درخت و حداقل اندازه گره‌های پایانی نیز اثر تعیین‌کننده بر صحت پیش‌بینی دارد (۶۱). مطالعات نشان می‌دهد که با افزایش درخت از ۱۰۰۰ به ۱۵۰۰ و سپس به ۲۰۰۰ نه‌تنها خطا افزایش پیدا می‌کند، بلکه صحت پیش‌بینی نیز کاهش می‌یابد (۲۳). همچنین در این حالت، زمان محاسبات نیز به‌طور محسوس (تا دو برابر) افزایش یافته است (۲۳). بنابراین در این مطالعه برای تعیین میزان بهینه تعداد یادگیرنده ضعیف در جنگل تصادفی بر اساس آماره جذر میانگین مربعات خطا عمل شد. در این مطالعه اندازه درخت را تا به مرور افزایش داده تا جایی که بیش‌ترین قدرت پیشگویی و کم‌ترین خطا در درخت‌ها به‌دست آید. هرچه تعداد درخت‌های یک جنگل افزایش می‌یابد، خطای پیشگویی جنگل همگرا به یک مقدار می‌شود (۱۳).



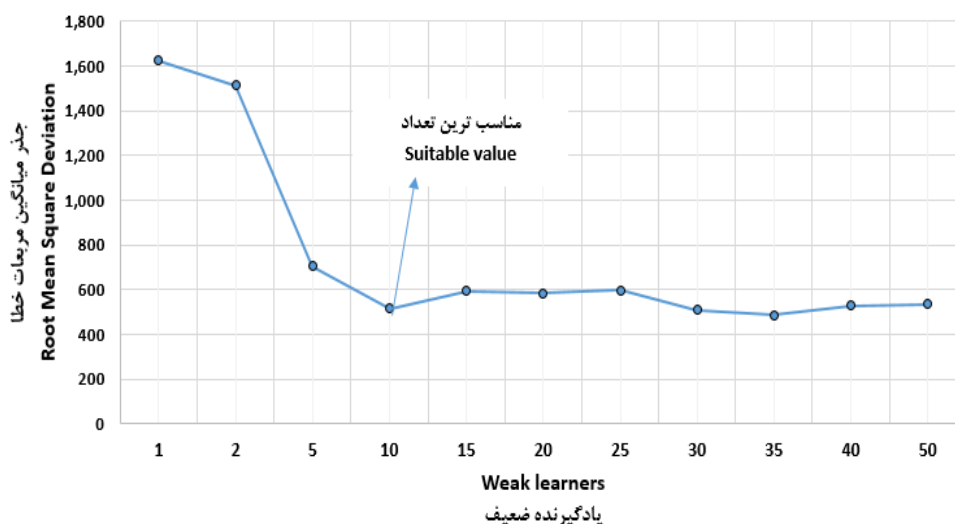


شکل ۲- تعیین میزان بهینه k در الگوریتم k نزدیک ترین همسایگی بر اساس آماره جذر میانگین مربعات خطا.

Fig. 2. Detirminig the optimal value of k in KNN algorithm by RMSE.

جنگل تصادفی است. نتایج نشان داد که تعداد ۱۰ یادگیرنده ضعیف دارای کم ترین آماره جذر میانگین مربعات خطا با ۵۱۱ کیلوگرم در هکتار است و این تعداد برای مدل انتخاب شد.

در این مطالعه تعداد یادگیرنده ضعیف بین ۱ تا ۵۰ در الگوریتم جنگل تصادفی تنظیم گردید و تعداد مناسب یادگیرنده ضعیف بر اساس آماره جذر میانگین مربعات خطا تعیین شد. شکل ۳ نشان دهنده اثر تعداد یادگیرنده ضعیف بین ۱ تا ۵۰ در الگوریتم

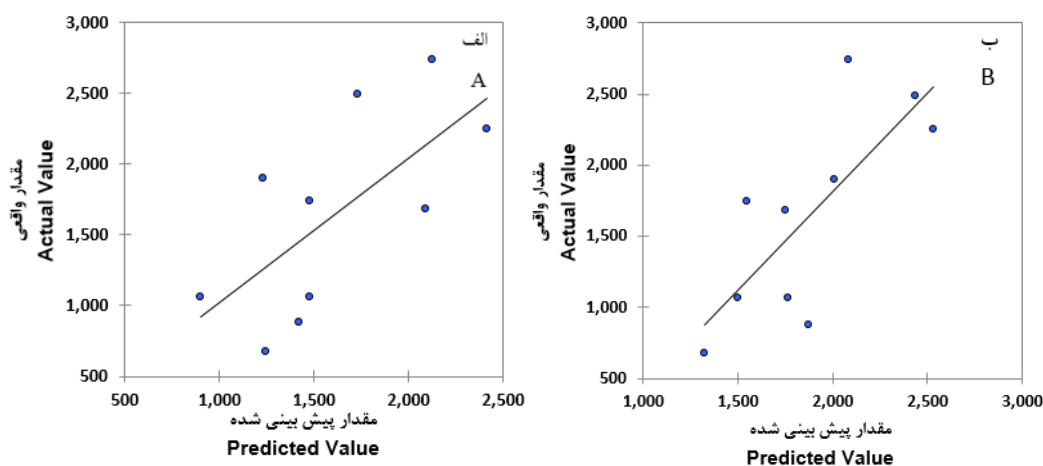


شکل ۳- تعیین میزان بهینه تعداد یادگیرنده ضعیف در الگوریتم جنگل تصادفی بر اساس آماره جذر میانگین مربعات خطا.

