مقایسه روش‌های رگرسیون و شبکه عصبی مصنوعی در برآورد پتانسیل تولید منطقه خواجه برای تیپ بهره‌وری جو

مریم زینی [۱]، علی اصغر جعفرزاده [۲]، محمد علی قربانی [۳] و نصرت اله نجفی [۳]
۱-دانشجوی کارشناسی ارشد دانشکده کشاورزی دانشگاه تبریز ، ۲-استاد گروه علوم خاک دانشکده کشاورزی دانشگاه تبریز، ۳-دانشیار دانشکده کشاورزی دانشگاه تبریز

چکیده
از اهداف کشاورزی دقیق بهبود راندمان برداشت محصولات با اعمال مدیریت صحیح می‌باشد که لازمه آن درک کامل‌تر روابط بین میزان تولید محصول با خصوصیات خاک و محیط در کشاورزی پایدار می‌باشد. که در این زمینه مدل‌هایی با شکل ساده شده‌ای از واقعیت برای تعیین پتانسیل تولید و تناسب اراضی مورد استفاده قرار می‌گیرند. در این تحقیق، از دو روش رگرسیون و شبکه عصبی مصنوعی برای تخمین پتانسیل تولید جو در بخشی از منطقه خواجه واقع در استان آذربایجان شرقی استفاده شده است. عواملی همانند ویژگی‌های فیزیکی، شیمیایی و حاصلخیزی که بر روی پتانسیل تولید تأثیر داشته در ۸۰ نمونه از ۲۶ واحد مورد مطالعه قرار گرفته و با استفاده از رگرسیون چند متغیره خطی با روش Enter فقط دو متغیر ذرات درشت‌تر از شن و هدایت الکتریکی نقش معنی‌داری در را بر روی پتانسیل تولید جو داشتند. همچنین مقدار جذر میانگین مربعات خطا یا RMSE در رابطه با داده‌های آزمون جو ۳۱/۱۲۶۳ کیلوگرم بر هکتار و مقدار ضریب تبیین ۵۸% محاسبه گردید. ، در حالیکه نتایج آنالیز شبکه عصبی مصنوعی برای داده‌های صحت‌سنجی نشان دهنده بهبود ضریب تبیین (۷۳%) و کاهش RMSE از ۳۱/۱۲۶۳ به ۶۱/۱۲۱۴ می‌باشد. همچنین در این مدل آزمون نش-ساتکلیف از ۵۷% به ۶۱% افزایش یافته است. بنابراین شبکه های عصبی مصنوعی نتایج بهتری را در مقایسه با رگرسیون خطی در منطقه خواجه تشریح می‌کند.

واژه های کلیدی: آمار چند متغیره، پتانسیل تولید، جو، شبکه عصبی مصنوعی.

مقدمه

در عصر حاضر به دلیل افزایش جمعیت و تخریب اراضی کشاورزی، استفاده بهینه از آن‌ها ضروری می‌باشد. از نظر زارعان و اقتصاددانان در یک منطقه میزان تولید و سودآوری بیشتر از همه حائز اهمیت بوده و روش‌هایی که در این راستا بتواند میزان تولید محصولات مختلف را برآورد نماید از اهمیت ویژه‌ای برخوردار است. در روش کمی فائو تولید پتانسیل با توجه به پتانسیل ژنتیکی محصول و خصوصیات گیاهی آن بر اساس داده‌های اقلیمی نظیر تابش خورشیدی و درجه حرارت محاسبه می‌شود که از ویژگی‌های خاک و مدیریت تأثیرپذیر نیست. روش‌های مختلفی برای محاسبه این تولید پیشنهاد و توسعه یافته است که یکی از مهمترین آنها مدل فائو می‌باشد. در این مدل بر اساس روش پهنه‌بندی اکولوژی-زراعی[۲۴]، تولید محصول برای دامنه وسیعی از محصولات و ارقام، با توجه به شرایط بهینه از نظر آب، عناصر غذایی و عدم وجود بیماری‌ها و آفات، بر مبنای داده‌های تابش و درجه حرارت، تولید زیست توده برآورد می‌گردد (سایس، ۱۹۸۶). بنابراین نیاز به مدل‌سازی و روش‌های نوین در محاسبه پتانسیل تولید احساس می‌گردد و در این تحقیق از روش‌های آماری چند متغیره و شبکه عصبی مصنوعی در تعیین پتانسیل تولید جو استفاده قرار خواهد گرفت. روش‌های آماری چند متغیره از روش‌های آماری پردازش داده‌ها بوده که به منظور تلخیص داده‌ها، گروه‌بندی، برآورد، آزمون فرض‌ها و بررسی همبستگی بین متغیرهای مختلف به کار گرفته می‌شوند (محمدی، ۱۳۸۵). شبکه عصبی مصنوعی ایده‌ای است برای پردازش اطلاعات که از سیستم عصبی زیستی الهام گرفته‌شده و مانند مغز به پردازش اطلاعات می‌پردازد. در واقع شبکه عصبی مصنوعی یک مدل ریاضی است که توانایی مدل‌سازی و ایجاد روابط ریاضی غیرخطی برای میان یابی را دارد(منهاج، ۱۳۸۴). لویی و همکاران(۱۹۹۹) الگوریتم پس انتشار شبکه‌های عصبی مصنوعی را به منظور برآورد بازده محصولات بر روی قطعات کوچک با در نظر گرفتن ویژگی های خاک و هوا به کار بردند. نتایج این کارشناسان امید بخش بود. هرچند نتایج دقت بالای را نشان نمی داد. وین و همکاران(۱۹۹۴)، اوریق و همکاران(۱۹۹۲)، جردینق و همکاران (۱۹۹۲)، فالمن و همکاران (۱۹۹۸)، اظهار داشته اند که شبکه های عصبی پیش خور کاربردی ترین نوع شبکه‌های عصبی مصنوعی می باشند و این نوع شبکه های عصبی پیش خور قادرند هر تابعی را با دقت دلخواه تقریب بزنند. هدف از از این مطالعه بر آورد پتانسیل تولید با روش آماری چند متغیره و شبکه عصبی مصنوعی مقایسه دقت روش‌ها از طریق ایجاد روابط رگرسیونی بین تولید برآورد شده از هر روش و تولید مشاهده شده می‌باشد.

