

## شناسایی ویژگی‌های مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای با استفاده از روش‌های

### انتخاب ویژگی

مسعود پورغلام آمیجی، خالد احمدآلی<sup>۱</sup> و عبدالمجید لیاقت

دانشجوی دکتری گروه مهندسی آبیاری و آبادانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. Mpourgholam6@ut.ac.ir

استادیار گروه احیاء مناطق خشک و کوهستانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. Khahmadauli@ut.ac.ir

استاد، گروه مهندسی آبیاری و آبادانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. Aliaghat@ut.ac.ir

دریافت: دی ۱۴۰۱ و پذیرش: اسفند ۱۴۰۱

#### چکیده

این پژوهش با هدف انتخاب ویژگی‌های مهم برای مدل‌سازی هزینه سامانه‌های آبیاری تحت فشار با استفاده از داده‌های ۵۱۵ پروژه آبیاری قطره‌ای در چهار بخش شامل هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (TCP)، هزینه لوازم داخل مزرعه (TCF)، هزینه نصب و اجرای داخل مزرعه و ایستگاه پمپاژ (TCI) و هزینه کل (TCT) انجام شد. در مرحله اول بانک اطلاعاتی شامل ۳۹ متغیر تأثیرگذار در هزینه بخش‌های یادشده، تهیه و قیمت تمام پروژه‌ها (۱۳۸۵ تا ۱۳۹۸) برای سال پایه ۱۴۰۰ به‌روزرسانی شد. سپس انتخاب ویژگی با الگوریتم‌های مختلف در محیط MATLAB و در دو بخش شامل (۱) کل ویژگی‌ها (ویژگی‌های قبل از طراحی و ویژگی‌های بعد از آن شامل ۳۹ ویژگی) و (۲) ویژگی‌های قبل از مرحله طراحی (شامل ۱۸ ویژگی) انجام شد. نتایج انتخاب ویژگی نشان داد که مقادیر RMSE و  $R^2$  برای بخش کل ویژگی‌ها به ترتیب برابر با ۰/۰۰۷ و ۰/۹۲ و برای بخش ویژگی‌های قبل از طراحی به ترتیب برابر ۰/۰۰۳ و ۰/۸۹ است. از بین الگوریتم‌های مختلف برای انتخاب ویژگی، ماشین بردار پشتیبان (SVM) و الگوریتم‌های بهینه‌سازی (Wrapper) به ترتیب عنوان بهترین یادگیرنده و روش انتخاب ویژگی شناسایی شدند. نتایج معیارهای ارزیابی نشان داد که دو الگوریتم FOA و LCA برآورد مناسبی را به دست دادند و معیار خطای آن در بخش کل ویژگی‌ها به ترتیب ۰/۰۰۲۰ و ۰/۰۰۱۸ و همبستگی آن ۰/۹۴ و ۰/۹۴ به دست آمد. در بخش ویژگی‌های قبل از طراحی نیز این معیارها به ترتیب ۰/۰۰۰۶ و ۰/۹۵ و برای دو الگوریتم بود. در نهایت در بخش کل ویژگی‌ها، ۱۰ مورد از ۳۹ ویژگی و در بخش ویژگی‌های قبل از طراحی، ۸ مورد از ۱۸ ویژگی به‌عنوان مؤثرترین ویژگی‌ها انتخاب شد. نتایج انتخاب مؤثرترین ویژگی‌ها که بر هزینه بخش‌های مختلف سامانه آبیاری قطره‌ای اثرگذارند، می‌تواند مدل‌سازی هزینه سامانه‌ها را ساده‌تر و سریع‌تر کرده و ضمن کاربرد در کارهای پژوهشی، در عمل نیز برآورد و مدیریت هزینه‌ها را قبل از طراحی و اجرای این طرح‌ها ممکن کند.

واژه‌های کلیدی: مدل‌سازی اقتصادی، سامانه‌های نوین آبیاری، تحلیل حساسیت، شناخت الگو، الگوریتم‌های فراکاوشی

<sup>۱</sup> - آدرس نویسنده مسئول: گروه احیاء مناطق خشک و کوهستانی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

کشور است. در مجموع طبق آمار ارائه شده در آمارنامه‌های وزارت جهاد کشاورزی، از سال ۱۳۶۹ تا پایان سال ۱۳۹۸ مجموعاً ۱۹۸۷۵۴۸ هکتار سامانه نوین آبیاری تحت فشار در کشور اجرا شده است. اگرچه روند توسعه کمی سامانه‌های نوین آبیاری تحت فشار به علت نوسانات اقتصادی و معادلات تجاری دست‌خوش تغییراتی شده است، اما در مجموع روند کلی سطوح اجرا شده در کشور صعودی است. طبق بررسی‌ها چنانچه آمار منتهی به سال ۱۴۰۰-۱۳۹۹ اجرا شده به وسیله معاونت آب‌و خاک و دیگر معاونت‌ها در نظر گرفته شود، سطوح اجرا شده به بیش از ۲/۴ میلیون هکتار افزایش یافته است (کیانی و شاکر، ۲۰۲۲).

انتخاب ویژگی (FS<sup>۱</sup>) و پردازش داده‌ها یک مؤلفه اساسی بسیاری از مشکلات طبقه‌بندی، مدل‌سازی و رگرسیون است زیرا برخی داده‌ها اثر یکسانی دارند، برخی دارای اثر گمراه‌کننده و برخی دیگر تأثیری در طبقه‌بندی یا مشکلات رگرسیون ندارند و بنابراین انتخاب اندازه بهینه و حداقل برای ویژگی‌ها می‌تواند مفید باشد (چاندرشکار و ساهین، ۲۰۱۴؛ میائو و نیو، ۲۰۱۶؛ سولوریو-فرناندز و همکاران، ۲۰۲۰؛ التریچی و همکاران، ۲۰۲۳). از آنجایی که انتخاب ویژگی به‌عنوان یکی از مهم‌ترین مشکلات پردازش داده‌ها، یک موضوع تحقیق مهم و به‌روز در تشخیص الگو (PR<sup>۲</sup>)، یادگیری ماشین (ML<sup>۳</sup>) و داده‌کاوی (DM<sup>۴</sup>) است، لذا بسیار پرکاربرد بوده و پروژه‌ها زمانی ارزشمند خواهند بود که از تکنیک انتخاب ویژگی و آنالیز حساسیت (SA<sup>۵</sup>) در آن‌ها استفاده شود (متل و همکاران، ۲۰۲۲). با توسعه ذخیره‌سازی اطلاعات، داده‌های ورودی دارای تعداد زیادی ویژگی هستند که ممکن است شامل انبوهی از ویژگی‌های بی‌ربط و یا کم‌اهمیت باشند. ویژگی‌های غیرضروری اغلب منجر به راندمان پایین یادگیری الگوریتم و مشکل می‌شوند؛

در کشور ایران کمتر از ۳۰ درصد اراضی فاریاب تحت پوشش سامانه‌های نوین آبیاری قرار گرفته‌اند (نعلبندان و همکاران، ۲۰۲۳) و این موضوع، اهمیت بررسی نقاط ضعف و قوت، تعیین عوامل اثرگذار بر هزینه‌های سامانه‌های مختلف آبیاری و سپس مدل‌سازی اقتصادی سامانه‌ها و به‌ویژه برآورد هزینه یک سامانه آبیاری پیش از احداث را نشان می‌دهد. از سویی یکی از مهم‌ترین نیازهای متولیان صنعت آب و همچنین کارفرمایان، مشاوران و پیمانکاران پیش‌بینی هزینه‌های پروژه‌ها در مرحله اولیه طراحی است. از آنجایی که توسعه سامانه‌های مختلف آبیاری تحت فشار جزء سیاست‌های راهبردی وزارت جهاد کشاورزی است، بنابراین داشتن دانش، اطلاعات و پیش‌آگاهی از عوامل مؤثر بر هزینه‌های یک سامانه آبیاری در مناطق مختلف و قبل از اجرا، کمک مؤثری به مدیریت هزینه خواهد نمود. همچنین با توجه به وسعت قابل توجه اراضی تحت پوشش آبیاری تحت فشار و همچنین پتانسیل توسعه آن در کشور، این شناسایی ویژگی‌های تأثیرگذار بر هزینه و سپس مدل‌سازی آن، در بودجه‌بندی سالانه کشور نیز نقش بسزایی ایفا خواهد کرد و توجه به آن از اهم امور است؛ بنابراین نداشتن آگاهی کامل از هزینه تمام‌شده پروژه در ابتدای کار، وجه اجتناب‌ناپذیر بیشتر پروژه‌های توسعه و ساختمانی بوده است، به‌گونه‌ای که زبده‌ترین مجریان و پیمانکاران نیز در این رابطه با مشکلاتی مواجه هستند. این‌گونه تخمین‌ها به مالکان و طراحان امکان‌پذیری انجام پروژه با اطمینان بالاتر و نیز کنترل هزینه‌های مؤثر در طراحی و اجرای پروژه را ارائه می‌دهد (پورغلام آمیجی و همکاران، ۲۰۲۱). به‌عنوان یک نتیجه‌گیری کلی در مورد روند توسعه سامانه‌های آبیاری تحت فشار در برنامه‌های اول تا ششم توسعه کشور می‌توان مطرح کرد که مسائل اقتصادی و در واقع پرداخت تسهیلات از سوی دولت برای اجرای این سامانه‌ها به‌عنوان مهم‌ترین و همچنین تأثیرگذارترین عامل تغییر روند در مساحت سامانه‌های آبیاری اجرا شده در

<sup>1</sup> -Feature Selection

<sup>2</sup> -Pattern Recognition

<sup>3</sup> -Machine Learning

<sup>4</sup> -Data Mining

<sup>5</sup> -Sensitivity Analysis

ورودی در توسعه مدل II دخیل بودند. یافته‌ها نشان داد که دو مدل ANN به‌طور خوبی در مرحله آموزش یاد گرفته‌اند و در مرحله آزمون قابلیت تعمیم خوبی به دست آوردند. مدل I و II به ترتیب به میانگین درصد دقت ۷۹/۳ و ۸۲/۲ درصد دست یافتند. همچنین ونکتاچالم (۱۹۹۳) در برآورد هزینه نرم‌افزار با استفاده از شبکه‌های عصبی مصنوعی به چنین نتایجی دست یافتند.

مطالعه‌ای توسط کیم (۲۰۱۳) با هدف مقایسه نکات مثبت و منفی استفاده از روش‌های مختلف برآورد هزینه ساخت از جمله تحلیل رگرسیون، ANN و ماشین بردار پشتیبان ( $SVM^2$ )، در برآورد هزینه ساخت ساختمان مدارس انجام شد و نتیجه گرفتند که مدل ANN برآورد دقیق‌تری نسبت به مدل‌های تحلیل رگرسیون و ماشین بردار پشتیبانی ایجاد کرده و نتیجه مدل برآورد تحلیل رگرسیون نیز مناسب‌تر از مدل برآورد SVM بوده است. الفکی و همکاران (۲۰۱۴) نیز یک بررسی ۱۰ ساله از فنون هوشمند در برآورد هزینه پروژه ساخت‌وساز انجام دادند و بر قابلیت بالایی ANN در شناخت، بررسی، وزن‌دهی معیارهای مهم و در نهایت تخمین اولیه و نهایی پروژه‌ها تأکید کردند. در همین رابطه، آهیاگا-داگبوی و اسمیت (۲۰۱۲) از ANN برای مدل‌سازی هزینه نهایی پروژه‌های آب استفاده کردند. بدین منظور، از داده‌های مربوط به ۹۸ پروژه ساخت‌وساز مربوط به آب در بین سال‌های ۲۰۰۷ تا ۲۰۱۱ در اسکاتلند استفاده شد. مدل‌های جداگانه هزینه برای هزینه‌های هدف نرمالیزه شده و ورود به هزینه‌های هدف توسعه داده شد. سپس وزن متغیرها تعیین و به‌منظور بهبود عملکرد مدل نهایی مورد بررسی قرار گرفت. به‌عنوان نمونه اولیه یک تحقیق گسترده، عملکرد مدل نهایی بسیار رضایت‌بخش بود و نتایج حاکی از توانایی بالای ANN در ثبت تعامل بین متغیرهای پیش‌بینی‌کننده و هزینه نهایی بود. همچنین روکساس و اونگ‌پنگ (۲۰۱۴) از رویکرد شبکه عصبی مصنوعی برای برآورد هزینه ساختاری پروژه‌های

بنابراین، انتخاب ویژگی‌های مرتبط و لازم برای یک کار یادگیری معین، یک قدم بسیار مهم است (احمدآلی و همکاران، ۲۰۱۳؛ لیو و همکاران، ۲۰۱۷؛ پازوکی و همکاران، ۲۰۲۰؛ تالوکدر و همکاران، ۲۰۲۲؛ تاخار و لوهیا، ۲۰۲۳).

