

## The application of the new intelligent Adaptive Nero Fuzzy Inference System (ANFIS) in prediction of human papilloma virus oncogenicity potency

Kenar Koohi A(MSc)<sup>1</sup>, Soleimanjahi H(PhD)<sup>1\*</sup>, Falahi Sh(MSc)<sup>2</sup>, Riahi M H(PhD)<sup>1</sup>, Meshkat Z(PhD)<sup>1</sup>

1- Department of Virology, Tarbiat Modares University, Tehran, Iran

2- Department of Water Engineering, Shahid Bahonar University, Kerman, Iran

Received 22 Dec 2009 Accepted 14 Apr 2010

### Abstract

**Background:** Based on the severity and prognostic condition of respective cancers caused by them, papilloma viruses are classified into high, medium, and low risk groups using E6 and E7 viral proteins. Nowadays, different methods of modeling in clinical medicine are used for diagnosis of diseases and evaluation of their molecular characteristics. Among the new methods of modeling, fuzzy systems are of particular importance in various fields of science. The aim of this study was to use a new intelligent Adaptive Nero Fuzzy Inference System (ANFIS) for predicting human papilloma virus oncogenicity based on a number of biochemical properties of E7 protein.

**Materials and Methods:** In this study, using ANFIS model, a new model was developed for predicting oncogenicity of papilloma virus isolated from patients. The process of training and testing was performed using a set of available published filed data and several statistical and graphical criteria. Accordingly, through provision of needed biochemical and biophysical data on E7 gens from the existing data, this model was developed. The results of this model were, then, validated by the authentic published data.

**Results:** Based on the results, the developed model is capable of predicting papilloma virus oncogenicity efficiently. R2 and RMSE values in training stage were 0.99 and 101.18, respectively. In the testing stage, however, they stood at 0.94 and 173.8, respectively.

**Conclusion:** Based on the findings, the use of ANFIS model significantly improves the accuracy of estimating virus oncogenicity phenomenon. The methodology presented in this study is a new approach in estimating viral oncogenicity and can successfully be combined with other mathematical models for model updating in real conditions.

**Keywords:** ANFIS, Oncogenicity, Papilloma virus, Prediction

\*Corresponding author:

Email: soleim\_h@modares.ac.ir

Address: Department of Virology, Tarbiat Modares University, Ale-Ahamad Highway, Tehran, Iran

## استفاده از سیستم جدید هوشمند استنتاج فازی - عصبی تطابقی (ANFIS) برای پیش بینی قدرت سرطان زایی ویروس پاپیلوما انسانی

عذرا کنارکوهی<sup>1</sup>، دکتر حوریه سلیمان جاهی<sup>2</sup>، شهاب فلاحی<sup>1</sup>، دکتر حسین ریاحی مدوار<sup>3</sup>، زهرا مشکات<sup>4</sup>

1- دانشجوی کارشناسی ارشد ویروس شناسی پزشکی، گروه ویروس شناسی پزشکی، دانشگاه تربیت مدرس تهران، ایران

2- دانشیار، دکترای ویروس شناسی پزشکی، گروه ویروس شناسی پزشکی، دانشگاه تربیت مدرس تهران، ایران

3- دکترای سازه های آبی، گروه سازه های آبی، دانشگاه شهید باهنر کرمان، کرمان، ایران

4- دانشجوی دکترای ویروس شناسی پزشکی، گروه ویروس شناسی پزشکی، دانشگاه تربیت مدرس، تهران، ایران

تاریخ دریافت 88/10/1، تاریخ پذیرش 89/1/25

### چکیده

**زمینه و هدف:** ویروس های پاپیلوما بر اساس قدرت و پیش آگهی سرطان های ایجاد شده توسط آنها با استفاده از پروتئین های ویروسی E6 و E7 به سه گروه با خطر بالا، پایین و متوسط تقسیم بندی می گردند. امروزه از روش های مختلف مدل سازی در پزشکی بالینی، در تشخیص بیماری ها و بررسی ویژگی های مولکولی آنها استفاده می شود. در میان روش های نوین مدل سازی، سیستم های فازی از جایگاه ویژه ای در زمینه های مختلف علوم برخوردارند. هدف از این مطالعه به کارگیری یک مدل هوشمند ریاضی برای پیش بینی قدرت سرطان زایی ویروس پاپیلوما انسانی بر اساس تعدادی از ویژگی های بیوشیمیایی پروتئین E7 می باشد.

**مواد و روش ها:** در این تحقیق با استفاده از سیستم استنتاج فازی - عصبی تطابقی (ANFIS) روش جدیدی جهت تخمین میزان سرطان زایی ویروس های پاپیلوما جدا شده از بیماران ارائه شده است. فرآیند توسعه و ارزیابی مدل با استفاده از مجموعه داده های واقعی و معیارهای آماری و گرافیکی مختلفی صورت گرفته است. بدین منظور با تهیه داده های بیوشیمیایی و بیوفیزیکی مورد نیاز در مورد ژن E7 از اطلاعات موجود، اقدام به ایجاد مدل مورد نظر شد. در مرحله بعد نتایج حاصل از مدل با داده های واقعی اعتبار یابی شد.

**یافته ها:** طبق نتایج تحقیق، مدل ایجاد شده قادر به پیش بینی موفقیت آمیز سرطان زایی پاپیلوما ویروس ها است. مقادیر RMSE و R2 مربوط به مدل در مرحله آموزش به ترتیب برابر 101/18 و 0/99 و در مرحله ارزیابی 173/8 و 0/94 به دست آمد.

**نتیجه گیری:** طبق نتایج به دست آمده، استفاده از مدل تطبیقی استنتاج فازی - عصبی، دقت تخمین شدت پدیده سرطان زایی ویروس را به میزان قابل توجهی بهبود می بخشد. روش ارائه شده در این تحقیق رهیافت جدیدی در تخمین سرطان زایی محسوب شده و به خوبی قابلیت اتصال و ترکیب با مدل های بالینی و نیز بهنگام سازی با توجه به شرایط واقعی را دارد.

