تخمین پتانسیل تولید انگور با استفاده از مدل‌های ANN و ANN-GA در منطقه ملکان

حسن محمدی[۱*]، مسلم ثروتی[۲]، حسین بیرامی[۳]
۱*- دانشجوی دکتری گروه علوم و مهندسی خاک دانشگاه شهرکرد، ۲- استادیار مرکز آموزش عالی شهید باکری میاندوآب، دانشگاه ارومیه، ۳- استادیار مرکز ملی تحقیقات شوری، سازمان تحقیقات، آموزش و ترویج کشاورزی، یزد

چکیده

هدف از این تحقیق، ارزیابی کارایی الگوریتم ژنتیک (GA[۱]) در مدل‌سازی تخمین پتانسیل تولید انگور با استفاده از شبکه عصبی مصنوعی (ANN[۲]) می‌باشد. بدین منظور ۱۶۰ نمونه از منطقه ملکان اخذ و سپس مدل‌سازی با دو مدل ANN و شبکه عصبی مصنوعی-ژنتیک (ANN-GA) انجام شد. در این تحقیق شبکه‌های عصبی با ساختار ۵ نرون در لایه پنهان و تابع محرک سیگموئید آکسون مناسب‌ترین ساختار بوده است. مدل ANN-GA دارای کارایی بیشتری بوده، به‌طوری‌که ضریب تبیین و میانگین انحراف مربعات خطا و ضریب نش-ساتکلیف به ترتیب در مرحله آموزش ۰/۹۲ ، ۱/۶۱ و ۰/۶۷ و در مرحله آزمون ۰/۹۶ ، ۱/۵۶ و ۰/۶۹ است. الگوریتم ژنتیک به‌عنوان روشی مؤثر در بهینه‌سازی اوزان ویژگی‌های ورودی در شبکه عصبی مصنوعی با به حداقل رساندن خطای شبکه ایفا می‌کند. نهایتاً می‌توان گفت مدل ترکیبی ژنتیک-عصبی به‌عنوان یک مدل کارا در تخمین پتانسیل تولید انگور می‌باشد.

واژه‌های کلیدی: بهینه‌سازی، الگوریتم ترکیبی، ویژگی‌های زودیافت

مقدمه

انگور (*Vitis vinifera L.*) یکی از محصولات باغی مهم است، پیش‌بینی پتانسیل تولید انگور (Rotter *et al.*, 2011) اهمیت زیادی دارد. با این وجود پتانسیل یک محصول مستقیماً تحت تأثیر ویژگی‌های خاکی قرار گرفته که برای تأثیر بهتر برنامه‌های مدیریت باغی شناسایی آن‌ها ضروری است (Yemefack *et al.*, 2005). شناسایی ویژگی‌های خاکی، به دلیل اثرات متقابل بین آن‌ها فرآیند پیچیده‌ای است (Ayoubi *et al.*, 2002). راه‌کارها و روش‌های مختلفی برای حصول به این منظور ارائه شده است که تکنیک‌های محاسباتی نرم (روش‌های فراکاوشی) از جمله شبکه‌های عصبی مصنوعی، عصبی-فازی و الگوریتم‌های ترکیبی از این جمله هستند (Shukla & Ebinger., 2004). پتانسیل تولید انگور تحت تأثیر عوامل اقلیمی، فیزیولوژیکی، ذخیره عناصر غذایی، مدیریت باغی، مدیریت اراضی و وضعیت زمین قرار می‌گیرد (Sadras *et al.*, 2012). نوروزی و همکاران (Norouzi *et al.*, 2010) از شبکه‌های عصبی مصنوعی به‌منظور پیش‌بینی عملکرد گندم دیم در مناطق نیمه‌خشک و کوهستانی غرب ایران استفاده نمودند و بیان داشتند که شاخص انتقال رسوب مهم‌ترین عامل توپوگرافی بر میزان عملکرد این گیاه بوده است و میزان پروتئین موجود در دانه‌ها تحت تأثیر میزان نیتروژن کل خاک می‌باشد. باقری و همکاران (Bagheri *et al.*, 2012) به‌منظور تخمین پتانسیل تولید ذرت علوفه‌ای، از شبکه‌های عصبی مصنوعی بهره بردند و نشان دادند که زمانی که آب آبیاری، مقدار کود نیتروژنی و میزان درجه حرارت روزانه به‌عنوان ورودی به مدل معرفی شوند، مدل می‌تواند عملکرد خشک گیاه را با دقت نسبتاً بالایی برآورد نماید. دراموند و سادث (Drummond & Sudduth., 1998) جهت تعیین روابطی میان ویژگی‌های خاک، توپوگرافی و عملکرد غلات در منطقه میسوری آمریکا مطالعاتی را انجام دادند. این پژوهشگران دریافتند که روش شبکه‌های عصبی مصنوعی نسبت به روش‌های رگرسیونی دارای خطای کمتری می‌باشد. امامقلی زاده (Emamgholizadeh *et al.*, 2015) نیز با مطالعات خود در منطقه اصفهان و عملکرد محصول کنجد گزارش نمودند که شبکه‌های عصبی مصنوعی کارایی بیشتری نسبت به رگرسیون چندگانه دارد. از آنجایی‌که تخمین منطقه‌ای پتانسیل تولید انگور یکی از نیازهای اساسی

