

بررسی امکان استفاده از مدل انفیس و رگرسیون خطی چند متغیره در پیش بینی نوسانات جمعیت سن گندم (*Eurygaster integriceps* Puton (Hemiptera, Scutelleridae) با استفاده از متغیرهای محیطی در شهرستان چادگان

زهرا دوستی^{۱*}، ناصر معینی نقده^۲، عباسعلی زمانی^۳ و لیلا ندرلو^۴

۱- دانش آموخته‌ی دکتری حشره شناسی، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران،

۲- استادیار گروه گیاهپزشکی دانشکده کشاورزی دانشگاه رازی، کرمانشاه ایران،

۳- دانشیار گروه گیاهپزشکی دانشکده کشاورزی دانشگاه رازی، کرمانشاه ایران،

۴- استادیار گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران،

* مسئول مکاتبه zahradustiy@gmail.com

تاریخ پذیرش: ۱۳۹۸/۳/۱۲

تاریخ دریافت: ۱۳۹۷/۵/۲

چکیده

سیستم‌های هوشمند به عنوان یکی از روش‌های نوین مدل‌سازی در سال‌های اخیر مورد توجه ویژه قرار گرفته‌اند. این مدل‌ها برای پیش‌بینی و طبقه‌بندی در مواردی که روش‌های کلاسیک آماری به خاطر محدودیت‌هایشان قابل استفاده نیستند، کاربرد دارند. هدف از این مطالعه، مقایسه‌ی توانایی مدل‌های انفیس و رگرسیون خطی چند گانه جهت پیش‌بینی تراکم مراحل مختلف رشدی سن گندم *Eurygaster integriceps* است. داده‌های مربوط به نوسانات جمعیت سن گندم در دو مزرعه به مساحت یک هکتار طی سال‌های ۱۳۹۴ و ۱۳۹۵ در شهرستان چادگان بدست آمد. در این مدل‌ها از متغیرهای تاریخ نمونه برداری، متوسط دما، میانگین رطوبت نسبی، سرعت باد، جهت باد، بارش، روز-درجه و ارتفاع از سطح دریا به عنوان متغیرهای ورودی و تغییرات مجموع مراحل رشدی سن گندم به عنوان متغیر خروجی استفاده شد. در مدل انفیس ۷۰ درصد داده‌ها به آموزش و ۳۰ درصد آن‌ها به تعیین اعتبار مدل اختصاص یافت. سپس با آموزش شبکه و تعیین ساختار مطلوب بر اساس نوع، تعداد تابع عضویت و قوانین مربوطه به کمک نرم افزار MATLAB، مناسب‌ترین مدل بر اساس شاخص‌های آماری مجذور میانگین مربعات خطا (RMSE) و ضریب تبیین (R^2) بدست آمد. در نهایت پس از آنالیز حساسیت، نتایج نشان داد که روش انفیس ($R^2=0.97$, $RMSE=0.051$) از صحت و دقت بالاتری نسبت به روش رگرسیون خطی چند متغیره ($R^2=0.47$, $RMSE=0.26$) برخوردار است و عملکرد بهتری در پیش‌بینی نوسانات جمعیت سن گندم دارد.

واژه‌های کلیدی: انفیس، آب و هوا، رگرسیون خطی چند متغیره، سن گندم، نوسانات جمعیت.

مقدمه

همکاران (۱۳۸۱). مدل رگرسیون خطی چند متغیره^۱ یکی از مدل‌های آماری می‌باشد که برای مدل‌سازی و پیش‌بینی نوسانات جمعیت آفات بکار برده می‌شود. یکی از محدودیت‌های مهم برای این نوع مدل‌ها یادگیری کم آن‌ها

سن گندم از مهم‌ترین آفات غلات دانه ریز در ایران و بسیاری از کشورهای منطقه است و هر ساله خسارت فراوانی به گندم وارد می‌کند. پویایی جمعیت سن گندم نیز مانند حشرات دیگر تحت تأثیر دو گروه عوامل زنده (دشمنان طبیعی) و غیر زنده محیط (غذا، اقلیم، دما، بارندگی، رطوبت نسبی، باد) قرار می‌گیرد (ایرانی‌پور و

¹ Multi layer regression (MLR)

نتایج حاکی از وجود رابطه‌ی غیرخطی بین جمعیت حشره و عوامل آب‌وهوایی بود (یانگ و همکاران ۲۰۰۹). شبکه‌ی عصبی مصنوعی برای پیش‌بینی انبوهی جمعیت پروانه‌ی پشت الماسی، *Plutella xylostella* Linnaeus و پارازیتوئید مرحله‌ی لاروی آن *Diadegma semiclausum* Forster مورد استفاده قرار گرفته است. ضریب تبیین، نشان‌دهنده‌ی توانایی این مدل در پیش‌بینی انبوهی جمعیت میزبان - پارازیتوئید بوده است (تُنانگ و همکاران ۲۰۱۰). شبکه‌های عصبی مختلف برای برآزش و شناسایی الگوی توزیع فضایی حشرات مراتع بسیار کارآمد بود (ژانگ و همکاران ۲۰۰۸). کارآمدی شبکه عصبی مصنوعی برای پیش‌بینی دینامیسم جمعیت شته‌ی گال‌زای کاج سوزنی *Thecodiplosis Rkiefje japonensis* (چُن و همکاران ۲۰۰۰)، پیش‌بینی مهاجرت شته *Rhopalosiphum padi* Linnaeus در غلات (لانکین و همکاران ۲۰۰۱)، پیش‌بینی دینامیسم زنجره *Ball Scaphoideus titanus* (لسیوو و همکاران ۲۰۱۱) نشان داده شده است. هدف اصلی از این تحقیق تدوین مدلی است که بتواند بین نوسانات جمعیت سن گندم *Eurygaster integriceps* Puton (Heteroptera: Scutelleridae) با پارامترهای هواشناسی (میانگین دمای روزانه، رطوبت نسبی، سرعت و جهت باد، بارش)، ارتفاع از سطح دریا و تاریخ نمونه‌برداری ارتباط برقرار کند. در این تحقیق از انفیس و رگرسیون خطی چندگانه برای پیش‌بینی نوسانات جمعیت مجموع مراحل مختلف رشدی سن گندم با استفاده از داده‌های هواشناسی استفاده شده است