استفاده از فنون و فرمول‌هایی که با استفاده از یک سری اطلاعات اولیه از طرح مورد نظر بتواند هزینه آن را برآورد کند، کمک شایانی به شرکت‌های مهندسی مشاور فعال در مهندسی آب خواهد نمود و می‌تواند به‌عنوان یک دستورالعمل جامع برای برآورد هزینه‌های احداث سامانه‌های آبیاری تحت فشار قبل از اجرای آن مورد استفاده قرار گیرد (پورغلام آمیچی و همکاران، ۲۰۲۱a)؛ بنابراین برآورد هزینه یک کار مبتنی بر تجربه است که شامل ارزیابی شرایط ناشناخته و روابط پیچیده عوامل مؤثر بر هزینه است. شبکه عصبی مصنوعی ( $ANN^1$ )، یک فرآیند مبتنی بر قیاس است که به بهترین وجه برای برآورد هزینه مناسب است. مزایای اصلی ANN شامل توانایی آن‌ها در یادگیری با مثال (پروژه‌های گذشته) و تعمیم راه‌حل‌ها برای برنامه‌های کاربردی آینده (پروژه‌های آینده) است (آرورا و میسرا، ۲۰۱۸؛ مولک، ۲۰۲۱). به‌طورکلی مکانیزمی که فناوری‌های پیشرو را هدایت کرده و باعث ظهور ابزارهای نوآورانه به‌ویژه در بخش کشاورزی می‌شود، یادگیری ماشین (ML) است (آرورا و میسرا، ۲۰۱۷؛ شارما و همکاران، ۲۰۲۰؛ تسکین و همکاران، ۲۰۲۲). در همین باره، الهاگ و بوساباین (۱۹۹۸) به بررسی یک سیستم عصبی مصنوعی برای برآورد هزینه پروژه‌های ساخت‌وساز پرداختند. در این مطالعه، دو مدل ANN برای پیش‌بینی پایین‌ترین قیمت مناقصه ساختمان‌های مدارس ابتدایی و متوسطه توسعه یافت. ۳۰ پروژه در این بررسی شرکت داشتند و داده‌های مربوط به آن‌ها از پایگاه داده BCIS استخراج شد. در مدل I از ۱۳ ویژگی تعیین‌کننده هزینه استفاده شد اما در مقابل تنها چهار متغیر

<sup>2</sup>-Support Vector Machine

<sup>1</sup>-Artificial Neural Networks

مصنوعی برای برآورد کاربردی‌ترین فنون کنترل هزینه پس از قرارداد (PCCTs<sup>2</sup>) در پروژه‌های ساختمانی ارائه کردند. هدف از این مطالعه پیشنهاد یک روش ساختاریافته پشتیبانی تصمیم برای برآورد کاربردی‌ترین PCCTها با استفاده از ANN بود و بدین منظور از داده‌های ۱۳۵ نمونه استفاده شد. ابزاری بودن شبکه‌های عصبی مصنوعی در این مطالعه، توسعه یک متدولوژی پشتیبانی تصمیم ساختاریافته را برای تجزیه و تحلیل مناسب‌ترین PCCTها برای استقرار در مراحل مختلف فرآیند ساخت و ساز امکان‌پذیر کرد و معیارهای RMSE برابر ۰/۰۷۳ و RSquare برابر با ۰/۷۲۶ در مرحله صحت‌سنجی حاصل شد.

بنابراین روش‌های مختلفی برای تخمین هزینه‌ها در دسترس است. با افزایش قدرت محاسبات، اکنون تمایل به استفاده از روش‌های مبتنی بر یادگیری ماشین (ML) مانند شبکه‌های عصبی مصنوعی (ANN)، منطق فازی (FL<sup>3</sup>)، یادگیری عمیق (DL<sup>4</sup>)، الگوریتم ژنتیک (GA<sup>5</sup>) شامل برنامه‌ریزی بیان ژن (GEP<sup>6</sup>) و برنامه‌ریزی ژنتیک (GP<sup>7</sup>) و غیره برای برآورد دقیق‌تر مدت و هزینه‌های پروژه وجود دارد که می‌تواند با وجود جزئیات ناکافی در مرحله اولیه و با داده‌های کم اما همچنان قابل اعتماد باشند و روابط غیرخطی بین عوامل هزینه و هزینه پروژه را شناسایی کنند (چنگ و همکاران، ۲۰۱۰؛ بابایی و همکاران، ۲۰۲۲). درحالی‌که استفاده از ANN برای برآورد هزینه از دید پیمانکاران به وفور مورد بررسی قرار گرفته است، مطالعات محدودی در مورد توسعه و کاربرد روش‌های مبتنی بر ML برای شرکت‌های مهندسی مشاور وجود دارد. با توجه به اینکه ماهیت محصولات / خدمات ارائه‌شده توسط شرکت‌های مهندسی مشاور ذاتاً با پیمانکاران متفاوت است (آن‌ها انتزاعی‌تر و دارای ماده کمتری هستند) و همچنین با توجه به اینکه نوع و سطح

ساختمانی در فیلیپین استفاده کردند. هدف از این مطالعه، توسعه یک مدل شبکه عصبی مصنوعی (ANN) بود که می‌تواند کل هزینه ساختاری پروژه‌های ساختمانی در فیلیپین را پیش‌بینی کند. در این مطالعه از داده‌های ۳۰ پروژه ساختمانی استفاده شد. نتایج حاکی از آن بود که مدل ANN حاصل‌شده، به‌طور منطقی کل هزینه ساختاری پروژه‌های ساختمانی را با نتایج مطلوب مرحله آموزش و آزمایش پیش‌بینی کرد. در پژوهشی یاداو و همکاران (۲۰۱۶) نیز یک مدل برآورد هزینه (CEM<sup>1</sup>) را با استفاده از مدل شبکه عصبی مصنوعی (ANN) توسعه دادند که قادر به پیش‌بینی کل هزینه ساختاری مجتمع‌های مسکونی با در نظر گرفتن پارامترهای مختلف است. در این پژوهش، داده‌های ۲۳ سال اخیر جمع‌آوری شد. مدل ANN حاصل به‌طور منطقی کل هزینه ساختاری پروژه‌های ساختمانی را با فاکتور همبستگی برابر ۰/۹۹۶۰ و R<sup>2</sup> برابر با ۰/۹۹۰۵ پیش‌بینی کرد که نتایج فاز آموزشی و آزمایشی مطلوبی را ارائه می‌دهد.

برآورد هزینه ساخت و ساز پروژه‌های ساختمانی در مراحل اولیه با درجه دقت بالاتر، نقشی حیاتی در موفقیت هر پروژه ساختمانی دارد. به همین منظور، برآورد هزینه ساخت ساختمان با استفاده از شبکه‌های عصبی مصنوعی توسط چاندانشیو و کامبکار (۲۰۱۹) بررسی شد. بر اساس بررسی و بازخورد متخصصان طراحی و پیمانکاران ساخت و ساز، مجموعه داده‌ای از ۷۸ پروژه ساخت و ساز ساختمان از شهر بزرگ بمبئی (هند) و منطقه جغرافیایی نزدیک به آن جمع‌آوری شد. تأثیرگذارترین پارامترهای طراحی هزینه ساخت و ساز ساختمان‌ها به‌عنوان ورودی شناسایی و اختصاص داده شد و هزینه کل اسکلت سازه نشان‌دهنده خروجی مدل‌های شبکه عصبی بود. نتایج به‌دست‌آمده از مدل شبکه عصبی آموزش‌دیده نشان داد که قادر به برآورد هزینه پروژه‌های ساختمانی در مراحل اولیه ساخت و ساز بوده است. در تحقیق دیگری، اوموتایو و همکاران (۲۰۲۰) یک رویکرد شبکه عصبی

<sup>2</sup> -Post-Contract Cost Controlling Techniques

<sup>3</sup> -Fuzzy Logic

<sup>4</sup> -Deep Learning

<sup>5</sup> -Genetic Algorithm

<sup>6</sup> -Gene Expression Programming

<sup>7</sup> -Genetic Programming

<sup>1</sup> -Cost Estimation Model

اتوکید و فایل اکسل مربوط به محاسبات طراحی استخراج و هزینه سامانه‌ها به دو بخش کلی هزینه ایستگاه پمپاژ و هزینه مزرعه دسته‌بندی شد. انواع متغیرهای کاندید برای ورودی مدل برآورد هزینه‌های سامانه آبیاری قطره‌ای شامل متغیرهای هندسی زمین، متغیرهای مربوط به خاک، متغیرهای مربوط به منبع آبی، متغیرهای مربوط به گیاه و اقلیم، متغیرهای مدیریت آبیاری و متغیرهای هیدرولیکی هستند که متغیرهای یادشده به شرح زیر استخراج شد:

اطلاعات عمومی طرح (استان، شهرستان، مالک، نوع محصول، نوع منبع آب، نوع انرژی و سال اجرا)، اطلاعات مربوط به منبع آبی (میزان و وضعیت حقاچه، هدایت الکتریکی، اسیدیته، سدیم، پتاسیم، کلسیم، منیزیم، کربنات و بی‌کربنات)، اطلاعات مربوط به گیاه (فاصله ردیف‌های درختان، فاصله درختان روی ردیف، حداکثر تبخیر-تعرق روزانه گیاه، سطح سایه‌انداز و عمق توسعه ریشه)، اطلاعات مربوط به خاک (ظرفیت نگهداری آب در خاک، درصد تخلیه مجاز رطوبتی، درصد سطح خیس‌شده، نفوذپذیری نهایی و جرم مخصوص ظاهری)، اطلاعات مربوط به سامانه آبیاری (دبی متوسط کارکرد، فشار متوسط کارکرد، نوع آرایش لترال‌ها، تعداد گسیلنده برای هر گیاه و فاصله گسیلنده)، اطلاعات مربوط به مدیریت آبیاری مزرعه (دور آبیاری طراحی، نیاز خالص آبیاری، نیاز ناخالص آبیاری، مدت آبیاری، حداکثر ساعات آبیاری در یک شبانه‌روز، حداکثر تعداد نوبت آبیاری در هر دور، تعداد نوبت آبیاری، مساحت متوسط هر واحد آبیاری و دبی متوسط هر واحد آبیاری)، اطلاعات مربوط به ویژگی‌های مزرعه‌ای (شکل هندسی زمین، شیب متوسط، اختلاف ارتفاع منبع آبی تا بلندترین نقطه زمین، طول لترال‌ها، قطر و طول لوله‌های اصلی، نیمه‌اصلی و مانیفولدها و اتصالات)، اطلاعات ایستگاه پمپاژ (نوع پمپ، توان موتور، ارتفاع پمپاژ، دبی پمپاژ، اتصالات و لوازم سامانه کنترل مرکزی).

جزئیات اطلاعات موجود در مرحله مناقصه تفاوت دارد، بررسی کاربرد روش‌های مبتنی بر ML برای برآورد هزینه در شرکت‌های مشاوره مهم است (لستر، ۲۰۱۷؛ الشاهتی و رادیکا، ۲۰۱۸؛ درنتی و همکاران، ۲۰۱۹؛ گرنس‌برگ و روئدا، ۲۰۲۰؛ متل و همکاران، ۲۰۲۲؛ والیولو و آدی، ۲۰۲۲). با توجه به مرور منابع ملاحظه می‌شود که در خصوص انتخاب ویژگی‌های مؤثر بر هزینه سامانه‌های آبیاری، مطالعه‌ای همانند پژوهش حاضر به شناخت ویژگی‌های مؤثر بر هزینه‌های سامانه‌های آبیاری تحت فشار نپرداخته و تأثیر تک‌تک اجزای سامانه و شناخت مهم‌ترین ویژگی‌ها اعم از بخش مزرعه‌ای و ایستگاه پمپاژ بر هزینه اولیه اجرای یک سامانه تابه‌حال بررسی نشده است. پژوهش حاضر با هدف شناسایی مهم‌ترین ویژگی‌های مؤثر بر هزینه بخش‌های مختلف سامانه‌های آبیاری قطره‌ای در سطح کشور انجام شد. به عبارتی این تحقیق برای فرموله کردن هزینه نهایی سامانه‌های آبیاری قطره‌ای انجام شده است.

## مواد و روش‌ها

برای انجام این پژوهش ابتدا به جمع‌آوری داده‌های مورد نیاز، به‌روزرسانی هزینه تمام‌شده پروژه‌ها، پیش‌پردازش داده‌ها و انتخاب ویژگی‌های برتر پرداخته‌شده و سپس آموزش و آزمایش الگوریتم‌های مختلف انتخاب ویژگی‌های مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای شامل هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (TCP)، هزینه لوازم داخل مزرعه (TCF)، هزینه نصب و اجرای سامانه داخل مزرعه و ایستگاه پمپاژ (TCI) و هزینه کل (TCT) انجام شد.

## جمع‌آوری داده‌های مورد نیاز

در این پژوهش از آمار و اطلاعات ۵۱۵ سامانه آبیاری قطره‌ای اجراشده بین سال‌های ۱۳۸۵ تا ۱۳۹۸ در نقاط مختلف کشور استفاده شد. برای هر سامانه، آمار و اطلاعات مورد استفاده از دفترچه‌های طرح، نقشه‌های

## به‌روزرسانی هزینه پروژه‌ها و پیش‌پردازش داده‌ها

در پژوهش حاضر با استفاده از تورم سالانه (به‌صورت پله‌ای)، قیمت سال‌های پیشین به سال پایه تبدیل و هزینه پروژه‌های آبیاری قطره‌ای بر اساس نرخ‌های تورم برای سال ۱۴۰۰ به‌روز رسانی شد. برای پیش‌پردازش داده‌ها از روش‌های مختلف استانداردسازی، داده‌ها استاندارد شده و پس از اطمینان از تصادفی بودن داده‌ها، دسته‌بندی داده‌ها برای آموزش مدل و همچنین برای اعتبارسنجی انجام شد. در این پژوهش ۸۰ درصد از داده‌ها به‌عنوان داده‌های آموزش<sup>۱</sup> و ۲۰ درصد به‌عنوان داده‌های آزمایش<sup>۲</sup> یا آزمون در نظر گرفته شد (مسعودی سبحان‌زاده و همکاران، ۲۰۱۹). پس از استخراج متغیرهای مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای، در گام بعد اقدام انتخاب ویژگی‌های برتر که بیشترین تأثیر را بر مقدار خروجی مدل (هزینه) دارند، انجام شد. جدول (۱) متغیرهای کاندید برای تعیین رابطه بین متغیرهای مستقل و وابسته را نشان می‌دهد.