[۲۴] Agro-ecological zoning (AEZ)

مواد و روش‌ها

منطقه مورد مطالعه با وسعتی حدود ۷۳۳۵ هکتار در استان آذربایجان شرقی از منطقه خواجه را شامل می‌شود، که بین ″۳۰ ′۷ °۳۸ تا ″۳۰ ′۱۱ °۳۸ عرض شمالی و ″۳۰ ′۳۷ °۴۶ تا ″۳۰ ′۴۴ °۴۶ طول شرقی و با ارتفاع میانگین ۱۵۰۰ متر واقع شده است. خاک‌های این منطقه دارای رژیم رطوبتی Aridic border on Xeric و رژیم حرارتی Mesic بوده و بر اساس تقسیمات اقلیمی به روش گوسن، منطقه مورد مطالعه جزو نواحی نیمه‌خشک سرد با زمستان‌های سرد و تابستان‌های نسبتاً گرم می‌باشد. در این تحقیق ۸۰ نمونه در به ۲۶ واحد نقشه بررسی و در نهایت میزان تولید محصول در هر یک از ۲۶ پهنه تعریف شده با استفاده از اطلاعات جمع آوری شده از مزارع منطقه محاسبه گردید. در ایجاد توابع انتقالی رگرسیونی از گزینه‌های پیشنهادی نرم‌افزار SPSS شامل گام‌به‌گام و Enter مورد استفاده قرار گرفت و با توجه به نتیجه محاسبات از نظر ضریب تبیین رگرسیون تابع ایجاد شده با انتخاب هر گزینه، بهترین گزینه از بین روش‌های فوق برای تخمین پتانسیل تولید، انتخاب و معرفی گردیدند. ویژگی‌های زود یافت خاک شامل فراوانی نسبی ذرات، درصد ذرات درشت‌تر از شن، درصد گچ، درصد ظرفیت تبادل کاتیونی ظاهری، درصد کربن آلی، درصد کربنات کلسیم معادل، درصد شیب، pH، EC و عملکرد (تولید واقعی) به عنوان متغیرهای مستقل (ورودی) و پتانسیل تولید شده به عنوان متغیر وابسته (خروجی) به نرم‌افزار معرفی شدند. به منظور ایجاد یک شبکه جهت تخمین پتانسیل تولید به روش شبکه عصبی مصنوعی از امکانات و توابع موجود در محیط برنامه‌نویسی MATLAB استفاده گردید. در این پژوهش از شبکه پرسپترون چندلایه (Feed Forward Back Propagation) یا شبکه پیشخور با الگوریتم پس انتشار خطا با ده نرون و با تابع محرک تانژانت سیگموییدی استفاده شده است.

جهت اجرای دقیق شبکه پرسپترون چندلایه در محیط برنامه‌نویسی MATLAB موارد زیر به عنوان اصول کلی اجرا در نظر گرفته شد. شبکه عصبی مصنوعی از ۳ لایه ورودی، میانی، خروجی تشکیل می شود و لایه ورودی هر شبکه را نرون‌های مختلف که شامل تعدادی از خصوصیات زودیافت خاک اندازه‌گیری شده هستند تشکیل داده و لایه خروجی شامل پتانسیل تولید می‌باشد. ۸۰ درصد از داده‌ها یعنی حدود ۶۴ مجموعه داده برای آموزش شبکه و ۲۰% بقیه یعنی حدود ۱۶ مجموعه داده برای شبیه سازی به طور تصادفی در نظر گرفته شدند.