[1] Genetic Algorithm
[2] Artificial Neural Network

برنامه‌ریزی‌های اقتصادی برای موازنه تولید، واردات و صادرات بوده، انتظار می‌رود که مدل ترکیبی فازی-الگوریتم ژنتیک از طریق بهینه کردن پارامترهای تابع عضویت بر دقت مدل‌سازی افزوده شود. هدف از این تحقیق بررسی کارایی روش ترکیبی عصبی-الگوریتم ژنتیک (ANN-GA) در برآورد پتانسیل تولید انگور در شهرستان ملکان استان آذربایجان شرقی می‌باشد.

مواد و روش‌ها

منطقه مطالعاتی به وسعت حدود ۳۸۰۰ در جنوب‌غرب استان آذربایجان‌شرقی در محدوده بین ″۳۰ ′۰ °۴۶ تا ″۱۰ ′۵ °۴۶ طول شرقی و ″۲۹ ′۸ °۳۷ تا ″۵۲ ′۱۱ °۳۷ عرض شمالی واقع شده است. میانگین درجه حرارت سالانه دما ۱۵/۷ درجه سلسیوس و میانگین بارندگی سالانه ۲۸۵/۳ میلی‌متر بر اساس آمار هواشناسی ایستگاه کلیماتولوژی ملکان بین سال‌های ۲۰۰۶ تا ۲۰۱۵ می‌باشد. با توجه به میانگین درجه حرارت سالانه هوا و اضافه نمودن یک درجه سلسیوس به آن (۱۶/۷ درجه سلسیوس) رژیم حرارتی خاک‌ها در منطقه ترمیک می‌باشد. رژیم رطوبتی خاک نیز بر اساس نرم‌افزار نیوهال زریک برآورد گردید. برای نیل به اهداف تحقیق ۱۶۰ نمونه خاک از عمق ۰ تا ۲۵ سانتی‌متری (با فواصل تقریبی ۲۵۰ متر) اخذ گردید. سپس ویژگی‌های فیزیکی و شیمیایی این خاک‌ها برای مدل‌سازی، شامل مقدار نسبی سه ذره (بافت)، ذرات درشت‌تر از شن، کربن‌آلی، گچ، آهک و pH با روش‌های استاندارد اندازه‌گیری شدند. برای مدل‌سازی تخمین پتانسیل تولید انگور با روش شبکه‌های عصبی مصنوعی و عصبی-ژنتیک ابتدا با استفاده از ماتریس همبستگی پیرسون ویژگی‌های مؤثر در پتانسیل تولید انتخاب گردید.