**واژگان کلیدی:** پاپیلوما ویروس، ANFIS، پیش بینی، سرطان زایی

\* نویسنده مسئول: تهران، بزرگراه جلال آل احمد، دانشگاه تربیت مدرس، گروه ویروس شناسی پزشکی

Email: soleim\_h@modares.ac.ir

www.SID.ir

## مقدمه

فازی، غیرخطی بودن و قابلیت تطبیق پذیری این سیستم‌ها و دقت بهتر آنها در مقایسه با سایر روش‌ها در شرایط محدودیت داده‌ها، مهم‌ترین ویژگی‌های این سیستم‌ها است (14). یک سیستم فازی، سیستمی منطبق بر قواعد منطقی شرط - نتیجه است. نقطه شروع ساخت یک سیستم فازی به دست آوردن مجموعه‌ای از قواعد اگر - آنگاه فازی با استفاده از دانش فرد خبره یا دانش حوزه مورد نظر است (15)، (16). داشتن روشی که با استفاده از آن بتوان از اطلاعات عددی موجود برای ساخت قواعد منطقی فازی استفاده کرد ابزار مفیدی در این زمینه بشمار می‌رود. یکی دیگر از روش‌های نوین مدل‌سازی، شبکه‌های عصبی مصنوعی است که مهم‌ترین دلیل قدرتمندی آنها قابلیت آموزش پذیری آنها از روی الگوهای آموزشی (ورودی‌ها و خروجی‌های متناسب) با استفاده از الگوریتم‌های مختلف آموزش است که بر این اساس رابطه بین متغیرهای ورودی و خروجی را شناسایی می‌نمایند. طرح اولیه این مدل‌ها براساس الگوی یادگیری سیستم عصبی انسان پایه گذاری شده است (17-19). اصطلاح ANFIS مخفف عبارت "سیستم استنتاج فازی تطبیق‌پذیر مبتنی بر شبکه‌های عصبی مصنوعی" (Adaptive Neuro Fuzzy Inference System) است که به طور گسترده برای بررسی پدیده‌های با معادلات غیر خطی به کار گرفته شده است (20، 21). بنابر این، ترکیب سیستم‌های فازی که بر قواعد منطقی استوار هستند و روش شبکه‌های عصبی مصنوعی که توان استخراج دانش از اطلاعات عددی را دارند، ما را قادر می‌سازد تا بتوانیم در کنار استفاده از دانش بشری از اطلاعات موجود نیز در ساخت مدل استفاده کنیم. روشی که بر این اساس ارائه شده است روش استنتاج فازی عصبی تطابقی است (22، 23). از مدل‌های بر پایه سیستم‌های فازی و شبکه‌های عصبی و نیز مدل‌های ترکیبی این دو با استفاده از ANFIS در مطالعه جنبه‌های مختلف عفونت‌های ویروسی استفاده شده است (20، 24-29). هدف از این مطالعه به کارگیری یک مدل هوشمند ریاضی برای پیش بینی قدرت سرطانی ویروس پاپیلوما انسانی بر اساس ویژگی‌های بیوشیمیایی پروتئین E7 است. پارامترهای

ویروس پاپیلوما انسانی، ویروسی DNA دار با ژنوم حلقوی کوچک می‌باشد که به عنوان عامل ایجاد کننده عفونت‌هایی از قبیل زگیل‌های تناسلی، پاپیلوما تنفسی و فاکتور موثر در ایجاد انواعی از بدخیمی‌ها معرفی شده است (1، 2). عفونت‌ها و سرطان‌های ایجاد شده توسط پاپیلوما ویروس امروزه یک مشکل بهداشتی مهم در کلیه جوامع به شمار می‌روند (3). این ویروس‌ها بر اساس قدرت و پیش آگهی سرطان‌های ایجاد شده توسط آنها به سه گروه با خطر بالا، پایین و متوسط تقسیم‌بندی می‌گردند (3). ژن‌های E7 و E6 ویروس نقش اساسی در سرطان‌زایی آن دارند (4). پروتئین E7 ویروس به بسیاری از پروتئین‌های تنظیم کننده سیکل سلولی مانند p107، pRb و p130 که همگی در یک خانواده قرار دارند متصل می‌شود و باعث رها سازی فاکتور نسخه برداری EF2 شده و در نتیجه چرخه سلولی به سمت فاز S رفته و تکثیر DNA سلولی افزایش می‌یابد (5). مطالعات نشان داده‌اند که پروتئین‌های E7 و E6 ویروس‌های پاپیلوما پر خطر با قدرت بیشتری به پروتئین‌های مهار کننده تومور نسبت به دیگر انواع این ویروس متصل می‌شوند (1، 6). بهترین شواهد در مورد درگیری پروتئین‌های E6 و E7 در بدخیمی‌ها از مطالعات بیوشیمیایی آنها به دست آمده است (7). درک عملکرد یک ژن در محیط بدن یکی از مهم‌ترین اهداف این مطالعات است. E7 کد شده توسط ویروس‌های پر خطر در قدرت ترانسفورمسیون در شرایط آزمایشگاهی و نیز طیفی از ویژگی‌های بیوشیمیایی و بیولوژیکی با انواع ضعیف تر تفاوت دارند (8). یکی از کاربردهای رایج مدل‌های ریاضی از قبیل شبکه‌های عصبی ساختگی و سیستم‌های فازی در پزشکی بالینی، در بررسی ویژگی‌ها و تشخیص بیماری‌ها می‌باشد (9، 10). در میان روش‌های نوین مدل‌سازی، سیستم‌های فازی از جایگاه ویژه‌ای در زمینه‌های مختلف علوم برخوردارند و به صورت روز مره برای بررسی انواع پدیده‌های بیولوژیک و غیر بیولوژیک مورد استفاده قرار می‌گیرند (11-13). توانایی پیاده سازی دانش بشری با استفاده از مفاهیم خاص زبانی و قواعد

هدف اصلی مطالعه حاضر، توسعه یک روش جدید براساس ANFIS برای پیش بینی سرطان‌زایی پاپیلوما ویروس و مقایسه نتایج حاصل از آن با داده‌های واقعی است. چندین پایگاه داده حاوی اطلاعاتی در مورد توالی‌های پاپیلوما ویروس وجود دارد. برای گردآوری اطلاعات مورد استفاده در این پژوهش، از پایگاه‌های (www.expasy.ch/tools)، (www.ddbj.nig.ac.jp/DBJ) Expasy برای توسعه استراتژی مورد نظر استفاده شد. از این پایگاه‌ها 31 ویژگی بیوشیمیایی و بیوفیزیکی برای پروتئین‌های مورد بحث محاسبه و استخراج شد که با توجه به ویژگی‌های مدل مورد استفاده و مطالعات صورت گرفته تعداد 7 ویژگی برای سنجش در مدل مورد استفاده قرار گرفتند. در جدول 1 پارامترهای مورد استفاده نشان داده شده است.