لایه خروجی نیز همانند متغیر وابسته عمل کرده و تعداد نرون‌های آن بستگی به تعداد متغیر وابسته دارد، و چون در اینجا متغیر وابسته میزان عملکرد محصول می‌باشد، لایه خروجی شبکه دارای یک نرون بوده و خروجی شبکه میزان عملکرد محصول را تعیین می‌کند. بر خلاف لایه های ورودی و خروجی، لایه پنهان هیچ مفهومی را نشان نمی‌دهد و صرفا یک نتیجه میانی در فرآیند محاسبه ارزش خروجی است. برای تعیین تعداد نرون‌های لایه میانی و همچنین تعداد لایه‌های میانی از روش آزمون و خطا استفاده می‌شود که این کار ابتدا با یک لایه میانی و یک نرون شروع شده و بتدریج نرون‌های لایه میانی اول زیاد می‌گردد و همین عمل برای لایه‌های میانی بعدی انجام می‌شود، برای هر مرحله رگرسیون خطی میان مقادیر اندازه گیری شده و برآورد شده اعمال می گردد و ضریب تبیین (R^2) هر مرحله نیز محاسبه می شود که ضریب تبیین بالا (R^2 =۱) به معنای دست یافتن به آرایشی مناسب‌تر جهت برآورد پارامتر خروجی می‌باشد، علاوه بر تعیین مقدار ضریب تبیین از معیار ارزیابی آزمون نش-ساتکلیف و RMSE نیز استفاده شده است. به این ترتیب پس از انجام مرحله آزمون و خطا بهترین ساختار شبکه جهت پیش بینی مقدار محصول معرفی می‌گردد که این سه پارامتر آماری که برای مقایسه شبکه‌های عصبی مصنوعی مورد استفاده قرار گرفتند به صورت زیر بیان می‌شوند:

$$RMSE = \sqrt{\frac{\sum_1^N (y_i - \hat{y}_i)^2}{N}}$$

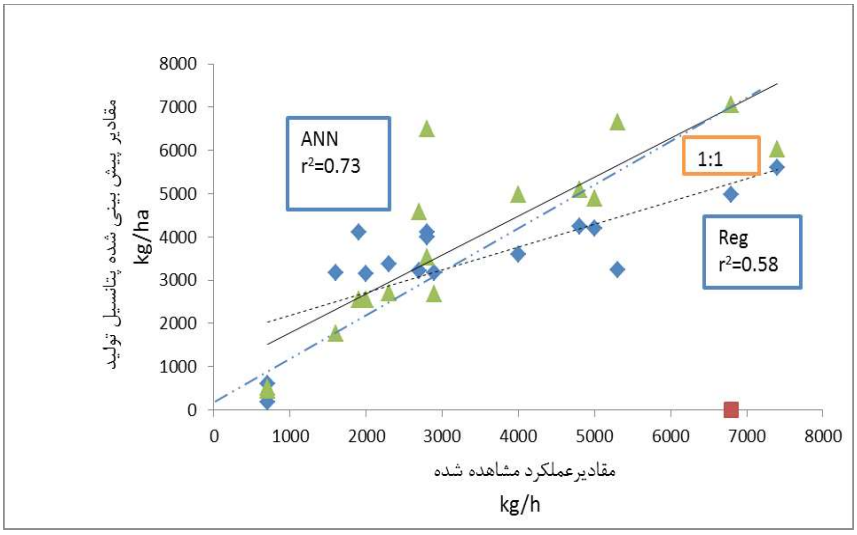
$$NSE = 1 - \left[\frac{\sum_1^N (y_i - \hat{y_i})^2}{\sum_1^N (y_i - \bar{y_i})^2}\right]$$

$$R^2 = \left(\frac{\sum_{i=1}^n (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}}\right)^2$$

در روابط بالا y_i، $\hat{y}_i$، $\bar{\hat{y}}_i$ و $\bar{y}_i$ به ترتیب برابر مقادیر مشاهده شده ، مقادیر برآورد شده و میانگین مقادیر برآورد شده و مشاهده شده، N تعداد کل نمونه‌ها می‌باشند.