مواد و روش‌ها

منطقه‌ی مورد بررسی

بررسی تغییرات جمعیت سن گندم در سال‌های ۱۳۹۴ تا ۱۳۹۵ در دو مزرعه‌ی آبی گندم به مساحت یک هکتار یکی در ارتفاع ۲۱۳۵ و دیگری در ارتفاع ۲۲۹۰ در شهرستان چادگان صورت گرفت (جدول ۱). جهت تخمین جمعیت تخم و پوره‌های سنین یک و دو از کادر مربعی یک در یک متر

در سیستم‌های غیر خطی می‌باشد (گاردنر و دورلینگ ۱۹۹۸). برای بدست آوردن نتایج مطلوب از طریق MLR، نمونه‌ها باید زیاد و دقیق باشند، بنابراین، این روش، حساسیت بالایی دارد و ممکن است منجر به بروز خطا در نتایج بدست آمده شود. علاوه بر این، برای استفاده از این روش، متغیرها باید توزیع نرمال داشته باشند و تغییر آن‌ها از یک رابطه‌ی خطی پیروی کند (بالان و همکاران ۱۹۹۵). اخیراً تلاش‌هایی برای به‌کارگیری روش‌ها و فنون جدید در مدل‌سازی توابع صورت گرفته است که می‌تواند برای نمونه به استفاده از روش‌های هوش مصنوعی از جمله سیستم استنتاج فازی - عصبی تطبیقی^۱ (ANFIS) اشاره کرد که در حل مسائل غیر خطی در کاربردهای مهندسی کشاورزی مانند سیستم‌های درگیر با خاک، هوا و گیاه که بسیار پیچیده هستند، مفید می‌باشد (آرخیوف و همکاران ۲۰۰۸). استفاده از سه مدل شبکه عصبی مصنوعی شامل پرسپترون چند لایه^۲، پایه‌ی شعاعی^۳ و سامانه‌ی استنتاج تطبیقی فازی - عصبی و روش رگرسیون خطی چند گانه برای پیش‌بینی بقا حشرات کامل *Chrysomya megacephala* Fabricius نشان داد هر سه مدل شبکه عصبی نسبت به روش رایج رگرسیون خطی عملکرد بهتری دارند (بیانکی و همکاران ۲۰۰۹). استفاده از مدل‌های شبکه عصبی در فرایند مدل‌سازی بقا و توزیع مرگ و میر حشره برگ‌خوار *Spodoptera litura* Fabricius در دماهای مختلف انجام شد که نتایج نشان داد شبکه عصبی در این زمینه می‌تواند مفید باشد (ژانگ و ژانگ ۲۰۰۸). جهت تجزیه و تحلیل داده‌های مربوط به جمعیت کرم ساقه خوار برنج *Scirpophaga incertulas* Walker (Lepidoptera: Crambidae) و پی بردن به رابطه‌ی موجود بین ظهور آفت و عوامل آب‌وهوایی شبکه عصبی به خوبی توانست جمعیت آفت را پیش‌بینی کند.

¹ Artificial neural network-based fuzzy inference system (ANFIS)

² Multi-layer perceptron (MLP)

³ Basis function radial (RBF)

داده‌های هواشناسی

متغیرهای آب و هوایی شامل میانگین دمای روزانه، رطوبت روزانه و بارش روزانه، جهت و سرعت باد از ایستگاه هواشناسی شهرستان چادگان گرفته شد.

تعیین مدل روز-درجه

داده‌های دمایی شامل کمینه و بیشینه دمای روزانه از ایستگاه هواشناسی شهرستان برای تعیین مدل روز-درجه آفت دریافت شد. محاسبه‌ی روز-درجه‌های مورد نیاز برای ظهور و اوج مراحل مختلف زیستی آفت با استفاده از نرم‌افزار DEGDAY.xls و روش مورد استفاده در آن Single Sine Method بود. آستانه‌ی پایین ۱۲/۲۶ درجه و آستانه‌ی بالا ۲۶/۵ درجه‌ی سلسیوس در نظر گرفته شد (معینی نقه، ۱۳۸۱).

رگرسیون خطی چند گانه (MLR)

رگرسیون چندگانه در حقیقت، ارتباط بین یک سری از متغیرهای پیشگو را با متغیر پاسخ مورد نظر بیان می‌کند (بالان و همکاران ۱۹۹۵). در صورت وجود متغیرهای مستقل X_1, X_2, \dots, X_n اگر بخواهیم ارتباط خطی بین آنها و متغیر Y که وابسته به آنهاست ایجاد کنیم، رابطه‌ی زیر باید بین آنها برقرار باشد:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n + e \quad \text{[معادله ۳]}$$

که در این رابطه، از مقادیر a_1, a_2, \dots, a_n با عنوان ضرایب رگرسیون یاد می‌شود. این ضرایب، ضرایب نامشخصی هستند که در حقیقت، مسئول برآورد پارامتر پاسخ هستند (بالان و همکاران ۱۹۹۵). تاریخ نمونه‌برداری، میانگین دما در روز نمونه‌برداری، میانگین رطوبت نسبی در روز نمونه‌برداری، میانگین بارش، ارتفاع از سطح دریا، سرعت باد، جهت باد، روز-درجه به عنوان متغیرهای مستقل و مجموع مراحل رشدی سن گندم به عنوان متغیر وابسته در نظر گرفته شدند. تجزیه و تحلیل داده‌ها به روش گام به گام، با استفاده از نرم افزار SPSS 16 انجام شد.