## روش‌های انتخاب ویژگی

انتخاب برترین ویژگی‌ها که بیشترین تأثیر را بر مقدار خروجی مدل که همان هزینه هست، بسیار حائز اهمیت است (پورغلام آمیجی و همکاران، ۲۰۲۱؛ a ۲۰۲۱). انتخاب زیرمجموعه‌ای از ویژگی‌ها، یکی از حوزه‌های جدید و فعال تحقیقات در مبحث یادگیری ماشین بوده که برای مسائل رگرسیون و دسته‌بندی استفاده می‌شود. فنون و روش‌های مختلفی برای انتخاب زیرمجموعه بهینه از ویژگی‌ها ارائه شده که دلیل آن‌هم اهمیت و تأثیر بسزایی است که بر کارایی الگوریتم‌های یادگیری ماشین دارد. انتخاب و استخراج ویژگی‌ها یک از دو مرحله اصلی در برنامه‌های یادگیری ماشین و مدل‌سازی‌ها است؛ یعنی باید مهم‌ترین ویژگی‌های با استفاده از مدل، روش و الگوریتم-های مختلف شناسایی شود (غدار و نائوم-ساواوا، ۲۰۱۸؛ مسعودی سبحان‌زاده و همکاران، ۲۰۱۹). یک طبقه‌بندی یا

یک مدل رگرسیون می‌تواند با استفاده از سه روش ایجاد شود (شیخ‌پور و همکاران، ۲۰۱۷؛ متین، ۲۰۱۸): (۱) روش تحت نظارت<sup>۳</sup> که در آن یک یادگیرنده از نوع و روند داده‌ها آگاه است. (۲) روش بدون نظارت<sup>۴</sup> که در آن یک یادگیرنده از روند تغییرات داده‌ها بی‌خبر است و سعی می‌کند رابطه بین داده‌ها را پیدا کند. (۳) روش نیمه نظارت<sup>۵</sup> که در آن روند و رابطه برخی از داده‌ها تعیین می‌شود؛ درحالی‌که سایر موارد مشخص نشده‌اند. در این روش، یک یادگیرنده معمولاً با استفاده از هر دو نمونه تعیین شده و نامعین آموزش داده می‌شود. فنون انتخاب ویژگی با استفاده از مدل‌های نظارت‌شده عمدتاً به سه دسته اصلی و یا در برخی ادبیات پیشین به پنج دسته طبقه‌بندی می‌شوند (چاندرشکار و ساهین، ۲۰۱۴؛ شیخ‌پور و همکاران، ۲۰۱۷؛ رحمانی‌نیا و مرادی، ۲۰۱۸؛ پازوکی و همکاران، ۲۰۲۰؛ سولوریو-فرناندز و همکاران، ۲۰۲۰) که شامل روش‌های فیلتر (FM<sup>۶</sup>)، روش‌های لفاف (WM<sup>۷</sup>)، روش‌های تعبیه‌شده (EM<sup>۸</sup>)، روش‌های آنلاین (OM<sup>۹</sup>) و روش‌های ترکیبی (HM<sup>۱۰</sup>) است.

<sup>3</sup> -Supervised  
<sup>4</sup> -Unsupervised  
<sup>5</sup> -Semi-Supervised  
<sup>6</sup> -Filter Methods  
<sup>7</sup> -Wrapper Methods  
<sup>8</sup> -Embedded Methods  
<sup>9</sup> -Online Methods  
<sup>10</sup> -Hybrid Methods

<sup>1</sup> -Training Data  
<sup>2</sup> -Testing Data

جدول ۱- متغیرهای کاندید مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای

| نماد                      | توضیح   | نماد                   | توضیح   |
|---------------------------|---|------------------------|---|
| A (ha)*                   | مساحت زمین  | L16mm (m)*             | طول کل لترال  |
| P/A (1/m)*                | شکل زمین (نسبت محیط به مساحت)                     | L32mm (m)              | طول لوله رابط یک                                    |
| N <sub>PL</sub> *         | تعداد قطعات زمین                                  | L40mm (m)              | طول لوله رابط دو                                    |
| DSF (m)*                  | فاصله منبع آبی تا زمین                            | L50mm (m)              | طول لوله مانیفولد                                   |
| ΔH <sub>SF</sub> (m)*     | اختلاف ارتفاع منبع آبی و زمین (بحرانی)            | L63mm (m)              | طول لوله فرعی                                       |
| Q <sub>T</sub> (l/s)*     | مقدار کل دبی آب قابل دسترس                        | L75mm (m)              | طول لوله نیمه اصلی                                  |
| N <sub>HO</sub> *         | نسبت ساعتی در دور آبیاری که آب در اختیار مالک است | L90mm (m)              | طول لوله اصلی                                       |
| S <sub>R</sub> (m)*       | فاصله ردیف گیاهان                                 | L110mm (m)             | طول لوله جانبی                                      |
| S <sub>P</sub> (m)*       | فاصله گیاه روی هر ردیف                            | L125mm (m)             | طول لوله جانبی                                      |
| ET <sub>P</sub> (mm/day)* | تبخیر-تعرق گیاه                                   | L160mm (m)             | طول لوله جانبی                                      |
| F <sub>SP</sub> (mm/hr)*  | نفوذپذیری نهایی خاک                               | L200mm (m)             | خط انتقال طولانی                                    |
| AW (mm/m)*                | ظرفیت نگهداری آب در خاک (FC-PWP)                  | W <sub>Q</sub> *       | کیفیت آب از نظر نیاز به فیلتراسیون                  |
| W <sub>A</sub> (%)        | سطح خیس شده                                       | Q <sub>S</sub> (l/s)   | دبی سامانه  |
| Q <sub>E</sub> (l/s)*     | دبی گسیلنده                                       | P <sub>P</sub> (kw)    | توان پمپ مورد نیاز                                  |
| N <sub>E</sub> *          | تعداد کل گسیلنده                                  | H <sub>P</sub> (m)     | ارتفاع پمپاژ  |
| T (h)                     | تعداد ساعات کاری در شبانه‌روز                     | D <sub>FE</sub> (m)*   | فاصله زمین تا خطوط برق                              |
| N <sub>IT</sub>           | تعداد نوبت‌های آبیاری                             | S (%)*                 | نوع منطقه یا شیب غالب (کوهستانی، دشت)               |
| A <sub>IT</sub> (ha)      | مساحت نوبت‌های آبیاری                             | TC <sub>P</sub> (Rial) | هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (ریال)     |
| N <sub>IU</sub>           | تعداد واحدهای آبیاری                              | TC <sub>F</sub> (Rial) | هزینه لوازم داخل مزرعه (ریال)                       |
| A <sub>IU</sub> (ha)      | مساحت واحدهای آبیاری                              | TC <sub>I</sub> (Rial) | هزینه نصب و اجرای داخل مزرعه و ایستگاه پمپاژ (ریال) |
| F (day)                   | دور آبیاری  | TC <sub>T</sub> (Rial) | هزینه کل (ریال)                                     |
| N <sub>IUS</sub>          | تعداد واحدهای آبیاری که همزمان آبیاری می‌شوند     |                        |   |

\*: ویژگی‌های فاز قبل طراحی (۱۸ متغیر) که قبل از طراحی یک سامانه آبیاری می‌توان به آن‌ها دست یافت

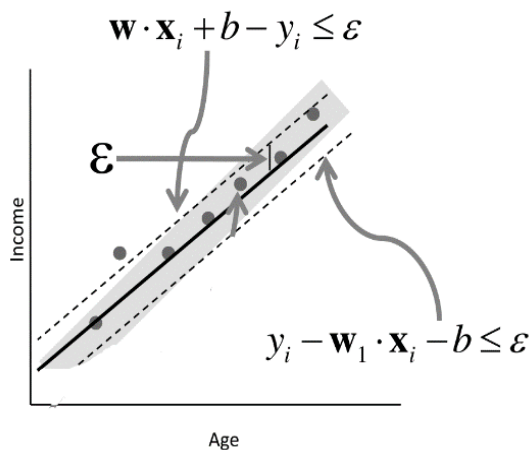
انتخاب ویژگی دو دسته کلی برای (۱) مسائل رگرسیون و (۲) مسائل طبقه‌بندی<sup>۱</sup> دارد که با توجه به نوع داده و خروجی‌های مورد انتظار، در این مطالعه از دسته اول استفاده شد. از آنجایی که نرم‌افزار MATLAB یک زبان برنامه‌نویسی متن‌باز و رایج در علوم مختلف بوده و همه پژوهشگران می‌توانند برخی از قابلیت‌های جدید را به آن اضافه کنند، برای به‌کارگیری روش‌های فوق‌الذکر از این نرم‌افزار استفاده شد.

#### انتخاب نوع یادگیرنده

برای حل مسائل رگرسیونی برای انتخاب ویژگی، سه نوع یادگیرنده موجود است. اولین مورد SVM است. دومی ANN هست که فقط شامل یک پارامتر (تکرار آموزش) است. سومین یادگیرنده درخت

از بین پنج روش انتخاب ویژگی فوق و بر اساس نتایج مطالعه مسعودی سبحانزاده و همکاران (۲۰۱۹)، سه نوع آن بهترین نتایج را به دست می‌دهند که به‌صورت زیر هستند: (۱) روش Wrapper (الگوریتم بهینه‌سازی)، (۲) روش فیلتر: این نوع انتخاب ویژگی از پنج روش رایج تشکیل شده است. نتایج تجربی نشان می‌دهد که هر یادگیرنده و هر روشی دیدگاه خاص خود را نسبت به مجموعه داده دارد، اما در حالت کلی، روش‌های Wrapper نسبت به روش‌های فیلتر می‌توانند منجر به نتایج بهتری شوند. (۳) روش ترکیبی (Hybrid-Ensemble): کاربر می‌تواند از انتخاب ویژگی دومرحله‌ای با استفاده از ترکیب روش‌های فیلتر و Wrapper استفاده کند. بر اساس کارایی مناسب روش الگوریتم‌های بهینه‌سازی و تأیید آن در سایر مطالعات (مسعودی سبحانزاده و همکاران، ۲۰۱۹)، در این پژوهش از روش انتخاب ویژگی Wrapper استفاده شده است.

<sup>1</sup>-Classification



شکل ۱- شماتیک یک رگرسیون بردار پشتیبان

### الگوریتم‌های قابل استفاده

۱۱ الگوریتم برای انتخاب بهترین ویژگی (ها) از مجموعه ویژگی‌ها در بخش روش wrapper توسعه داده شده است. توصیه شده که الگوریتم‌های بهینه‌سازی بیش از ۳۰ بار به دلیل ماهیت تصادفی آن‌ها تکرار شوند. همچنین می‌توان تعداد تکرارها را در هنگام استفاده از هر الگوریتم، تنظیم کرد. تمامی پارامترهای الگوریتم‌ها شامل تعداد تکرار، تیم، فصل، تولید و ... قابل تغییر است. الگوریتم‌های موجود برای این کار که می‌توان از آن‌ها بهره گرفت، شامل WCC<sup>۸</sup>، LCA<sup>۹</sup>، GA<sup>۱۰</sup>، PSO<sup>۱۱</sup>، AC<sup>۱۱</sup>، JCA<sup>۱۲</sup>، LA<sup>۱۳</sup>، HTS<sup>۱۴</sup>، FOA<sup>۱۵</sup>، DSOS<sup>۱۶</sup> و CUK<sup>۱۷</sup> می‌باشند. در خصوص نوع و تعداد الگوریتم‌های انتخابی در بخش نتایج و بحث توضیح داده شده است.

### معیارهای ارزیابی

برای ارزیابی و مقایسه نتایج مدل‌ها و الگوریتم‌ها با داده‌های واقعی در مرحله آموزش و آزمایش از ضریب تعیین یا تبیین<sup>۱۸</sup> (R<sup>2</sup>) و جذر میانگین مربعات

تصمیم (DT<sup>۱</sup>) است. این زبان آموزش نیازی به تنظیم پارامتر ندارد. درصد داده‌های آموزشی و حداکثر تعداد ویژگی‌هایی که برای هر پروژه مطلوب است را می‌توان در هر شرایطی تنظیم کرد. با توجه به انتخاب یادگیرنده SVM بر اساس نتایج مطالعه مسعودی سبحانزاده و همکاران (۲۰۱۹)، در زیر توضیحاتی ارائه شده است.

### ماشین بردار پشتیبان (SVM)

به کمک SVM می‌توان مدل‌های خطی و غیرخطی را ایجاد و پارامترهای آن را محاسبه کرد. این کار با به‌کارگیری یک تابع هسته<sup>۲</sup> غیرخطی (مانند چندجمله‌ای) حاصل می‌شود. محاسبه پارامترهای این تابع به این شکل است که خطا کمینه شود؛ به طوری که فاصله بین صفحاتی که عمل جداسازی بین دسته‌ها را ایجاد می‌کنند، بیشینه شود. SVM برای حل مسائل غیرخطی، ابعاد مسئله را از طریق توابع کرنل تغییر می‌دهند. انتخاب کرنل برای رگرسیون بردار پشتیبان (SVR<sup>۳</sup>)، به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد (کرتس و وپنیک، ۱۹۹۵)؛ به عبارت دیگر، با توجه به این پارامترها باید تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل چهار نوع کرنل خطی<sup>۴</sup>، کرنل چندجمله‌ای<sup>۵</sup>، کرنل تانژانت هیپربولیک<sup>۶</sup> و کرنل گوسی<sup>۷</sup> به کار گرفته می‌شوند. شماتیک مسئله SVR اغلب با استفاده از مثال یک بعدی در شکل (۱) نشان داده شده است (اواد و خانا، ۲۰۱۵):

$$y = f(x) = \langle w, x \rangle + b = \sum_{i=1}^M w_i x_i + b, \quad y, b \in \mathbb{R}, \quad x, w \in \mathbb{R}^M \quad (1)$$

$$f(x) = \begin{bmatrix} w \\ b \end{bmatrix}^T \begin{bmatrix} x \\ 1 \end{bmatrix} = w^T x + b \quad x, w \in \mathbb{R}^{M+1} \quad (2)$$

<sup>8</sup> -World Competitive Contest

<sup>9</sup> -League Championship Algorithm

<sup>10</sup> -Particle Swarm Optimization

<sup>11</sup> -Ant Colony Optimization

<sup>12</sup> -Imperialist Competitive Algorithm

<sup>13</sup> -Learning Automata

<sup>14</sup> -Heat Transfer Optimization Algorithm

<sup>15</sup> -Forest Optimization Algorithm

<sup>16</sup> -Discrete Symbiotic Organisms Search

<sup>17</sup> -Cuckoo Optimization

<sup>18</sup> -Coefficient of Determination

<sup>1</sup> -Decision Tree

<sup>2</sup> -Kernel

<sup>3</sup> -Support Vector Regression

<sup>4</sup> -Linear Kernel

<sup>5</sup> -Polynomial Kernel

<sup>6</sup> -Hyperbolic Tangent Kernel

<sup>7</sup> -Radial Base Function Kernel



به ترتیب برابر با ۰/۵۵۰، ۰/۶۳۱ و ۰/۶۳۸ بود. با توجه به آماره P-value، همگی این متغیرها در سطح اطمینان یک درصد معنی دار شد. با افزایش مساحت و بالطبع افزایش تعداد کل گسیلنده و لوله‌های آبد، هزینه نهایی لوازم داخل مزرعه افزایش یافته و به عبارتی این متغیرها، نماینده ۳۹ متغیر مورد مطالعه برای بررسی پیوند بین این پارامترها با هزینه نهایی لوازم داخل مزرعه (TCF) هستند. البته متغیرهای NPL، DSF، NIT، L63mm، L75mm و L90mm ضریب همبستگی بالایی از خود نشان دادند که همگی آن‌ها در سطح اطمینان یک درصد معنی دار شد (جدول ۲).