نتایج و بحث

نتایج رگرسیون چند متغیره خطی با روش Enter برای جو با توجه به جدول آماری و شکل ۱ مقدار ضریب تبیین برای داده‌های صحت‌سنجی ۵۸/۰می‌باشد. در حالیکه نتایج آنالیزهای شبکه عصبی مصنوعی برای داده‌های صحت‌سنجی نشان می‌دهد که استفاده از این روش باعث بهبود ضریب تبیین و کاهش RMSE یا خطای تخمین شده است. در این مدل‌سازی ضریب تبیین از ۵۸/۰ به ۷۳/۰ افزایش یافته و مقدار RMSE از ۳۱/۱۲۶۳ به ۶۱/۱۲۱۴ کاهش یافته و مقدار آزمون نش-ساتکلیف از ۵۷/۰ به ۶۱/۰ افزایش یافته است. این تحقیق نشان داد که شبکه‌های عصبی مصنوعی پیشخور با الگوریتم پس انتشار از قابلیت بالای در برآورد تولید محصولات زراعی با استفاده از خصوصیات اراضی برخوردار است. نتایج سایر محققین از جمله لویی و همکاران(۱۹۹۹)، وین و همکاران(۱۹۹۴)، اوریق و همکاران(۱۹۹۲)، جردینق و همکاران (۱۹۹۲)، فالمن و همکاران (۱۹۹۸) و عسگری و همکاران (۱۳۸۸)، نیز برتری مدل شبکه عصبی مصنوعی را بر روش رگرسیون خطی نشان می‌دهد و این مسئله ناشی از این است که این مدل می‌تواند روابط خطی و غیر خطی را به طور همزمان مدل‌سازی کند. در حالت کلی اهمیت پارامترهای زمین‌نما و خاک برای پتانسیل تولید جو در منطقه خواجه عبارت است از: درصد کربن آلی، گچ، آهک، ذرات درشت‌تر از شن و نیز هدایت الکتریکی و ظرفیت تبادل کاتیونی می‌باشد.



شکل ۱: نمودار پراکنش پتانسیل پیش‌بینی شده جو برای داده‌های آزمون با مدل رگرسیون خطی Enter و شبکه عصبی مصنوعی

منابع

عسگری، م.ص.، خدادادی، م.، سرمدیان، ف. و گزنی، ر. ۱۳۸۸.کارائی شبکه‌های عصبی مصنوعی در برآورد محصولات گندم، جو و ذرت دانه‌ای. نشریه زراعت، شماره ۸۵. صفحات ۷۱-۶۲.

محمدی، ج. ۱۳۸۵. پدومتری- آمار کلاسیک، جلد اول، انتشارات پلک.

منهاج م ب، ۱۳۸۴. مبانی شبکه‌های عصبي (هوش محاسباتی)، چاپ سوم، انتشارات دانشگاه صنعتی امیرکبیر.

Fahlman S. E. ۱۹۸۸. An empirical study of learning speed inback-propagation networks, Technical Report, Carnegie- Mellon University, Pittsburgh, PA.

Joerding, W. H. and Li, Y., Hu, S. and Meador, J. ۱۹۹۲. Approximating production technologies with feed forward neural networks. Forthcoming Advances in Artificial Intelligence in Economics, Finance, and Management. Greenwich : AI Press

Liu, J. and Goering, C. E. ۱۹۹۹ . Neural network for setting target corn yields. ۴۴(۳): ۷۰۵-۷۱۳.
Uhrig, J. W., Engle, B. A., and Baker,W. L. ۱۹۹۲. An Applicationof neural networks: Predicting Corn Yields." Paper presentedat the Fourth International Congress for Computer Technology in Agriculture, Paris-Versailles, France, June ۱-۳.
Sys, C. ۱۹۸۶. Land evaluation part I-III. International training center for post graduate soil scientists. State University of Ghent. Ghent.

Abstract

Precision farming attempts to improve cropping efficiency by correct planning ,which requisite fully understanding of relationships between crop yield and soil and sit properties in sustainable agriculture Therefor finding of suitable methods for identifiy functional relationships between soil and site and production potential is estimation of production potential , which in this context, simplified models of reality were used to determine production potential and land suitability. In this study two regression and artificial neural network methods are used to estimate the potential production of barley in part of khajeh region located East Azarbaijan. Factors such as physical, chemical and fertility characteristics which affect on production potential were studied in ۸۰ sample from ۲۶ mappping units based on linear regression analysis with Enter just two variables of particals>۲mm and Ec had significant effect on barley production potential and the root mean squre error (RMSE) of ۱۲۶۳.۳۱ kg/ha for experiment datas with coefficient of determination or $R^۲$ of ۰.۵۸ were calculated but results of artificial neural networks analysis according to RMSE= ۱۲۱۴.۱۴ kg/ha and $R^۲$ =۰.۷۳ is one best network and Nash-satcliff test was increased from ۰.۵۷ to ۰.۶۱.Therefore artificial neural network analysis in comperssion with linear regression can be revealed thebest result in Khajeh regresion in this manner.