جدول ۱- مختصات مزارع نمونه برداری شده با استفاده از دستگاه جی پی اس

Table 1- Field coordinates sampled using GPS device

UTM	Sea from level	
0465581 S 39 3624273	2135	Farm 1
0458060 S 39 3636298	2290	Farm 2

S: South

مربع و برای پوره‌های سنین سه، چهار، پنج و حشرات کامل نسل جدید از تور حشره گیری استفاده شد (هر پنج تور معادل یک متر مربع).

اندازه‌ی نمونه

به منظور تعیین تعداد مناسب نمونه، ابتدا یک نمونه‌برداری اولیه با تعداد ۳۰ واحد نمونه انجام شد، سپس با استفاده از داده‌های به دست آمده مقدار خطای نسبی^۱ تعیین شد. مقدار خطای نسبی از طریق معادله زیر محاسبه شد:

$$RV = \frac{SE}{\bar{X}} \times 100 \quad \text{[معادله ۱]}$$

در این معادله، \bar{X} میانگین داده‌های نمونه‌برداری اولیه و SE خطای استاندارد می‌باشد. مقدار خطای نسبی قابل پذیرش در تعیین الگوی توزیع فضایی ۲۵ درصد در نظر گرفته شد. سپس تعداد نمونه مورد نیاز با استفاده از معادله زیر محاسبه شد:

$$N = \left[\frac{t \times SD}{D \times x} \right]^2 \quad \text{[معادله ۲]}$$

N برابر تعداد نمونه مورد نیاز، t مقدار جدول t-student بر حسب درجه‌ی آزادی تعداد نمونه، SD انحراف معیار داده‌های نمونه‌برداری اولیه و D میزان خطای قابل قبول (۰/۲۵) است (پدیگو ۱۹۹۴).

¹ Relative variation

مدل سازی به کمک ANFIS

ترکیب سیستم‌های استنتاج فازی که مبتنی بر قواعد منطقی بوده و روش شبکه‌های عصبی مصنوعی که توان استخراج دانش از اطلاعات عددی را دارند، منجر به ارائه سیستم استنتاج تطبیقی عصبی فازی می‌شود (احمدزاده و همکاران، ۱۳۸۹). از آنجایی که این روش اصول شبکه‌های عصبی و سیستم استنتاج فازی را ادغام می‌کند، می‌تواند مزایای هر دو را هم‌زمان داشته باشد. سیستم استنتاج آن مطابق با مجموعه‌ای از قوانین فازی است که توانایی یادگیری برای تقریب توابع غیرخطی را دارند (آبراهام ۲۰۰۵). پارامترهای توابع عضویت این سامانه، از طریق الگوریتم پس انتشار^۱ یا ترکیب آن با روش حداقل مربعات تنظیم می‌شوند. اساس الگوریتم پس انتشار بر پایه قانون یادگیری اصلاح خطاست که از دو مسیر اصلی رفت و برگشت تشکیل می‌شود (خانا، ۱۹۹۰). این عملیات به سامانه‌های فازی اجازه می‌دهد تا ساختار خود را از مجموعه داده‌ها فرا بگیرد (کیا، ۱۳۸۹).

ساختار ANFIS شامل پنج لایه است در لایه اول، ورودی‌ها از توابع عضویت عبور می‌کنند و فازی سازی^۲ می‌شوند:

$$O_{1,i} = \mu A_i(x), \quad \text{For } i=1,2$$

$$O_{1,i} = \mu B_i(x), \quad \text{For } i=3,4$$

در لایه دوم، خروجی این لایه ضرب سیگنال‌های ورودی است که در واقع وزن قانون‌ها به دست می‌آید:

$$O_{2,i} = w_i = \mu A_i(x) \mu B_i(y), \quad i=1,2$$

در لایه سوم عمل محاسبه وزن نسبی قوانین در هر

گره انجام می‌شود:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2$$

در لایه ۴ غیرفازی سازی انجام می‌گیرد، این لایه به لایه قوانین معروف بوده و از جمع عملیات‌ها بر روی سیگنال‌های ورودی به این لایه بدست می‌آید:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i)$$

و خروجی لایه ۵، خروجی کلی سیستم است:

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

در نهایت شبکه‌ی تولید شده معادل سیستم استنتاج فازی سوگنو^۳ است (کیسی و همکاران، ۲۰۰۹). بنابراین ساختار انفیس به صورت زیر خواهد بود (شکل ۱).

برای آموزش سامانه‌های استنتاج عصبی - فازی از قسمتی از داده‌ها استفاده می‌گردد که به آنها اصطلاحاً داده‌های آموزش گفته می‌شود. در مرحله‌ی آموزش با اصلاح پارامترهای درجه‌ی عضویت بر اساس میزان خطای قابل قبول، مقادیر ورودی به مقادیر واقعی نزدیک‌تر می‌شوند. سایر داده‌ها که در فرایند آموزش استفاده نشده برای بررسی صحت عملکرد سامانه استنتاج استفاده می‌شوند که به آنها داده‌های آزمون گفته می‌شود. انتخاب داده‌های آزمون و آموزش باید طبق روال کلی و به صورت تصادفی باشد. از مجموعه داده‌های آموزشی برای یادگیری و از داده‌های آزمایشی برای بررسی دقت و اثر بخشی بودن مدل ANFIS در طبقه‌بندی استفاده می‌شود.