Fig. 3. Detirminig the optimal number of learning in Random Forest algorithm by RMSE.

همسایه و ۰/۸۲ برای روش جنگل تصادفی است. شاخص کارایی نش ساتکلیف از  $\infty$  تا +۱ است و هرچه کارایی مدل به ۱ نزدیک‌تر باشد، صحت مدل بیش‌تر است. شکل ۴ پراکنش بین داده‌های پیش‌بینی و داده‌های واقعی برای داده‌های آزمون را نمایش می‌دهد، همان‌طور که در شکل مذکور نشان مشخص است، روش جنگل تصادفی شبیه‌سازی مناسب‌تری در مقایسه با روش K نزدیک‌ترین همسایه در برآورد عملکرد کلزا داشته است. بنابراین نتایج نشان‌دهنده آن است که روش جنگل تصادفی خطای کم‌تری نسبت به روش K نزدیک‌ترین همسایه در برآورد عملکرد کلزا در منطقه مورد مطالعه دارد.

مقایسه آماره‌های ریشه دوم میانگین مربعات خطا برای روش K نزدیک‌ترین همسایه و جنگل تصادفی به ترتیب برابر با ۵۲۴ و ۵۴۵ و میانگین قدرمطلق خطا به ترتیب برابر ۴۵۶ و ۴۱۵ کیلوگرم در هکتار به دست آمد. نتایج میانگین درصد خطای مطلق در روش‌ها مورد استفاده نشان داد که روش K نزدیک‌ترین همسایه با ۲۶ درصد خطا و جنگل تصادفی با ۱۱ درصد خطا است. مقدار ضریب تبیین ( $R^2$ ) نشان داد که روش K نزدیک‌ترین همسایه با ۶۵ درصد و جنگل تصادفی با ۸۲ درصد قادر به توضیح برآورد عملکرد کلزا با استفاده از متغیرهای مورد استفاده در این پژوهش است. نتایج شاخص کارایی نش ساتکلیف برای داده‌های آزمون نشان‌دهنده مقدار ۰/۶۵ برای روش K نزدیک‌ترین

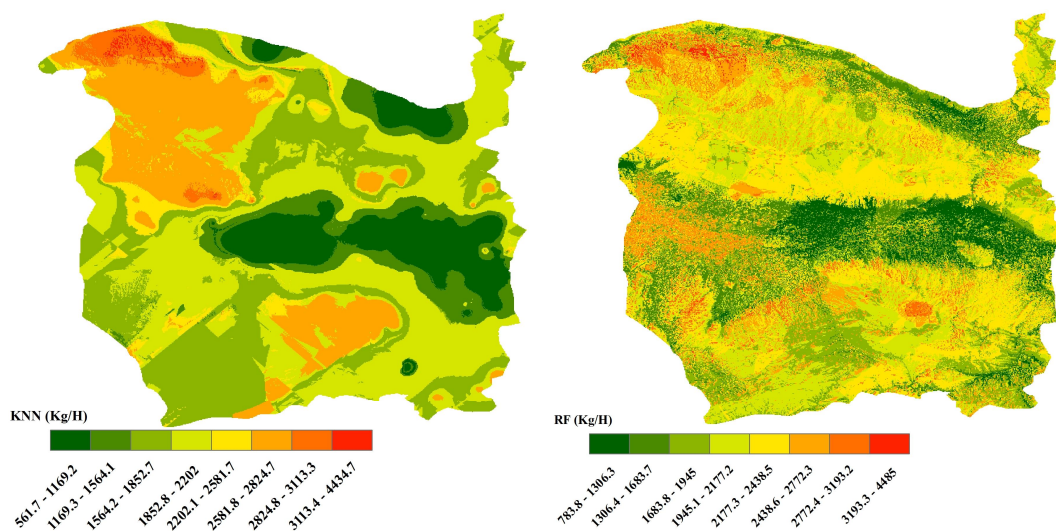


شکل ۴- عملکرد اندازه‌گیری شده و برآورد شده توسط الف) روش K نزدیک‌ترین همسایه و ب) روش جنگل تصادفی برای داده‌های آزمون.

Fig. 4. Observed vs estimated crop yield by A) KNN and B) Random Forest algorithm for testing dataset.

پایین دانه کلزا در هر دو روش، در مناطق مرکزی (نواحی سبز رنگ) و عملکرد بالا در نواحی شمالی و عمده در شمال غربی (نواحی نارنجی- قرمز رنگ) منطقه مورد مطالعه برآورد شد؛ اما به نظر می‌آید که روش جنگل تصادفی قادر به ارائه جزئیات مکانی بیش‌تر نسبت به روش K نزدیک‌ترین همسایه است.

شکل ۵، نتایج حاصل از اجرای روش‌های K نزدیک‌ترین همسایه و جنگل تصادفی برای برآورد عملکرد کلزا در منطقه مورد مطالعه را نشان می‌دهد. مقایسه الگوی مکانی برآورد عملکرد کلزا در هر دو روش نشان می‌دهد که الگوی مکانی کلی تا حدودی مشابه یکدیگر است، به عنوان مثال نواحی با عملکرد

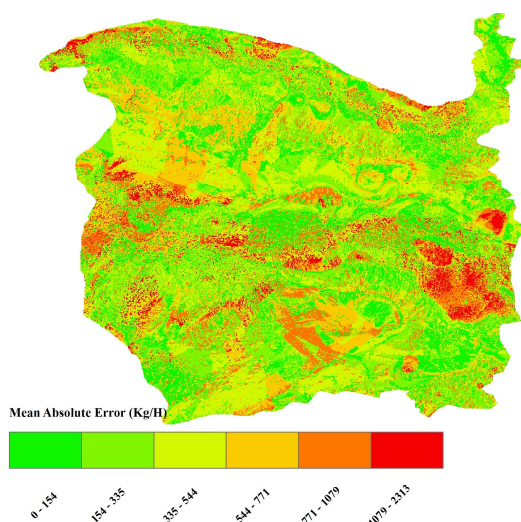


شکل ۵- برآورد عملکرد محصول کلزا با روش جنگل تصادفی (RF) و K نزدیک‌ترین همسایه (KNN).