در بخش هزینه نصب و اجرای سامانه داخل مزرعه و ایستگاه پمپاژ (TC<sub>1</sub>)، متغیرهای ΔH<sub>SF</sub> (اختلاف ارتفاع منبع آبی و زمین در بحرانی‌ترین حالت)، S<sub>R</sub> (فاصله ردیف گیاهان) و L<sub>40mm</sub> (طول لوله رابط دو) به ترتیب با مقدار ۰/۲۹۴، ۰/۳۵۵ و ۰/۳۳۵ دارای بیشترین ضریب همبستگی و در سطح اطمینان یک درصد هستند (جدول ۲). علت کاهش مقدار عددی ضریب همبستگی در بخش TC<sub>1</sub> نسبت به بخش‌های دیگر هزینه به این دلیل است که همه متغیرها، به گونه‌ای تأثیر مستقیم خود را بر TC<sub>F</sub> و TC<sub>P</sub> گذاشته‌اند و اکنون چون تأثیر غیرمستقیم خود را بر متغیر وابسته TC<sub>1</sub> می‌گذارند، مقدار عددی ضریب همبستگی نسبت به دو بخش قبل کاهش یافته است. بیشترین مقدار ضریب همبستگی منفی که رابطه عکس بین متغیرها و TC<sub>1</sub> را نشان می‌دهد، مربوط به متغیرهای ET<sub>P</sub> (تبخیر-تعرق گیاه)، W<sub>A</sub> (سطح خیس-شده) و L<sub>200mm</sub> (خط انتقال طولانی) به ترتیب برابر با -۰/۲۸۲، -۰/۴۷۸ و -۰/۲۴۷ است که هر سه متغیر در سطح اطمینان یک درصد معنی دار شد. A (مساحت زمین)، P<sub>P</sub> (توان پمپ مورد نیاز) و H<sub>P</sub> (ارتفاع پمپاژ) به ترتیب با ضریب همبستگی ۰/۴۰۰، ۰/۴۲۸ و ۰/۴۲۹ بیشترین همبستگی را با هزینه کل (TC<sub>T</sub>) داشتند. سه مورد از متغیرهایی که ضریب همبستگی منفی و به عبارتی رابطه عکس با TC<sub>T</sub> دارند، متغیرهایی هستند که در بخش

خطا (RMSE) استفاده شد. این معیارها به صورت زیر تعریف می‌شوند:

$$R^2 = \frac{[\sum_{i=1}^n (o_i - \bar{o})(p_i - \bar{p})]^2}{\sum_{i=1}^n (o_i - \bar{o})^2 \times \sum_{i=1}^n (p_i - \bar{p})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - o_i)^2}{n}} \quad (4)$$

که در آن‌ها O<sub>i</sub> مقادیر مشاهداتی، P<sub>i</sub> مقادیر برآوردشده،  $\bar{O}$  میانگین مقادیر مشاهداتی،  $\bar{P}$  میانگین مقادیر برآوردشده و n تعداد داده‌ها می‌باشند (وینستون، ۱۹۹۲؛ نورویگ و اینتلیجنس، ۲۰۰۲). هر مدل که دارای R<sup>2</sup> بیشتر و RMSE کمتر باشد، از مطلوبیت بیشتری برخوردار است.

## نتایج و بحث

### همبستگی بین متغیرها

نتایج همبستگی بین ۳۹ متغیر مستقل یادشده با هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (TC<sub>P</sub>)، هزینه لوازم داخل مزرعه (TC<sub>F</sub>)، هزینه نصب و اجرای داخل مزرعه و ایستگاه پمپاژ (TC<sub>1</sub>) و هزینه کل (TC<sub>T</sub>) در جدول (۲) ارائه شده است. بیشترین ضریب همبستگی (R) مثبت بین پارامترهای Q<sub>S</sub> (دبی سامانه)، P<sub>P</sub> (توان پمپ مورد نیاز) و H<sub>P</sub> (ارتفاع پمپاژ) با TC<sub>P</sub> به ترتیب با مقدار ۰/۶۱۸، ۰/۷۶۶ و ۰/۶۶۶ به دست آمد و هر سه در سطح یک درصد دارای اختلاف معنی‌داری بود. از آنجایی که عمده هزینه بخش ایستگاه پمپاژ و سامانه کنترل مرکزی مربوط به تجهیزات ایستگاه پمپاژ است، لذا سه متغیر یادشده بیشترین همبستگی را با TC<sub>P</sub> داشته و مهم‌ترین پارامترهای این بخش به حساب می‌آیند و با افزایش دبی سامانه، توان پمپ و ارتفاع پمپاژ، هزینه نهایی ایستگاه پمپاژ و سامانه کنترل مرکزی افزایش می‌یابد. سه پارامتری که بیشترین همبستگی را با هزینه TC<sub>F</sub> داشتند عبارت‌اند از A (مساحت زمین)، N<sub>E</sub> (تعداد کل گسیلنده) و L<sub>16mm</sub> (طول کل لترال) که ضریب همبستگی آن‌ها

شرح بود که در بخش هزینه لوازم داخل مزرعه (TC<sub>F</sub>): ۱۱ متغیر در سطح معنی‌داری یک درصد و ۲۸ متغیر با عدم تفاوت معنی‌دار، در بخش هزینه نصب و اجرای داخل مزرعه و ایستگاه پمپاژ (TC<sub>I</sub>)؛ هفت متغیر در سطح معنی‌داری یک درصد، چهار متغیر در سطح معنی‌داری پنج درصد و ۲۸ متغیر با عدم تفاوت معنی‌دار و در نهایت در بخش هزینه کل (TC<sub>T</sub>)؛ نه متغیر در سطح معنی‌داری یک درصد، شش متغیر در سطح معنی‌داری پنج درصد و ۲۴ متغیر با عدم تفاوت معنی‌دار به ثبت رسید (جدول ۲).

TC<sub>P</sub> و TC<sub>I</sub> تکرار شدند و این‌ها شامل P/A (شکل زمین، نسبت محیط به مساحت)، ET<sub>P</sub> (تبخیر-تعرق گیاه) و W<sub>A</sub> (سطح خیس‌شده) به ترتیب با ضریب همبستگی ۰/۲۱۱، -۰/۳۶۸ و -۰/۴۵۴ هر سه در سطح یک درصد دارای اختلاف معنی‌داری هستند (جدول ۲). به‌طورکلی نتایج همبستگی را می‌توان به این صورت جمع‌بندی کرد که ۱۲ متغیر با سطح معنی‌داری یک درصد، چهار متغیر با سطح معنی‌داری پنج درصد و ۲۳ متغیر عدم تفاوت معنی‌دار با هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (TC<sub>P</sub>) داشتند. این موضوع برای دیگر بخش‌ها بدین

جدول ۲- نتایج ضریب همبستگی و معنی‌داری (P-Value) متغیرهای مستقل با متغیر وابسته

| پارامتر                  | TC <sub>T</sub>     |        | TC <sub>I</sub>     |        | TC <sub>F</sub>     |        | TC <sub>P</sub>     |        |
|--------------------------|---------------------|--------|---------------------|--------|---------------------|--------|---------------------|--------|
|                          | P-Value             | R      | P-Value             | R      | P-Value             | R      | P-Value             | R      |
| A (ha)                   | ۰/۰۰۰**             | ۰/۴۰۰  | ۰/۱۵۸ <sup>NS</sup> | ۰/۱۴۲  | ۰/۰۰۰**             | ۰/۵۵۰  | ۰/۰۰۲**             | ۰/۳۰۱  |
| P/A (l/m)                | ۰/۰۳۵*              | -۰/۲۱۱ | ۰/۷۹۰ <sup>NS</sup> | -۰/۰۲۷ | ۰/۰۵۴ <sup>NS</sup> | -۰/۱۹۳ | ۰/۰۰۲**             | -۰/۳۰۴ |
| N <sub>PL</sub>          | ۰/۲۶۲ <sup>NS</sup> | ۰/۱۱۳  | ۰/۷۰۶ <sup>NS</sup> | ۰/۰۳۸  | ۰/۰۰۱**             | ۰/۳۳۰  | ۰/۷۴۳ <sup>NS</sup> | -۰/۰۳۳ |
| D <sub>SF</sub> (m)      | ۰/۰۹۷ <sup>NS</sup> | ۰/۱۶۷  | ۰/۲۴۳ <sup>NS</sup> | ۰/۱۱۸  | ۰/۰۰۰**             | ۰/۴۵۲  | ۰/۲۲۰ <sup>NS</sup> | -۰/۱۲۴ |
| ΔH <sub>SF</sub> (m)     | ۰/۰۰۱**             | ۰/۳۳۶  | ۰/۰۰۳**             | ۰/۲۹۴  | ۰/۵۸۶ <sup>NS</sup> | ۰/۰۵۵  | ۰/۰۱۰*              | ۰/۲۵۷  |
| Q <sub>T</sub> (l/s)     | ۰/۷۶۶ <sup>NS</sup> | -۰/۰۳۰ | ۰/۵۶۳ <sup>NS</sup> | ۰/۰۵۹  | ۰/۶۷۹ <sup>NS</sup> | -۰/۰۴۲ | ۰/۱۸۴ <sup>NS</sup> | -۰/۱۳۴ |
| N <sub>HO</sub>          | ۰/۳۵۶ <sup>NS</sup> | ۰/۰۹۳  | ۰/۸۸۷ <sup>NS</sup> | ۰/۰۱۴  | ۰/۳۱۲ <sup>NS</sup> | -۰/۱۰۲ | ۰/۰۰۹**             | ۰/۲۶۱  |
| S <sub>R</sub> (m)       | ۰/۰۰۱**             | ۰/۳۲۲  | ۰/۰۰۰**             | ۰/۳۵۵  | ۰/۵۴۳ <sup>NS</sup> | -۰/۰۶۲ | ۰/۰۳۹*              | ۰/۲۰۷  |
| S <sub>P</sub> (m)       | ۰/۰۱۱*              | ۰/۲۵۲  | ۰/۰۱۲*              | ۰/۲۵۱  | ۰/۳۲۸ <sup>NS</sup> | -۰/۰۹۹ | ۰/۰۱۶*              | ۰/۲۴۰  |
| ET <sub>P</sub> (mm/day) | ۰/۰۰۰**             | ۰/۳۶۸  | ۰/۰۰۵**             | -۰/۲۸۲ | ۰/۷۵۳ <sup>NS</sup> | ۰/۰۳۲  | ۰/۰۰۰**             | -۰/۴۰۹ |
| F <sub>SP</sub> (mm/hr)  | ۰/۷۸۳ <sup>NS</sup> | -۰/۰۲۸ | ۰/۹۵۷ <sup>NS</sup> | -۰/۰۰۵ | ۰/۸۸۵ <sup>NS</sup> | ۰/۰۱۵  | ۰/۵۱۹ <sup>NS</sup> | -۰/۰۶۵ |
| AW (mm/m)                | ۰/۹۶۲ <sup>NS</sup> | -۰/۰۰۵ | ۰/۹۱۲ <sup>NS</sup> | ۰/۰۱۱  | ۰/۶۴۹ <sup>NS</sup> | -۰/۰۴۶ | ۰/۹۷۶ <sup>NS</sup> | ۰/۰۰۳  |
| W <sub>A</sub> (%)       | ۰/۰۰۰**             | -۰/۴۵۴ | ۰/۰۰۰**             | -۰/۴۷۸ | ۰/۲۴۴ <sup>NS</sup> | ۰/۱۱۸  | ۰/۰۰۰**             | -۰/۳۵۰ |
| Q <sub>E</sub> (l/s)     | ۰/۷۴۴ <sup>NS</sup> | ۰/۰۳۳  | ۰/۹۱۳ <sup>NS</sup> | ۰/۰۱۱  | ۰/۲۲۷ <sup>NS</sup> | ۰/۱۲۲  | ۰/۷۸۶ <sup>NS</sup> | -۰/۰۲۸ |
| N <sub>E</sub>           | ۰/۵۱۰ <sup>NS</sup> | ۰/۰۶۷  | ۰/۱۴۱ <sup>NS</sup> | -۰/۱۴۸ | ۰/۰۰۰**             | ۰/۶۳۱  | ۰/۶۲۶ <sup>NS</sup> | -۰/۰۴۹ |
| T (h)                    | ۰/۸۲۸ <sup>NS</sup> | ۰/۰۲۲  | ۰/۴۵۸ <sup>NS</sup> | -۰/۰۷۵ | ۰/۹۱۴ <sup>NS</sup> | -۰/۰۱۱ | ۰/۰۷۵ <sup>NS</sup> | ۰/۱۷۹  |
| N <sub>IT</sub>          | ۰/۱۷۳ <sup>NS</sup> | ۰/۱۳۷  | ۰/۷۳۲ <sup>NS</sup> | ۰/۰۳۵  | ۰/۰۰۰**             | ۰/۳۵۸  | ۰/۹۳۷ <sup>NS</sup> | ۰/۰۰۸  |
| A <sub>IT</sub> (ha)     | ۰/۰۲۰*              | ۰/۲۳۲  | ۰/۳۸۸ <sup>NS</sup> | ۰/۰۸۷  | ۰/۰۸۸ <sup>NS</sup> | ۰/۱۷۲  | ۰/۰۰۷**             | ۰/۲۶۹  |
| N <sub>IU</sub>          | ۰/۱۹۱ <sup>NS</sup> | ۰/۱۳۲  | ۰/۶۵۷ <sup>NS</sup> | ۰/۰۴۵  | ۰/۰۰۰**             | ۰/۵۲۰  | ۰/۱۸۳ <sup>NS</sup> | -۰/۱۳۴ |
| A <sub>IU</sub> (ha)     | ۰/۲۵۰ <sup>NS</sup> | ۰/۱۱۶  | ۰/۹۰۵ <sup>NS</sup> | -۰/۰۱۲ | ۰/۷۸۳ <sup>NS</sup> | -۰/۰۲۸ | ۰/۰۰۲**             | ۰/۳۰۴  |
| F (day)                  | ۰/۲۱۷ <sup>NS</sup> | ۰/۱۲۴  | ۰/۰۴۳*              | ۰/۲۰۲  | ۰/۸۸۴ <sup>NS</sup> | -۰/۰۱۵ | ۰/۷۵۶ <sup>NS</sup> | -۰/۰۳۱ |
| N <sub>IUS</sub>         | ۰/۴۵۰ <sup>NS</sup> | ۰/۰۷۶  | ۰/۵۴۲ <sup>NS</sup> | ۰/۰۶۲  | ۰/۰۰۹**             | ۰/۲۵۹  | ۰/۲۹۶ <sup>NS</sup> | -۰/۱۰۶ |
| L <sub>16mm</sub> (m)    | ۰/۴۰۹ <sup>NS</sup> | ۰/۰۸۳  | ۰/۱۳۹ <sup>NS</sup> | -۰/۱۴۹ | ۰/۰۰۰**             | ۰/۶۳۸  | ۰/۸۸۱ <sup>NS</sup> | -۰/۰۱۵ |
| L <sub>32mm</sub> (m)    | ۰/۳۶۲ <sup>NS</sup> | ۰/۰۹۲  | ۰/۲۲۸ <sup>NS</sup> | ۰/۱۲۲  | ۰/۳۹۲ <sup>NS</sup> | -۰/۰۸۶ | ۰/۴۵۷ <sup>NS</sup> | ۰/۰۷۵  |
| L <sub>40mm</sub> (m)    | ۰/۰۰۷**             | ۰/۲۶۶  | ۰/۰۰۱**             | ۰/۳۳۵  | ۰/۲۹۳ <sup>NS</sup> | -۰/۱۰۶ | ۰/۱۵۶ <sup>NS</sup> | ۰/۱۴۳  |
| L <sub>50mm</sub> (m)    | ۰/۲۹۹ <sup>NS</sup> | ۰/۱۰۵  | ۰/۱۹۸ <sup>NS</sup> | ۰/۱۳۰  | ۰/۴۵۸ <sup>NS</sup> | ۰/۰۷۵  | ۰/۸۳۱ <sup>NS</sup> | -۰/۰۲۲ |
| L <sub>63mm</sub> (m)    | ۰/۲۷۹ <sup>NS</sup> | ۰/۱۰۹  | ۰/۷۰۶ <sup>NS</sup> | -۰/۰۳۸ | ۰/۰۰۰**             | ۰/۳۶۲  | ۰/۵۶۴ <sup>NS</sup> | ۰/۰۵۸  |
| L <sub>75mm</sub> (m)    | ۰/۴۷۱ <sup>NS</sup> | -۰/۰۷۳ | ۰/۰۴۰*              | -۰/۲۰۵ | ۰/۰۰۱**             | ۰/۳۲۸  | ۰/۵۲۱ <sup>NS</sup> | -۰/۰۶۵ |
| L <sub>90mm</sub> (m)    | ۰/۰۲۹*              | ۰/۲۱۸  | ۰/۱۷۳ <sup>NS</sup> | ۰/۱۳۷  | ۰/۰۰۰**             | ۰/۴۶۴  | ۰/۶۴۳ <sup>NS</sup> | -۰/۰۴۷ |
| L <sub>110mm</sub> (m)   | ۰/۳۴۹ <sup>NS</sup> | -۰/۰۹۵ | ۰/۴۴۸ <sup>NS</sup> | -۰/۰۷۷ | ۰/۰۵۷ <sup>NS</sup> | ۰/۱۹۱  | ۰/۰۲۴*              | -۰/۲۲۶ |
| L <sub>125mm</sub> (m)   | ۰/۳۴۷ <sup>NS</sup> | -۰/۰۹۵ | ۰/۶۷۸ <sup>NS</sup> | -۰/۰۴۲ | ۰/۰۸۳ <sup>NS</sup> | ۰/۱۷۴  | ۰/۰۰۶**             | -۰/۲۷۱ |
| L <sub>160mm</sub> (m)   | ۰/۳۳۵ <sup>NS</sup> | -۰/۰۹۷ | ۰/۵۶۴ <sup>NS</sup> | -۰/۰۵۸ | ۰/۳۶۸ <sup>NS</sup> | ۰/۰۹۱  | ۰/۰۵۵ <sup>NS</sup> | -۰/۱۹۲ |
| L <sub>200mm</sub> (m)   | ۰/۳۱۳ <sup>NS</sup> | ۰/۲۴۷  | ۰/۰۱۳*              | -۰/۲۴۷ | ۰/۶۰۷ <sup>NS</sup> | ۰/۰۵۲  | ۰/۲۰۲ <sup>NS</sup> | ۰/۱۲۹  |
| W <sub>Q</sub>           | ۰/۰۸۳ <sup>NS</sup> | -۰/۱۷۴ | ۰/۰۹۳ <sup>NS</sup> | -۰/۱۶۹ | ۰/۳۸۸ <sup>NS</sup> | -۰/۰۸۷ | ۰/۵۱۶ <sup>NS</sup> | -۰/۰۶۶ |
| Q <sub>S</sub> (l/s)     | ۰/۰۰۴**             | ۰/۲۸۷  | ۰/۸۳۴ <sup>NS</sup> | ۰/۰۲۱  | ۰/۹۵۰ <sup>NS</sup> | ۰/۰۰۶  | ۰/۰۰۰**             | ۰/۶۱۸  |
| P <sub>P</sub> (kw)      | ۰/۰۰۰**             | ۰/۴۲۸  | ۰/۱۲۵ <sup>NS</sup> | ۰/۱۵۴  | ۰/۵۹۷ <sup>NS</sup> | -۰/۰۵۴ | ۰/۰۰۰**             | ۰/۷۶۶  |
| H <sub>P</sub> (m)       | ۰/۰۰۰**             | ۰/۴۲۹  | ۰/۰۰۷**             | ۰/۲۶۹  | ۰/۰۹۰ <sup>NS</sup> | -۰/۱۷۱ | ۰/۰۰۰**             | ۰/۶۶۶  |
| D <sub>FE</sub> (m)      | ۰/۰۲۹*              | ۰/۲۱۸  | ۰/۵۷۳ <sup>NS</sup> | ۰/۰۵۷  | ۰/۰۸۸ <sup>NS</sup> | ۰/۱۷۱  | ۰/۰۰۴**             | ۰/۲۸۷  |
| S (%)                    | ۰/۰۲۷*              | ۰/۲۲۲  | ۰/۰۰۸**             | ۰/۲۶۴  | ۰/۹۶۱ <sup>NS</sup> | -۰/۰۰۵ | ۰/۴۰۲ <sup>NS</sup> | ۰/۰۸۵  |