در این تحقیق ۷۰ درصد داده‌ها برای آموزش و ۳۰ درصد برای آزمون در نظر گرفته شد. در پژوهش حاضر از نرم‌افزار Matlab 2013a برای انجام مدل‌سازی و برآورد میزان تقلب استفاده شد. ANFIS یکی از پنجره‌های موجود در نرم‌افزار متلب است.

قبل از بکارگیری داده‌های خام اولیه در آموزش شبکه، داده‌ها باید در دامنه‌ی مناسبی استاندارد شوند زیرا الگوریتم یادگیری همراه با داده‌های خام نمی‌تواند عملکرد مناسبی داشته باشد. برای تبدیل داده‌ها از روش استانداردسازی [معادله ۴] استفاده شد که ورودی‌ها را

³ Sugeno

¹ Back Propagation

² Fuzzifier

مربعات خطا (RMSE) معادله [۶]، که از روابط زیر قابل محاسبه می‌باشند، استفاده گردید.

نتایج و بحث

الف) واحد نمونه برداری

نتایج حاصل از مقایسه دو روش نمونه برداری مطلق استفاده از کادر یک در یک متر با روش نمونه برداری نسبی تور زدن برای نمونه برداری از جمعیت سن گندم در دو سال ۱۳۹۴ و ۱۳۹۵ نشان داد که بصورت میانگین تعداد پنج تور معادل یک متر مربع می‌باشد. لذا پنج تور به عنوان یک واحد نمونه برداری انتخاب شد.

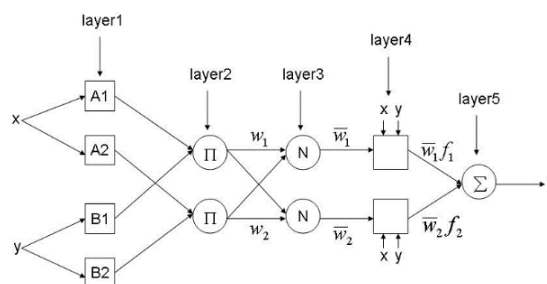
ب) اندازه نمونه

نتیجه حاصل از اولین نمونه برداری نشان داد که تعداد نمونه‌ی مورد نیاز جهت برنامه‌ی نمونه برداری با خطای قابل قبول ۲۵٪ (D=0.25) برابر با ۳۱ واحد نمونه برداری می‌باشد. مقدار RV نمونه برداری اولیه برابر ۱۰/۱۷ درصد بود که در محدوده‌ی قابل قبولی قرار داشت.

هشت ورودی انتخاب شده به همراه کل جمعیت آفت توسط آنالیز حساسیت مورد بررسی قرار گرفت که نمودار حاصله در شکل ۲ نشان داده شده است. بر اساس شکل ۲ چهار پارامتر انتخاب شده برای مدل سازی عبارتند از:

- ۱- میانگین دما در روز نمونه برداری ۲- میانگین رطوبت در روز نمونه برداری ۳- سرعت باد ۴- شماره روز نمونه برداری.

در تکنیک انفیس، مسئله‌ی اصلی تعیین نوع تابع عضویت و درجه‌ی عضویت آن‌هاست که برای این امر راه حل مشخصی وجود ندارد و با بهره گیری از سعی و خطا، بهترین ساختار مدل تعیین می‌گردد. بدین منظور با بهره گیری از داده‌های آموزش و آزمون و انتخاب تابع عضویت ورودی و خروجی از میان توابع مختلف و درجه‌ی عضویت آن‌ها، شبکه را اجرا کرده و با تعیین معیارهای آماری RMSE و R^2 بین داده‌های مشاهده‌ای و برآورد شده، بهترین ساختار شبکه بدست آمد. جدول ۲ نتایج



شکل ۱- ساختار مدل ANFIS

Figure 1- ANFIS structure model

بین صفر و یک استاندارد می‌کند (ایراه‌اقی و همکاران، ۱۹۹۳). در این معادله، X داده‌ی خام اولیه، X_n داده‌ی استاندارد شده و X_{min} و X_{max} به ترتیب مقادیر بیشینه و کمینه داده‌های اولیه می‌باشد.