مدل‌سازی با شبکه‌های عصبی مصنوعی

در این شیوه از مدل‌سازی، از ساختاری نرونی و هوشمند با الگوبرداری مناسب از نرون‌های موجود در مغز انسان سعی می‌گردد تا از طریق توابع تعریف‌شده ریاضی رفتار درون سلولی نرون‌های مغز شبیه‌سازی شود و از طریق وزن‌های محاسباتی موجود در خطوط ارتباطی نرون‌های مصنوعی، عملکرد سیناپسی در نرون‌های طبیعی به مدل تبدیل شود. بدین منظور الگو، با دسته‌ای از داده‌های واقعی اندازه‌گیری‌شده، آموزش داده شده تا برای ورودی‌های جدید با توجه به رابطه پیدا شده در مرحله آموزش، خروجی مناسب را محاسبه نماید. در این پژوهش جهت شبیه‌سازی و تخمین پتانسیل تولید انگور از بین ساختارهای گوناگون شبکه‌های عصبی مصنوعی، از روش آموزش تحت نظارت یا سرپرستی با ساختار پرسپترون چند لایه (MLP) با الگوریتم پس انتشار خطا (BP) (DuBose and Klimasauskas, 1989) استفاده شد. برای این منظور ابتدا داده‌ها دسته‌بندی و با رسم باکس پلات مربوط به هر ویژگی، داده‌های پرت شناسایی و حذف شد. اصولا وارد کردن داده‌ها به‌صورت خام باعث کاهش سرعت شبکه‌های عصبی می‌شود. برای اجتناب از چنین شرایطی داده‌ها استاندارد شدند. برای استانداردسازی داده‌های اولیه از رابطه (۱) استفاده شد.

$$x_n = (x_r - x_{min})/(x_{max} - x_{min}) \quad (۱)$$

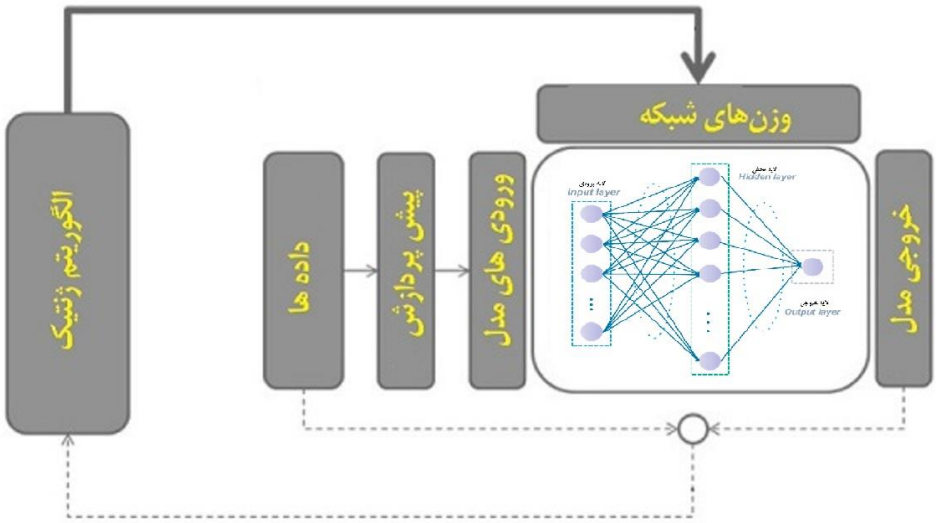
در تابع فوق xr، xn، xmax و xmin به ترتیب نشان‌دهنده مقادیر واقعی، استاندارد شده، حداکثر و حداقل داده‌های تحت بررسی هستند.