علامت \* و \*\* به ترتیب معنی‌داری در سطح پنج و یک درصد و <sup>NS</sup> عدم تفاوت معنی‌دار را نشان می‌دهد

ارزیابی الگوریتم‌ها در مرحله آموزش  
 سامانه آبیاری قطره‌ای تأثیر می‌گذارند که تعداد آن ۳۹  
 است. بخش دوم ویژگی‌هایی که مربوط به فاز قبل  
 طراحی بوده که می‌توان قبل از طراحی و اجرای یک

نتایج این مقاله به دو بخش جداگانه تقسیم شده  
 است. بخش اول تمامی ویژگی‌هایی که بر هزینه‌های یک

سامانه آبیاری به آن دسترسی پیدا کرد. تعداد این ویژگی‌ها ۱۸ هست و از بین آن انتخاب ویژگی صورت گرفت. آموزش مدل با ۸۰ درصد داده‌ها و پیش‌پردازش اولیه انجام شد و معیارهای ارزیابی در بخش کل ویژگی‌ها به دست آمد. نتایج آموزش مدل در بخش کل ویژگی‌ها که شامل ۳۹ ویژگی بود، نشان داد که معیار  $RMSE$ ، زمان سپری شده  $SVM$  بر حسب ثانیه و  $R^2$  به ترتیب برابر  $۰/۰۰۷$ ،  $۱/۳۰$  و  $۰/۹۲$  به دست آمد. در تمامی منابع مطالعاتی مطرح شده که این مقدار عددی این معیارها کاملاً معقول بوده و نشان از دقت داده‌های ورودی و آموزش صحیح الگوریتم‌ها است (مسعودی سبحان‌زاده و مطیع‌قادر، ۲۰۱۶؛ مسعودی سبحان‌زاده و همکاران، ۲۰۱۹). همچنین آموزش مدل و دستیابی به پارامترهای ارزیابی اولیه در بخش ویژگی‌های فاز قبل طراحی که ۱۸ ویژگی بود نیز انجام شد. در این بخش نیز معیار  $RMSE$ ، زمان سپری شده  $SVM$  بر حسب ثانیه و  $R^2$  به ترتیب برابر  $۰/۰۰۳۰۵$ ،  $۲/۰۱۳۲$  و  $۰/۸۸۷۹۵$  به دست آمد. تغییر ناچیز مقدار معیارهای ارزیابی در این بخش نسبت به قبل در این است که از مجموعه ویژگی‌های تأثیرگذار بر هزینه‌های سامانه آبیاری قطره‌ای (هزینه ایستگاه پمپاژ و سامانه کنترل مرکزی (TCP)، هزینه لوازم داخل مزرعه و  $(TC_F)$ ، هزینه نصب و اجرای سامانه داخل مزرعه و ایستگاه پمپاژ  $(TC_I)$  و هزینه کل  $(TC_T)$ )، ۱۸ ویژگی جدا شده و سپس از بین آن انتخاب ویژگی صورت گرفت. لذا کمی دقت کار افت کرده اما این معیارها کاملاً مجاز و در محدوده بسیار عالی قرار گرفته است. دیگر محققین این معیارها را در مطالعات خود گزارش کرده و این را دال بر آموزش صحیح دانستند (کاشان، ۲۰۱۴).

### نتایج الگوریتم‌های منتخب

نتایج الگوریتم‌های منتخب برای مسائل رگرسیون در بخش کل ویژگی‌های مؤثر بر هزینه‌های سامانه آبیاری قطره‌ای در شکل (۲) و ویژگی‌های فاز قبل طراحی در شکل (۳) نشان داده شده است. شایان ذکر

است که از بین ۱۱ الگوریتم، چهار الگوریتم WCC، LA، LCA و FOA برای شناسایی مهم‌ترین ویژگی‌ها انتخاب شدند. انتخاب این‌ها دو دلیل داشت: نخست اینکه مسعودی سبحان‌زاده و همکاران (۲۰۱۹) که توسعه‌دهندگان نرم‌افزار FeatureSelect هستند، بعد از بررسی‌های فراوان و آزمایش‌های بسیار بیان نمودند که این الگوریتم‌ها بهترین نتایج را حاصل خواهند کرد. ثانیاً اجرای همین الگوریتم‌ها، ساعت‌ها و گاهی چند شبانه‌روز به طول انجامیده و لذا منطقی نیست که زمان بیشتری صرف الگوریتم‌های دیگر شود. با توجه به این مقدمه، نمودارهای تولیدشده با استفاده از مدل یادگیرنده SVM و روش انتخاب ویژگی الگوریتم‌های بهینه‌سازی (شکل‌های ۲ و ۳)، عملکرد الگوریتم‌ها را بر اساس خطا،  $RMSE$  و نمرات همبستگی با یکدیگر مقایسه می‌کند. برای هر امتیاز، نمودارهای همگرایی<sup>۱</sup>، میانگین همگرایی<sup>۲</sup> و پایداری<sup>۳</sup> نشان داده شده است. هدف از ارائه این شکل‌ها این است که آیا الگوریتم‌ها به درستی پیاده‌سازی شده‌اند یا خیر؟ معیار میانگین همگرایی برای این است که وقتی تعداد تکرارها<sup>۴</sup> یا زمان اجرای<sup>۵</sup> اختصاص داده شده به الگوریتم‌ها افزایش می‌یابد، پاسخ‌ها باید بهبود یابند. علاوه بر همگرایی، مفهوم میانگین همگرایی نیز وجود دارد. تفاوت این دو در این است که همگرایی با استخراج بهترین پاسخ در پایان هر تکرار به دست می‌آید، درحالی‌که میانگین همگرایی بر اساس میانگین نمرات راه‌حل بالقوه در پایان هر تکرار محاسبه می‌شود. پایداری نیز نتایج الگوریتم‌ها در گذر زمان و بعد اجراهای بیشتر را بازگو می‌کند.