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{[معادله ۴]}$$

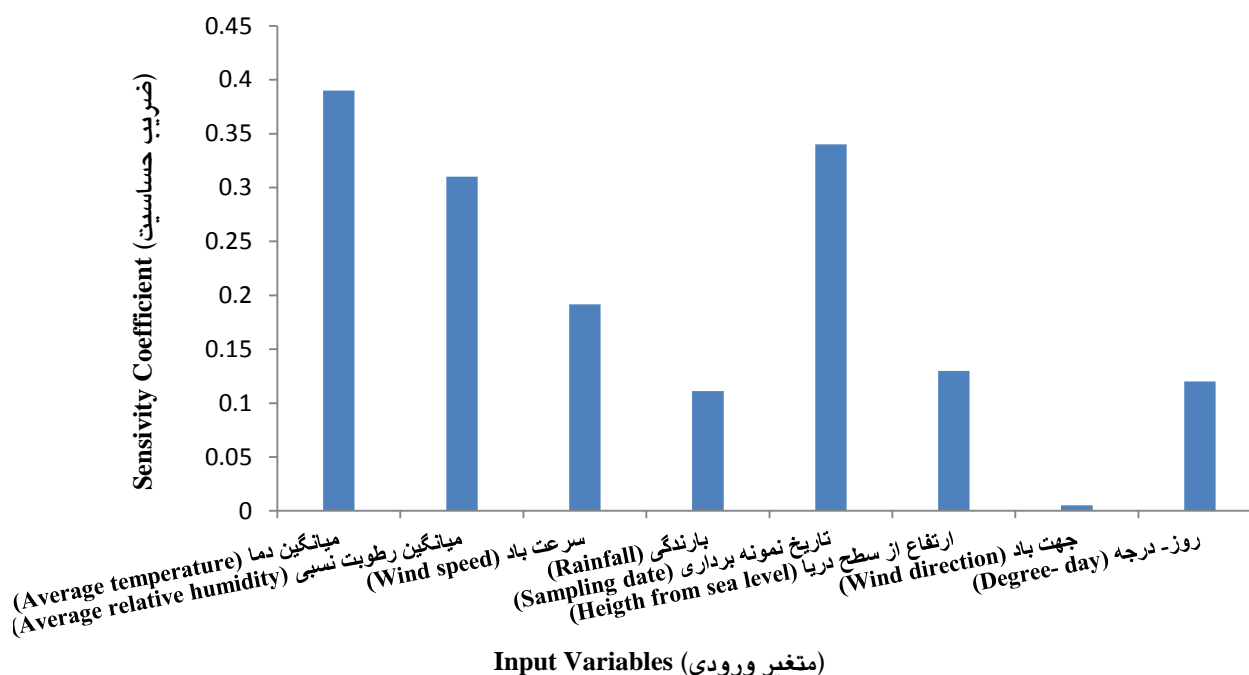
آنالیز حساسیت^۱

آنالیز حساسیت برای نشان دادن واکنش شبکه به تغییرات هر یک از متغیرها بکار می‌رود. بنابراین تأثیر ورودی‌های قوی‌تر دخیل می‌تواند مورد بررسی قرار گیرد. در این تحقیق، پس از بدست آوردن بهترین شبکه‌ی عصبی با استفاده از کل ورودی‌ها (هشت ورودی) آنالیز حساسیت با مقید کردن هر ورودی، به جز یکی که در حال بررسی است، در مقدار متوسط آن برای مجموعه‌ی داده‌ها انجام شد. سپس مقادیر ورودی در حال بررسی در محدوده‌ی صفر تا یک تغییر کرد و شبکه برای هر مرحله فراخوانی شد تا در نهایت از بین هشت ورودی عوامل مؤثرتر استخراج گردید.

ارزیابی آماری مدل‌ها

برای ارزیابی و مقایسه‌ی دقت مدل‌ها از پارامترهایی مانند ضریب تبیین (R^2) معادله [۵] و مجذور میانگین

¹ Sensivity analysis



شکل ۲- آنالیز حساسیت هشت پارامتر منتخب

Figure 2- Sensitivity analysis 8 selected parameter

پارامترهای تابع عضویت است که ترکیبی از روش حداقل مربعات و روش کاهش شیب پس انتشار می باشد (کیا، ۱۳۸۹). در آموزش سیستم استنتاج فازی، تعداد دوره های آموزش و خطای مطلوب صفر و تعداد دوره های آموزش، ۱۰۰ در نظر گرفته شد. از این مرحله به بعد، آموزش مدل و پیش بینی مولفه ی مورد نظر، قابل اجرا است. طی ساز و کارهای تعیین شده در مراحل قبل، آموزش مدل صورت گرفت که داده های مدل سازی شده در مقایسه با داده های آموزش در شکل ۴ مشخص می باشد. در این شکل، داده های آموزش به شکل دایره و داده های مدل سازی با علامت ستاره مشخص شده اند.

مطابق شکل های ۵ و ۶ همبستگی بین داده های خروجی از شبکه انفیس، با داده های هدف، برای داده های آموزش و آزمون به ترتیب ۰/۹۷۰ و ۰/۹۶۷ بدست آمد.

حاصل از آموزش مدل های گوناگون را نشان می دهد. با توجه به این جدول، مدل ۱۲ در مقایسه با سایر مدل ها نوسانات مجموع مراحل رشدی سن گندم را با دقت بالاتری پیش بینی می کند. همان طور که مشخص است تعداد توابع عضویت، سه تابع بوده و نوع تابع به شکل گوسی انتخاب شد. بهترین همبستگی بین داده های پیش بینی شده و داده های واقعی برای مراحل آموزش و آزمون به ترتیب با ضریب تعیین ۰/۹۷۱ و ۰/۹۶۸ و مجذور میانگین مربعات خطای ۰/۰۵۱ و ۰/۰۵۰ به ترتیب برای داده های آموزش و آزمون به وسیله مدل ۱۲ بدست آمد (جدول ۳).

پس از ایجاد FIS، بر اساس تعداد و شکل توابع، ساختار مدل به صورت شکل ۳ تعیین شد که ساختار انفیس با لایه ی پنج گانه که در مطلب قبلی بیان شد، در این ساختار قابل مشاهده می باشد. از این ساختار برای پیش بینی مجموع مراحل رشدی سن گندم استفاده شد.

برای آموزش ساختار انفیس از روش هیبرید استفاده شد. این روش، از روش های بهینه سازی برای آموزش

جدول ۲- نتایج حاصل از مدل سازی به روش ANFIS، برای پیش بینی جمعیت مجموع مراحل رشدی سن گندم.

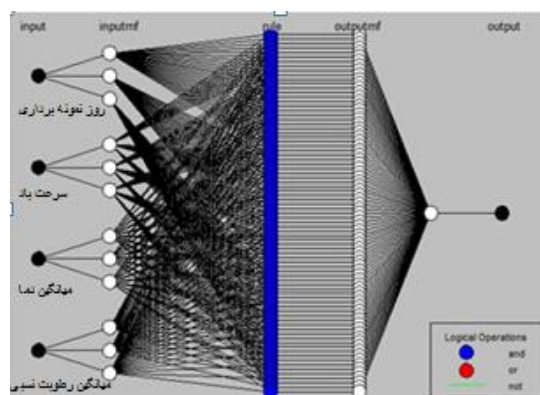
Table 2- Statistical performance of Anfis, for predicted total growth stage population of Sunn pest.