پس از استانداردسازی داده‌ها، کل داده‌ها به دو بخش آموزش و آزمون تفکیک شد. ۸۰ درصد مجموعه داده‌ها برای آموزش و ۲۰ درصد مجموعه داده‌ها برای آزمون در نظر گرفته شدند. برای کنترل و نظارت بر یادگیری صحیح شبکه، ۲۵ درصد مابقی داده‌ها به‌عنوان داده‌های آزمون مدل معرفی گردید. برای مدل‌سازی شبکه‌های عصبی مصنوعی پس از وارد کردن داده‌ها به نرم‌افزار، جهت تعیین شبکه بهینه از کلیه توابع محرک تعریف شده در نرم‌افزار (تانژانت هیپربولیک آکسون، سیگموئید آکسون، تانژانت هیپربولیک خطی آکسون، سیگموئید خطی آکسون، بایاس آکسون، آکسون خطی و آکسون) به شکل سعی و خطا تا حصول بهترین نتیجه استفاده شد. برای آموزش شبکه‌های طراحی شده با معماری متفاوت، از توابع

آموزشی بایزین[1] استفاده شد. این الگوریتم آموزش، بهترین روش برای آموزش شبکه‌های عصبی مصنوعی طراحی شده می‌باشد و پایداری بیشتری نسبت به دیگر الگوریتم‌های پس انتشار خطا دارد (Mirzaee et al, 2017).

مدل‌سازی با الگوریتم شبکه‌های عصبی مصنوعی-ژنتیک

در این روش الگوریتم ژنتیک به تنظیم اوزان اتصالات شبکه می‌پردازد. شایان ذکر است که در شبکه‌های عصبی مصنوعی بایستی تعداد و نوع ورودی‌ها، لایه‌های میانی، نرون‌ها و الگوریتم مناسب آموزش شبکه درست انتخاب شوند، تا شبکه کارآمد و دقیق طراحی گردد. شکل ۱ چگونگی تعامل شبکه عصبی و الگوریتم ژنتیک را نمایش می‌دهد. پس از فرآیند پیش‌پردازش، مجموعه داده‌ها به‌عنوان ورودی شبکه عصبی مورد استفاده قرار گرفت. وزن‌های اولیه متناظر با ورودی‌های شبکه عصبی (کروموزم‌های الگوریتم ژنتیک) در تکرار اول به صورت تصادفی شدند. همچنین، رابطه میانگین مجذور خطا[2] به عنوان تابع برازش هر کروموزم از الگوریتم ژنتیک انتخاب شد و از این رو به کارگیری عملگر تقاطع[3] امکان‌پذیر شد. در پایان هر تکرار، ۱۰ درصد از بهترین کروموزم‌ها به همراه ۹۰ درصد کروموزم تولید شده تصادفی جدید، به نسل بعد منتقل گردید. فرآیند مذکور تا رسیدن به شرایط توقف الگوریتم ادامه یافت. در پایان بهترین کروموزم به عنوان وزن‌های اولیه به شبکه عصبی مصنوعی اعمال شد. تعداد کل تکرار الگوریتم ژنتیک برابر با ۲۰۰ در نظر گرفته شد و اگر بعد از ۴۰ تکرار بهینه‌شدن افزایش نیافت، الگوریتم متوقف شد. به منظور انتخاب پارامترهای الگوریتم ژنتیک شامل تعداد جمعیت اولیه، احتمال تقاطع و احتمال جهش، الگوریتم از روش سعی و خطا استفاده شد. فرآیند آموزش و تست شبکه عصبی پیشنهادی و همچنین الگوریتم ژنتیک در نرم‌افزار متلب نسخه ۲۰۱۲ کدنویسی شده و نتایج محاسباتی آن‌ها توسط سیستم یکسان با مشخصات پردازنده ۲/۲۷ گیگاهرتزی و رَمِ ۴ گیگا بیتی به‌دست آمد.



شکل ۱- بهینه‌سازی شبکه‌های عصبی مصنوعی با الگوریتم ژنتیک

نتایج و بحث

جدول ۱ میانگین ویژگی‌های فیزیکی و شیمیایی مورد استفاده را در ایجاد توابع انتقالی نشان می‌دهد. این ویژگی‌ها به همراه آماره‌های توزیع برای مراحل آموزش و آزمون جداگانه مورد بررسی قرار گرفتند. چنانچه ملاحظه می‌شود، خاک‌های مورد بررسی در دامنه وسیعی از نظر ویژگی‌های مورد بررسی قرار دارند. نرمال بودن داده‌ها نیز با استفاده از روش کولموگروف-اسمیرنوف مورد بررسی قرار گرفت که نتایج مؤید نرمال بودن داده‌است.