Fig. 5. Estimated rapeseed crop yield by Random Forest and KNN algorithm.

در اکثر نواحی شمالی مشاهده می‌شود. با توجه به شکل ۶ می‌توان نتیجه گرفت که در نواحی که قدرمطلق خطای هر دو روش کم است (نواحی سبزرنگ) انتظار صحت بالای در برآورد عملکرد می‌رود و در نواحی با اختلاف بالا (نواحی قرمز رنگ) صحت کم در برآورد عملکرد کلزای پاییزه وجود دارد.

شکل ۶ نشان‌دهنده قدرمطلق خطای عملکرد دانه کلزا به منظور شناسایی مناطق با حداقل خطا در برآورد عملکرد بین روش‌های K نزدیک‌ترین همسایه و جنگل تصادفی در منطقه مورد مطالعه است. بیش‌ترین اختلاف بین دو روش مورد استفاده در نواحی مرکزی منطقه مورد مطالعه با اختلاف ۲۳۱۳-۱۰۷۹ کیلوگرم در هکتار (نواحی قرمز رنگ) و کم‌ترین اختلاف بین ۱۵۴-۰ کیلوگرم در هکتار



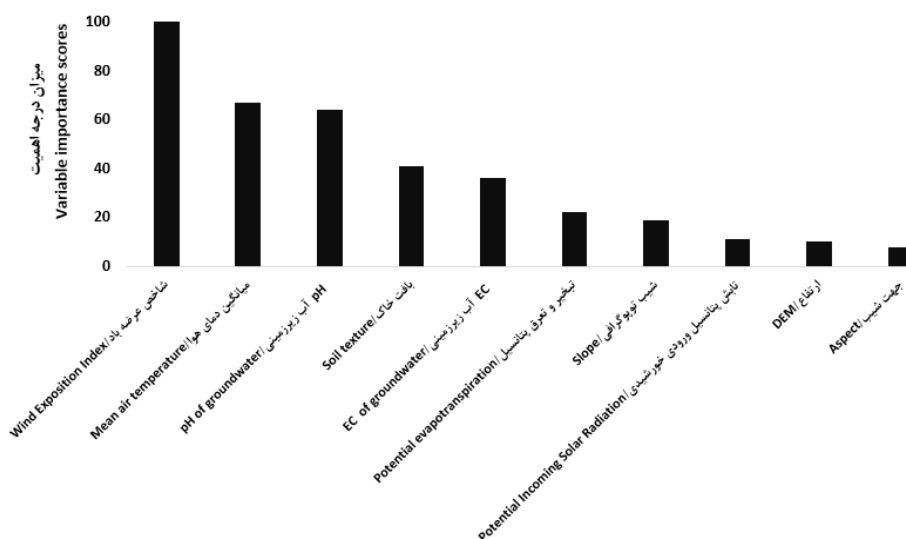
شکل ۶- قدرمطلق خطای برآورد عملکرد دانه کلزا (کیلوگرم در هکتار) بین دو روش K نزدیک‌ترین همسایه و جنگل تصادفی.

Fig. 6. Mean Absolute Error (MAE) between estimated crop yield by KNN and Random Forest algorithm.

شکل ۷ نشان‌دهنده نرخ تأثیرات عوامل دهگان محیطی بر عملکرد واقعی دانه کلزا در مدل جنگل تصادفی است. نتایج نشان داد که از میان عوامل مورد بررسی، شاخص عرضه باد و میانگین دما بیش‌ترین تأثیر و عوامل توپوگرافی که مانند جهت شیب و ارتفاع از کم‌ترین تأثیر برخوردار هستند (شکل ۷). طبق نتایج پژوهش‌های قبلی در دیگر مناطق نشان داده شد که باد و جریان هوا در سطح زمین تأثیر زیادی در گرده‌افشانی گیاه کلزا دارد (۱۵، ۵۸ و ۷۹). گرده‌های رها شده گل‌های کلزا به وسیله باد و حشرات کمک مؤثری بر کمیت و کیفیت آن است (۱۶). سرعت باد در دوره روزت این گیاه بیش‌ترین تأثیر بر عملکرد کلزا دارد (۵۹) و افزایش نسبی سرعت باد موجب افزایش عملکرد می‌شود (۷۸). نتایج مطالعه حاضر نشان می‌دهد که با افزایش شاخص عرضه باد، میزان عملکرد دانه به‌طور معناداری افزایش نسبی پیدا می‌کند ( $p\text{-value}=0/01$ )، اما کاهشی در مقدار عملکرد در نواحی با شاخص عرضه باد بالا مشاهده نشد. لازم به ذکر است که حضور کندوهای زنبورعسل تنها در ۱ مزرعه از میان ۲۴ مزرعه زنبور مشاهده شد و می‌توان افزایش عملکرد دانه را در اثر گرده‌افشانی توسط حشرات را در نظر نگرفت. البته مشاهدات آزمایشگاهی در نرخ رشد گیاهان جوان کلزا در تونل باد نشان داد که نرخ رشد نسبی با سرعت باد کم افزایش می‌یابد، اما زمانی که سرعت باد بیش‌تر می‌شود، نرخ رشد کاهش پیدا می‌کند.