R ² (ضریب تیین)	RMSE (مجذور میانگین مربعات خطا)	Membership function of output variable (تابع عضویت متغیر خروجی)	Optimal model structure (ساختار بهینه مدل)		ANFIS model (مدل انفیس)
			Number of membership function (تعداد تابع عضویت)	membership function of input variable (تابع عضویت متغیر ورودی)	
0.942	0.07078	Linear (خطی)	2	Trimf (مثلثی)	1
0.954	0.0627	Linear (خطی)	2	Trapmf (ذونقه ای)	2
0.951	0.0653	Linear (خطی)	2	Gbellmf (زنگوله ای)	3
0.944	0.0694	Linear (خطی)	2	Gaussmf (گوسی)	4
0.952	0.0648	Linear (خطی)	2	Gaussmf (گوسی)	5
0.952	0.0647	Linear (خطی)	2	Psigmf (سیگموئید ترکیبی)	6
0.950	0.0655	Linear (خطی)	2	Sigmf (سیگموئیدی)	7
0.950	0.0655	Linear (خطی)	2	Pimf (شبه پی)	8
0.969	0.0516	Linear (خطی)	3	Trimf (مثلثی)	9
0.968	0.0525	Linear (خطی)	3	Gbellmf (زنگوله ای)	10
0.970	0.0510	Linear (خطی)	3	Trapmf (ذونقه ای)	11
0.970	0.0510	Linear (خطی)	3	Gaussmf (گوسی)	*12
0.969	0.0512	Linear (خطی)	3	Gauss2mf (گوسی)	13
0.969	0.0517	Linear (خطی)	3	Pimf (شبه پی)	14
0.970	0.0509	Linear (خطی)	3	Sigmf (سیگموئیدی)	15
0.970	0.0509	Linear (خطی)	3	Pimf (شبه پی)	16

جدول ۳- خروجی بهترین مدل ANFIS برای تخمین کل جمعیت سن گندم.

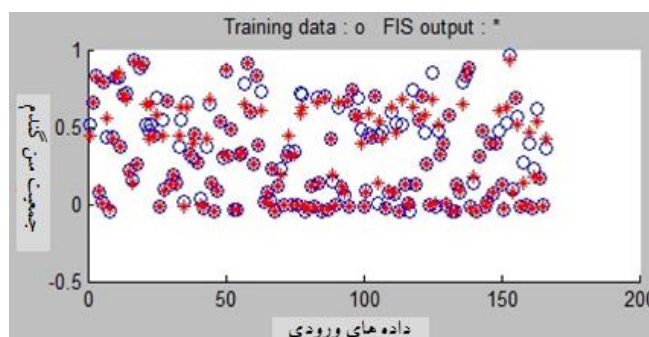
Table 3- Output of the best ANFIS for estimation of total population of the Sunn pest.

P (معنا داری)	R ² (ضریب تبیین)	SSE (مجموع مربعات خطا)	RMSE (مجذور میانگین مربعات خطا)	Modeling steps (مراحل مدل سازی)
0>0.001	0.971	0.307	0.051	Train (آموزش)
0>0.001	0.968	0.125	0.050	Test (آزمون)
0>0.001	0.970	0.431	0.051	Total (کل)



شکل ۳- ساختار مدل ۱۲ انفیس برای پیش‌بینی نوسانات مجموع مراحل رشدی سن گندم.

Figure 3- ANFIS 12 model structure for prediction of seasonal total growth stage of the Sunn pest.



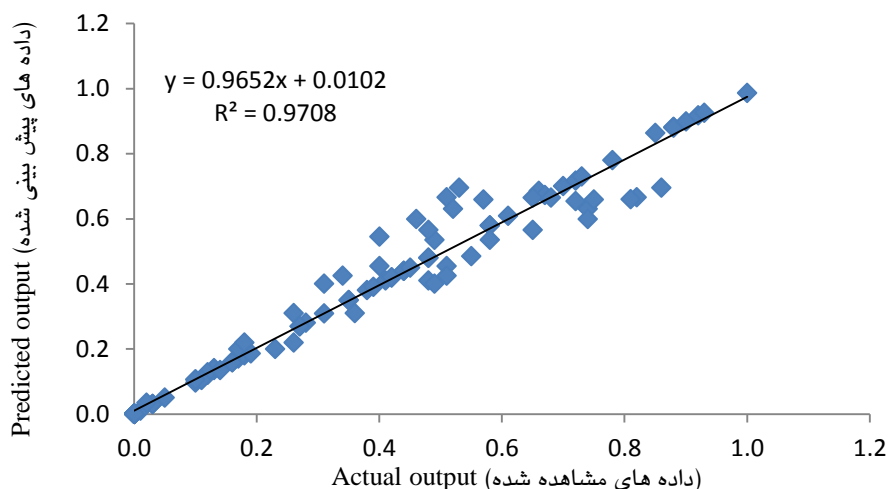
شکل ۴- مقایسه داده‌ی مدل‌سازی شده و داده‌ی آموزشی برای پیش‌بینی مجموع مراحل رشدی سن گندم.

Figure 4- Compare modeling and training data for prediction total growth stage of Sunn pest.