1-Bayesian (Ba)
2-Mean squared error (MSE)
3-Crossover

نتایج مدل‌سازی با شبکه‌های عصبی مصنوعی مؤید این مطلب است که این روش دارای دقت بالایی بوده به‌طوری‌که ضریب تبیین و میانگین انحراف مربعات خطا به ترتیب در مرحله آموزش ۰/۸۱ و ۲/۱۱ و در مرحله آزمون ۰/۸۴ و ۱/۹۳ برآورد گردید. استفاده از معیار نش و ساتکلیف در مرحله آموزش ۰/۳۷ و در مرحله آزمون ۰/۴۸ نشان‌دهنده دقت نسبتاً بالایی مدل می‌باشد. میانگین هندسی نسبت خطا در این روش ۱/۷۹ برآورد گردید که نشان‌دهنده کم‌برآوردی مدل است.

جدول ۱- پارامترهای آماری ویژگی‌های فیزیکی و شیمیایی خاک

مرحله	آماره	گچ (٪)	کربن آلی (٪)	pH	شن (٪)	سیلت (٪)	رس (٪)	سنگریزه (٪)	آهک (٪)	پتانسیل تولید (تن بر هکتار)
آموزش train	میانگین	2.11	0.67	7.59	43.28	32.2	24.38	17.5	23.2	21.5
	بیشینه	12.34	1.00	8.21	86	44	58	71	36	25
	کمینه	0.0	0.00	7.73	8	4	6	4	4	17
	ضریب تغییرات	0.32	1.33	7.81	23.28	22.4	4.16	0.87	2.16	12.1
	کشیدگی	0.81	0.75	0.34	0.22	0.37	0.91	0.43	0.75	0.74
	چولگی	0.67	0.53	0.41	0.68	0.34	0.52	0.45	0.39	0.63
	تست نرمال بودن	0.11*	0.42**	0.13**	0.12**	0.07*	0.12**	0.02**	0.17**	0.23**
آزمون test	میانگین	0.77	0.44	7.71	49.81	35.36	14.95	23.5	25.4	21.3
	بیشینه	7.5	1.11	8.3	84	40	59	53	38.5	26.4
	کمینه	0.05	0.00	7.5	5	3	7	3	5.5	16.8
	ضریب تغییرات	0.42	1.26	17.34	3.92	4.09	3.19	1.75	1.93	11.1
	کشیدگی	3.24	1.42	0.53	0.94	1.49	0.48	0.35	0.76	0.69
	چولگی	10.11	1.68	0.43	1.67	1.96	0.39	0.88	0.43	0.48
	تست نرمال بودن	0.11*	0.24**	0.19**	0.08**	0.05*	0.18**	0.19**	0.18**	0.24**

* :در سطح احتمال ۵ درصد، **: در سطح احتمال ۱ درصد

نتایج مدل‌سازی با شبکه‌های عصبی مصنوعی-الگوریتم ژنتیک مؤید این مطلب است که این روش نیز دارای دقت بالایی بوده به‌طوری‌که ضریب تبیین و میانگین انحراف مربعات خطا به ترتیب در مرحله آموزش ۰/۹۲ و ۱/۶۱ و در مرحله آزمون ۰/۹۶ و ۱/۵۶ برآورد گردید. استفاده از معیار نش و ساتکلیف در مرحله آموزش ۰/۶۷ و در مرحله آزمون ۰/۶۹ نشان‌دهنده دقت نسبتاً بالایی مدل می‌باشد. میانگین هندسی نسبت خطا در این روش ۱/۴۳ برآورد گردید که نشان‌دهنده کم‌برآوردی مدل است.