ورودی مدل در این تحقیق شامل میزان PH ایزوالکتریک، مقدار درصد حضور هر کدام از آمینواسیدهای سیستئین، لیزین، تیروزین، تریپتوفان، گلیسن و مقدار سولفور موجود در ساختار این پروتئین می باشد.

### مواد و روش‌ها

ابتدا DNA از نمونه‌های پارافینه سرطان رحم جداسازی و برای حضور ویروس پاپیلوما مورد غربالگری قرار گرفت. برای انجام این مرحله، در ابتدا پرایمرهای اختصاصی ویروس طراحی شدند و بعد از انجام PCR، محصولات به دست آمده برای توالی یابی ارسال شدند. سپس توالی‌های به دست آمده مورد آنالیز فیلوژنتیک قرار گرفتند که نتایج آن تحت شماره DQ323401 در پایگاه ژن بانک به ثبت رسید (1).

جدول 1. پارامترهای آماری داده‌های جمع‌آوری شده و ضریب تاثیر معادله ی رگرسیون خطی چندگانه

ضریب تاثیر رگرسیون چندگانه خطی	Statistics									
	مقدار بیشینه	کم مقدار	Kurtosis	(Skewness) چولگی	انحراف از واریانس	مد	میانگین	میانگین	میانگین	پارامتر
-0/2288	9/29	0/9240	-1/62735	-0/38756	5/393062	2/322297	9/1	8/83	7/030527	pI
0/829662	10/4	6/1	0/257657	1/0839	0/991671	0/995827	7/1	7/1	7/470916	Cys (C)
0/821101	8/4	.	-1/0207-	-0/51855	5/979654	2/445333	1/9	5/4	4/728685	Lys (K)
0/459704	3/5	0	0/12705	0/662265	0/581926	0/762841	0	0/9	0/760159	Trp (W)
0/059088	8/4	0	-0/80344	-0/22741	5/349415	2/31288	4	4/3	4/357371	Tyr (Y)
0/57303	6/1	1/3	0/042264	1/044484	1/447916	1/203294	1/9	2/7	3/066534	Gly (G)
0/56577	18	7	0/629666	0/9532	6/454056	2/540483	12	11	11/61355	Sulfur
-28/3179	Constan	t								

در مدل مورد نظر، سه سطح سرطان‌زایی کم، متوسط و زیاد به عنوان استاندارد در نظر گرفته شد که به ترتیب با اعداد 1 و 2 و 3 نمایش داده شدند. از مجموعه داده‌های موجود 71 درصد (180 مورد) آنها برای آموزش مدل و 29 درصد (73 مورد) برای ارزیابی آن استفاده شده است. قابل ذکر است که تقسیم‌بندی داده‌ها به دو گروه آموزش و ارزیابی به

جهت طراحی مدل با جمع‌آوری مجموعه‌ای وسیع از داده‌های موجود در پایگاه‌های داده مولکولی که در آنها مقادیر تمامی پارامترهای مورد نیاز محاسبه شده بود، اقدام به توسعه مدل استنتاج فازی - عصبی تطابقی و ارزیابی دقت آن شد. در جدول 1 پارامترهای استفاده شده در توسعه مدل فازی - عصبی و محدوده تغییرات آنها نشان داده شده است.

برای ارزیابی توانمندی مدل‌های مورد بررسی باید از معیارهای دیگری استفاده گردد، بهمین دلیل دو روش شاخص میانگین قدرمطلق خطای نسبی (AARE) و تحلیل آستانه (TS) نیز مورد استفاده قرار گرفت (30). این دو معیار نه تنها شاخص عملکرد را به صورت ترم‌هائی از مقادیر پیش‌بینی شده ارائه می‌دهند بلکه توزیع خطا را نیز نشان می‌دهند. شاخص  $TS_x$  برای مقدار  $x\%$  از پیش‌بینی‌ها، نشان دهنده توزیع خطا در مقادیر پیش‌بینی شده برای هر مدل می‌باشد. این شاخص که برحسب درصد تعریف می‌شود برای مقادیر مختلف قدرمطلق خطای نسبی ارائه می‌شود.