نتیجه‌گیری کلی

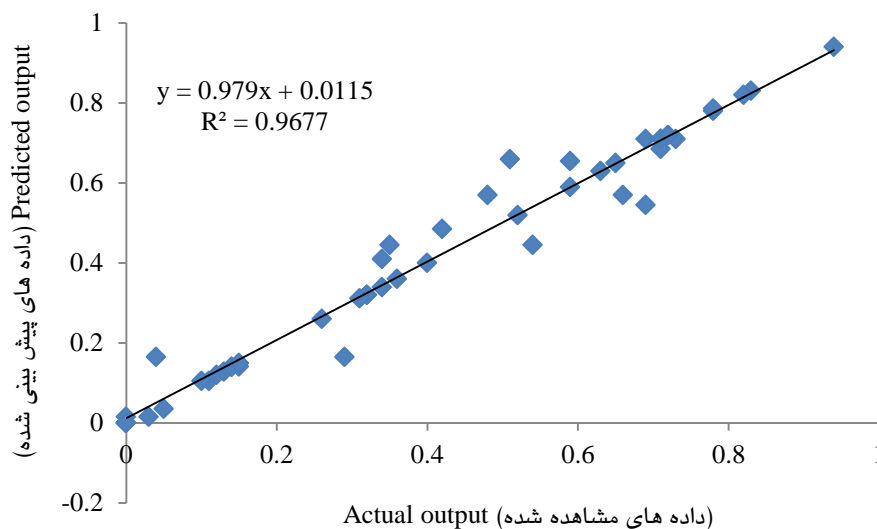
در بررسی پیش‌بینی نوسانات جمعیت آفات، مدل‌های تجربی دارای پیچیدگی کمتری در مقایسه با مدل‌های انفیس هستند، در مقابل، مدل‌های انفیس با وجود پیچیدگی دارای دقت بالاتری هستند. به علاوه امکان بکارگیری بیش از دو متغیر در ورودی وجود دارد. در این مطالعه، مدل انفیس توانست با ضریب تبیین بالایی (۰/۹۷) نسبت به مدل رگرسیون خطی (۰/۴۷) تغییرات جمعیت مجموع مراحل رشدی سن گندم را پس از انجام تحلیل حساسیت و انتخاب بهترین ورودی‌ها، با استفاده از چهار عامل محیطی شماره روز سال، میانگین دما، میانگین رطوبت و سرعت باد به عنوان ورودی پیش‌بینی کند. یعنی محقق می‌تواند در هر زمان و هر روز سال به راحتی با داشتن اطلاعات قابل

نتایج رگرسیون خطی چند متغیره با ورودی‌های منتخب که در تحلیل حساسیت به دست آمد، در جدول ۴ نشان داده شده‌است. مقدار ضریب تبیین و جذر میانگین مربعات خطا به ترتیب ۰/۴۷ و ۰/۲۶ به دست آمد. از آنجایی که مقدار p آزمون برای متغیر میانگین دمای روزانه ۰/۰۰ بوده و کمتر از ۰/۰۵ بود، فرض صفر با اطمینان ۹۵ درصد رد می‌شود یعنی مدل موجود با اطمینان ۹۵ درصد معنی دار بوده و بر اساس داده‌های میانگین دمای روزانه این مدل قادر به بیان تغییرات مراحل مختلف رشدی سن گندم می‌باشد، اما برای سه متغیر دیگر (میانگین رطوبت نسبی، سرعت باد و روز نمونه برداری) مقدار p آزمون کمتر از ۰/۰۵ بوده و در نتیجه معنی دار نمی‌باشد.



شکل ۵- همبستگی بین داده‌های پیش بینی شده و مشاهده شده توسط مدل انفیس در مرحله‌ی آموزش.

Figure 5- Correlation between the predicted and observed data by the ANFIS model at the training stage.



شکل ۶- همبستگی بین داده‌های پیش بینی شده و مشاهده شده توسط مدل انفیس در مرحله آزمون.

Figure 6- Correlation between the predicted and observed data by the ANFIS model at the testing stage.

جدول ۴- رابطه‌ی رگرسیونی بین جمعیت مجموع مراحل رشدی با متغیرهای منتخب.

Table 4- Regression relationship between the population of the total growth stage with selected variables

RMSE (مجذور میانگین مربعات خط)	R ² (ضریب تبیین)	P value (معناداری)	Coefficient (ضرایب)	Input variable (متغیرهای ورودی)
0.26	0.47	0.00	0.71	Average temperature (میانگین دما)
		0.26	-0.11	Average relative humidity (میانگین رطوبت نسبی)
		0.66	-0.03	Wind speed (سرعت باد)
		0.30	-0.20	Sampling date (تاریخ نمونه برداری)

دسترس از سازمان هواشناسی و قرار دادن آن‌ها در مدل انفیس تراکم جمعیت آفت را پیش‌بینی و اقدام لازم را در زمان مناسب جهت رسیدن به عملکرد بالاتر انجام دهد. این تحقیق بخشی از رساله دکتری نگارنده اول می‌باشد که با حمایت دانشگاه رازی کرمانشاه انجام شده است و بدینوسیله تشکر و قدردانی می‌گردد.