سرعت باد ۰/۳ متر بر ثانیه باعث رشد مطلوب شده است هر چند که تغییرات در نرخ رشد کوچک بود (۷۸).

در این مطالعه، میانگین دمای هوا به‌عنوان یکی از عوامل مهم در عملکرد دانه در مدل جنگل تصادفی شناخته شد. نتایج مطالعه حاضر نشان داد که میزان عملکرد دانه با کاهش دما همبستگی منفی دارد، که این امر نشان‌دهنده سرمادوست بودن این گیاه است (۳۱). اگرچه میزان معناداری قابل توضیح نبود ( $p\text{-value}=0/09$ ). نتایج سایر پژوهشگران نیز بیانگر این بحث است که مناسب‌ترین درجه دما برای گیاه کلزا دمای کم‌تر است و میانگین سالانه دما کم‌تر از ۹ درجه سلسیوس باعث محدودیت شدید کشت می‌شود (۴۱)، در مطالعه حاضر، مزارع انتخابی با عملکرد کم‌تر از ۱۰۰۰ کیلوگرم در هکتار در نواحی با میانگین دمای بیش‌تر از ۱۹ درجه سلسیوس واقع بودند. این امر موجب تنش گرما در طول و دوره پرشدن دانه است که در نتیجه آن کاهش تعداد و وزن دانه به‌دلیل تسریع پیری در گیاه را به دنبال دارد (۶). بیان شده که دمای بالای ۲۷ درجه سلسیوس موجب عقیمی گیاه کلزا می‌شود که این امر به‌دلیل عدم جوانه‌زنی دانه گرده هست (۵۳ و ۸۳). نتایج پژوهشگران نشان داد که دمای آستانه گیاه کلزا ۲۹ درجه سلسیوس است و حساس‌ترین مراحل به گرما یکتا دو هفته بعد از گلدهی آن است (۵۴).



شکل ۷- اهمیت نسبی تأثیر متغیرهای مستقل مورد استفاده در برآورد عملکرد کلزا در مدل مستخرج از جنگل تصادفی.

Fig. 7. Relative importance of independent variables in estimated crop yield by derived model from Random Forest algorithm.

شده است. سایر نتایج نشان داد که درصد سبز شدن و استقرار گیاه در مزارعی که درصد شن بالایی داشتند، بیش تر است و درصد سبز شدن با افزایش درصد رس خاک کاهش یافت (۸۳). خاک‌های رسی سنگین به دلیل سلب بندی موجب ایجاد فشار فیزیکی بر نشاء و سپس بر سبز شدن کلزا در حین خروج از خاک تأثیر منفی می‌گذارد (۸۱). دشت‌های دامنه‌ای و رودخانه‌ای با خاک عمیق و نفوذپذیر مناسب‌ترین مکان برای کشت کلزا است (۴۱).

یکی دیگر از عوامل مؤثر در عملکرد دانه کلزا در مدل جنگل تصادفی، EC آب زیرزمینی با همبستگی منفی است ( $p\text{-value}=0/0005$ ). در مطالعه‌ای تغییرات عملکرد کلزا در برابر شوری آب آبیاری نشان‌دهنده ارتباط منفی معناداری بود (۸۱) به طوری که نتایج نشان داد که عملکرد دانه به شدت تحت تأثیر شوری قرار گرفت و عملکرد به میزان قابل توجهی کاهش می‌یابد (۸۲)، که در پژوهش‌هایی تأکید بر خطی بودن تابع عملکرد و شوری آب

به دلیل کشت آبی این گیاه، pH آب زیرزمینی یکی دیگر از عوامل مهم در عملکرد مدل شده دانه کلزا در این مطالعه است. pH خاک تحت تأثیر آب آبیاری است و آبیاری با pH بیکربنات و یا کربنات بالا، باعث می‌شود که به مرور زمان pH خاک افزایش یابد (۴۹). نتایج مطالعه حاضر نشان داد که میزان عملکرد دانه با افزایش pH تا حدود ۸ با معناداری ( $p\text{-value}=0/06$ ) همبستگی مثبت و قابل توضیح است. عملکرد گیاه در pH بیش تر از ۸/۵ یا EC بالاتر از ۴ دسی‌زیمنس بر متر کاهش می‌یابد (۸۲). افزایش pH خاک موجب کاهش جذب عناصر ریزمغذی و کاهش عملکرد می‌گردد (۴ و ۵۰).

بافت خاک به عنوان چهارمین عامل مؤثر بر عملکرد کلزا در مدل جنگل تصادفی شناخته شد. نتایج مطالعه حاضر نشان می‌دهد که متوسط عملکرد برآورد شده در خاک‌هایی با بافت متوسط و سبک بیش تر از بافت سنگین است، همچنین کم‌ترین عملکرد در بافت خیلی سنگین پیش‌بینی