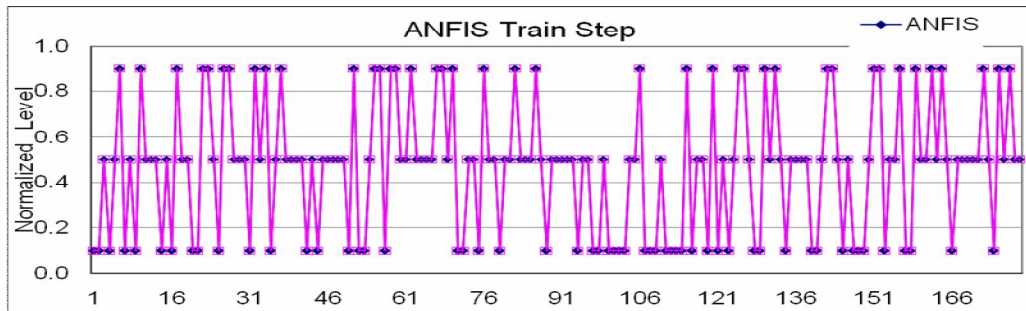
#### یافته ها

#### نتایج مدل ANFIS در تعیین سرطانزایی و وروس

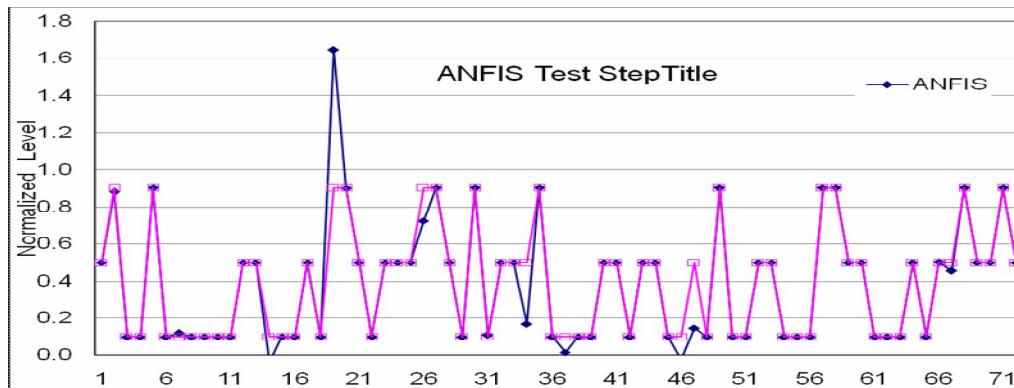
با استفاده از مجموعه داده‌های گردآوری شده اقدام به توسعه مدل جدیدی برای تخمین سرطانزایی مبتنی بر قابلیت‌های سیستم‌های استنتاج فازی - عصبی تطابقی گردید. در نمودارهای 1 و 2 نتایج مدل فازی - عصبی تطابقی در مراحل آموزش و ارزیابی آمده است. همچنین مشخصات آماری مدل در این مراحل در جدول 2 آمده است. متغیرهای ورودی در مدل تطابقی فازی - عصبی شامل میزان PH ایزوالکتریک، مقدار درصد حضور هرکدام از آمینواسیدهای تعیین کننده، سیستمین، لیزین، تیروزین، تریپتوفان، گلیسن و مقدار سولفور موجود در ساختار این دو پروتئین می‌باشد. مشاهده نمودارهای 3 تا 6 نشان می‌دهد که مدل به خوبی و با دقت مطلوب توانسته است الگوی حاکم بر سرطانزایی را درک نموده و آن را در هر دو مرحله آموزش و ارزیابی برآورد نماید. مقایسه نتایج آماری مدل تطابقی فازی - عصبی جدول 3 با نتایج حاصل از اطلاعات قبلی نشان می‌دهد که مدل فازی - عصبی دقت بسیار خوبی داشته و موفقیت قابل ملاحظه‌ای در این زمینه داشته است.

صورت تصادفی انجام شده است. معماری بهینه هر مدل با استفاده از پیش فرض‌های جعبه ابزار نرم‌افزار مطلب (MATLAB 7.0.4) و آزمون و خطا بر ساختار مدل و نظر فرد خبره تعیین گردیده است. با آزمون ترکیبات مختلفی از معماری مدل، حالت بهینه با توجه به ضریب همبستگی و مقادیر خطای حاصله به دست آمد. برای استفاده از سیستم فازی - عصبی، دو روش گسسته سازی شبکه‌ای (Grid Partitioning) و خوشه‌بندی جزئی (Sub-Clustering) وجود دارد. تفاوت عمده این دو روش در چگونگی تعیین تابع عضویت فازی است. در روش جداسازی شبکه‌ای نوع و تعداد تابع عضویت بردار اطلاعات ورودی توسط کاربر تعیین می‌شود و در روش خوشه‌بندی جزئی نوع تابع عضویت با توجه به خصوصیات بردار اطلاعات ورودی و دسته‌بندی‌های موجود در آنها توسط خود مدل استنتاج فازی - عصبی تطابقی تعیین می‌گردد. در این تحقیق از روش گسسته‌سازی شبکه‌ای و از تابع عضویت (Gaussmf Gaussian Membership Function) با هفت متغیر ورودی و سه تابع عضویت برای هر کدام و 30 تکرار، استفاده شده است که منجر به تولید خروجی مورد نظر می‌شود که در اینجا میزان سرطانزایی و وروس می‌باشد. ترکیبات مختلفی از تعداد توابع و نوع آنها در توسعه مدل استفاده و در نهایت مدل بهینه انتخاب شد. برای آموزش مدل از روش هیبرید استفاده شد که ترکیبی از روش‌های شبکه عصبی و فازی است. جزئیات و چگونگی توسعه سیستم فازی - عصبی تطابقی را در منابع 12 و 22 می‌توان مشاهده نمود.

نتایج روابط تجربی و مدل فازی - عصبی با استفاده از معیارهای ضریب همبستگی  $R^2$ ، میانگین خطای مطلق (MAE)، ریشه میانگین خطای مطلق (RMSE) و میانگین مربعات خطا (MSE) ارزیابی شده است. این شاخص‌ها مبین متوسطی از خطا در عملکرد مدل و آماره‌های جهانی (Global Statistics) هستند که اطلاعاتی در مورد توزیع خطا در نتایج را ارائه نمی‌کنند و



نمودار 1. عملکرد مدل فازی - عصبی در مرحله آموزش مدل



نمودار 2. عملکرد مدل فازی - عصبی در مرحله ارزیابی مدل

**نتیجه آزمایش مدل با داده‌های واقعی**

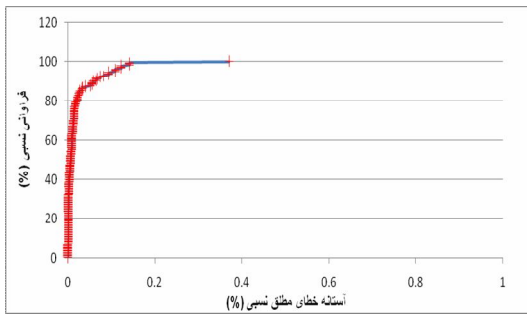
جدول 3. مقایسه مشخصات آماری مدل تطبیقی فازی - عصبی 1 و 2 در مراحل آموزش و ارزیابی

پارامتر آماری					مرحله
توسعه مدل					
AARE (%)	MAE	MSE	RMS E	R <sup>2</sup>	
60/43	9/50	185/43	101/18	0/99	مرحله آموزش مدل
140/68	104/77	286/14	173/8	0/94	مرحله ارزیابی مدل
78/13	10/26	323/28	107/98	0/99	مرحله آموزش مدل (2)
192/28	112/98	397/49	199/46	0/71	مرحله ارزیابی مدل (2)