درکل می‌توان چنین بیان کرد که روش‌های شبکه‌های عصبی مصنوعی- الگوریتم ژنتیک و شبکه‌های عصبی مصنوعی به دلیل ضریب تبیین بیشتر، ریشه میانگین مربعات خطای کمتر، ضریب نش‌ساتکلیف نزدیک‌تر به یک به ترتیب بیشترین کارایی را دارند. همچنین این روش علاوه بر کارایی بیشتر از سرعت بالایی در رسیدن به جواب تخمین برخوردار است. شایان ذکر است که دو مدل فوق دارای کم‌برآوردی هستند. نهایتاً نتایج بهینه‌سازی حاکی از امکان افزایش کارایی شبکه‌های عصبی مصنوعی با استفاده از الگوریتم ژنتیک دارد.

منابع

Ayoubi SH., Givi J., Jalalian A.. and Amini A.M. 2002. Quantitative land suitability evaluation in north Baron Region (Esfahan province) for wheat, barley, maize and rice. Journal of Agricultural and Natural resource research and Technology. 3(6): 105-118. (In Persian)

Bagheri S., Gheysari M., Ayoubi Sh. and Lavaee N. 2012. Silage maize yield prediction using artificial neural networks. Journal of Plant Production. 19(4): 96-77.

Drummond S.T., Joshi A. and Sudduth. K.A. 1998. Application of neural networks: Precision farming. Proceeding of the 26th IEEE World Congress on Computational Intelligence, Anchorage, England.

DuBose P.. and Klimasauskas C. 1989. Introduction to Neural Networks with Examples and Applications. Neural Ware Inc, Pittsburgh.

Emamgholizadeh S., Parsaeian M.. and Baradaran M. 2015. Seed yield prediction of sesame using artificial neural network. European Journal of Agronomy. 68: 89-96.

Mirzaee S., Ghorbani-dashtaki Sh., Mohammadi J., Asadzadeh F. and Kerry R. 2017. Modeling WEPP erodibility parameters in calcareous soils in northwest Iran. Ecological Indicators. 74:302-310.

Norouzi M., Ayoubi SH.A., Jalalian A., Khademi H. and Dehghani, A. 2010. Predicting rainfed wheat quality and quantity by artificial neural network using terrain and soil characteristics. Acta Agriculture Scandinavia Section B–Soil and Plant Science. 60(4): 341-352.

Sadras V.O. and McDonald G. 2012. Water use efficiency of grain crops in Australia: principles, benchmarks and management. CSIRO, Australia.

Shukla M., Lal R. and Ebinger M. 2004. Principle component analysis for predicting corn biomass and grain yields. Journal of Soil Science. 169: 215-224.

Yemefack M., Rossiter D.G. and Njomgang R. 2005. Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon. Geoderma. 125:117-14.

Grape potential production prediction by artificial neural network in Malekan region

H. Mohammadi[1*], M. Servati[2] and H. Beyrami[3]
1*-PhD candidate of Soil Science and engineering at the Shahrekord University
2-Assistant prof, Shahid Bakeri High Education Center of Miandoab, Urmia University
3-Assistant prof, National Salinity Research Center, Agricultural Research, Education and Extension Organization, Yazd

Abstract

The aim of this study was to assess performance of Genetic Algorithm (GA) in optimization of Artificial Neural Network (ANN) methods in Grape potential production prediction. To achieve this goal, 160 soil Collection data was detected from Malekan region and then modeling with ANN and ANN-GA methods was performed. The optimum number of neurons in hidden layer was estimated 5. Additionally, the most efficient activity function in hidden layer was Axon sigmoid. ANN-GA model have highest R^2 (0.92), lowest RMSE (1.61 $Cmol^+ Kg^{-1}$) and highest Nash–Sutcliffe coefficient (0.67) in training stage and high R^2 (0.96), lowest RMSE (1.56 $Cmol^+ Kg^{-1}$) and highest Nash–Sutcliffe coefficient (0.69) in test stage. Genetic algorithm as new known method is utilized to optimize the amount of the weights by minimizing the network error. Finally, results revealed that this suggested technique leads to performance modeling.

Keywords: Optimization, Hybrid algorithm, Easy-measured properties