آبیاری بود (۴۶ و ۴۸). در این مطالعه متغیرهای تبخیر و تعرق پتانسیل، تابش دریافتی و سیمای توپوگرافی (ارتفاع، شیب و جهت شیب توپوگرافی) از اهمیت کمتری در مدل جنگل تصادفی برخوردار بودند. این امر می‌تواند به دلیل تأثیر این متغیرها بر سایر متغیرهای مؤثر در مدل مورد استفاده باشد. به‌عنوان مثال، ارتفاع و شیب توپوگرافی بر میانگین دما و شیب توپوگرافی بر خصوصیات خاک مؤثر است. همچنین تبخیر و تعرق پتانسیل تحت تأثیر مستقیم دما است به طوری که کمبود آب در مرحله گلدهی و اوایل غلاف‌بندی دانه و نیاز شدید گیاه به آب به‌خاطر افزایش دمای و مصرف بالای روزانه و تبخیر و تعرق گیاه، موجب کاهش فوق‌العاده عملکرد دانه می‌شود (۵۵)؛ اما آنچه عملکرد دانه را صفتی پیچیده می‌سازد، تأثیر مجموعه‌ای از عوامل ساختار ژنتیکی گیاه و شرایط محیطی است (۶۷) که در این مطالعه بر عوامل محیطی آن تأکید شد.

### نتیجه‌گیری کلی

یکی از مشکلات اصلی در تولید گیاهان زراعی وجود شرایط نامناسب محیط کشت از نظر آب، خاک و هوا است. هدف از انجام این پژوهش بررسی تأثیر برخی از عوامل محیطی بر عملکرد کلزا در مزارع منطقه سبزوار است. نتایج مطالعه حاضر نشان داد که روش جنگل تصادفی خطای کمتری نسبت به روش K نزدیک‌ترین همسایه در برآورد عملکرد کلزا در منطقه مورد مطالعه دارد. میانگین درصد خطای مطلق نشان داد که روش K نزدیک‌ترین همسایه با ۲۶ درصد خطا و جنگل

تصادفی با ۱۱ درصد خطا است. مقدار ضریب تبیین ( $R^2$ ) نشان داد که روش K نزدیک‌ترین همسایه با ۶۵ درصد و جنگل تصادفی با ۸۲ درصد قادر به توضیح برآورد عملکرد کلزا با استفاده از متغیرهای مورد استفاده در این پژوهش است. نتایج این بررسی نشان داد که از بین متغیرهای مورد استفاده، دو متغیر شاخص عرضه باد و میانگین دما بیش‌ترین تأثیر را نسبت به سایر متغیرها بر عملکرد دانه کلزا در مدل دارند. مدل مکانی طراحی شده در سامانه اطلاعات جغرافیایی توانایی به‌روزرسانی داده‌های محیطی پویا (هواشناسی و کیفیت آب) که دائماً در حال تغییر هستند را دارد و می‌توان با وارد کردن سایر متغیرهای ایستا مانند خاک و سیمای توپوگرافی، عملکرد دانه کلزا در منطقه با توجه به شرایط همان سال انجام داد. بر اساس نقشه نهایی تهیه‌شده مشخص شد که نواحی مناسب کشت گیاه کلزای در مناطق شمالی و شمال‌غربی شهرستان سبزوار واقع شده است و عملکرد پایین در مناطق مرکزی عمدتاً به دلیل شوری بیش‌ازحد آب به دلیل وجود سازندهای گچی، نمکی در این مناطق و همچنین وجود زهکش کال شور و وجود نمکزارها در این منطقه است. از این رو توصیه می‌شود که برای توسعه کشت این گیاه این مناطق در اولویت کشت قرار گیرد. نتایج این مطالعه نشان می‌دهد که موفقیت در طرح‌های آمایش کشاورزی و همچنین اصلاح الگوی کشت نیاز به اجرای مدل‌های مکانی برآورد عملکرد محصولات به‌منظور شناسایی تفاوت‌های مکانی است.