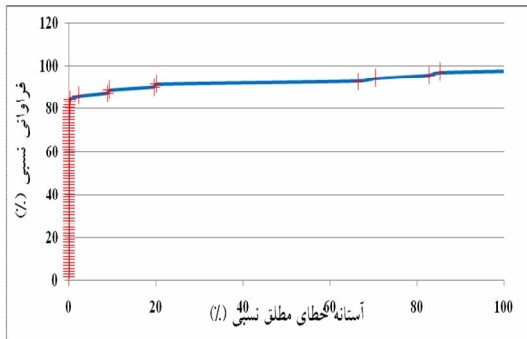
از نمودارهای 1 و 2 مشخص می‌شود که عملکرد مدل در هر دو مرحله آموزش و ارزیابی نزدیک به صد درصد است. همچنین در مرحله ارزیابی نیز علاوه بر این که مطابق داده‌های جدول 2 و 3، دقت مدل ANFIS قابل قبول بوده است، در نمودار 6 نیز مشاهده می‌شود که در اکثر موارد مدل ANFIS در مرحله ارزیابی نیز با شرایط واقعی قابل رقابت است.

جدول 2. مشخصات آماری مدل تطبیقی فازی - عصبی شماره 1 در مراحل آموزش و ارزیابی

پارامتر آماری					مرحله
توسعه مدل					
AARE (%)	MAE	MSE	RMS E	R <sup>2</sup>	
60/43	9/50	185/43	101/18	0/99	مرحله آموزش مدل
140/68	104/77	286/14	173/8	0/94	مرحله ارزیابی مدل



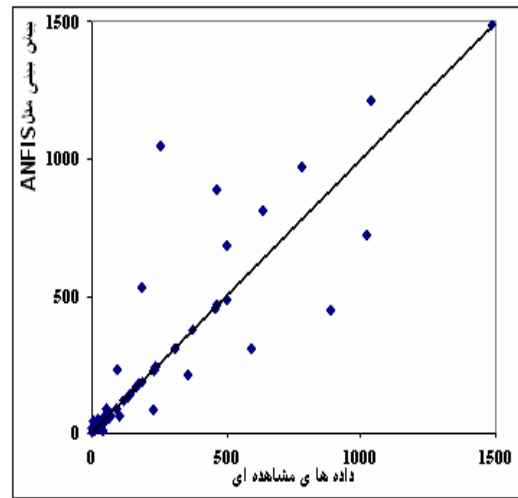
نمودار 5. توزیع خطای پیش‌بینی مدل ANFIS در مراحل آموزش و ارزیابی



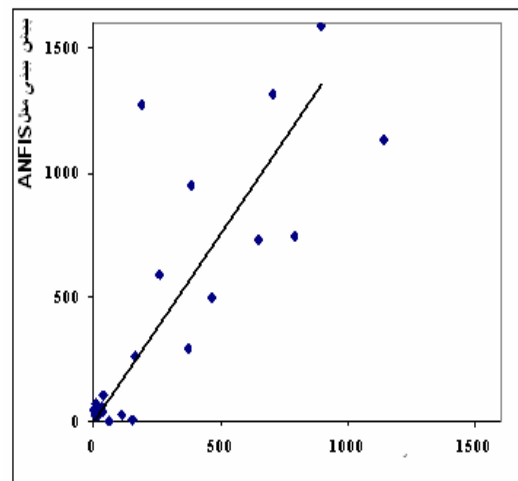
نمودار 6. توزیع خطای پیش‌بینی مدل ANFIS در مراحل آموزش و ارزیابی

همان طور که در بخش قبل بیان شد ساختار مدل ANFIS نهائی مورد استفاده در این تحقیق دارای هفت متغیر ورودی است که هر کدام سه تابع عضویت گوسین دارند. بنابراین تعداد قوانین فازی و نیز تعداد پارامترهای خطی برابر  $3^7=2401$  و تعداد پارامترهای غیرخطی برابر با تعداد پارامترهای توابع گوسین مورد استفاده است. از آنجا که هفت ورودی و هر ورودی سه تابع عضویت گوسین و هر تابع گوسین دو پارامتر دارد، پارامترهای غیرخطی برابر  $42=7 \times 3 \times 2$  خواهد بود. بنابراین تعداد کل پارامترهای مدل ANFIS برابر 2443 خواهد بود و سری داده‌های مورد استفاده نیز برابر 251 است.

در مرحله اعتبار یابی، با وارد کردن اطلاعات واقعی مدل به طور موفقیت آمیز قادر به تعیین سرطان‌زایی ویروس مورد نظر بود. برای این منظور ژن E7 از یک بیمار مبتلا به سرطان دهانه رحم توسط PCR (polymerase chain reaction) جداسازی، توالی یابی و کلون شد. سپس ویژگی‌های مورد نظر برای این توالی محاسبه و به مدل



نمودار 3. مقایسه نتایج مدل ANFIS با داده‌های واقعی در مرحله آموزش مدل



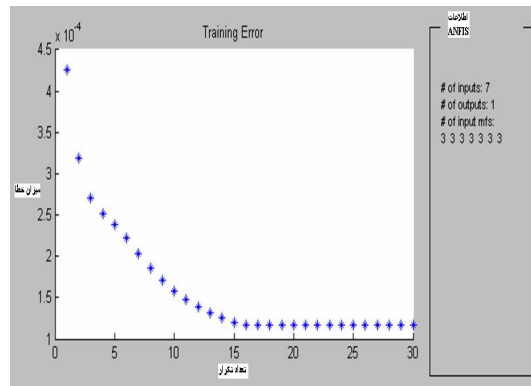
نمودار 4. مقایسه نتایج مدل ANFIS با داده‌های واقعی در مرحله ارزیابی مدل

دقت قابل قبول مدل در پیش‌بینی علی رغم محدود بودن تعداد داده‌های مورد استفاده و محدوده تغییرات بسیار وسیع آنها، سادگی آموزش و توسعه آن، به خوبی قابلیت بالای آن را در فراگیری و پیش‌بینی فرآیند پیچیده و ناشناخته حاکم بر سرطان‌زایی را نشان می‌دهد. نتایج این مدل امکان جایگزینی آن برای تخمین قدرت سرطان‌زایی به جای روش‌های آزمایشگاهی پیچیده، وقت گیر و پرهزینه، در مطالعات غربالگری را نشان می‌دهد.