همان‌طور که قابل مشاهده است، تمام پاسخ‌های بالقوه تولیدشده توسط الگوریتم‌ها (WCC، LCA و FOA) به جز LA با افزایش تکرار بهبود می‌یابند. در شکل (۲) مشخص است که برای معیار همگرایی خطا، با

<sup>1</sup>-Convergence

<sup>2</sup>-Average Convergence

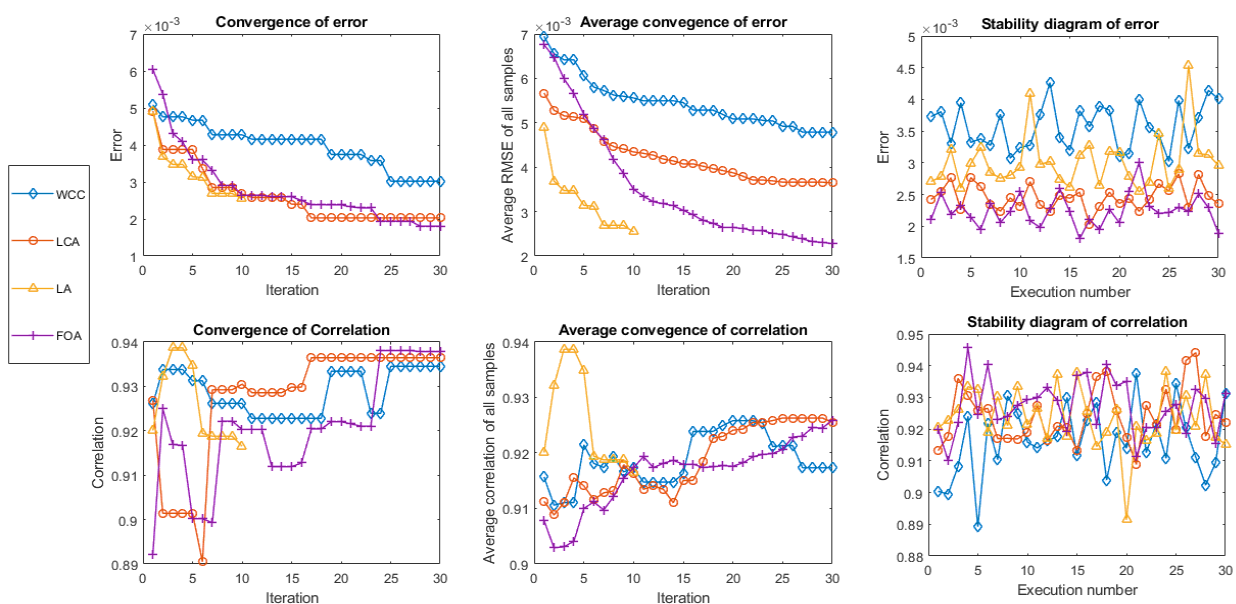
<sup>3</sup>-Stability

<sup>4</sup>-Iteration

<sup>5</sup>-Execution

داد. نمودارهای پایداری نیز نشان می‌دهد که اگر الگوریتمی بهتر از سایر الگوریتم‌ها باشد، نتایج آن در نمودار به جلو بوده و میانگین نتایج آن بهتر از الگوریتم‌های دیگر است. در بخش پایداری می‌توان یافت که الگوریتم FOA و LCA کمترین نوسان را داشته و با اجرای بیشتر، مقداری خطا-همبستگی آن کاهش و در نهایت ثابت شده است. آماره خطا و همبستگی بهترین الگوریتم‌ها به ترتیب در حدود  $0.002-0.0025$  و  $-0.94$  است. لازم به ذکر است که هرچند الگوریتم WCC نتایج بهتری نسبت به LA گزارش کرده است اما نسبت به دو الگوریتم منتخب دارای ضعف‌هایی است. یافته‌های سایر محققین نیز همسو با نتایج این پژوهش است (کاشان، ۲۰۱۴؛ الوشاه، ۲۰۲۱).

تکرار زیاد (۳۰ بار) مقدار خطا برای همه الگوریتم‌ها کاهش داشته و به یک حد ثابتی می‌رسد اما الگوریتم LA روندی غیر از این را طی کرده و از یک مقدار مشخص تکرار به بعد، دیگر کارایی نداشته است. در بخش همگرایی، بهترین مدل‌ها خطای زیر  $0.002$  و همبستگی بالای  $0.93$  را به ثبت رساندند. در معیار همگرایی همبستگی نیز همین روند طی شده است. در بخش میانگین همگرایی خطا و همبستگی، با تکرار زیاد الگوریتم LA به نتیجه معقول و منطقی نرسیده و الگوریتم WCC در عین اینکه به حد ثابتی رسیده اما مقدار RMSE آن بیشتر از الگوریتم LCA و FOA است. در بخش میانگین همگرایی، بهترین مدل (FOA) معیار RMSE کمتر از  $0.003$  و همبستگی بالای  $0.92$  را نشان



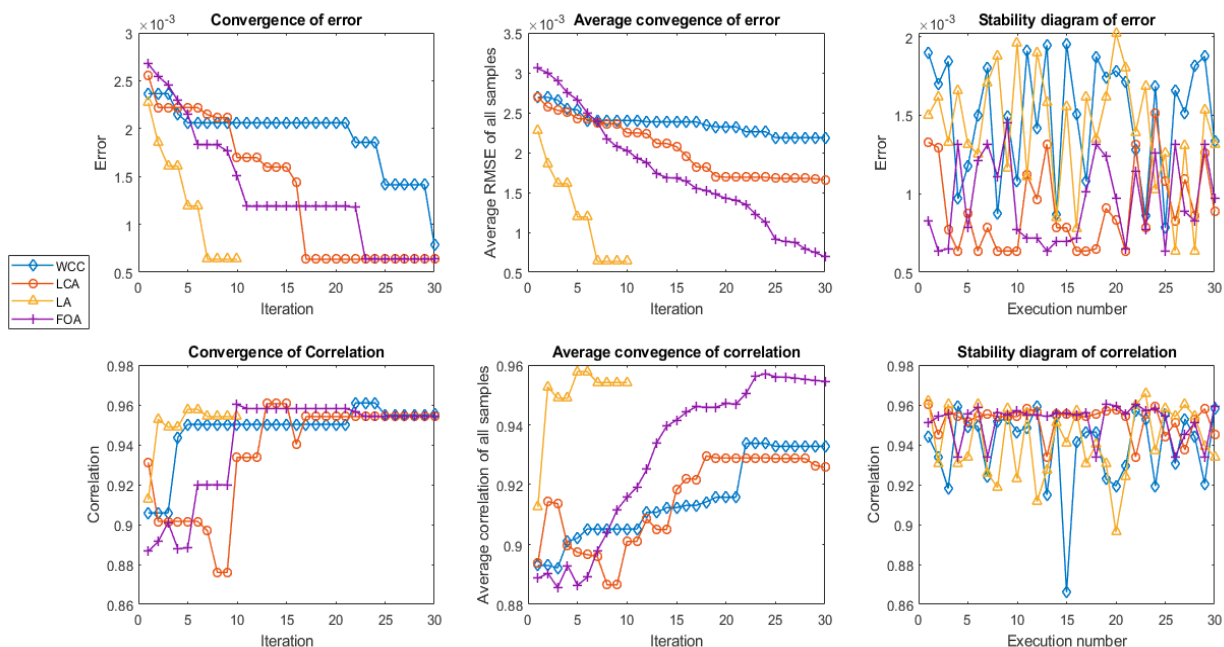
شکل ۲- نتایج کارایی الگوریتم‌های منتخب برای مسائل رگرسیون در بخش کل ویژگی‌ها

و FOA) در حدود  $0.0006$  بوده و همبستگی بهترین الگوریتم بالای  $0.95$  به دست آمده که نشان از آموزش خوب الگوریتم و دقت برآورد/تخمین آن در طول تکرار زیاد است. در بخش ویژگی‌های فاز قبل طراحی، الگوریتم FOA با اختلاف بهتر از سایرین بوده و با LCA و WCC نیز رقابتی ندارد. نتایج میانگین همگرایی خطا-همبستگی نیز حاکی از آن است که بهترین الگوریتم

در شکل (۳) که نتایج الگوریتم‌های منتخب برای ویژگی‌های فاز قبل طراحی مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای به نمایش در آمده است، همانند شکل (۲)، معیارهای همگرایی، میانگین همگرایی و پایداری تحلیل و تفسیر شده است. در بخش همگرایی خطا و همبستگی، به‌غیر از نتیجه الگوریتم LA، بقیه همگرا شده و به حد ثابت رسیده‌اند. خطای الگوریتم‌های برتر (LCA، WCC)

نتایج این بخش مشخص است که الگوریتم‌ها آموزش مناسبی دیده و این خود را در آماره‌های همگرایی، میانگین همگرایی و پایداری در دو قسمت خطا و همبستگی نشان داده است (شکل‌های ۲ و ۳). این بخش با نتایج سایر محققین مطابقت دارد (رستگار و همکاران، ۲۰۰۵؛ قائمی و فیضی-درخشی، ۲۰۱۶).

آموزش دیده شده FOA بوده و خطای آن نسبت به دیگر الگوریتم‌ها حداقل (۰/۰۰۰۷) و همبستگی آن حداکثر (۰/۹۶) شد. آماره پایداری نیز بعد از ۳۰ بار اجرا نشان داد که الگوریتم‌های FOA و LCA با اختلاف بهتر از دو الگوریتم LA و WCC بوده و ضمن حداقل بودن مقدار خطا، همبستگی بالای ۰/۹۵ را ثبت کردند (شکل ۳).



شکل ۳- نتایج کارایی الگوریتم‌های منتخب برای مسائل رگرسیون در بخش ویژگی‌های فاز قبل طراحی

معیار اصلی و چندین زیر معیار دارد. معیارهای اصلی شامل آماره خطا (ER)، همبستگی (CR)، تعداد ویژگی انتخابی (NOF) و زمان سپری شده (ET) بوده و زیرمعیارهای انحراف استاندارد (STD)، فاصله اطمینان (CI)، مقدار احتمال (P-Value) و رد یا پذیرش فرضیه (TS) برای معیارهای خطا و همبستگی تکرار شده است. در جدول (۳)، تعداد ویژگی‌هایی (NOF) که هر الگوریتم آن‌ها را شناسایی کرده، نشان داده شده و برای این پژوهش پارامتر بسیار مهمی است. تعداد ویژگی‌های انتخاب شده بر اساس نتایج الگوریتم‌های WCC، LCA، LA و FOA به ترتیب برابر با ۱۸، ۱۱، ۱۵ و ۸ به دست آمده است. پس از ۳۰ بار اجرای الگوریتم‌ها، بهترین نتایج برای هر کدام انتخاب شد. ET نشان می‌دهد که چه مدت زمان بر حسب ثانیه در اجرای آن سپری شده

### ارزیابی الگوریتم‌ها در مرحله آزمایش

نتایج نشان داده شده در جداول (۳) و (۴) به ترتیب بر اساس نتایج پایداری شکل‌های (۲) و (۳) محاسبه شده است. بعد از بررسی چگونگی آموزش مدل‌ها و همچنین بررسی کارایی الگوریتم‌های مختلف و شناسایی آن در دو بخش قبلی، در نهایت نوبت به آزمایش (صحت‌سنجی یا آزمون) و انتخاب ویژگی با این الگوریتم‌ها و ارائه نتایج معیارهای ارزیابی رسیده است. در جدول (۳)، نتایج معیارهای ارزیابی انتخاب مهم‌ترین ویژگی‌ها از بین کل ویژگی‌های مؤثر بر هزینه کل سامانه‌های آبیاری قطره‌ای (۳۹ ویژگی) در مرحله صحت‌سنجی ارائه شده است. چون این جدول برگرفته از نتایج پایداری شکل‌های قبلی هست، بر این اساس دو

تصادفی بیان می‌کند. الگوریتمی با حداقل مقدار P-Value، قابل اعتمادتر از بقیه است. در نتایج نیز مشخص است که هر چهار الگوریتم قابل اعتماد بود و دلیل آن تمایل به صفر مقدار این آماره است. زیرمعیار خطا-رد یا پذیرش فرضیه (ER-TS) معمولاً برای رد یا پذیرش یک فرضیه صفر استفاده می‌شود. هنگامی که TS حداکثر باشد، مقدار P حداقل است. نکته جالب توجه، معیار همبستگی (CR) نزدیک و بالای همه الگوریتم‌ها بوده و مقدار آن برای چهار الگوریتم WCC، LCA، LA و FOA به ترتیب برابر با ۰/۹۳۴۵، ۰/۹۳۶۵، ۰/۹۱۶۵ و ۰/۹۳۷۸ به دست آمد و نشان از توانایی و دقت این الگوریتم‌ها به ویژه دو الگوریتم LCA و FOA دارد. دیگر زیرمعیارها که در بخش خطا مورد بررسی و تحلیل قرار گرفت، برای همبستگی (CR) نیز تکرار شده و روند نتایج همانند زیرمعیارهای خطا (ER) است. به دلیل حجم بیش از حد مجاز مقاله، از ارائه توضیحات بیشتر خودداری می‌شود (جدول ۳). نتایج این بخش با یافته‌های شوبرت و همکاران (۲۰۱۷)، پاندی و همکاران (۲۰۱۸)، مسعودی سبحان‌زاده و همکاران (۲۰۱۹) و متل و همکاران (۲۰۲۲) مطابقت دارد.

تا بهترین نتیجه برای یک الگوریتم به دست آید. الگوریتم‌ها به دلیل طی مراحل مختلف در اجرا، دارای ET‌های متفاوتی هستند. معیار خطا (ER) نیز مهر تأییدی بر الگوریتم‌های شناسایی شده در مرحله قبل بوده و مقدار آن برای الگوریتم‌های برتر LCA و FOA به ترتیب ۰/۰۰۲۰ و ۰/۰۰۱۸ به دست آمده و از بقیه کمتر است. زیرمعیار خطا-انحراف استاندارد (ER-STD) نشان می‌دهد که نتایج چقدر با میانگین نتایج متفاوت است؛ بنابراین مطلوب است که این معیار حداقل باشد و در نتایج نیز مشخص است که برای الگوریتم‌های WCC و LA برابر با ۰/۰۰۰۴ و الگوریتم‌های LCA و FOA برابر با ۰/۰۰۰۲ حاصل شد. زیرمعیار خطا-فاصله اطمینان (ER-CI)، محدوده‌ای از مقادیر را نشان می‌دهد و انتظار می‌رود نتایج با حداکثر احتمال خاص در این محدوده قرار گیرند و برای حصول افزایش دقت، این معیار دو بار تکرار شد. در هر دو بار، نتیجه تقریباً مشابهی گزارش و بازهم دو الگوریتم برتر LCA و FOA کمترین مقدار را داشتند. زیرمعیار خطا-مقدار احتمال (ER-P) یکی از مهم‌ترین آماره‌ها در ارزیابی نتایج مدل‌ها و الگوریتم‌هاست. P-Value میزان شباهت نتایج به دست آمده را با مقادیر