## منابع

1. Abtew, W. and Melesse, A. 2013. Crop yield estimation using remote sensing and surface energy flux model, evaporation and evapotranspiration: measurements and estimations. Springer Netherlands, Dordrecht, Pp: 161-175.
2. Adab, H. 2007. Modeling of the colza yield by multiple regression technique in gis a case study: sabzevar township, M.Sc. Thesis, Tarbiat Modarres University, Tehran, Iran. (In Persian)
3. Adab, H., Farajzadeh, M., Filekesh, A. and Ismaili, R. 2014. Preparation of autumn rapeseed yield map using perceptron neural network case study: Sabzevar City. Geogr. Space, 13: 41. 171-180. (In Persian)
4. Alhendawi, R.A., Römheld, V., Kirkby, E.A. and Marschner, H. 1997. Influence of increasing bicarbonate concentrations on plant growth, organic acid accumulation in roots and iron uptake by Barley, Sorghum, and Maize. J. Plant Nutr. 20: 12. 1731-1753.
5. Amirabadizadeh, H. 1997. Plant coverage in sabzevar region, Forest and Rangeland Research Institute of Ministry of Jihad-e Sazandegi, Tehran. Pp: 21-31. (In Persian)
6. Angadi, S., Cutforth, H., McConkey, B. and Gan, Y. 2003. Yield adjustment by canola grown at different plant populations under semiarid conditions. Crop Sci. 43: 4. 1358-1366.
7. Azarakhshi, M., Farzadmehr, J., Eslah, M. and Sahabi, H. 2013. An investigation on trends of annual and seasonal rainfall and temperature in different climatologically regions of iran. J. Range Water. Manage. 66: 1. 1-16.
8. Bairagi, G.D. and Hassan, Z.U. 2002. Wheat crop production estimation using satellite data. J. Ind. Soc. Rem. Sens. 30: 4. 213-219.
9. Bakhsh, A., Kanwar, R.S. and Malone, R.W. 2007. Role of landscape and hydrologic attributes in developing and interpreting yield clusters. Geoderma, 140: 3. 235-246.
10. Balakrishnan, P., Saleem, A. and Mallikarjun, N. 2011. Groundwater quality mapping using geographic information system (GIS): a case study of gulbarga city, karnataka, india. Afric. J. Environ. Sci. Technol. 5: 12. 1069-1084.
11. Baniyasi, M., Babaei, Gh.R., Zeraati, H.A. and Memari, F. 2007. Application of bootstrap open sampling method in logistic regression and its application in analyzing data related to patients with breast cancer. Quarterly J. Sch. Publi. Health Inst. Publi. Health Res. 4: 1. 9-18. (In Persian)
12. Böhner, J. and Antonic, O. 2009. Chapter 8 land-surface parameters specific to topo-climatology. In: H. Tomislav and I.R. Hannes (Editors), Developments in soil science. Elsevier, Pp: 195-226.
13. Böhner, J., McCloy, K.R. and Strobl, J. 2006. SAGA-analysis and modelling applications, University of Goettingen.
14. Breiman, L. 2001. Random Forests. Mach. Learn. 45: 1. 5-32.
15. Choong, J. 2012. Powerful forecasting with ms excel. Kindle Edition.
16. Crane, E. 1972. Bees in the pollination of seed crops. Bee Res. Assoc. 163p.
17. Delaplane, K.S., Mayer, D.R. and Mayer, D.F. 2000. Crop Pollination by Bees. Cabi, 332p.
18. Dehshiri, A. and Modares-Sanavy, S.A.M. 2017. Effects of salinity on yield quantity and quality of three rapeseed (*brassic napus*) cultivars under different atmospheric carbon dioxide concentrations. J. Crop Prot. 9: 4. 1-16.
19. Ebrahimi-Pak, N. 2010. Determination on potential evapotranspiration canola using lysimeter method. Proceedings of 10th National Seminar on Irrigation and Evapotranspiration. Shahid Bahonar University of Kerman, 7p. (In Persian)
20. Edmeades, G.O., Bolanos, J., Lafitte, H.R. and Rajaram, S. 1989. Traditional approaches to breeding for drought resistance in cerea. 1435-0653, ICSU and CAB1, Wallingford, UK.
21. Efron, B. 1981. Nonparametric standard errors and confidence intervals. Can. J. Stat. 9: 2. 139-158.

22. Farajzadehasl, M. and Zarrin, A. 2002. Rainfed wheat yield modeling according to agrometeorological parameters in west azarbayjan provinde. *Modares J.* 6: 71-96. (In Persian)
23. Farshad, M.D. and Sadeh, J. 2014. Short-circuit fault location in high-voltage direct current transmission lines using a generalized regression neural network and random forest algorithm. *Computational Intel. Elect. Eng. Smart Sys. Elec. Eng.* 4: 2. 1-14.
24. Ghafouri-Kesbi, F., Rahimi Mianji, Gh., Honarvar, M. and Nejati, J.A. 2017. The regulation and application of random forest algorithm in genomic evaluation. *Res. Anim. Prod.* 7: 13. 185-178. (In Persian)
25. Ghorbani, Kh. and Soltani, A. 2014. The effect of climate change on soybean yield in gorgan. *J. Plant Prod. Res.* 21: 2. 67-85. (In Persian)
26. Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. and Toulmin, C. 2010. Food security: the challenge of feeding 9 billion people. *Sci.* 327: 5967. 812-818.
27. Hanjra, M.A. and Qureshi, M.E. 2010. Global water crisis and future food security in an era of climate change. *Food Policy*, 35: 5. 365-377.
28. Häntzschel, J., Goldberg, V. and Bernhofer, C. 2005. GIS-based regionalisation of radiation, temperature and coupling measures in complex terrain for low mountain ranges. *Meteorol. Appl.* 12: 1. 33-42.
29. Hatamvand, M., Hasanloo, T., Dehghan Nayeri, F., Shiranirad, A.H., Tabatabaei, S.A. and Hosseini, S.M. 2015. Evaluation of some physiological and biochemical indices of canola cultivars in response to drought stress. *J. Environ. Stresses Crop Sci.* 7: 2. 173-185.
30. Hodgson, A. 1979. Rapeseed adaptation in Northern New South Wales. III.\* Yield, yield components and grain quality of *Brassica campestris* and *Brassica napus* in relation to planting date. *Austr. J. Agric. Res.* 30: 1. 19-27.
31. Honar, T., Sabet, S.A., Kamgar, H.A. and Shams, S. 2011. Calibration of crop system model for growth simulation and yield estimation of canola. *J. Water Soil.* 25: 3. 593-605. (In Persian)
32. Honar, T., Sabet-Sarvestani, A., Sepaskhah, A., Kamgar-Haghighi, A. and Shams, S. 2012. Simulation of soil water content and yield of canola using CRPSM. *J. Sci. Technol. Agric. Nat. Resour.* 16: 59. 45-57. (In Persian)
33. Hosaini, M.T., Siosemarde, A., Fathi, P. and Siosemarde, M. 2008. Application of artificial neural network (ann) and multiple regression for estimating assessing the performance of dry farming wheat yield in Ghorveh Region, Kurdistan Province. *Agric. Res.* 7: 1. 41-54. (In Persian)
34. Iqbal, J., Read, J.J., Thomasson, A.J. and Jenkins, J.N. 2005. Relationships between soil-landscape and dryland cotton lint yield. The National Aeronautical and Space Administration-funded Remote Sensing Technology Center at Mississippi State University (NASA grant number NCC13-99001). *Soil Sci. Soc. Amer. J.* 69: 3. 872-882.
35. Jan, Z., Abrar, M., Bashir, S. and Mirza, A.M. 2009. Seasonal to inter-annual climate prediction using data mining knn technique. In: D.M.A. Hussain, A.Q.K. Rajput, B.S. Chowdhry and Q. Gee (Editors), *Wireless Networks, Information Processing and Systems: International Multi Topic Conference, IMTIC 2008 Jamshoro, Pakistan, April 11-12, 2008 Revised Selected Papers.* Springer Berlin Heidelberg, Berlin, Heidelberg, Pp: 40-51.
36. Jay, P. 2000. *Weather and yield.* Iranian Students Booking Agency. Translated by Kafi, M. Ganjali, M. Nezami, A. and Shariatmadar, F.: 311p. (In Persian)
37. Jey, P. 2014. *Climate and yield of crops.* J. Jihad University, Mashhad, 311p. (In Persian)
38. Kamkar, B., Dorri, M.A. and Teixeira, J.A. 2014. Assessment of land suitability and the possibility and performance of a canola (*Brassica Napus* L.) – Soybean



