

## خوشه‌بندی خودکار داده‌های حاصل از نمونه‌برداری و ارزیابی شبکه فازی - عصبی جهت تخمین پراکندگی سفیدبالک پنبه (*Bemisia tabaci* (Hem.:Aleyrodidae)

علیرضا شعبانی‌نژاد<sup>۱</sup> و بهرام تفقدی‌نیا<sup>۲\*</sup>

۱- دانشکده کشاورزی، دانشگاه شاهرود، شاهرود و ۲- گروه گیاه‌پزشکی سازمان پژوهش‌های علمی و صنعتی ایران، تهران،

\*مسئول مکاتبات، پست الکترونیکی: tafaghodinia@gmail.com

### چکیده

پژوهش حاضر با هدف پیش‌بینی تراکم سفیدبالک پنبه با روش شبکه‌ی فازی - عصبی مصنوعی در شهرستان بهبهان انجام گرفت. داده‌های مربوط به تراکم جمعیت سفیدبالک پنبه از طریق نمونه‌برداری بر روی یک شبکه علامت‌گذاری شده مربعی با ابعاد ۱۰×۱۰ متر و در مجموع از ۱۰۰ نقطه از سطح مزرعه به دست آمد. مختصات طول و عرض نقاط علامت‌گذاری شده سطح مزرعه به عنوان ورودی شبکه فازی - عصبی تعریف شد. خروجی نیز تعداد این آفت در آن نقاط بود. برای بررسی میزان حساسیت این روش به سطوح مختلف این آفت پس از جمع‌آوری نمونه‌ها از روش خوشه‌بندی اتوماتیک برای تعیین تعداد خوشه‌ها، واز شاخص دیویس و بولدین به عنوان معیار ارزیابی استفاده شد. به منظور جست و جوی فضای جواب خوشه‌بندی از الگوریتم ژنتیک استفاده شد. نتایج خوشه‌بندی بر اساس شاخص دیویس و بولدین (۰.۴۶) نشان داد که داده‌ها باید به سه خوشه تقسیم شود. نتایج نشان داد که در فازهای آموزش و آزمایش بین مقادیر ویژگی‌های آماری واریانس، توزیع آماری و میانگین مجموعه داده‌های واقعی و پیش‌بینی شده مکانیافتتوسط شبکه فازی - عصبی، تفاوت معنی‌داری وجود نداشت. نقشه‌های ترسیم شده نشان داد که پراکندگی این آفت به صورت تجمعی است و امکان کنترل متناسب با مکان را در مزرعه مورد مطالعه دارد.

واژگان کلیدی: الگوریتم ژنتیک، خوشه‌بندی اتوماتیک، سفید بالک پنبه، شبکه فازی - عصبی

## Automatic clustering of data from sampling and evaluating of neuro-fuzzy network forestimatinge the distribution of *Bemisia tabaci* (Hem.:Aleyrodidae)

Alireza Shabani nejad<sup>1</sup> & Bahram Tafaghodiniya<sup>2&\*</sup>

1. Faculty of Agriculture, Shahrood university, Shahrood & 2. Iranian Research Organization for Science and Technology.

\*Corresponding author, E-mail: tafaghodinia@gmail.com

### Abstract

In this study, Neuro Fuzzy network was used to estimate the spatial distribution of *Bemisia tabaci* in cucumber field in Behbahan. Pest density assessments were performed based on a 10 m × 10 m grid pattern and a total of 100 sampling units. In this method latitude and longitude information was used the input data and output of method showed the number of pest. To determine the sensitivity of this method to different levels of the pest after collecting samples, automatic clustering method was used to determine the number of clusters. Davies and Bouldin index was used to evaluate criterion. In order to find the answer, Clustering Search Space Genetic Algorithm was used. Davies and Bouldin index (0.46) showed that the data should be divided into three clusters. Results indicated average, variance, statistical distribution and also coefficient of determination in the observed and the estimated *Bemisia tabaci* density were not significantly different. Our map showed that patchy pest distribution offers large potential for using site-specific pest control on this field.

**Keywords:** Automatic clustering, *Bemisia tabaci*, Genetic Algorithm, Neuro Fuzzy network

Received: 24 April, Accepted: 3 June 2017

## مقدمه

خیار (*Cucumis sativus* L.) با ۱۴ کروموزوم، گیاه یک‌ساله دو جنسی، و متعلق به تیره کدوئیان است. این گیاه بومی کشور هندوستان است و به عنوان اقتصادی‌ترین گیاه تیره کدوئیان، یکی از محصولات ارزشمند جالیزی در خاورمیانه به شمار می‌رود (Nario et al., 2010). سطح زیر کشت و میزان تولید این محصول در ایران، به ترتیب ۸۰۰۰۰ هکتار و ۱۴۰۰۰۰۰ تن می‌باشد که اغلب در سیستم‌های مزرعه و گلخانه کشت می‌شود (Anonymous, 2011). خیار دارای آفات و بیماری‌های متعددی است که یکی از آفات مهم آن، سفیدبالک پنبه (*Bemisia tabaci* (Gennadius) (Hem.: Aleyrodidae) می‌باشد که علاوه بر خیار به حدود ۲۰۰ محصول زراعی و صیفی در اکثر مناطق جهان به‌ویژه در نواحی گرم، معتدل و خشک خسارت وارد می‌کند. حشرات بالغ و پوره‌های این آفت با تغذیه از شیر گیاهی سبب کاهش عملکرد محصول می‌گردند (Shishehbor, 2001). این آفت به صورت غیرمستقیم نیز با تولید عسلک، سبب چسبندگی ذرات گرد و غبار به سطح برگ‌ها شده و میزان فتوسنتز را کاهش می‌دهد. عسلک ناشی از این آفت موجب جلب قارچ‌های ساپروفیت موسوم به کپک سیاه یا دوده (مانند *Capnodium* spp.) می‌گردد که منجر به تضعیف گیاه و ایجاد لکه‌های رنگ پریده روی اندام‌های مختلف گیاه می‌شود (Gerling, 1990). سفیدبالک پنبه ناقل بیماری‌های ویروسی متعددی مانند موزاییک خیار است که خسارت جبران‌ناپذیری به گیاه وارد می‌کند (Yang et al., 2010). برای تدوین برنامه‌های مدیریت تلفیقی این آفت مهم، آگاهی از تغییرات جمعیت آن در مزرعه و در طی زمان و مکان ضروری است (Garzia et al., 2011). به همین خاطر امروزه تحقیقات زیادی در مورد تهیه و استفاده از نقشه‌های توزیع آفات با استفاده از مدل‌سازی پویایی جمعیت آفات به منظور بکارگیری در مدیریت متناسب با مکان در حال انجام است