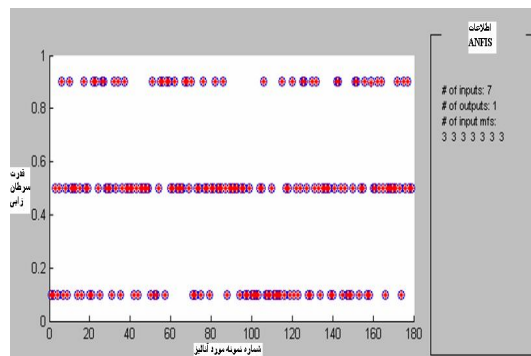
هنگام موجب کاهش میزان بروز و نیز مرگ و میر ناشی از سرطان سرویکس در جوامع بشری به ویژه جوامعی که دارای برنامه غربالگری فعال می‌باشند گشته است (32، 33). با این وجود هیچ یک از این روش‌ها اطلاعاتی در مورد آینده بالینی افراد آلوده به دست نمی‌دهد. از آنجایی که پروتئین‌های E6 و E7 به طور پیوسته در بدخیمی‌های سرویکس وابسته به پاپیلوما تولید می‌شود (34، 35)، ردیابی اثر نهایی آنها، در غربالگری و کنترل عفونت موثر خواهد بود. این پروتئین‌ها با اتصال و غیر فعال سازی پروتئین‌های مهار کننده تومور قادرند موجب ترانسفورمسیون سلول یا نئوپلازی شوند. خطر سرطان‌زایی انواع مختلف پاپیلوما ویروس‌های انسانی وابسته به شدت تمایل پروتئین‌های فوق جهت اتصال به p53 و Rb و خانواده‌های وابسته می‌باشد (34، 35). از آنجا که این تمایل اتصال به پروتئین‌های مهار کننده تومور می‌تواند به ویژگی‌های بیوشیمیایی و بیوفیزیکی آنها ارتباط داشته باشد، لذا بر آن شدیم تا با مطالعه ساختار انکو پروتئین E7 و مدل‌های عددی به پیش بینی انکوژنیسیته ویروس‌های پاپیلوما بپردازیم. همان‌طور که در نتایج مشاهده شد، مدل فازی عصبی آموزش دیده، در ارزیابی نهایی مدل، وجود ارتباط بین ساختار بیوشیمیایی با انکوژن بودن را نشان می‌دهد. محققین این پروژه با این پیش فرض که برخی از اسیدهای آمینه و خصوصیات بیوشیمیایی مانند حضور اسید آمینه سیستئین و سولفور که در ساختار موتیف‌های اتصال به روی نقش دارند (36، 37)، در شکل‌گیری انکوژنیسیته مهم هستند را مورد آزمایش قرار دادند. مطالعات پیشین نشان داده که این موتیف‌ها در اتصال انکو پروتئین‌ها به پروتئین‌های مهار کننده تومور نقش اساسی دارند (38). علاوه بر این پارامترها حضور و تعداد اسید آمینه‌های مهم دیگر مانند لیزین، گلیسین، تیروزین و تریپتوفان و همچنین pH ایزو الکتريک هم مد نظر قرار گرفت.

علیرغم این که تعداد سری داده‌های مورد استفاده کم است ولی تاثیری بر نتایج نهایی مدل نخواهد داشت. چرا که در این شرایط که سری داده‌ها کمتر از مجموعه

وارد شد که توسط مدل، در رده با سرطان‌زایی بالا قرار گرفت. این نتیجه با شرایط بالینی بیمار و ژنوتیپ ویروس که در آنالیز فیلوژنتیک از نوع 16 بود، همخوانی داشت. میزان خطای مدل با افزایش تعداد تکرار و میزان صحت نهایی مدل در نمودارهای 7 و 8 آورده شده است.



نمودار 7. کاهش میزان خطای مدل با افزایش تعداد تکرار



نمودار 8. مرحله ی تست مدل بر اساس تعداد نمونه مورد استفاده در این مرحله

## بحث

رایج‌ترین روش‌ها برای بررسی بیماری‌های شایع استفاده از مطالعات مورد شاهدهی، در برخی موارد همگروهی و یا حتی مقطعی است (31). روش‌های غربالگری معمول برای پاپیلوما ویروس و شناسایی روند انکوژنز با انجام روش‌های پاپ اسمیر یا شناسایی حضور ژن‌های مربوط به ویروس با روش‌های PCR یا هیبریداسیون انجام می‌گیرد (32). همه روش‌های ذکر شده به علت تشخیص زود



مطلوبی در برون‌یابی ندارند. سیر نزولی میزان خطای مدل با افزایش تعداد تکرارها از یک تا 30 در نمودار 7 نشان دهنده قدرت یادگیری مناسب مدل با افزایش تعداد تکرار و ارتباط مناسب بین داده‌های انتخاب شده به عنوان ورودی و سرطان‌زایی و ویروس می‌باشد. میانگین خطای مدل در مرحله آموزش (نمودار 7 و 8) به مقدار 0/00011717 است، که عددی بسیار کوچک و قابل چشم پوشی است. با توجه به نمودار 8 میزان صحت نهایی مدل در مرحله تست حدود 90 درصد و میانگین خطا در این نمودار 0/10966 است. تاکنون مورد گزارش شده مشابهی از این مطالعه برای ویروس پاپیلوما انجام نشده است، اما در مطالعات مختلفی برای طراحی دارو، پیش بینی مقاومت دارویی، پیش بینی سیر بیماری و دیگر جنبه‌های عفونت و ویروس‌های مختلف به ویژه HIV و HCV از این روش‌ها استفاده شده است.

### نتیجه گیری

به طور خلاصه به نظر می‌رسد که با ترکیب روش‌های ریاضی و بیوانفورماتیکی با مطالعات میدانی، می‌توان سیستم هوشمندی را طراحی کرد تا بتواند ویژگی‌های مولکولی انکوپروتئین ناشناخته در رابطه با سرطان‌های مختلف انسانی از جمله سرطان گردن رحم را شناسایی نماید. با استفاده از مدل حاضر قادر به تخمین شدت احتمالی سرطان ایجاد شده توسط ویروس موجود در فرد آلوده خواهیم بود، که این امر دارای اهمیت فراوانی در تعیین اقدامات درمانی و نگه دارنده بعدی برای بیمار می‌باشد.

### سپاسگزاری

بدین وسیله از کمیته تحقیقات دانشکده پزشکی دانشگاه تربیت مدرس که ما را در انجام این پروژه یاری دادند قدردانی می‌کنیم. لازم به ذکر است که قسمتی از اطلاعات این مقاله مستخرج از پایان نامه رشته ویروس شناسی پزشکی دانشگاه تربیت مدرس می‌باشد.

پارامترهاست، مدل ANFIS با توجه الگوریتم آموزش کمترین مربعات خطا و شیب نزولی خطا، تنها پارامترهای موثرتر را اصلاح می‌نماید و بقیه پارامترها را همان مقادیر پیش فرض که براساس اطلاعات ورودی تعیین شده‌اند، فرض نموده و تغییر نمی‌دهد. در این زمینه بایلاز و همکاران (2008) با استفاده از 72 سری داده راندمان هوادهی سرریزها را توسط مدل ANFIS شبیه سازی نموده‌اند (39). در مدل توسعه یافته توسط ایشان تعداد پارامترهای خطی 108 مورد و تعداد پارامترهای غیرخطی 27 مورد بوده است که مطابق نظر این محقق محدود بودن تعداد داده‌ها در مقایسه با تعداد کل پارامترها تأثیری بر نتیجه نهایی نخواهد داشت. علاوه بر این به منظور بررسی اثر تعداد داده‌های مورد استفاده در مرحله آموزش مدل بر نتایج مدل در مرحله ارزیابی، با استفاده از نصف داده‌های مورد استفاده در مرحله آموزش، یک بار دیگر مدل ANFIS آموزش داده شد (مدل 2) و نتایج آن در مراحل آموزش و ارزیابی با نتایج مدل قبلی که با استفاده از کل داده‌ها آموزش داده شده بود (مدل 1) در جدول 3 مقایسه شد. در انتخاب داده‌های مورد استفاده برای آموزش مدل دوم، داده‌ها به صورتی انتخاب شده است که کل محدوده پارامترهای موجود در پدیده را دربر داشته باشند. شایان ذکر است که ساختار مدل ANFIS مورد استفاده در این مرحله همان ساختار مدل بهینه مرحله قبل بوده است. مطابق نتایج ارائه شده در جدول 3 کاهش داده‌های مورد استفاده در مرحله آموزش مدل تأثیر چندانی بر نتایج نهایی مدل نداشته است و این امر به علت قابلیت ویژه مدل‌های ANFIS در استخراج دانش از اطلاعات عددی است. در این زمینه تنها و مهم‌ترین نکته استفاده از داده‌هایی است که در مرحله آموزش کل محدوده پدیده مورد بررسی را در برداشته باشند. این نکته یک بار دیگر قدرت ویژه مدل ANFIS در مقایسه با دیگر مدل‌های مرسوم خصوصاً شبکه‌های عصبی مصنوعی را نشان می‌دهد که حساسیت خاصی به تعداد داده‌های موجود ندارد بلکه داده‌های موجود تنها باید محدوده مورد بررسی را پوشش دهند، چرا که مدل‌های هوشمند قابلیت چندان

## منابع

1. Meshkat Z, Mahmoudi M, Mirshahabi H, Zuhair M, Hassan, Saeed R, Sabokbar T. Determination of human papillomavirus type 16 genotype and construction of cloning vector ptz57r encoding hpv16 e7 gene. Saudi Med J. 2007; 28(10):1511-5.
2. H. Soleimanjahi, Nategh MJ, Falahi S. A performance appraisal of neural networks developed for response prediction across heterogeneous domains. WCSET 2009: World Congress on Science, Engineering and Technology. 2009; rome - Italy.
3. Roden RB, Yutzy WH, Fallon R, Inglis S, Lowy DR, Schiller JT. Minor capsid protein of human genital papillomaviruses contains subdominant, cross-neutralizing epitopes. Virology. 2000; 270(2): 254-7.
4. Oishee SK. Molecular interactions of 'high risk' human papillomaviruses e6 and e7 oncoproteins: Implications for tumour progression. J Biosci. 2003;28(3):337-48
5. Hausen H. Papillomaviruses causing cancer: Evasion from host-cell control in early events in carcinogenesis. Journal of the National Cancer Institute. 2000;92(9): 690-8.
6. Collins AS, Nakahara T, Do A, Lambert PF. Interactions with pocket proteins contribute to the role of human papillomavirus type 16 e7 in the papillomavirus life cycle. J Virol. 2005; 79(23): 14769-80.
7. Teimoori HS, Fotouhi F, Meshkat Z. Isolation and cloning of human papillomavirus 16 L1 gene from Iranian isolate. Saudi Med J. 2008; 29: 1105-8.
8. Erica W. Wu KEC, Donald V, Karl M. The human papillomavirus e7 oncoprotein and the cellular transcription factor e2f bind to separate sites on the retinoblastoma tumor suppressor protein. J of Virol. 1993;67(4):2402-7.
9. Narayanan A, Wu X, Yang ZR. Mining viral protease data to extract cleavage knowledge. Bioinformatics. 2002; 18 Suppl 1:5-13.
10. Wei HY, Lu CS, Lin TH. Exploring the p2 and p3 ligand binding features for hepatitis c virus ns3 protease using some 3d qsar techniques. J Mol Graph Model. 2008; 26(7): 1131-44.
11. Roy Chowdhury S, Saha H. Development of a fpga based fuzzy neural network system for early diagnosis of critical health condition of a patient. Comput Biol Med. 2010; 40(2): 190-200.
12. Riahi-Madvar H, Ayoob zadeh SA. Estimating longitudinal dispersion coefficient of pollutants using adaptive neuro-fuzzy inference system. water and waste water. 2008;67:34-46.
13. Tao CW, Taur J, Chang JH, Su SF. Adaptive fuzzy switched swing-up and sliding control for the double-pendulum-and-cart system. IEEE Trans Syst Man Cybern B Cybern. 2010; 40(1): 241-52.
14. Dezfuli KA. Principal of fuzzy theory and application in whater engineering modeling. 1 ed. Tehran: Jahad Daneshgahi vahede Amirkabir. 2005.
15. Zadeh LA. Fuzzy sets. Information Control. 1965; 8(3): 338-53.
16. Jamshidi M. Tools for intelligent control: Fuzzy controllers, neural networks and genetic algorithms. Philos Transact A Math Phys Eng Sci. 2003; 361(1809): 1781-808.
17. Perez-Gandia C, Facchinetti A, Sparacino G, Cobelli C, Gomez EJ, Rigla M, et al. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. Diabetes Technol Ther. 2010; 12(1): 81-8.
18. Jung E, Kim J, Choi SH, Kim M, Rhee H, Shin JM, et al. Artificial neural network study on organ-targeting peptides. J Comput Aided Mol Des. 2010;24(1):49-56.
19. Committee AT. Artificial neural network in hydrology ii. J of Hydrol Engin. 2000; 5(2): 124-44.
20. Goodarzi M, Freitas MP. Mia-qsar coupled to principal component analysis-adaptive neuro-fuzzy inference systems (pca-anfis) for the modeling of the anti-hiv reverse transcriptase activities of tibo derivatives. Eur J Med Chem. 2010; 10(2): 132-8.
21. Buyukbingol E, Sisman A, Akyildiz M, Alparslan FN, Adejare A. Adaptive neuro-fuzzy inference system (anfis): A new approach to predictive modeling in qsar applications: A study of neuro-fuzzy modeling of pcp-based nmda receptor antagonists. Bioorg Med Chem. 2007; 15(12): 4265-82.

22. Riahi-Madvar H, Ayob zadeh SA, Khadangi E, Ebadzadeh MM. An expert system for predicting longitudinal dispersion coefficient in natural streams by using anfis. *Expert Systems with Applications*. 2009; 36(4): 8589-96.
23. Polat K, Yosunkaya S, Gunes S. Pairwise anfis approach to determining the disorder degree of obstructive sleep apnea syndrome. *J Med Syst*. 2008; 32(5): 379-8.
24. Woolfson AD, Umrethia ML, Kett VL, Malcolm RK. Freeze-dried, mucoadhesive system for vaginal delivery of the hiv microbicide, dapivirine: Optimisation by an artificial neural network. *Int J Pharm*. 2010; 388(1-2): 136-43.
25. Huang ML, Chen HY, Huang WC, Tsai YY. Linear discriminant analysis and artificial neural network for glaucoma diagnosis using scanning laser polarimetry-variable cornea compensation measurements in taiwan chinese population. *Graefes Arch Clin Exp Ophthalmol*. 2010; 248(3): 435-41.
26. Xiaodong L, Hao Y, Feng L, MacArthur RDC, Barth J, Hong Y, Crane LR. A fuzzy discrete event system for hiv/aids treatment. *The 14th IEEE International Conference on Fuzzy Systems*. 2005; 167-72.
27. Andonie RF-A, Collar L, Abdul-Wahid CJ, Salim S. Neuro-fuzzy prediction of biological activity and rule extraction for hiv-1 protease inhibitors. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2005; 2008; Sun Valley, Idaho: Institute of Electrical and Electronics Engineers
28. Jaime De La Torre FL, Martínez J, Padilla A, Mora M. SERS and ANFIS: Fast identification of the presence of retrovirus in cd4 cells, cause of aids: Springer Berlin / Heidelberg; 2008.
29. Jalali-Heravi AKM. Comparison of shuffling-adaptive neuro fuzzy inference system (shuffling-anfis) with conventional anfis as feature selection methods for nonlinear systems. *QSAR Comb Sci*. 2007;00:1-14.
30. Baylar H, Hanbay D, Ozpolat E. An expert system for predicting aeration performance of weirs by using anfis. *Expert Systems with Applications*. 2007;23(8):1214-22
31. Eshrati B ZS, Gooya MM, Soroush M, Masoomi Asl H, Afshani A, Ramezani M, Azimi MS. Use of meta-anaglysis to determining the associated factors of an outbreak occurred in the summer 2005 in iran. *Journal of Arak University of Medical Sciences*. 2008;11(3):99-108.
32. Li J, Lee JY, Yeung ES. Quantitative screening of single copies of human papilloma viral DNA without amplification. *Anal Chem*. 2006; 78(18): 6490-6.
33. Moscicki AB. Human papilloma virus, papanicolaou smears, and the college female. *Pediatric Clinics of North America* 2005; 52(1): 163-77.
34. Narisawa-Saito M, Kiyono T. mechanisms of high-risk human papillomavirus-induced cervical carcinogenesis. *Nippon Rinsho*. 2009;67(1):53-61.
35. Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: Roles of e6 and e7 proteins. *Cancer Sci*. 2007; 98(10): 1505-11.
36. Margaret E. McLaughlin-Drubin K-WH, and Karl Mu'nger. Human papillomavirus type 16 e7 oncoprotein associates with e2f6. *J of Virol*. 2008; 82(17):8695-705.
37. Eugene Klimov SV, Moisjak E, n Rakhmanaliev E, Kobseva V, Laimins L, Kissel'jov F. Galina Sulimova. Human papilloma viruses and cervical tumours: Mapping of integration sites and analysis of adjacent cellular sequences. *BMC cancer*. 2002; 2(24):183-92.
38. Longworth MS, Laimins LA. The binding of histone deacetylases and the integrity of zinc finger-like motifs of the e7 protein are essential for the life cycle of human papillomavirus type 31. *J Virol* 2004; 78(7): 3533-41.
39. Baylar H, Hanbay D., Ozpolat, E. An expert system for predicting aeration performance of weirs by using anfis. *Expert Systems with Applications*. 2007; 23(8):352-64.