- (*Glycine Max* L.) Rotation in Four Basins of Golestan Province, Iran. The Egypt. J. Rem. Sens. Space Sci. 17: 1. 95-104.
39. Kandiannan, K., Karthikeyan, R., Krishnan, R., Kailasam, C. and Balasubramanian, T.N. 2002. A crop-weather model for prediction of rice (*Oryza sativa* L.) yield using an empirical-statistical technique. J. Agron. Crop Sci. 188: 1. 59-62.
40. Karimbeigi, H., Nazarian-Firouzabadi, F., Khademi, M. and Mousav, E. 2016. Assessment of genetic diversity among some oilseed rape (*Brassica napus* L.) plants, using single sequence repeats (SSR) molecular markers. J. Plant Genet. Res. 3: 1. 45-56. doi:10.29252/pgr.3.1.45.
41. Kazemi, H. 2014. Agroecological zoning of gorgan agricultural lands for hullless barley cropping based on boolean logic. J. Crop Prod. 6: 4. 165-185. (In Persian)
42. Khorshid Dost, E.M., Hoseini, A.V. and Pour, K. 2012. Suitable Areas for canola cultivation in Kurdistan province using GIS. Water Soil Knowl. 21: 3. 48-37. (In Persian)
43. Lall, U. and Sharma, A. 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour. Res. 32: 3. 679-693.
44. Lashkari, H., Keikhosravi, B.C. and Ghiyasi, D. 2011. Modeling of topography effects on estimation of potential evapotranspiration using Geographic Information System (Case study- Sabzevar, Iran. Earth Sci. Res. 1: 2. 87-102. (In Persian)
45. Liaw, A. and Wiener, M. 2002. Classification and regression by randomforest. R news, 2: 3. 18-22.
46. Lobell, D.B. 2013. The use of satellite data for crop yield gap analysis. Field Crops Res. 143: 56-64.
47. Loeppert, R.H. 1986. Reactions of iron and carbonates in calcareous soils. J. Plant Nutr. 9: 3-7. 195-214.
48. Maas, E. 1986. Salt tolerance of plants. Applied Agric. Res. (USA).
49. Maas, E.V. and Hoffman, G. 1977. Crop salt tolerance-current assessment. J. Irrig. Drain. Div. 103: 2. 115-134.
50. Mathers, H. 2003. What you should know about water, Organ State University, USA.
51. Matlabefard, R., Mjmalakouti, M.J. and Kafee, M. 2009. The effect of pH of irrigation on quantity and quality characteristics of dianthus *caryophyllus* L. Soil Water Sci. 22: 1. 103-111. (In Persian)
52. Mazhari, M. and Parsapoor, K. 2012. Factors influencing the adoption of rapeseed cultivation (Case study of Khorasan Razavi Province). J. Econ. Agric. Dev. 25: 4. 410-419.
53. Mo, X., Liu, S., Lin, Z., Xu, Y., Xiang Y. and McVicar, T.R. 2005. Prediction of crop yield, water consumption and water use efficiency with a svat-crop growth model using remotely sensed data on the north china plain, Ecol. Model. 183: 2-3. 301-322.
54. Morrison, M. 1993. Heat stress during reproduction in summer rape. Can. J. Bot. 71: 2. 303-308.
55. Morrison, M.J. and Stewart, D.W. 2002. Heat stress during flowering in summer brassica. Crop Sci. 42: 3. 797-803.
56. Najafi, A.B. 2010. Evaluation of the effect of climatic parameters on canola in Pars-Abad, Iran to Provide agricultural calendar in the year of 2008/2009, M.Sc. Thesis. University of Ardabil. (In Persian)
57. Nellemann, C., MacDevette, M., Manders, T., Eickhout, B., Svihus, B., Prins, A.G. and Kaltenborn, B.P. 2009. The environmental food crisis: the environment's role in averting future food crises: A UNEP Rapid Response Assessment. UNEP/Earthprint, Norway, 101p.
58. Noori, K. and Jahan Nama, F. 2008. Study of comparative advantage of spring soybean in iran. Research Construct. Agron. Hort. 21: 2. 26-35. (In Persian)
59. Olsson, G. 1955. Wind pollination of cruciferous oil plants. Sverig. Utsadesforen. Tidskr, 65: 6. 418-422.