(Dille et al., 2003). اما آنچه رسیدن به این اهداف را محقق می‌سازد افزایش صحت و دقت روش‌های درون‌یابی و تهیه نقشه‌های کاربردی از جمعیت آفات است. درون‌یابی مکانی شامل تخمین متغیرهایی مانند تراکم آفت مورد نظر در نقاط نمونه‌برداری نشده، با استفاده از داده‌های بدست آمده از نقاط نمونه‌برداری شده می‌باشد. از روش‌های درون‌یابی مورد استفاده در مطالعات حشره‌شناسی روش‌های مثل ثبندی خطی و کریجینگرا می‌توان نام برد. یکی از نارسایی‌های روش‌های متداول در آمار کلاسیک عدم توجه به سودمندی اطلاعات مربوط به موقعیت جغرافیایی مشاهدات بوده و متعاقب آن در تجزیه و تحلیل چنین داده‌هایی هیچگونه ارتباط ریاضی بین تغییرات مکانی نمونه‌ها به‌عنوان تابعی از فاصله بین آنها برقرار نمی‌شود (Makarjian et al., 2007). مطالعات متعددی نشان داده است که شبکه‌های عصبی توانایی بالایی در تشخیص الگوهای توزیع مکانی جمعیت‌های اکولوژیکی دارند (Goel et al., 2003 ; Irmak et al., 2006). در پژوهشی مدل شبکه عصبی بردار چندی ساز یادگیر برای مطالعه توزیع مکانی حشرات در اراضی مرتعی استفاده شد و کارایی مطلوبی نشان داد (Zhang et al., 2008). از شبکه عصبی LVQ4 در تخمین توزیع پراکندگی شب‌پره مینوز گوجه‌فرنگی استفاده شد و کارایی بسیار بالایی از خود نشان داد (Shabani nejad et al., 2016). بنا به بررسی‌های انجام شده، تاکنون پژوهشی به منظور ارزیابی عملکرد شبکه فازی - عصبی به عنوان یک روش نوین جایگزین در پیش‌بینی و خوشه‌بندی توزیع سفیدبالک پنبه صورت نگرفته است. به نظر می‌رسد که با تعیین مسیرهای ورود و خروج این حشره به مزرعه بهتر بتوان از عوامل کنترل‌کننده طبیعی و یا سموم در کنترل آفت در کانون‌های آلودگی بهره گرفت.

## مواد و روش‌ها

به منظور انجام این پژوهش یک مزرعه خیار به مساحت یک هکتار در اطراف شهرستان بهبهان در محدوده جغرافیایی بین ۴۷ درجه و ۴۱ دقیقه تا ۵۰ درجه و ۳۹ دقیقه طول شرقی از نصف‌النهار گرینویچ و ۲۹ درجه و ۵۸ دقیقه تا ۳۳ درجه و ۴ دقیقه عرض شمالی از خط استوا، انتخاب شد. جهت علامت‌گذاری موقعیت مکانی نقاط نمونه‌برداری پس از مشخص کردن جهت شمال زمین، نقطه‌ای در حاشیه جنوبی زمین به عنوان مبدا مختصات مشخص شد. مزرعه به شبکه‌های ۱۰ متر مربعی تقسیم، در تمامی نقاط نمونه‌برداری یک بلوک ۲×۲ متر مربع انتخاب و درون آن تعداد ۴ بوته به صورت تصادفی به عنوان واحد نمونه‌برداری انتخاب، و تعداد حشرات بالغ موجود پشت برگ گیاه شمارش و ثبت شد.

### سیستم استنتاج تطبیقی فازی - عصبی

سیستم استنتاج تطبیقی فازی - عصبی (ANFIS) اولین بار در سال ۱۹۹۳ توسط ژانگ معرفی شد (Jang, 1993). این سیستم ابزاری جهانی جهت تخمین توابع پیوسته حقیقی در دامنه‌ای محدود می‌باشد که به ازای هر درجه‌ای از صحت قابل استفاده است (Jang, et al., 1997).

شبکه فازی - عصبی از نظر کارایی معادل سیستم استنتاج فازی می‌باشد. در ادامه الگوریتم‌های آموزش مختلط که از ترکیب روش گرادیان کاهشی و حداقل مربعات استفاده می‌کند، معرفی شده و چگونگی تطبیق و آموزش سریع سیستم استنتاج فازی معادل با این الگوریتم بحث شده است.

اگر یک سیستم استنتاج فازی با دو ورودی X و Y و یک خروجی Z باشد مدل فازی مرتبه اول سونگو با دو قانون فازی If-then می‌تواند به صورت زیر بیان شود:

قانون ۱ - اگر X برابر A<sub>1</sub> و Y برابر B<sub>1</sub> باشد آنگاه:

$$z_1 = p_1x + q_1y + r$$

قانون ۲ - اگر X برابر A<sub>2</sub> و Y برابر B<sub>2</sub> باشد آنگاه (Amini, et al., 2005):

$$z_2 = p_2x + q_2y + r$$

لایه اول (ورودی): هرگاه عضوی از این لایه، مقادیر عضویتی را که به هر یک از مجموعه‌های فازی مناسب تعلق دارند، با استفاده از توابع عضویت آن را تولید می‌کند.

$$O_{1,i} = \mu_{A_i}(x) \quad \text{for } i = 1, 2 \text{ or}$$

$$O_{1,i} = \mu_{B_{i=2}}(y) \quad \text{for } i = 3, 4$$

که در آن X (و یا Y) ورودی‌های گره i ام بوده و A<sub>i</sub> (و یا B<sub>i=2</sub>) برچسب‌های زبانی (مانند کوچک یا بزرگ) مربوط به این گروه هستند. O<sub>1,i</sub> درجه عضویت فازی (A=(A<sub>1</sub>, A<sub>2</sub>, B<sub>1</sub>, B<sub>2</sub>) می‌باشد و درجه‌ای را مشخص می‌کند که متغیر ورودی X (و یا Y) با کمیت A مطابقت دارد. معمولاً تابع عضویت A و B توسط توابع زنگوله‌ای بیان می‌شود.

لایه دوم (ورودی تابع عضویت): این لایه از گره‌ای با عنوان II تشکیل شده که سیگنال‌های ورودی را درهم ضرب نموده و به خروجی ارسال می‌کند. به عنوان مثال:

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

به عبارت دیگر در این لایه عملگر (و) به کار برده می‌شود.  
 لایه سوم (قانون): در این لایه گره‌ها دارای برچسب N نسبت i امین قانون قوه آتش را به مجموع قوه اشتعال‌های همه قوانین محاسبه می‌کند.

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

خروجی این لایه قوه اشتعال نرمال شده نامیده می‌شود.  
 لایه چهارم (خروجی تابع عضویت): گره‌های این لایه با گره‌های توابع مطابقت دارند.

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

که در آن  $\bar{w}_i$  خروجی لایه سوم بوده و  $\{p_i, q_i, r_i\}$  مجموعه پارامترها هستند. پارامترهای این لایه به عنوان مجموعه پارامترهای بخش تالی مدل فازی می‌باشند.  
 لایه پنجم (خروجی): این لایه با گره ثابت با نام  $\sum$ ، خروجی کلی را با جمع کردن همه سیگنال‌های ورودی محاسبه می‌کند. بنابراین در این لایه فرآیند غیر فازی‌سازی، نتایج هر قانون فازی را به خروجی غیر فازی تبدیل می‌کند (Afrous et al., 2007).

$$O_{5,i} = \sum_{i=1} \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

### مدل یادگیری ترکیبی

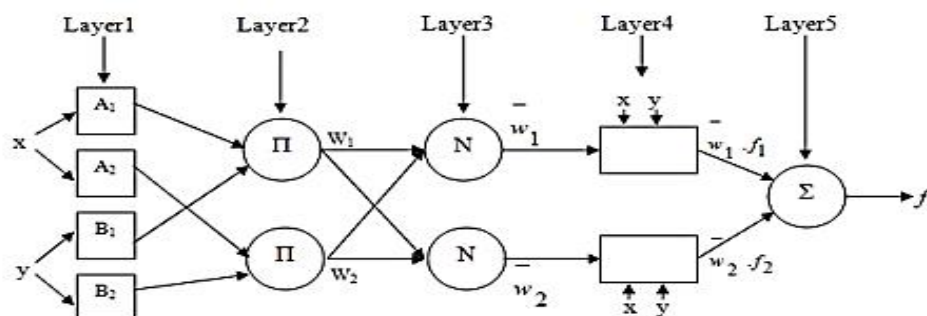
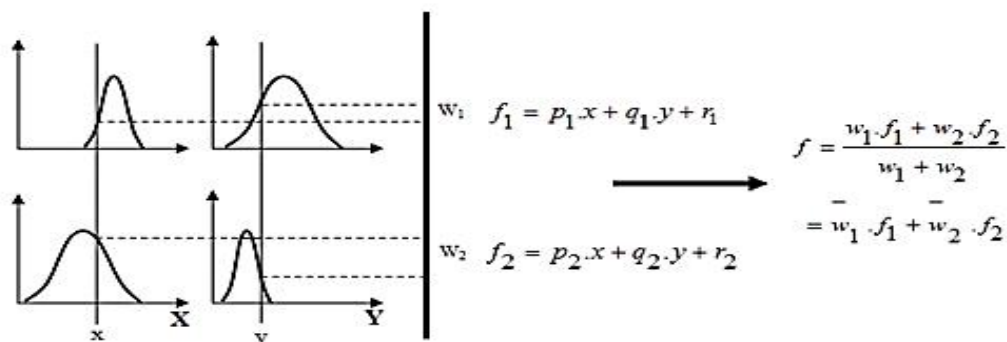
مدل فازی - عصبی اجازه می‌دهد که سیستم‌های فازی در مباحث آموزش پارامترها از الگوریتم آموزش برگشتی تطابقی استفاده نمایند. در این روش با استفاده از الگوریتم شیب نزولی خطا، مقدار خطا به سمت ورودی‌ها پخش شده و پارامترها تصحیح می‌شوند. این روش آموزش دقیقاً همانند روش پس‌انتشار خطای مورد استفاده در شبکه عصبی مصنوعی است (Mertens & Huwe, 2002). طبق ساختار شبکه‌های فازی - عصبی که در شکل ۱ ملاحظه می‌گردد، می‌توان خروجی کل (f) را به صورت ترکیب خطی پارامترهای نتیجه به صورت زیر نوشت:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 = (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_1 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2$$

به این ترتیب پیاده‌سازی یک سیستم فازی به گونه‌ای صورت می‌گیرد که قابلیت یادگیری داشته باشد. بنابراین مقادیر پارامترهای نتیجه شده با استفاده از روش کمترین مجموع مربعات خطا بدست می‌آیند. با ترکیب این روش و روش پس‌انتشارخطای یک روش آموزشی ترکیبی ایجاد می‌شود که به صورت زیر عمل می‌کند:  
 در هر دو آموزش همگام حرکت رو به جلو خروجی‌های گره‌ها به صورت عادی تا لایه چهارم محاسبه می‌شوند و سپس پارامترهای نتیجه توسط روش‌های محاسبه خطا نظیر کمترین مجموع مربعات خطا محاسبه می‌شوند. در ادامه پس از محاسبه خطا در بازگشت رو به عقب نسبت خطا بر روی پارامترهای شرط پخش شده و با استفاده از روش گرادینان نزولی خطا مقدار آنها تصحیح می‌شوند (Mousavi & Amiri, 2012).

### انتخاب تابع عضویت

برای در نظر گرفتن درجه عضویت از توابع عضویت استفاده می‌شود که شامل توابع عضویت مثلثی، ذورنقه‌ای، گوسی، گوسی دو طرفه، سیگموئیدی است (Shrestha et al., 2007). با توجه به عدم وجود پیش فرض برای انتخاب مناسب این توابع تمامی این توابع بررسی شدند. بر اساس جدول ۱ تابع سیگموئیدی با کمترین مربعات خطا به عنوان تابع عضویت انتخاب شد.



شکل ۱- سیستم استنتاج فازی (A) مدل انفیس معادل (B). (Shrestha et al., 2007)

Fig. 1. Fuzzy inference systems (A) Equivalent ANFIS model(B)

### پیش پردازش داده‌ها

در ابتدا داده‌ها بطور تصادفی به دو دسته مجموعه آموزش با ۷۰ عضو (۷۰ درصد کل داده‌ها) و مجموعه آزمایش با ۳۰ عضو (۳۰ درصد کل داده‌ها) تقسیم بندی شد. قبل از بکارگیری داده‌های خام اولیه در آموزش شبکه، باید داده‌ها در دامنه مناسبی نرمال‌سازی شوند زیرا الگوریتم یادگیری همراه با داده‌های خام نمی‌تواند عملکرد مناسبی داشته باشد و شبکه در طول فاز آموزش همگرا نخواهد شد. نتیجتاً نتایج مطلوبی هم به دست نمی‌آید (Yuxin et al, 2006). برای تبدیل داده‌ها از روش نرمال‌سازی خطی معادله (۱) استفاده شد:

$$X_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \times (\Gamma_{\max} - \Gamma_{\min}) + \Gamma_{\min} \quad (1)$$

در اینجا  $X$  داده‌ی خام اولیه،  $X_n$  داده‌ی نرمالیزه شده،  $x_{\min}$  و  $x_{\max}$  به ترتیب مقادیر بیشینه و کمینه داده‌های اولیه،  $r_{\min}$  و  $r_{\max}$  هم به ترتیب حد بالایی و پایین دامنه تغییرات داده‌های تبدیل شده است. برای ارزیابی قابلیت شبکه عصبی - فازی به تغییرات پراکندگی این آفت در سطح زمین که در نهایت منجر به ترسیم نقشه‌های پراکندگی دقیق‌تر می‌شود می‌بایست چند سطح آستانه فرضی تعریف کرد، با توجه به این موضوع که هیچ پیش‌زمینه برای تعیین تعداد این حدود آستانه وجود ندارد لذا از روش خوشه‌بندی استفاده شد. خوشه‌بندی یکی از شاخه‌های یادگیری بدون نظارت می‌باشد و فرآیند خودکاری است که در طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر می‌باشند تقسیم می‌شوند که به این دسته‌ها خوشه گفته می‌شود. بنابراین خوشه مجموعه‌ای از اشیاء می‌باشد که در آن اشیاء با یکدیگر مشابه بوده و با اشیاء موجود در خوشه‌های دیگر غیرمشابه می‌باشند (Murphy, 2012). در خوشه‌بندی هیچ دسته از پیش تعیین شده‌ای وجود ندارد و داده‌ها صرفاً براساس تشابه گروه‌بندی می‌شوند و عناوین هر گروه نیز توسط کاربر تعیین می‌گردد.

جدول ۱- میانگین مربعات خطا هر تابع

Table 1. Mean square error of each function

Kind of function	goals for clusters	MSE
Trapezoidal	*BD=0 Bd≤5	0.1323
	BD =0 BW≤10	0.1412
	BD=0 BD≤15	0.1345
Gaussian	BD=0 Bd≤5	0.1342
	BD =0 BW≤10	0.1219
	BD=0 BD≤15	0.0921
Triangular	BD=0 Bd≤5	0.1003
	BD =0 BW≤10	0.0994
	BD=0 BD≤15	0.0803
Sigmodial	BD=0 Bd≤5	0.00331
	BD =0 BW≤10	0.00221
	BD=0 BD≤15	0.00189

\*BD=Bemisia Density

برای یک خوشه‌بندی نخستین موضوعی که باید به آن پرداخت شود سنجش فاصله میان داده‌ها و انتخاب رویکردی جهت جست و جوی فضای جواب و بهینه‌سازی تابع هدف است. یکی از تکنیک‌های جست و جوی فضای جواب و بهینه‌یابی، روش‌های فرا ابتکاری هستند که از آن جمله می‌توان به الگوریتم ژنتیک، الگوریتم کلونی مورچگان، روش جست و جو ممنوع اشاره نمود (Velmurugan, 2010). از میان روش‌های فراابتکاری، الگوریتم ژنتیک با توجه به قدرت و قابلیت بالایی که در جست و جوی فضای جواب دارد، از اهمیت ویژه‌ای برخوردار است و در این تحقیق از این الگوریتم برای جست و جوی فضای جواب از این الگوریتم استفاده شد.

### الگوریتم خوشه‌بندی

هدف الگوریتم‌های خوشه‌بندی به حداقل رساندن عدم تشابه نمونه‌ها در یک گروه است بنابراین نمونه‌ها با تشابه بالا در یک خوشه کلاسه‌بندی می‌شوند (Nugraha, 2011). اجرای این روش شامل مراحل زیر است:

#### محاسبه مراکز خوشه‌ها برای داده‌های جمع‌آوری شده از سطح زمین

برای تعریف مراکز خوشه‌ها، مقدار مرکزی به ازای مشخصه‌های عددی (تعداد آفت شمارش شده) با مقدار میانگین نمایش داده می‌شوند.

#### فاصله بین یک داده و مرکز خوشه متناظرش

در مورد مشخصه‌های عددی، فاصله اقلیدسی میان مقدار مشخصه عددی و میانگین مقادیر آن مشخصه در خوشه مورد نظر مورد استفاده قرار می‌گیرد (Ahmad & Dey, 2007).

#### اجزای الگوریتم ژنتیک در خوشه‌بندی

در این خوشه‌بندی کروموزوم‌ها از اعداد حقیقی تشکیل شده‌اند و مقادیر مربوط به مراکز خوشه‌ها را در خود جای داده‌اند. طول کروموزوم‌ها ثابت و برابر مقدار  $K_{max}$  است. مقدار  $K$  یعنی تعداد خوشه‌ها به صورت تصادفی از بازه  $[K_{min}, K_{max}]$  انتخاب می‌شوند که این متغیر جز ورودی‌های مسأله است که توسط شخص به صورت پیش فرض وارد می‌شود. پس از مشخص شدن مقدار  $K$ ، تعداد  $K$  ژن، مراکز خوشه‌ها را در خود جای می‌دهند و به مابقی ژن‌ها یک عدد خاص (به صورت فرض  $-1$ ) تخصیص داده می‌شود تا مشخص شود که ژن مربوطه خالی است و مرکز خوشه‌ای در آن قرار نگرفته است.

#### مقداردهی اولیه جمعیت

به ازای هر کروموزوم  $i$  در جمعیت ( $P, \dots, 1, I = P$  و  $P$  برابر با اندازه جمعیت)، یک مقدار تصادفی  $K_i$  در بازه تعریف شده تولید می‌شود. سپس  $K_i$  نقطه به صورت تصادفی از میان داده‌ها انتخاب می‌شود و به صورت تصادفی در میان ژن‌های کروموزوم قرار داده می‌شود. در نهایت به ژن‌های خالی کروموزوم مقدار  $(-1)$  تخصیص داده می‌شود.

#### عملگر تقاطع و جهش

به منظور تقاطع کروموزوم‌های منتخب والد و تولید کروموزوم‌های فرزند، از عملگر تقاطع یک نقطه‌ای استفاده شده است. عملگر تقاطع به منظور تولید نسل در هر تکرار مورد استفاده قرار می‌گیرد و هدف اصلی از به کارگیری این عملگر، تولید نسلی با برازندگی بهتر است، مشروط به اینکه هر یک از فرزندان، مشخصه‌های مطلوب والدین خور را به ارث می‌برند (Kumar et al. 2006). برای انتخاب کروموزوم‌های والد از رویکرد چرخ رولت بهره گرفته شد (Goldberg, 1999). پس از به کارگیری عملگر تقاطع، جهت بررسی گسترده‌تر فضای مسئله، از عملگر جهش نیز استفاده می‌شود.

### تابع برازندگی

یکی دیگر از مواردی که در طراحی الگوریتم‌های خوشه‌بندی باید مد نظر قرار گیرد، انتخاب مقیاس اعتبار مناسب جهت انتخاب به عنوان تابع برازندگی است. شاخص‌های اعتبار مختلفی نظیر شاخص دان، شاخص (چی-بنی)، شاخص (دیویس و بولدین) در این زمینه ارائه شده است. شاخص دیویس و بولدین به صورت تابعی از نسبت مجموع پراکندگی نقاط در داخل خوشه به جدایی بین خوشه‌ها تعریف می‌شود، در مقایسه با سایر شاخص‌ها نتایج دقیق‌تری به دست می‌دهد. نکته مهم درباره این شاخص این است که مقادیر کوچکتر این شاخص نشان دهنده خوشه بندی داده‌ها به نحوی بهتری است (Paasche et al., 2010).

### شرط توقف

تعداد کل تکرار الگوریتم‌های ژنتیک برابر با ۲۰۰ در نظر گرفته شده است و اگر بعد از ۵۰ تکرار بهبودی در مقدار برازندگی به وجود نیاید، الگوریتم متوقف می‌شود.

### تنظیم پارامترهای ورودی الگوریتم

مقدار پارامترهای ورودی الگوریتم شامل اندازه جمعیت (تعداد کروموزوم‌ها)، نرخ کراسینگ آور، نرخ جهش، حداکثر تعداد نسل‌ها و نیز  $K_{min}$  و  $K_{max}$  تعیین گردید که مقدار این پارامترها در جدول ۲ آورده شده است.

minimum number of clusters ( $K_{min}$ )	2	
maximum number of clusters ( $K_{max}$ )	20	
maximum number of generations (Max - gen)	100	جدول ۲- مقادیر پارامترهای ورودی الگوریتم
Crossover rate (Pc)	0.5	<b>Table 2</b> .The values input parameters of Algorithm
Mutation rate (Pm)	0.02	
Population size	50	

### تجزیه و تحلیل آماری

تست نرمال‌سازی داده‌های مربوط به نمونه‌برداری به کمک نرم‌افزار SPSS 19 و آزمون کولموگروف-اسمیرنوف، بررسی شد و به دلیل عدم نرمال بودن آنها، داده‌ها توسط روش کاکس باکس خطی نرمال شدند. کد کامپیوتری شبکه عصبی فازی و خوشه‌بندی اتوماتیک در محیط نرم افزار Matlab نسخه ۸/۱ تهیه شد.

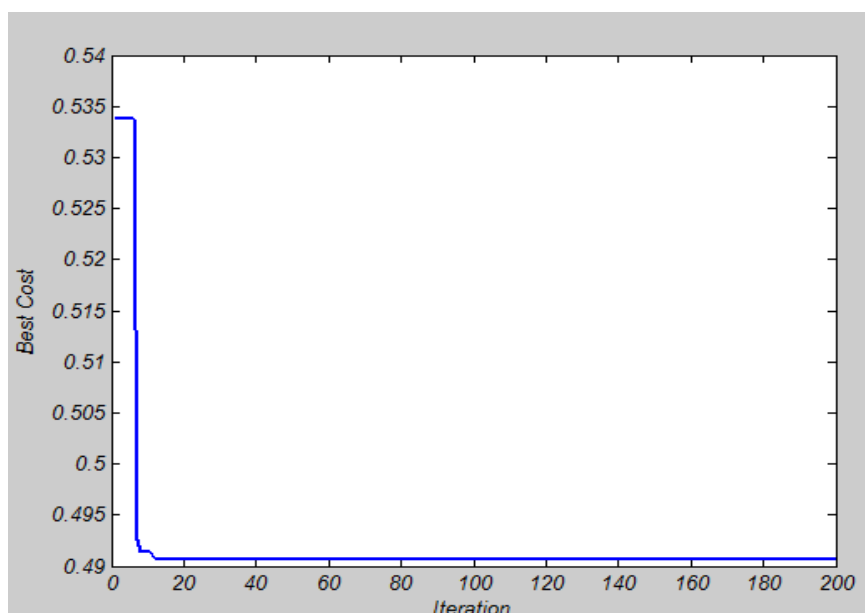


## نتایج و بحث

### نتیجه خوشه‌بندی روی داده‌ها

با توجه به شکل شماره ۲ پس از تکرار ۱۸ الگوریتم ژنتیک روند کاهش شاخص دیویس و بولدین ثابت مانده مجموعه داده‌های جمع‌آوری شده از طریق نمونه‌برداری با توجه به شاخص دیویس و بولدین که (۰/۴۶) به دست آمده به سه خوشه باید تبدیل شوند. لذا برای ارزیابی قابلیت شبکه نو - فازی داده‌ها باید به سه خوشه با مراکز ( $BD \leq 5$ ,  $BD \leq 10$ ,  $BD \leq 15$ ) تقسیم شدند.

برای اطمینان از یادگیری شبکه عصبی آموزش دیده برای پیش‌بینی الگوی پراکنش سفید بالک پنبه، داده‌های واقعی و پیش‌بینی شده توسط شبکه از نظر آماری با هم مقایسه شدند. در اینجا فرض صفر بر تساوی میانگین، واریانس و توزیع آماری دلالت دارد. هر فرضیه در سطح احتمال ۹۵ درصد به کمک پارامتر  $p$  آزمون شد. به ترتیب برای مقایسه میانگین، واریانس و توزیع آماری از آزمون  $F$ ،  $T$  و کولموگروف - اسمیرنو استفاده شد. مقادیر  $P$  محاسبه شده برای هر مورد در جدول ۳ نشان داده شده است. نتایج نشان می‌دهد که میانگین و واریانس برای شبکه عصبی - فازی اختلاف معنی داری را نشان نمی‌دهد ( $p > 0.60$ ) و بین توزیع آماری مقادیر واقعی و پیش‌بینی شده توسط شبکه فازی - عصبی اختلاف معنی داری در سطح ۹۵ درصد وجود ندارد ( $p > 0.84$ ). وجود  $p=1$  در مورد توزیع آماری بین مقادیر پیش‌بینی شده و مقادیر واقعی تراکم سفید بالک پنبه در فاز آموزش و همچنین وجود  $p=1.00$  در فاز آزمایش در سطح مزرعه، نشان از دقت بالا و نیز قابلیت این شبکه نو - فازی، به منظور خوشه‌بندی سطح مزرعه از نظر تراکم این آفت با هر مقدار تراکم بحرانی می‌باشد.



شکل ۲- همگرایی شاخص دیویس و بولدین توسط الگوریتم ژنتیک

Fig. 2. Davis and Bouldin convergence by Genetic Algorithm

ضرایب تبیین و رابطه خطی رگرسیونی بین مقادیر واقعی هر کلاس در مقابل مقادیر پیش‌بینی شده توسط شبکه نو - فازی در جدول شماره ۴ نشان داده شده است. بهترین نتایج بر اساس این دو معیار وقتی به دست

می آید که معادله خطی مابین تراکم آفت واقعی و تراکم آفت پیش بینی شده توسط شبکه نرو - فازی علاوه بر داشتن ضریب تبیین بالا دارای عرض از مبدا کم و شیب نزدیک به یک باشد. کاملاً واضح است که ضرایب تبیین بین داده های واقعی و پیش بینی شده بسیار بالا می باشند (۰/۸). نتایج جدول ۴ نشان می دهد که قابلیت تعمیم پذیری شبکه عصبی آموزش دیده بسیار بالا می باشد زیرا فاز آزمایش در مقایسه با فاز آموزش دارای ضریب تبیین بالاتر و نیز معادله خطی رگرسیونی مابین مقادیر واقعی و پیش بینی شده از شیب نزدیکتر به واحد و عرض از مبدا نزدیک به صفر برخوردار است.

### نقشه های توزیع مکانی سفید بالک پنبه

نقشه پراکندگی مکانی سفید بالک پنبه توسط شبکه نرو- فازی تفکیک و ترسیم شده و در شکل ۳ نشان داده شده است. در این شکل ابتدا تراکم این آفت به دو خوشه فاقد، و دارای ۵ آفت، ترسیم و نقشه های آن رسم شد (شکل ۳ a و b). در شکل ۳ c و d و نیز در شکل ۳ e و f به ترتیب، جمعیت این آفت بر اساس دو مرکز خوشه ۵ و ۱۵ عدد ترسیم شده است. این حدود انتخابی یک حد آستانه محاسبه است تا به وسیله آنها قابلیت شبکه نرو - فازی نشان داده شود. اما اکنون با مقایسه بین نقشه های حاصله از داده های واقعی و پیش بینی شده توسط شبکه متوجه خواهیم شد بین نقاط متناظر در نقشه در برخی از مکان ها تفاوت هایی وجود دارد که ناشی از خطای مدل شبکه عصبی - فازی است. ولی در مورد صحت کارایی یک مدل نمی توان تنها با مقایسه یک یا چند نقطه قضاوت کرد (Zhang & Fuh, 1998).

جدول ۳- مقایسات آماری مقادیر واقعی و پیش بینی تراکم سفید بالک پنبه توسط شبکه نرو- فازی

**Table 3.** Statistical comparisons between the observed and estimated to *Bemisiatabaci* by Neuro Fuzzy networks.

Clusters	Members of cluster 1	Members of cluster 2	Network Utilization phase	Comparisons of means	Comparisons of variance	Comparisons of distribution
BD=0 Bd≤5	51	49	Training Phase	0.614	0.906	0.84
			Test Phase	0.978	0.999	0.95
BD=0 BW≤10	80	20	Training Phase	0.745	0.976	1.00
			Test Phase	0.986	1.00	1.00
BD=0 BD≤15	91	9	Training Phase	0.799	0.984	1.00
			Test Phase	0.994	1.00	1.00

در مقایسه نقشه های تفکیک شده توسط مدل شبکه عصبی و نقشه های واقعی بخش های شمالی و جنوبی زمین شکل ۳ c و d دارای آفت بودند که توسط شبکه عصبی قابل پیش بینی نبود. در مقایسه سه مدل شبکه عصبی برای تعیین الگوهای پراکندگی حشرات در سطح یک چمنزار، بیان شد که شبکه عصبی MLP الگوریتم قوی تری در تشخیص الگوی پراکندگی حشرات بود، آنها رفتار اکولوژیک حشرات در کارایی مدل های شبکه عصبی موثر دانستند (Zhang et al., 1998). لذا برای بالابردن کارایی یک شبکه عصبی برای تشخیص پراکندگی

تعداد لایه‌های مخفی، توابع تحریک و حتی مرحله زیستی حشره موثر است (Filippi & Jensen, 2006 Zhang *et al.*, 2008). در پژوهش دیگری کارایی بالای شبکه عصبی مصنوعی MLP در پیش‌بینی پویایی جمعیت مگس گالزا (Diptera: Cecidomyiida) در جنگل‌های سوزنی‌برگ آمریکا به اثبات رسید (Yeong *et al.*, 2000). در پژوهشی روی خیار پاییزه در منطقه ملاتانی استان خوزستان با استفاده از قانون تیلور پراکندگی سفید بالک پنبه مورد بررسی قرار گرفت که نتایج نشان از تجمع بودن تراکم این آفت داشت (zandisohani *et al.*, 2012). همچنین در پژوهشی مستقل که بر روی گیاه طالبی انجام شد نیز پراکندگی این آفت به صورت تجمعی گزارش شد (tonhasca *et al.*, 1994). در بررسی دیگری که با استفاده از قانون تیلور و آیواو پراکندگی سفید بالک پنبه را روی خیار مورد بررسی قرار داد، بیان شد که پراکنش این آفت به صورت تجمعی است (Naranio & Flint, 1994). مطالعه‌ای که روی پراکندگی سفید بالک پنبه روی بادنجان منطقه ورامین انجام شد نشان داد که پراکندگی این آفت روی این گیاه نیز تجمعی است (Kianpouret *et al.*, 2009). نیز نتایج به دست آمده از این پژوهش‌ها با نتایج حاصل از این پژوهش هماهنگی داشت. آستانه اقتصادی به صورت تراکمی از آفت تعریف می‌شود که در آن هزینه کنترل آفت معادل سود حاصل از کنترل است (Seraj, 2011). واضح است که نقاط دارای تراکم بیشتر از حد آستانه در مدیریت متناسب با مکان، مورد سم‌پاشی قرار می‌گیرند و سایر نقاط که زیر حد آستانه قرار دارند نیاز به مصرف سم ندارند. نقشه‌های تهیه شده به وسیله شبکه عصبی چنانچه از دقت بالایی برخوردار باشند می‌توانند به مدیر مزرعه در جهت اعمال کنترل محدوده‌ای آفت‌کش کمک کنند. علاوه بر این چنانچه نقشه‌ها منطبق بر مختصات جغرافیایی هر مکان تهیه شوند، راهنمای خوبی برای فعال شدن نازل‌ها در نقاط بالای حد آستانه آفت در عملگرهای نرخ متغیر خواهند بود. نقشه‌های ترسیم شده نشان دهنده توزیع به صورت تجمعی است. با توجه به مناسب بودن پارامترهای آماری می‌توان با اطمینان به نقشه‌های شبکه عصبی محل‌های مناسب سم‌پاشی را انتخاب نمود و به صورت موضعی نقاط آلوده را سم‌پاشی کرد و از سم‌پاشی سراسری مزرعه و آلودگی بیشتر محیط زیست جلوگیری کرد.

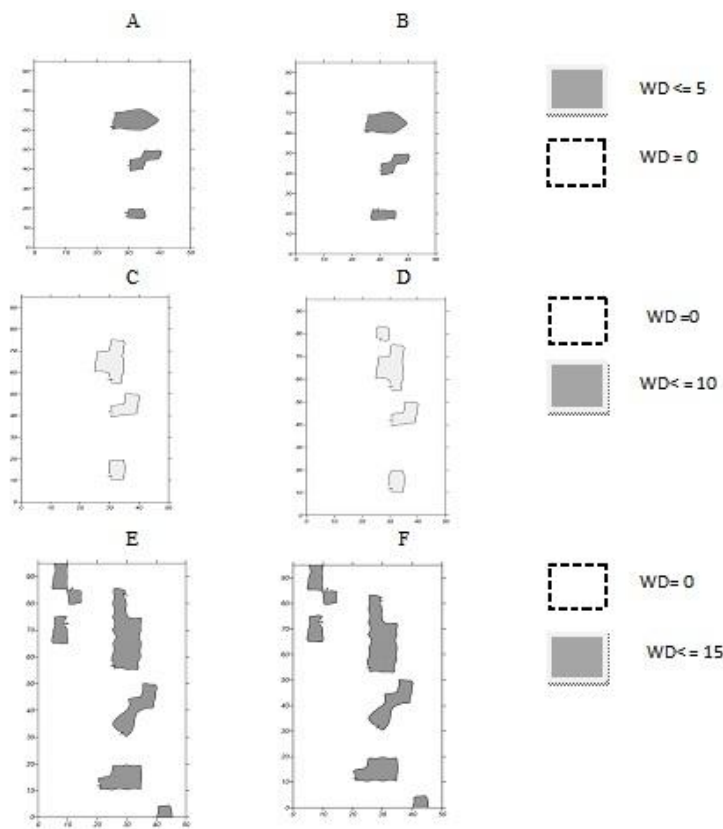
جدول ۴- رابطه خطی رگرسیونی و ضریب تبیین بین av (مقادیر واقعی داده‌ها) و Pv (مقادیر پیش‌بینی شده) توسط مدل

**Table 4.** Linear regression relationship and coefficient of determination between dv (actual value) and pv (predicted value by model)

Clusters	Members of cluster 1	Members of cluster 2	Network Utilization phase	Linear regression	Coefficient of Determination ( $r^2$ )
BD=0 Bd≤5	51	49	Training Phase	pv= 0.9405 av +0.0149	0.9387
			Test Phase	pv= 0.8606 dv +0.0108	0.899
BD =0 BW≤10	80	20	Training Phase	pv= 0.9795 av +0.0116	0.9671
			Test Phase	pv= 0.9035 av +0.0108	0.9119
BD=0 BD≤15	91	9	Training Phase	pv= 0.9852 av +0.0041	0.9898
			Test Phase	pv= 0.9089 av +0.0054	0.9333

## نتیجه‌گیری

در این پژوهش با استفاده از روش خوشه‌بندی اتوماتیک، داده‌ها جهت بررسی دقت شبکه عصبی-فازی دسته‌بندی شدند و این شبکه توانست با دقت مناسبی علاوه بر پیش‌بینی، نقشه پراکنش سفید بالک پنبه را ترسیم نماید. نقشه حاصله بیان‌کننده پراکنش تجمعی این آفت است. بنابراین تنها با سم‌پاشی محل‌هایی با تراکم بالا می‌توان به مدیریت مناسب مزرعه و کاهش مصرف سموم که در نهایت سبب افزایش سلامت جامعه شود دست یافت.



شکل ۳- نقشه توزیع سفید بالک پنبه در دو حالت واقعی (شکل‌های b, d, f) و تفکیک شده توسط مدل شبکه عصبی (شکل‌های a, c, e). نقشه‌های (a, b) بر اساس حد آستانه پنج آفت، نقشه‌های (c, d) حد آستانه ده آفت و نقشه‌های (e, f) بر اساس حد آستانه پانزده آفت.

**Fig. 3.** *Bemisia tabaci* distribution maps in actual (b, d and f) and classified conditions by Neuro Fuzzy neural networks (c, e and a). The maps of a, c, e and b, d, f have been drawn according to economic threshold of 5, 10 and 15 respectively.

## سپاسگزاری

بدینوسیله از نمایندگی محترم شرکت بازرگان کالا جناب مهندس امرالله فرج‌پور به دلیل در اختیار نهادن زمین کشاورزی برای این پژوهش قدردانی می‌گردد.