جدول ۳- نتایج معیارهای ارزیابی انتخاب مهم‌ترین ویژگی‌ها از بین کل ویژگی‌های مؤثر بر هزینه کل سامانه‌های آبیاری قطره‌ای

| الگوریتم‌های منتخب |          |          |          | نماد    | معیار                     |
|--------------------|----------|----------|----------|---------|---------------------------|
| FOA                | LA       | LCA      | WCC      |         |                           |
| ۸                  | ۱۵       | ۱۱       | ۱۷       | NOF     | تعداد ویژگی               |
| ۱۵۶/۶۰۶۴           | ۷۵/۰۹۸۸  | ۶۲۷/۲۷۸۱ | ۱۴۴/۳۵۸۵ | ET      | زمان سپری شده             |
| ۰/۰۰۱۸             | ۰/۰۰۲۵   | ۰/۰۰۲۰   | ۰/۰۰۳۰   | ER      | خطا                       |
| ۰/۰۰۰۲             | ۰/۰۰۰۴   | ۰/۰۰۰۲   | ۰/۰۰۰۴   | ER_STD  | خطا-انحراف استاندارد      |
| ۰/۰۰۲۲             | ۰/۰۰۲۹   | ۰/۰۰۲۴   | ۰/۰۰۳۴   | ER_CI_1 | خطا-فاصله اطمینان ۱       |
| ۰/۰۰۲۳             | ۰/۰۰۳۲   | ۰/۰۰۲۵   | ۰/۰۰۳۷   | ER_CI_2 | خطا-فاصله اطمینان ۲       |
| ۰/۰۰۰۰             | ۰/۰۰۰۰   | ۰/۰۰۰۰   | ۰/۰۰۰۰   | ER_P    | خطا-مقدار احتمال          |
| ۴۹/۷۶۱۴            | ۳۸/۴۵۵۷  | ۶۹/۱۱۵   | ۵۵/۵۲۳۶  | ER_TS   | خطا-رد یا پذیرش فرضیه     |
| ۰/۹۳۷۸             | ۰/۹۱۶۵   | ۰/۹۳۶۵   | ۰/۹۳۴۵   | CR      | همبستگی                   |
| ۰/۰۰۸۶             | ۰/۰۰۹۶   | ۰/۰۰۸۸   | ۰/۰۱۱۲   | CR_STD  | همبستگی-انحراف استاندارد  |
| ۰/۹۲۴۰             | ۰/۹۲۰۱   | ۰/۹۲۰۸   | ۰/۹۱۲۳   | CR_CI_1 | همبستگی-فاصله اطمینان ۱   |
| ۰/۹۳۰۴             | ۰/۹۲۷۲   | ۰/۹۲۷۳   | ۰/۹۲۰۷   | CR_CI_2 | همبستگی-فاصله اطمینان ۲   |
| ۰/۰۰۰۰             | ۰/۰۰۰۰   | ۰/۰۰۰۰   | ۰/۰۰۰۰   | CR_P    | همبستگی-مقدار احتمال      |
| ۵۸۹/۳۲۳۲           | ۵۲۸/۹۵۶۳ | ۵۷۶/۷۵۷۸ | ۴۴۶/۳۴۷۹ | CR_TS   | همبستگی-رد یا پذیرش فرضیه |

۰/۰۰۰۳ برای هر دو الگوریتم، زیرمعیار خطا-فاصله اطمینان (ER-CI) به ترتیب ۰/۰۰۰۸ و ۰/۰۰۰۹ در تکرار اول و ۰/۰۰۱۰ و ۰/۰۰۱۱ در تکرار دوم، زیرمعیار خطا-مقدار احتمال (ER-P) برابر با صفر برای هر دو الگوریتم، زیرمعیار خطا-رد یا پذیرش فرضیه (ER-TS) به ترتیب ۱۸/۷۷۰۶ و ۱۹/۲۵۱۱ و معیار همبستگی (CR) برابر با ۰/۹۵۴۱ برای هر دو الگوریتم به دست آمد. همین روند که در زیرمعیارهای خطا طی شد، برای همبستگی نیز صدق می‌کند (جدول ۴). نتایج این بخش با یافته‌های دیگر پژوهشگران همخوانی دارد (قائم و فیضی-درخشی، ۲۰۱۶؛ مسعودی سبحان‌زاده و همکاران، ۲۰۱۹).

همچنین در جدول (۴)، نتایج معیارهای ارزیابی انتخاب مهم‌ترین ویژگی‌ها از بین ویژگی‌های فاز قبل طراحی مؤثر بر هزینه کل سامانه‌های آبیاری قطره‌ای (۱۸ ویژگی) نشان داده شده است. به غیر از معیار تعداد ویژگی انتخابی (NOF) که برای همه الگوریتم‌ها (WCC، LCA، LA و FOA) برابر ۶ انتخاب از بین ۱۸ ویژگی شد، در بقیه معیار و زیرمعیارها، الگوریتم‌های LCA و FOA دقت و همبستگی بالاتر و خطای کمتری داشتند. برای دو الگوریتم منتخب LCA و FOA، معیار زمان سپری‌شده (ET) به ترتیب ۳۴۴/۵۷۴۰ و ۱۵۳/۷۳۸۶ ثانیه، خطا (ER) برابر با ۰/۰۰۰۶ برای هر دو الگوریتم، زیرمعیار خطا-انحراف استاندارد (ER-STD) برابر با

جدول ۴- نتایج معیارهای ارزیابی انتخاب مهم‌ترین ویژگی‌ها از بین ویژگی‌های فاز قبل طراحی مؤثر بر هزینه کل سامانه‌های آبیاری قطره‌ای

| الگوریتم‌های منتخب |         |          |          | نماد    | معیار                     |
|--------------------|---------|----------|----------|---------|---------------------------|
| FOA                | LA      | LCA      | WCC      |         |                           |
| ۶                  | ۶       | ۶        | ۶        | NOF     | تعداد ویژگی               |
| ۱۵۳/۷۳۸۶           | ۷۳/۶۷۹۴ | ۳۴۴/۵۷۴۰ | ۱۴۷/۹۲۹۷ | ET      | زمان سپری‌شده             |
| ۰/۰۰۰۶             | ۰/۰۰۰۶  | ۰/۰۰۰۶   | ۰/۰۰۰۸   | ER      | خطا                       |
| ۰/۰۰۰۳             | ۰/۰۰۰۴  | ۰/۰۰۰۳   | ۰/۰۰۰۴   | ER_STD  | خطا-انحراف استاندارد      |
| ۰/۰۰۰۹             | ۰/۰۰۱۳  | ۰/۰۰۰۸   | ۰/۰۰۱۴   | ER_CI_1 | خطا-فاصله اطمینان ۱       |
| ۰/۰۰۱۱             | ۰/۰۰۱۵  | ۰/۰۰۱۰   | ۰/۰۰۱۶   | ER_CI_2 | خطا-فاصله اطمینان ۲       |
| ۰/۰۰۰۰             | ۰/۰۰۰۰  | ۰/۰۰۰۰   | ۰/۰۰۰۰   | ER_P    | خطا-مقدار احتمال          |
| ۱۹/۲۵۱۱            | ۲۰/۷۷۱  | ۱۸/۷۷۰۶  | ۲۱/۷۰۷۵  | ER_TS   | خطا-رد یا پذیرش فرضیه     |
| ۰/۹۵۴۱             | ۰/۹۵۴۱  | ۰/۹۵۴۱   | ۰/۹۵۵۲   | CR      | همبستگی                   |
| ۰/۰۰۸۷             | ۰/۰۱۷۶  | ۰/۰۰۷۰   | ۰/۰۱۹۸   | CR_STD  | همبستگی-انحراف استاندارد  |
| ۰/۹۴۸۷             | ۰/۹۳۴۴  | ۰/۹۴۹۷   | ۰/۹۳۱۵   | CR_CI_1 | همبستگی-فاصله اطمینان ۱   |
| ۰/۹۵۵۳             | ۰/۹۴۷۶  | ۰/۹۵۴۹   | ۰/۹۴۶۳   | CR_CI_2 | همبستگی-فاصله اطمینان ۲   |
| ۰/۰۰۰۰             | ۰/۰۰۰۰  | ۰/۰۰۰۰   | ۰/۰۰۰۰   | CR_P    | همبستگی-مقدار احتمال      |
| ۵۹۷/۱۹۹۳           | ۲۹۲/۸۶  | ۷۴۹/۳۱۳۶ | ۲۵۹/۴۲۲۳ | CR_TS   | همبستگی-رد یا پذیرش فرضیه |

سامانه‌های آبیاری تحت فشار و بالأخص قطره‌ای را انجام داد. در بخش انتخاب ویژگی از بین کل ویژگی‌ها (۳۹ عدد)، نتایج الگوریتم‌های WCC، LCA، LA و FOA به ترتیب برابر با ۱۷، ۱۱، ۱۵ و ۸ به دست آمد و برای انتخاب ویژگی از بین ویژگی‌های فاز قبل طراحی (۱۸ عدد)، الگوریتم‌های WCC، LCA، LA و FOA همگی به اتفاق دست روی شش ویژگی گذاشتند. بر همین اساس

#### انتخاب برترین ویژگی‌ها

بعد از صحت‌سنجی (یا آزمایش) نتایج الگوریتم‌ها و بررسی معیارهای ارزیابی، ضمن تأیید این نتایج، ویژگی‌های مدنظر استخراج شد (جدول ۵). آنچه این پژوهش به دنبال آن بود، انتخاب مجموعه‌ای از مفیدترین، بهترین، مؤثرترین، مهم‌ترین و دقیق‌ترین ویژگی‌هاست که بتوان با آن مدل‌سازی هزینه اولیه



و یا پیمانکاری نیز می‌تواند بر روی این ادعا صحه بگذارد. آنچه در این تحقیق مهم بود، ابتدا به‌کارگیری روش‌های جدید انتخاب ویژگی بود که بدین‌منظور از کدنویسی در نرم‌افزار MATLAB استفاده شد. سپس الگوریتم‌ها، مدل‌ها و روش‌های مختلف انتخاب ویژگی، آموزش و آزمایش شد. در نهایت ویژگی‌های برتر از دل آن استخراج و در جدول (۵) ارائه شد.

و در این ایستگاه آخر، ویژگی‌های برتر در هر بخش مشخص شد. ویژگی‌های برتر از ترکیب کردن ویژگی‌های مشترک در هر بخش، حاصل شد. با یک نگاه کارشناسانه مشخص است که ویژگی‌های منتخب در هر دو بخش، در نوع طراحی و اجرای سامانه‌ها بسیار حیاتی بوده و نقش اساسی در هزینه‌ها ایفا می‌کند. حتی هر فرد غیرحرفه‌ای و یا یک کارشناس کم‌تجربه در شرکت‌های مهندسی مشاور

#### جدول ۵- انتخاب ویژگی‌های برتر از بین کل ویژگی‌ها و ویژگی‌های فاز قبل طراحی مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای

| بخش   | ویژگی‌ها و نماد آن   |
|---|--|
| کل ویژگی‌ها<br>(انتخاب ۱۰ از ۳۹ ویژگی)            | توان پمپ مورد نیاز (PP (kw))، طول لوله جانبی (L125mm (m))، مساحت زمین (A (ha))، مقدار کل دبی آب قابل‌دسترس (QT (l/s))، طول لوله جانبی (L160mm (m))، طول لوله جانبی (L110mm (m))، دبی سامانه (QS (l/s))، ظرفیت نگهداری آب در خاک (AW (mm/m))، خط انتقال طولانی (L200mm (m)) و مساحت واحدهای آبیاری (AIU (ha)) |
| ویژگی‌های فاز قبل طراحی<br>(انتخاب ۷ از ۱۸ ویژگی) | مقدار کل دبی آب قابل‌دسترس (QT (l/s))، مساحت زمین (A (ha))، شکل زمین (نسبت محیط به مساحت) (P/A (1/m))، طول کل لترال (L16mm (m))، فاصله منبع آبی تا زمین (DSF (m))، تعداد قطعات زمین (NPL) و دبی گسیلنده (QE (l/s))   |

#### نتیجه‌گیری

هزینه بالا می‌شد. (۲) نتایج کاربرپسند نبوده و هر فرد نمی‌توانست از روابط مستخرج بهره کافی را ببرد. (۳) از همه مهم‌تر اینکه استفاده از کل ویژگی‌ها، منجر به نتایج بسیار پیچیده و غیرکاربردی می‌شد. حال با پیشرفت فناوری اطلاعات، رویکرد انتخاب ویژگی به دنبال تسهیل در همین روند است.

استفاده از روش‌های دانش‌مبنا و عددی که دسترسی به تمام پارامترهای مورد نیاز آن‌ها، مشکل و یا اندازه‌گیری آن‌ها نیازمند صرف هزینه و زمان زیادی است، کمتر مورد توجه پژوهشگران قرار گرفته است. در عوض از مدل‌های هوش محاسباتی داده‌مبنا که از دقت و اعتبار بالایی برخوردار هستند و پارامترهای ورودی کمتر و در دسترس‌تر نیاز دارند، استفاده می‌شود. به‌طور سنتی، یک مدل‌ساز برای ساخت مدل‌های ریاضی مانند شبکه عصبی مصنوعی برای ترکیبات مختلف ورودی باید از آزمون‌وخطا استفاده کند که این بسیار وقت‌گیر است زیرا مدل‌ساز نیاز به آموزش و آزمایش مدل‌های مختلف با تمام ترکیبات ورودی احتمالی دارد. الگوریتم‌های مختلف یادگیری ماشین این امکان را فراهم می‌سازد که بتوان با صرف کمترین زمان و هزینه، بهترین داده‌ها را با بالاترین معیارهای ارزیابی شناسایی کرده و یک مدل‌سازی دقیق و کم‌خطا حاصل کرد. از سویی، در گذشته و هنگام استفاده از روابط خطی و رگرسیونی، مدل‌های هوش مصنوعی و به‌طورکلی الگوریتم‌های یادگیری ماشین، از تمام ویژگی‌ها و متغیرها استفاده می‌شد. این کار چند عیب اساسی داشت؛ (۱) باعث طولانی شدن زمان اجرا و صرف زمان-

به‌طور خلاصه، نتایج این مطالعه حاکی از آن بود که مدل ماشین بردار پشتیبان (SVM) به‌عنوان بهترین یادگیرنده شناسایی شد. روش الگوریتم‌های بهینه‌سازی (Wrapper)، بهترین روش انتخاب ویژگی شد. الگوریتم‌های LCA و FOA نیز بهترین راه برای رسیدن به نتایج مطلوب (خطای کم و همبستگی بالا) برای انتخاب مهم‌ترین ویژگی‌ها بودند. نتایج انتخاب ویژگی نیز نشان داد که از بین کل ویژگی‌های مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای (۳۹ ویژگی)، ۱۰ ویژگی شامل توان پمپ مورد نیاز (PP (kw))، طول لوله جانبی (L125mm (m))، مساحت زمین (A (ha))، مقدار کل دبی آب قابل‌دسترس (QT (l/s))، طول لوله جانبی (L160mm (m))، طول لوله جانبی (L110mm (m))، دبی سامانه (QS (l/s))، ظرفیت نگهداری آب در خاک (AW (mm/m))، خط انتقال طولانی (L200mm (m)) و مساحت واحدهای

آبیاری پیش از انجام کار می‌توان درک درستی از میزان هزینه‌ها داشت و بودجه‌بندی برای تخصیص اعتبار را به‌درستی مدیریت کرد.