60. Omidvar, K. and Dastmoradi, S. 2014. Study of the relationship between climatic elements and canola performance in Kermanshah Province, First National Conference on Climatology. University of Advanced Industrial Technology and Advanced Technology, Kerman, Pp: 32-53. (In Persian)
61. Oshiro, T.M., Perez, P.S. and Baranauskas, J.A. 2012. How many trees in a random forest? International Workshop on Machine Learning and Data Mining in Pattern Recognition. Springer, Pp: 154-168.
62. ÖZER, H., Oral, E. and Dogru, U. 1999. Relationships between yield and yield components on currently improved spring rapeseed cultivars. Turk. J. Agric. For. 23: 6. 603-608.
63. Ozkan, B. and Akcaoz, H. 2002. Impacts of climate factors on yields for selected crops in the southern turkey. Mitigation Adapt. Strateg. Glob. Chang. 7: 4. 367-380.
64. Parida, A.K. and Das, A.B. 2005. Salt tolerance and salinity effects on plants: a review. Ecotoxicol. Environ. Saf. 60: 3. 324-349.
65. Pasban-Eslam, B. 2013. Evaluation of seed and oil yields and their components and relationships in oilseed rape genotypes under east azarbaijan conditions in iran. Crop Breed. J. 3: 1. 53-59. (In Persian)
66. Reganold, J.P., Papendick, R.I. and Parr, J.F. 1990. Sustainable agriculture. Sci. Amer. 262: 6. 112-120.
67. Riffkin, P., Potter, T. and Kearney, G. 2012. Yield performance of late-maturing winter canola (*Brassica Napus* L.) Types in the high rainfall zone of southern australia. Crop Pasture Sci. 63: 1. 17-32.
68. Sadat Seyed Mohammadi, N., Allahdadi, A., Seyed Mohammadi, E., Sarafraz, E. 2013. The variety of some physiological and morphological characteristics of spring rapeseed varieties under irrigation intervals and regimes. Physiol. Plant. 4: 16. 5-17. (In Persian)
69. Sadeh, J., and Farshad, M. 2014. Fault locating in hvdc transmission lines using generalized regression neural network and random forest algorithm. Comput. Intel. Electr. Eng. 4: 2. 1-14. (In Persian)
70. Sairam, R.K. and Srivastava, G.C. 2002. Changes in antioxidant activity in sub-cellular fractions of tolerant and susceptible wheat genotypes in response to long term salt stress. Plant Sci. 162: 6. 897-904.
71. Savin, I.Y., Stathakis, D., Negre, T. and Isaev, V.A. 2007. Prediction of crop yields with the use of neural networks. Russ. Agric. Sci. 33: 6. 361-363.
72. Serele, C.Z., Gwyn, Q.H.J., Boisvert, J.B., Pattey, E., McLaughlin, N. and Daoust, G. 2000. Corn yield prediction with artificial neural network trained using airborne remote sensing and topographic data. Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE. International. IGARSS Honolulu (USA), Pp: 384-386.
73. Singh, H., Singh, K.P., Jarwal, S.D., Singh, T., Tonk, D.S. and Faroda, A.S. 1989. Water production function for indian rape. J. Oilseeds Res. 6: 2. 316-321.
74. Sudduth, K.A., Drummond, S.T., Birrell, S.J. and Kitchen, N.R. 1996. Analysis of spatial factors influencing crop yield, precision agriculture. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, Pp: 129-139.
75. Sun, R. and Zhang, B. 2016. Topographic effects on spatial pattern of surface air temperature in complex mountain environment. Environ. Earth Sci. 75: 7. 621.
76. Tachikawa, T., Hato, M., Kaku, M. and Iwasaki, A. 2011. Characteristics of aster gdem version 2. Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, Pp: 3657-3660.
77. Tikle, S., Saboori, M.J. and Sankpal, R. 2012. Spatial distribution of ground water quality in some selected parts of Pune City, Maharashtra, India using GIS. Curr. World Environ. 7: 2. 281-286.
78. van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K. and van der Werf, M.J. 2006. Centering,

- scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7: 142-142.
79. Wadsworth, R.M. 1959. An optimum wind speed for plant growth. *Annals of Botany*, 23: 1. 195-199.
80. Williams, I.H., Martin, A. and White, R. 1987. The effect of insect pollination on plant development and seed production in winter oil-seed rape (*brassica napus* L.). *J. Agric. Sci.* 109: 01. 135-139.
81. Wu, X., He, J., Yip, T. and Zhang, P. 2015. A two-stage random forest method for short-term load forecasting. *PowerTech, 2015 IEEE Eindhoven. IEEE*, Pp: 1-6.
82. Yanegh, A. and Khajeh-hosseini, M. 2014. Effects of field conditions on emergence of oilseed rape seed lots grown in Khorasan Province. *Iranian J. Field Crops Res.* 12: 1. 9-16. (In Persian)
83. Yazdani, V., Davari, K., Ghahreman, B. and Kafi, M. 2015. Assesment of the water-salinity production function models-canola application in the Mashhad Area. *J. Irrig. Water Eng.* 5: 18. 32-53. (In Persian)
84. Young, L.W., Wilen, R.W. and Bonham-Smith, P.C. 2004. High temperature stress of brassica napus during flowering reduces micro-and megagametophyte fertility, induces fruit abortion, and disrupts seed production. *J. Exp. Bot.* 55: 396. 485-495.
85. Zevenbergen, L.W. and Thorne, C.R. 1987. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* 12: 1. 47-56.
86. Zhao, Y. and Zhang, Y. 2008. Comparison of decision tree methods for finding active objects. *Adv. Space Res.* 41: 12. 1955-1959.