## References

- Amini, M., Afyuni, M., Fathianpour, N., Khademi, H & Fluhler, H.** (2005) Continuous soil pollution mapping using fuzzy logic and spatial interpolation. *Geoderma* 124, 223- 233.
- Afrous, A., Hosseini, S.M. & Goudarzi, Sh.** (2007) Assessment of the Ordinary Kriging and NeuroFuzzy approaches in interpolation of the groundwater level. *Journal of Groundwater* 13, 978-984.
- Ahmad, A., Dey, L.** (2007) A k-Mean Clustering Algorithm for Mixed Numeric and Categorical Data. *Data & Knowledge Engineering* 63, 503- 527.
- Anonymous.** (2011) Agricultural statistics, Department of Planning and Economy, The office of Statistics and Information Technology, Tehran
- Dille, J. A., Milner, M., Groeteke, J. J., Mortensen, D. A. & Williams, M. M.** (2003) How good is your weed map? A comparison of spatial interpolators. *Weed Science* 51, 44 – 55.
- Filippi, A. M. & Jensen, J. R.** (2006) Fuzzy learning vector quantization for hyper spectral coastal vegetation classification. *Remote Sensing Environment* 100, 512–530.
- Goldberg, D.** (1999) *Genetic algorithms in search optimization and machine learning*. 1th ed. 570 pp. Addison-Wesley Longman Publishing Company.
- Garzia, T. G., Siscaro, G., Biondi, A. & Zappala, L.** (2011) Distribution and damage of *Tuta absoluta*, an exotic invasive pest from South America. In: International symposium on management of *Tuta absoluta* (Tomato borer) Proceeding. Agadir, Morocco, November, 16-18.
- Goel, P. K., Prasher, S. O., Patel, R. M., Landry, J. A., Bonnell, R. B. & Viau, A. A.** (2003) Classification of hyper spectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Computers and Electronics in Agriculture* 39, 67–93.
- Gerling, D.** (1990). *Whiteflies: their bionomics, pest status and management*. 2<sup>th</sup>. 562pp. Oxford University Press.
- Irmak, A., Jones, J. W., Batchelor, W. D., Irmak, S., Boote, k. J. & Paz, J.** (2006) Artificial neural network model as a data analysis tool in precision farming. *Transactions of the American Society of Agricultural and Biological Engineers* 49, 2027-2037.
- Jang, J.S.R.** (1993) Anfis: adaptive-network-based fuzzy inference systems. *Journal of IEEE Transactions on System, Management and Cybernetics* 23, 665–685.
- Jang, J. S. R., Sun, C.T. & Mizutani, E.** (1997) *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence Upper Saddle River*. 4<sup>th</sup> ed. 560 pp. New Jersey University Press.
- Kumar, D. N., Srinivasa, R. K. & Ashok, B.** (2006) Optimal reservoir operation for irrigation of multiple crops using genetic algorithms. *Journal of Irrigation Drainage Engineering* 132, 123-129.
- Kianpour, R., Fathipour, Y & Kamali, K.** (2002). Population Fluctuation and spatial distribution patterns b. Tabasi on eggplants in varamin. *Journal of Plant protection* 77, 71-94.
- Makarjian, H., Rashed Mohassel, M. H., Bannayan, M. & Nassiri, M.** (2007) Soil seed bank and seedling populations of *Hordeum murinum* and *Cardaria draba* in saffron fields. *Agriculture Ecosystems and Environment* 120, 307- 312.
- Mertens, M. & Huwe B.** (2002) Fun-Balance: a fuzzy balance approach for the calculation of nitrate leaching with incorporation of data imprecision. *Geoderma* 109, 269 287.

- Mousavi, S.F. & Amiri, M.J.** (2012) Modelling nitrate concentration of groundwater using adaptive neural based fuzzy inference system. *Journal of Soil and water Research* 7, 73-83. [In Persian with English summary]
- Murphy, K.** (2012) *Machine learning a probabilistic perspective*. 2<sup>th</sup> ed. 875 pp. MIT Press.
- Nario, L. S., Oliver-Verel, J., & Stashenko, E. E.** (2010). Repellent activity of essential oils. A review. *Bioresource Technology* 101, 372-378.
- Nugraha, H.S.** (2011) Integration of stream sediment geochemical and airborne gamma-ray data for surficial lithological mapping using clustering methods, Master of Science Thesis, Twente University, Netherland, Supervisor: E.J.M. Caranza.
- Paasche, H. & Eberle, D.** (2010) automated integration of large geophysical data using three partitioning cluster algorithms: A Comparison. 11<sup>th</sup> SAGA Biennial Technical Meeting and Exhibition Swaziland, pp. 286-291.
- Shrestha, R.R, Bardossy, A., Rode, M.** (2007). A hybrid deterministic fuzzy rule based model for catchment scale nitrate dynamics. *Journal of Hydrology* 342, 143-156.
- Seraj A. A.** (2011) *Principle of Plant pest control*. 1<sup>th</sup> ed. 7745pp. Shahid chamran Press. [In Persian]
- Shishehbor, P.** (2001). *White fly*. 1<sup>th</sup> ed. 750pp. Shahid chamran Press. [In Persian]
- Tonhasca, A., Palumbo, J.C. & Byrne, D.N.** (1994). Distribution pattern of *bemisia tabaci* in cantaloupe fields in Arizona. *Environmental Entomology* 23, 949 – 954.
- Velmurugan, T.** (2010) Performance evaluation of K-Means and Fuzzy K-Means clustering algorithms for statistical distributions of input data points. *European Journal of Scientific Research* 46, 320-330
- Yuxin, M., Mulla, D. J. & Pierre, C. R.** (2006) Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. *Precision Agriculture* 7, 117–135.
- Young-S. P., K, Ja-Myung, L, Buom-Young, L, Yeong-Jin & YooShin, K.** (2010) Use of an Artificial Neural Network to Predict Population Dynamics of the Forest-Pest Pine Needle Gall Midge (Diptera: Cecidomyiida). *Environmental Entomology* 29, 1208-1215.
- Zhang, Y. F. & Fu, J. Y.H.** (1998) A neural network approach for early cost estimation of packaging products. *Computers & Industrial Engineering*, 34, 433-50.
- Zandi sohani, N., Shishehbor, P. & Kocheli, F.** (2012) Seasonal changes and spatial distribution of *Bemisia tabaci* on cucumbe in ahvaz. *Plant Protection* 35, 73-85. [In Persian with English summary]
- Naranjo, S.E. & Flint, H.M.** (1994) spatial distribution *bemisia tabaci* in cotton and development of fixed- precision sequential plant. *Environmental Entomology* 23, 245-266.
- Zhang, W. J., Zhong, X. Q. & Liu, G. H.** (2008) Recognizing spatial distribution patterns of grassland insects: neural network approaches. *Stochastic Environmental. Research and Risk Assessment* 22, 207–216.