منابع مورد استفاده

- احمدزاده قره‌گویزک، میرلطیفی س م، محمدی ک، ۱۳۸۹. مقایسه سیستم‌های هوش مصنوعی (ANFIS و ANN) در تخمین میزان تبخیر تعرق گیاه مرجع در مناطق بسیار خشک ایران، نشریه آب و خاک (علوم و صنایع کشاورزی)، جلد ۴، شماره ۲۴، صفحه‌های ۶۸۹-۶۷۹.
- ایرانی پور ش، خرازی پاکدل ع، رجبی غ، رسولیان غ. و مجنی، ح. ۱۳۸۱. تلفات ویژه سنی و تغییرات سرعت نشو و نمای مراحل نابالغ سن گندم *Eurygaster integriceps* در چهار دمای ثابت آزمایشگاهی، مجله آفات و بیماری‌های گیاهی، جلد ۲، شماره ۷۰، صفحه‌های ۱۷-۱.
- کیا م، ۱۳۸۹. شبکه‌های عصبی در متلب، خدمات نشر کیان رایانه سبز.
- معینی ن. ۱۳۸۱. مدل پیش‌آگاهی روز-درجه ناحیه‌ای برای پیش‌بینی مراحل رشدی سن گندم در شرایط متغیر دمایی در مزرعه. رساله دکتری، دانشگاه تربیت مدرس، تهران. ۹۰ صفحه.
- Abraham A, 2005. Adaptation of fuzzy inference system using neural learning. Fuzzy Systems Engineering, 914-914.
- Arkipov M, Kruger E and Kurtener D, 2008. Evaluation of ecological conditions using bioindicators: application of fuzzy modeling. Lecture Notes in Computer Science, 491-500.
- Balan B, Mohaghegh S and Ameri S, 1995. State- of- Art- in permeability determination from well log data: Part 1- A comparative study, Model development. Society of Petroleum Engineers, 17-25.
- Bianconi A, Von Zuben CJ, Scapiao ABS, and Govone J, 2009. Artificial neural networks: A novel approach to analysing the nutritional ecology of a blowfly species, *Chrysomya megacephala*. Journal of Insect Science, 10: 1-18.
- Chon TS, Kim JM, Lee BY, Chung YJ, and Kim Y, 2000. Use of an artificial neural network to predict population dynamics of the Forest-Pest pine needle gall midge (Diptera: Cecidomyiida). Journal of Environmental Entomology, 29: 1208-1215.
- Erahaghi I, Xuchai L, Mahnaz H. and Yusuf S, 1993. A robust neural network model for pattern recognition of pressure transient test data. Society petroleum engineering annual technical conference and exhibition, 3-6 October 1993. Houston, Texas.
- Gardner MW, Dorling SR, 1998. Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences. Atmospheric Environment. 32 (14/15): 2627-2636.
- Khanna TF, 1990. Foundation of Neural Networks. New York: Addison-Wesley.
- Kisi OT, Haktanir M, Ardiclioglu O, Ozturk E, Yalcin and Uludag S. 2009. Adaptive neurofuzzy computing technique for suspended sediment estimation. Advances in Engineering Software. 40: 438-444.

- Lankin GO, Worner GO, Samarasinghe S and Teulon DAJ, 2001. Can artificial neural network systems be used for forecasting aphid flight patterns. *New Zealand Plant Protection*, 54: 188-192.
- Lessio F, Mondino EB and Alma A. 2011. Spatial patterns of *Scaphoideus titanus* (Hemiptera: Cicadellidae): a geostatistical and neural network approach. *International Journal of Pest Management*, 57: 205-216.
- Pedigo LP, 1994. Introduction to sampling Arthropod population. *Hanbook of sampling methods for Arthropoda Agriculture*, ed. P. Pedigo and G.D. Buntin. CRC Boca Raton.
- Tonnangt HEZ, Nedorezov LV, Ochanda H, Owino JO and Lohr B, 2010. Host- parasitoid population density prediction using artificial neural networks: diamondback moth and its natural enemies. *Agricultural and Entomology*. 12: 233- 242.
- Yang LN, Peng L, Zhang LM, Zhang LL and Yang SS, 2009. A prediction model for population occurrence of paddy stem borer (*Scirpophaga incertulas*), based on Back Propagation Artificial Neural Network and Principal Components Analysis. *Computers and Electronics in Agriculture*, 200-206.
- Zhang WJ, Liu GH and Dia, HQ, 2008. Simulation of food intake dynamics of holometabolous insect using functional link artificial neural network. *Stochastic Environmental Research and Risk Assessment*. 22: 123-133.
- Zhang, WJ and Zhang XY, 2008. Network modeling of survival dynamics of holometabolous insects: A case study. *Ecological Modelling*. 211: 433-443

Investigation for Application of Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multi Linear Regression in Predicting Seasonal Total Development Stage Population Fluctuations of the Sunn Pest, *Eurygaster integriceps* Puton (Hemiptera, Scutelleridae) Using Environmental Variables in Chadegan

Z Doosti^{1*}, N Moeini-Naghadeh², A Ali Zaman³ and L Naderloo⁴

¹Graduated Ph.D. Student of Entomology, College of Agriculture, Razi University, Kermanshah, Iran.

²Assistant Professor, Department of Plant Protection, College of Agriculture, Razi University, Kermanshah, Iran.

³Associate Professor., Department of Plant Protection, College of Agriculture, Razi University, Kermanshah, Iran.

⁴Assistant Professor, Department of Biosystem Mechanization Engineering, College of Agriculture, Razi University, Kermanshah, Iran.

*Corresponding author: zahradustiy@gmail.com.

Received: 24 July 2018

Accepted: 2 June 2019

Abstract

Intelligent systems have received considerable attention as a modern modeling methods in recent years. These models is used for prediction and classification in situations where the classic statistical models are not able due to their constraints. This study is aimed to compare the ability of ANFIS and multi factor linear regression models for predicting density of all growing stages of Sunn pest. The data population fluctuation of Sunn pest in the years 2015 and 2016 on a farm with an area of one hectar in chadegan city was obtained. Predictor variables including variables sampling date, average temperature, average relative humidity, wind speed, wind direction, rainfall, height from sea level and degree- day were processed as input data to achive an output of number of developmental stages as response variable. In the ANFIS model, 70% of the data was assigned to training and 30% for validation. After network training and assessment of the best structure according to type, number of membership function and related rules with the use of MATLAB software, the appropriate model was selected based on statistical indices of, root mean square error (RMSE) and coefficient of determination (R^2). After sensivity analysis the results showed that ANFIS method (RMSE= 0.051, $R^2= 0.97$) had higher accuracy than multi linear regression (RMSE= 0.26, $R^2= 0.47$) and better predicts the population fluctuation of Sunn pest *Eurygaster integriceps*.

Keywords: ANFIS, Climate, Multi factor linear regression, Sunn pest, Population fluctuation.