#### تقدیر و تشکر

این مقاله مستخرج از رساله دکتری نویسنده اول است. بدین‌وسیله از شرکت مهندسی مشاور آب‌و‌خاک البرز و معاونت آب‌و‌خاک وزارت جهاد کشاورزی به دلیل در اختیار قرار دادن داده‌ها و همچنین گروه مهندسی آبیاری و آبادانی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران به دلیل تأمین امکانات لازم برای انجام این پژوهش و تهیه مقالات مربوطه تشکر و قدردانی می‌شود. از آقای دکتر یوسف مسعودی سبحان‌زاده و آقای مهندس ایمان حاجی‌راد نیز به دلیل کمک‌های بی‌دریغشان، کمال قدردانی به عمل می‌آید.

آبیاری (AIU (ha)) به‌عنوان برترین ویژگی‌ها شناسایی شد. در بخش ویژگی‌های مؤثر بر هزینه سامانه‌های آبیاری قطره‌ای در فاز قبل از طراحی (۱۸ ویژگی) نیز، هفت ویژگی شامل مقدار کل دبی آب قابل‌دسترس ( $Q_T (l/s)$ )، مساحت زمین (A (ha))، شکل زمین (نسبت محیط به مساحت) ( $P/A (1/m)$ )، طول کل لترال ( $L_{16mm} (m)$ )، فاصله منبع آبی تا زمین ( $D_{SF} (m)$ )، تعداد قطعات زمین (NPL) و دبی گسیلنده ( $Q_E (l/s)$ ) به‌عنوان مؤثرترین ویژگی‌ها بر هزینه سامانه‌های آبیاری قطره‌ای انتخاب شد. نتایج کلی این پژوهش نشان داد که برای سهولت مدل‌سازی هزینه‌های سامانه‌های آبیاری قطره‌ای می‌توان به نتایج این پژوهش تکیه کرد، مدل‌های بهینه را به دست آورد و مدل‌هایی را برای این کار توسعه داد. طبیعتاً آن مدل دینامیک بوده و با استفاده از تورم همان سال، اجرا خواهد شد. لذا تحلیل اقتصادی آن قابل استناد، دقیق و به‌روز خواهد بود. سپس با مدل‌سازی هزینه سامانه‌های

#### فهرست منابع

1. Ahiaga-Dagbui, D. D., and Smith, S. D. 2012. Neural networks for modelling the final target cost of water projects.
2. Ahmadaali, K., Liaghat, A., Heydari, N., and Bozorg-Haddad, O. 2013. Application of artificial neural network and adaptive neural-based fuzzy inference system techniques in estimating of virtual water. *International Journal of Computer Application*, 76: 12-19.
3. Alshahethi, A. A. A., and Radhika, K. L. 2018. Estimating the Final Cost of Construction Project Using Neural Networks: A Case of Yemen Construction Projects. *International Journal for Research in Applied Science & Engineering Technology*, 6(11): 2141-2151.
4. Altarabichi, M. G., Nowaczyk, S., Pashami, S., and Mashhadi, P. S. 2023. Fast Genetic Algorithm for feature selection—A qualitative approximation approach. *Expert Systems with Applications*, 211: 118528.
5. Alweshah, M. 2021. Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm. *Applied Intelligence*, 51(6): 4058-4081.
6. Arora, S., and Mishra, N. 2017. Software cost estimation using single layer artificial neural network. *International Journal of Advanced Engineering Research and Science*, 4(9): 237250.
7. Arora, S., and Mishra, N. 2018. Software cost estimation using artificial neural network. In *Soft Computing: Theories and Applications* (pp. 51-58). Springer, Singapore.
8. Awad, M., and Khanna, R. 2015. Support vector regression. In *efficient learning machines* (pp. 67-80). Apress, Berkeley, CA.
9. Babaei, M., Rashidi-baqhi, A., and Rashidi, M. 2022. Estimating Project Cost under Uncertainty Using Universal Generating Function Method. *Journal of Construction Engineering and Management*, 148(2): 04021194.

10. Chandanshive, V., and Kambekar, A. R. 2019. Estimation of building construction cost using artificial neural networks. *Journal of Soft Computing in Civil Engineering*, 3(1): 91-107.
11. Chandrashekar, G., and Sahin, F. 2014. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1): 16-28.
12. Cheng, M. Y., Tsai, H. C., and Sudjono, E. 2010. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction industry. *Expert Systems with Applications*, 37(6): 4224-4231.
13. Cortes, C., and Vapnik, V. 1995. Support-vector networks. *Machine learning*, 20(3): 273-297.
14. Drenthe, N. T., Zandbergen, B. T. C., Curran, R., and Van Pelt, M. O. 2019. Cost estimating of commercial smallsat launch vehicles. *Acta Astronautica*, 155: 160-169.
15. Elfaki, A. O., Alatawi, S., and Abushandi, E. 2014. Using intelligent techniques in construction project cost estimation: 10-year survey. *Advances in Civil Engineering*, 2014: 1-11.
16. Elhag, T. M. S., and Boussabaine, A. H. 1998. An artificial neural system for cost estimation of construction projects. In *14th Annual ARCOM Conference* (Vol. 1, pp. 219-226). University of Reading: Association of Researchers in Construction Management.
17. Ghaddar, B., and Naoum-Sawaya, J. 2018. High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, 265(3): 993-1004.
18. Ghaemi, M., and Feizi-Derakhshi, M. R. 2016. Feature selection using forest optimization algorithm. *Pattern Recognition*, 60: 121-129.
19. Gransberg, D. D., and Rueda, J. A. 2020. *Construction equipment management for engineers, estimators, and owners*. CRC Press.
20. Kashan, A. H. 2014. League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships. *Applied Soft Computing*, 16: 171-200.
21. Kiani, A., and Shaker, M. 2022. Evaluating the Effectiveness of Pressurized Irrigation Systems in Iran. *Water Management in Agriculture*, 8(2): 167-182. (In Persian)
22. Kim, G. H., Shin, J. M., Kim, S., and Shin, Y. 2013. Comparison of school building construction costs estimation methods using regression analysis, neural network, and support vector machine. *Journal of Building Construction and Planning Research*, 1(1): 1-7.
23. Lester, E. I. A. 2017. Estimating. In: *Project management, planning and control*. *The Netherlands: Elsevier*, 61-65.
24. Liu, J., Lin, Y., Lin, M., Wu, S., and Zhang, J. 2017. Feature selection based on quality of information. *Neurocomputing*, 225: 11-22.
25. Masoudi-Sobhanzadeh, Y., and Motieghader, H. 2016. World Competitive Contests (WCC) algorithm: A novel intelligent optimization algorithm for biological and non-biological problems. *Informatics in Medicine Unlocked*, 3: 15-28.
26. Masoudi-Sobhanzadeh, Y., Motieghader, H., & Masoudi-Nejad, A. 2019. FeatureSelect: a software for feature selection based on machine learning approaches. *BMC Bioinformatics*, 20(1): 1-17.
27. Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., and Voordijk, H. 2022. An artificial neural network approach for cost estimation of engineering services. *International Journal of Construction Management*, 22(7): 1274-1287.
28. Metin, S. K. 2018. Feature selection in multiword expression recognition. *Expert Systems with Applications*, 92: 106-123.
29. Mevellec, P. 2021. Cost systems: A new approach. *Academia Letters*, 2.
30. Miao, J., and Niu, L. 2016. A survey on feature selection. *Procedia Computer Science*, 91: 919-926.
31. Nalbandan, R. B., Delavar, M., Abbasi, H., and Zaghiyan, M. R. 2023. Model-based water footprint accounting framework to evaluate new water management policies. *Journal of Cleaner Production*, 382: 135220.

32. Norvig, P. R., and Intelligence, S. A. 2002. A modern approach. Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. *Knowledge-Based Systems*, 90: 33-48.
33. Omotayo, T., Bankole, A., and Olubunmi Olanipekun, A. 2020. An artificial neural network approach to predicting most applicable post-contract cost controlling techniques in construction projects. *Applied Sciences*, 10(15): 5171-5195.
34. Panday, D., de Amorim, R. C., and Lane, P. 2018. Feature weighting as a tool for unsupervised feature selection. *Information processing letters*, 129: 44-52.
35. Pazoki, M., Yadav, A., and Abdelaziz, A. Y. 2020. Pattern-recognition methods for decision-making in protection of transmission lines. In *Decision making applications in modern power systems* (pp. 441-472). Academic Press.
36. Pourgholam-Amiji, M., Ahmadaali, K., and Liaghat, A. 2021a. Sensitivity Analysis of Parameters Affecting the Early Cost of Drip Irrigation Systems Using Meta-Heuristic Algorithms. *Iranian Journal of Irrigation & Drainage*, 15(4): 737-756. (In Persian)
37. Pourgholam-Amiji, M., Liaghat, A., and Ahmadaali, K. 2021b. Early Stage Cost Modeling of Drip Irrigation Systems. *Irrigation and Drainage Structures Engineering Research*, 22(82): 1-22. (In Persian)
38. Rahmaninia, M., and Moradi, P. 2018. OSFSMI: online stream feature selection method based on mutual information. *Applied Soft Computing*, 68: 733-746.
39. Rastegar, R., Rahmati, M., and Meybodi, M. R. 2005. A clustering algorithm using cellular learning automata based evolutionary algorithm. In *Adaptive and Natural Computing Algorithms* (pp. 144-150). Springer, Vienna.
40. Roxas, C. L. C., and Ongpeng, J. M. C. 2014. An artificial neural network approach to structural cost estimation of building projects in the Philippines. *Proc. DLSU Res. Congr.*
41. Schubert, A. L., Hagemann, D., Voss, A., and Bergmann, K. 2017. Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology*, 77: 29-45.
42. Sharma, A., Jain, A., Gupta, P., and Chowdary, V. 2020. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9: 4843-4873.
43. Sheikhpour, R., Sarram, M. A., Gharaghani, S., and Chahooki, M. A. Z. 2017. A survey on semi-supervised feature selection methods. *Pattern Recognition*, 64: 141-158.
44. Solorio-Fernández, S., Carrasco-Ochoa, J. A., and Martínez-Trinidad, J. F. (2020): A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2): 907-948.
45. Talukdar, S., Naikoo, M. W., Mallick, J., Praveen, B., Sharma, P., Islam, A. R. M. T. ... and Rahman, A. 2022. Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning based sensitivity analysis for agricultural suitability mapping. *Agricultural Systems*, 196: 103343.
46. Teksin, S., Azginoglu, N., and Akansu, S. O. 2022. Structure estimation of vertical axis wind turbine using artificial neural network. *Alexandria Engineering Journal*, 61(1): 305-314.
47. Thakkar, A., and Lohiya, R. 2023. Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System. *Information Fusion*, 90: 353-363.
48. Venkatachalam, A. R. 1993. Software cost estimation using artificial neural networks. In *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)* (Vol. 1, pp. 987-990). IEEE.
49. Waliulu, Y. E. P. R., and Adi, T. J. W. 2022. A system dynamic thinking for modeling infrastructure project duration acceleration. *Procedia Computer Science*, 197: 420-427.
50. Winston, P. H. 1992. *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc.
51. Yadav, R., Vyas, M., Vyas, V., and Agrawal, S. 2016. Cost estimation model (CEM) for residential building using artificial neural network. *International Journal of Engineering Research & Technology (IJERT)*, 5(1): 430-432.

## Identifying the Features Affecting the Cost of Drip Irrigation Systems Using Feature Selection Methods

**M. Pourgholam-Amiji, K. Ahmadaali<sup>1</sup>, and A. Liaghat**

Ph.D. Candidate, Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. [Mpourgholam6@ut.ac.ir](mailto:Mpourgholam6@ut.ac.ir)

Assistant Prof., Department of Arid and Mountainous Regions Reclamation, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. [Khahmadauli@ut.ac.ir](mailto:Khahmadauli@ut.ac.ir)

Professor, Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. [Aliaghat@ut.ac.ir](mailto:Aliaghat@ut.ac.ir)

Received: December 2022 and Accepted: February 2023

### Abstract

This research aimed to select essential features for modeling the cost of pressurized irrigation systems using the data of 515 drip irrigation projects in four parts, including the cost of pumping station and central control system ( $TC_P$ ), cost of on-farm equipment ( $TC_F$ ), cost of installation and operation on-farm and pumping station ( $TC_I$ ), and total cost ( $TC_T$ ). In the first stage, a database including 39 features influencing the cost of the mentioned sectors was prepared and the price of all projects (2006 to 2019) was updated for the base year of 2021. Then, feature selection was done with different algorithms in MATLAB environment and in two parts including (1) all features (39 features before and after the design stage) and (2) 18 features before the design phase (BD). The results showed that the amounts of RMSE and  $R^2$  for all the features were equal to 0.007 and 0.92, respectively, and for the BD section, they were equal to 0.003 and 0.89, respectively. Among the different algorithms for feature selection, support vector machine (SVM) and optimization algorithms (Wrapper) were identified as the best learner and feature selection method, respectively. The results of the evaluation criteria showed that the two LCA and FOA algorithms achieved the best estimation, and their error criterion in all the features were 0.0020 and 0.0018, respectively, while their correlations were 0.94 and 0.94. In the BD features, these criteria were 0.0006 and 0.95 for both algorithms, respectively. Finally, in the all features section, 10 out of 39 features and for BD section, 8 out of 18 were selected as the most effective features. The results of choosing the most effective features that affect the cost of different parts of the drip irrigation system can make the cost modeling of the systems simpler and faster and, while being useful for research works, it facilitates estimation and management of costs before implementation of each project.

**Keywords:** Economic modeling, Modern irrigation systems, Sensitivity analysis, Pattern recognition, Meta-heuristic algorithms.

---

<sup>1</sup> - Corresponding author: Department of Arid and Mountainous Regions Reclamation, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran.