

مقایسه شبکه‌های عصبی نوع GMDH چند هدفه و شبکه خودباوری بیزین در پیش‌بینی کدورت آب تصفیه شده (مطالعه موردی: تصفیه‌خانه بزرگ آب گیلان)

الهیارد داغبندان^۱، فرشته علی طالشی^۲، مهران یعقوبی^۳

۱- استادیار، گروه مهندسی شیمی، دانشکده فنی و مهندسی، دانشگاه گیلان، رشت (نویسنده مسئول) ۳۳۸۲۴۳۹۲ (۰۱۳) daghbandan@guilan.ac.ir

۲- دانشجوی کارشناسی ارشد مهندسی شیمی، دانشگاه گیلان، رشت

۳- مسئول واحد بهره‌برداری تصفیه‌خانه بزرگ آب گیلان، رشت

(دریافت ۹۳/۳/۱ پذیرش ۹۳/۷/۲۹)

چکیده

در این پژوهش ابتدا متغیرهای تأثیرگذار بر فرایند حذف کدورت آب، با استفاده از روش شناسی سطح پاسخ شناسایی شد. در ادامه شبکه‌های عصبی نوع GMDH و شبکه خودباوری بیزین برای مدل‌سازی و پیش‌بینی کدورت آب تصفیه شده، با استفاده از مجموعه داده‌های ورودی- خروجی مورد مطالعه قرار گرفت. برای ارزیابی مدل پیشنهادی، تصفیه‌خانه بزرگ آب گیلان به صورت موردی بررسی و داده‌های مورد نیاز شامل ۷۰۰ سری داده به دست آمد. به منظور مدل‌سازی، داده‌های برگرفته از واحد بهره‌برداری به دو دسته (۷۰ درصد برای آموزش و ۳۰ درصد برای آزمایش) تقسیم شدند. نتایج حاصل از مدل‌سازی با داده‌های تجربی مقایسه شد که ضریب تعیین مقادیر آزمایشی برای دو الگوریتم شبکه خودباوری بیزین شامل EM و GD و برای مدل GMDH به ترتیب ۰/۹۳۸۸، ۰/۹۱۹۶ و ۰/۹۷۰۹۵ بود. بر اساس این نتایج مدل GMDH نسبت به مدل BBN کارایی بهتری برای پیش‌بینی میزان کدورت آب تصفیه شده دارد.

واژه‌های کلیدی: تصفیه آب، کدورت، مدل‌سازی، شبکه عصبی نوع GMDH، شبکه خودباوری بیزین

۱- مقدمه

پیچیده‌ای باشد [۱ و ۲]. در صنعت، مواد منعقدکننده برای کاهش کدورت آب به کار گرفته می‌شوند. مقدار مصرف این مواد به شرایط محیطی مثل دما، pH، میزان و نوع کدورت آب بستگی دارد. فرایند فیلتراسیون هم تأثیر قابل توجهی در حذف ذراتی دارند که موجب کدورت می‌شوند [۳].

فرایند تصفیه آب مجموعه‌ای پیچیده از چندین عملیات فیزیکی و شیمیایی پشت سر هم است. این فرایند امروزه در بیشتر تصفیه‌خانه‌ها عموماً از طریق انجام آزمایش‌های مختلف و تجربه شخصی اپراتور صورت می‌گیرد. به دلیل مداخله عوامل مختلف فیزیکی، شیمیایی و بیولوژیکی در عملیات تصفیه، کنترل این عملیات و پیش‌بینی رفتار سیستم بسیار مشکل است [۴]. یکی از مهم‌ترین بخش‌های عملیات تصفیه که پیش‌بینی آن بسیار سخت است، پیش‌بینی میزان کدورت باقیمانده در آب خروجی از تصفیه‌خانه به ازای مصرف بهینه منعقدکننده‌هاست که انجام مدل‌سازی در آن کمک شایانی به بهبود فرایند تصفیه می‌نماید.

در فرایند مدل‌سازی، مهم‌ترین قسمت مربوط به ایجاد یک

مواد جامد معلق و ذرات بسیار ریز کلوئیدی باقیمانده در آب باعث کدورت و رنگ آب می‌شوند. کدورت یک خاصیت فیزیکی نمونه است که باعث می‌شود نور تابیده شده به نمونه، متفرق یا جذب شود ولی عبور نکند. بنابراین برای اندازه‌گیری کدورت، به یک منبع نور و دستگاهی برای اندازه‌گیری نور متفرق شده نیاز است که واحد ^۱JTU معرف نور عبور کرده از نمونه، ^۲NTU معرف تفرق نور تابیده شده و ^۳FTU معرف واحد کدورت فرمازین می‌باشند. امروزه به منظور سنجش کدورت‌های پایین از روش‌های نفلومتری (NTU) و جذب سنجی استفاده می‌شود [۱].

به کاهش مواد معلق آب، زلال‌سازی گفته می‌شود. مواد معلق آب می‌توانند ذرات درشت و سنگینی باشند که به راحتی با ته‌نشینی ساده از آب جدا می‌شوند و یا ذرات بسیار ریزی باشند که طبیعت کلوئیدی داشته و جداسازی آنها از آب مستلزم تمهیدات بسیار

^۱ Jackson Turbidity Unit (JTU)

^۲ Neflometric Turbidity Unit (NTU)

^۳ Formazin Turbidity Unit (FTU)

تایوان استفاده نمودند و مدل ارائه شده به خوبی در مقابل تغییرات ناگهانی در کیفیت آب خام ورودی نیز انعطاف‌پذیر بود [۱۰]. علیائی و همکاران نیز در سال ۲۰۱۰ شاخص‌های مؤثر بر کیفیت آب، شامل BOD و COD را به کمک شبکه عصبی پیش‌بینی کردند که نتایج از کارایی خوب شبکه در شبیه‌سازی کیفیت آب خبر داد [۱۱]. جانتانن و همکاران در سال ۲۰۱۲، کیفیت آب شرب شامل کدورت و آلومینیوم باقیمانده را در تصفیه‌خانه‌ای در فنلاند به دو روش رگرسیون خطی چندگانه و شبکه عصبی پرسپترون چندلایه با متغیرهای ورودی pH، $KMnO_4$ ، سختی، رنگ، هدایت الکتریکی، سیلیکات و کدورت آب خام مدل‌سازی کردند که در نهایت برای هر دو خروجی، مدل غیر خطی (شبکه پرسپترون) نتایج بهتری را نشان داد [۱۲]. در سال ۲۰۱۳ دو محقق چینی کیفیت آب یک استخر پرورش ماهی را به ازای متغیرهای دما، اکسیژن محلول، میزان اشعه خورشید، pH، رطوبت هوا و سرعت باد با استفاده از سه نوع شبکه عصبی موجک^۳، المان^۴ و پایه شعاعی^۵ پیش‌بینی کردند که در نهایت شبکه نوع موجک پیش‌بینی دقیق‌تر، پیشرفته‌تر و با سرعت یادگیری بالاتری نسبت به دو مدل دیگر نتیجه داد [۱۳]. در ادامه این تحقیقات در سال ۲۰۱۳ مدل‌سازی و بهینه‌سازی دز مصرفی پلی‌الکترولیت به ازای دما، pH، هدایت الکتریکی، جامدات معلق، رنگ و کدورت آب خام ورودی توسط داغبندان و همکاران در تصفیه‌خانه بزرگ آب گیلان با استفاده از شبکه‌های عصبی نوع GMDH^۶ و الگوریتم‌های ژنتیک چند هدفه MOGA^۷ انجام شد [۱۴]. تحقیق دیگری نیز در سال ۲۰۱۴ در تصفیه‌خانه مذکور به ازای ورودی‌های مشابه، توسط داغبندان و همکاران بر روی مدل‌سازی دز بهینه منعقدکننده (سولفات آلومینیوم) به کمک شبکه‌های عصبی مصنوعی انجام شد که با توجه به نتایج به دست آمده از مدل‌سازی با هر دو نوع شبکه و مقایسه نتایج با داده‌های آزمایشگاهی، مشخص شد که استفاده از شبکه‌های عصبی نوع GMDH جایگزین بسیار مناسبی برای آزمایش جار است زیرا قابلیت پاسخگویی و تعیین میزان بهینه دز مصرفی در زمان بسیار کم و با دقت بسیار بالا را داراست. همچنین مزیت دیگر استفاده از این مدل، کمک به صرفه‌جویی در زمینه مصرف مواد شیمیایی است [۱۵].

در این پژوهش میزان کدورت باقیمانده در آب شرب شهر رشت - آب خروجی از تصفیه‌خانه بزرگ آب گیلان به عنوان

رابطه غیر خطی بین ورودی‌ها و خروجی است که به دلیل تأثیر عوامل مختلف در میزان کدورت آب خام ورودی به تصفیه‌خانه، پیش‌بینی میزان بهینه دز مصرفی مواد شیمیایی به منظور دستیابی به حد مطلوب کدورت باقیمانده در آب شرب، توسط اپراتور با خطا همراه است. این حد مطلوب در استاندارد جهانی، کمتر از 1^{NTU} و حداکثر مجاز آن 5^{NTU} ذکر شده است [۳]. در مجموع می‌توان گفت دستیابی به حداکثر راندمان کاهش کدورت معمولاً بر مبنای تجربه بهره‌برداران و یا به صورت دقیق‌تر بر اساس نتایج حاصل از آزمایش جار در آزمایشگاه صورت می‌پذیرد.

دانشمندان و محققان برای غلبه بر مشکلات حاصل از آزمایش جار و اپراتور شخصی که ناشی از تغییرات ناگهانی و غیر قابل پیش‌بینی در کیفیت آب ورودی است، انجام مدل‌سازی از طریق روش‌های هوشمند به ویژه شبکه‌های عصبی مصنوعی را بهترین راه برای کاهش هزینه‌ها و همچنین سریع‌تر انجام شدن فرایند و پایین آوردن میزان اشتباه دانسته‌اند [۴]. لذا مطالعات و پژوهش‌های بسیاری در این مورد صورت گرفته است که در زیر به برخی از آنها اشاره می‌شود.

نخستین مدل‌سازی نسبتاً موفق توسط گانگون و همکاران در سال ۱۹۹۷ انجام پذیرفت. ایشان برای نخستین بار یک مدل شبکه عصبی مصنوعی به منظور تعیین میزان بهینه دز مصرفی منعقدکننده برای تصفیه‌خانه استی-فوی^۱ در کبک کانادا طراحی کردند [۵]. کس در سال ۱۹۹۹ با کمک شبکه‌های بی‌زین^۲ کیفیت آب در مصب رودخانه‌ای در کارولینای شمالی را با دقت قابل قبولی پیش‌بینی کرد [۶]. جو نیز در سال ۲۰۰۰ یک مدل شبکه عصبی مصنوعی به منظور تعیین میزان بهینه دز مصرفی منعقدکننده برای تصفیه‌خانه چانگجو در کره جنوبی طراحی نمود [۷]. در سال ۲۰۰۴، پیک مدلی بی‌زینی به منظور مدل‌سازی خطای تعیین کیفیت آب آشامیدنی تصفیه‌خانه‌ای در پنسیلوانیا را با استفاده از مشخصات فرایندی و محیطی طراحی کرد که در نهایت نشان داد، تصمیمات اپراتور، دز منعقدکننده و تناوب شستشوی معکوس فیلتر در تعیین احتمال خطا بسیار مهم هستند [۸]. بناردوس و ووسکینوس در سال ۲۰۰۷ شبکه‌ای عصبی بر مبنای الگوریتم ژنتیک طراحی نمودند که در آن تعداد لایه‌ها و نرون‌های هر لایه برای رسیدن به حداقل مقدار تابع هدف، بهینه شد که این نوع مدل‌سازی علی‌رغم پیچیدگی، عملکرد بسیار خوب و قابل قبولی داشت [۹]. دو محقق تایوانی در سال ۲۰۰۸ از شبکه‌ای عصبی مصنوعی برای تعیین میزان بهینه دز مصرفی منعقدکننده برای آب‌های سطحی شمال

³ Wavelet-type NN

⁴ Elman-type NN

⁵ Radial Basis Function-type NN

⁶ Group Method of Data Handling

⁷ Multi Objective Genetic Algorithms

¹ Stefoy Water Treatment Plant in Quebec
² Bayesian Networks

میزان مصرف مواد شیمیایی بستگی دارد [۵]. داده‌های حاصل از این فرایند، نیل به هدف مطالعه که ارائه یک مدل به‌منظور پیش‌بینی میزان کدورت باقیمانده در آب صاف به ازای مصرف بهینه مواد شیمیایی است را سهولت می‌بخشد.

۲-۲- جمع‌آوری داده‌ها

در واحد بهره‌برداری با اندازه‌گیری برخی شاخص‌های مهم، میزان و نوع مواد مصرفی کنترل می‌شود. بهره‌برداران با توجه به گزارش‌های روزانه واحد آزمایشگاه، شرایط آب خام ورودی و تجربیات شخصی، برای رسیدن به حد مطلوب کیفیت آب خروجی، در هر لحظه نسبت به میزان مصرف مواد منعقدکننده و گندزدا تصمیم‌گیری می‌کنند. در این واحد به‌منظور اندازه‌گیری متغیرهای مورد نظر از دماسنج و pH متر مدل اندرس-هاوزر^۱، کدورت سنج مدل هچ^۲ و برای اندازه‌گیری میزان جامدات معلق از طیف‌سنج مدل هچ-لانگ^۳ استفاده می‌شود و مقادیر آن‌ها در هر ساعتی ثبت و گزارش می‌شود. بهره‌بردار با توجه به این داده‌ها نسبت به میزان مصرف مواد شیمیایی تصمیم‌گیری می‌کنند. از این روی ارائه مدل‌هایی برای پیش‌بینی مقدار کدورت باقیمانده در آب با استفاده از نتایج داده‌های تجربی در جهت کاهش هزینه، زمان و خطای ناشی از اپراتور ضروری به‌نظر می‌رسد.

این پژوهش به‌صورت تحقیق میدانی در مقیاس صنعتی با جمع‌آوری ۷۰۰ سری داده به‌صورت تصادفی از گزارش‌های ساعتی واحد بهره‌برداری تصفیه‌خانه بزرگ آب گیلان، طی مدت ۱۸ ماه از فروردین سال ۱۳۹۱ تا شهریور سال ۱۳۹۲ انجام پذیرفت. در این تصفیه‌خانه از سولفات آلومینیوم به‌عنوان منعقدکننده، پلی‌الکترولیت به‌عنوان کمک منعقدکننده و از گاز کلر به‌عنوان ماده گندزدا استفاده می‌شود.

۲-۳- شناسایی و انتخاب متغیرهای ورودی

به‌منظور مدل‌سازی، ابتدا متغیرهای تأثیرگذار بر فرایند شناسایی و ارتباط میان آنها مشخص شد. با بررسی فرایند تصفیه آب، عوامل مؤثر بر کاهش کدورت به سه دسته تقسیم شدند: دسته اول مربوط به مشخصات آب خام شامل دما، pH جامدات معلق؛ دسته دوم مربوط به میزان مصرف کلر در دو مرحله مقدماتی و ثانویه و دسته سوم مربوط به میزان مصرف مواد منعقدکننده شامل سولفات آلومینیوم و پلی‌الکترولیت.

شاخص کیفیت فیزیکی آب با استفاده از مشخصات آب خام و میزان مصرف مواد شیمیایی، با بهره‌گیری از روش‌های هوشمند به‌طور دقیق و به‌گونه‌ای که مقدار مواد مصرفی نیز بهینه باشد، پیش‌بینی شد. از آنجا که شرایط آب‌های سطحی به‌خصوص در شمال کشور در هر لحظه تغییر می‌کند، میزان مصرف مواد شیمیایی به‌صورت حدس و خطا توسط بهره‌برداران با توجه به نتایج آزمایشگاهی (که تنها یکبار در روز انجام می‌گیرد)، شرایط آب ورودی و به‌طور عمده از روی تجربه بهره‌برداران تعیین می‌شود، لذا همیشه میزان مناسب مواد شیمیایی را به‌دست نمی‌دهد و ضمن بالا بردن هزینه‌ها، کیفیت مناسب و استاندارد در آب خروجی مشاهده نمی‌شود. ارائه یک مدل با دقت قابل قبول می‌تواند به بهره‌برداران در تعیین میزان بهینه مواد مصرفی پیش از فرایند تصفیه، برای دستیابی به میزان کدورت مطلوب آب برای استفاده شرب کمک شایانی نماید. لذا از میان روش‌های هوشمند، شبکه‌های عصبی GMDH چند هدفه و نیز شبکه خودباوری بی‌زین برای حل مسئله مورد استفاده قرار می‌گیرند. این سیستم‌ها توانایی بالایی در مدل‌سازی این فرایند دارند و می‌توانند تغییرات خروجی را نسبت به تغییر پارامترهای مؤثر بر فرایند پیش‌بینی کنند.

۲- مواد و روش‌ها

۲-۱- مطالعه موردی

تصفیه‌خانه بزرگ آب گیلان که در فاصله بیست کیلومتری جنوب شرقی شهر رشت قرار گرفته است، وظیفه تأمین آب آشامیدنی شهرها و روستاهای مناطق مرکزی و بخشی از مناطق شرق و غرب استان گیلان را به عهده دارد. جمعیت تحت پوشش این طرح حداکثر دو میلیون و سیصد هزار نفر است. بهره‌برداری از مرحله اول تصفیه‌خانه از بهمن سال ۱۳۷۸ شروع شده است و در حال حاضر با حداکثر ظرفیت در مدار بهره‌برداری است و آب بخش مهمی از مناطق مورد نظر از این تصفیه‌خانه تأمین می‌شود. آب خام مورد نیاز این تصفیه‌خانه از رودخانه سفیدرود و رودخانه شهر بیجار تأمین می‌شود. وضعیت آب این رودخانه‌ها در فصل‌های مختلف بسیار متغیر است. این تصفیه‌خانه برای حداکثر غلظت مواد معلق ورودی آب خام تا پانزده گرم در لیتر طراحی شده و با افزایش غلظت تا بیست گرم در لیتر نیز قادر به تصفیه آب است.

فرایند تصفیه آب در این تصفیه‌خانه ۲ ساعت و ۵۵ دقیقه شامل ۱ ساعت زمان ته‌نشینی اولیه و ۱ ساعت و ۵۵ دقیقه زمان ته‌نشینی ثانویه است؛ زمان و درجه اختلاط و میزان تزریق آهک و لجن مقادیری ثابت می‌باشند. در این فرایند میزان کاهش کدورت به متغیرهای قابل اندازه‌گیری چون دما، pH، ذرات جامد معلق و

¹ Endress-Houser, CPA510

² HACH, 1720C

³ HACH-LANGE, Sali Tax-SC100

نرون‌ها از ساختاری مشابه برخوردارند که همگی آنها دارای دو ورودی و یک خروجی می‌باشند. متغیرهای ورودی هر نرون، خروجی و یا مقادیر تخمین زده شده توسط هر دو نرون انتخاب شده در لایه پیشین است که به ازای N خروجی سیستم اصلی مجدداً مدل می‌شوند [۲۰ و ۲۱].

به‌طور کلی مسائل شناسایی سیستم این گونه مطرح می‌شوند: فرض شود رابطه نامشخص تابع f میان متغیرهای $X=(x_1, x_2, \dots, x_m)$ و مقادیر متناظر y همانند رابطه زیر باشد

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{im}) \quad i = 1, 2, \dots, N \quad (1)$$

که در آن x_i متغیرهای مدل، m تعداد متغیرها، y_i خروجی مدل و N تعداد مشاهدات یا سری داده‌های تجربی است. رابطه f را می‌توان توسط تابع \hat{f} به‌گونه‌ای تقریب زد که مجموع مربعات خطا به ازای نمونه داده‌های ورودی و خروجی کمینه شود

$$\sum_{k=1}^N [(\hat{f}(x_{ki}, x_{kj}) - y_i)^2] \rightarrow \text{Min} \quad (2)$$

در این رابطه، \hat{f} می‌تواند به‌صورت تابعی خطی و یا غیر خطی از متغیرهای ورودی سیستم باشد. نگاشتی که بین متغیرهای ورودی و خروجی توسط شبکه عصبی GMDH برقرار می‌شود، یک تابع غیر خطی به صورت رابطه زیر است [۱۹ و ۲۲]

$$\hat{y} = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (3)$$

که به نام سری ولترا^۴ شناخته می‌شود و ضرایب a ، ضرایب سری ولترا در مدل GMDH هستند. الگوریتم GMDH بر اساس تجزیه سری توابع ولترا به چند جمله‌ای‌های درجه دوم و متغیره پایه‌ریزی شده است. در واقع هدف این الگوریتم یافتن ضرایب مجهول (a) در سری توابع ولترا است. به این منظور با تجزیه آن به چند جمله‌ای‌های درجه دوم و متغیره به‌گونه‌ای این ضرایب مجهول در عوامل تجزیه شده پخش می‌شوند، لذا ضرایب مجهول در این چند جمله‌ای‌های درجه دوم تنظیم می‌شوند

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i^2 + a_4 x_j^2 + a_5 x_i x_j \quad (4)$$

تابع y دارای شش ضریب مجهول است، لذا باید طوری تنظیم شود که به ازای تمام نمونه‌های دو متغیره وابسته به سیستم

به‌منظور تعیین متغیرهای مؤثر و چگونگی ارتباط بین آنها روش شناسی سطح پاسخ^۱ استفاده شد. این روش توسط باکس و ویلسون در سال ۱۹۵۱ به‌منظور تقریب متغیر پاسخ از یک مدل چند جمله‌ای درجه اول، معرفی شد [۱۶]. با استفاده از این روش می‌توان روابط موجود بین پاسخ و متغیرهای مستقل را مشخص کرد. به این منظور از نرم افزار مینی‌تب^۲ استفاده شد. برای این کار ابتدا داده‌های تجربی به نرم‌افزار معرفی شد. پس از حصول نتایج، یکی از پارامترهایی که برای هر اثر متقابل مشخص می‌شود، مقدار P ^۳ است. در صورتی که این مقدار از 0.05 بیشتر باشد، نشان می‌دهد که متغیر مربوطه اثر قابل توجهی بر خروجی ندارد که با این شرایط حذف می‌شود [۱۷ و ۱۸]. مقادیر P مربوط به متغیرهای تأثیرگذار در جدول ۱ نشان داده شده است، که در آن متغیرهای X_1 تا X_5 به ترتیب مربوط به جامدات معلق، مقدار سولفات آلومینیوم، pH، دما و مقدار پلی‌کترولیت است.

جدول ۱- متغیرها و مقادیر P به‌دست آمده از RSM

متغیر	P-Value	متغیر	P-Value
$X_1 X_1$	۰/۰۰۰	X_1	۰/۰۰۴
$X_1 X_2$	۰/۰۰۱	X_2	۰/۰۰۰
$X_1 X_3$	۰/۰۰۵	X_3	۰/۰۰۰
$X_1 X_4$	۰/۰۰۵	X_4	۰/۰۰۱
$X_2 X_3$	۰/۰۱۳	X_5	۰/۰۰۴۹

بر اساس جدول ۱ و مقادیر P به‌دست آمده، از تمام متغیرها به استثنای میزان کلر مقدماتی و ثانویه، در مدل‌سازی استفاده می‌شود و نیز اثر متقابلی بین برخی متغیرها مشاهده می‌شود که این برهمکنش‌ها در ساختار شبکه بی‌زین، با اتصال یال‌هایی بین متغیرهای دارای اثر متقابل اعمال شده است.

۲-۴- شبکه‌های عصبی چند هدفه نوع GMDH

در میان روش‌های شناسایی مختلف، الگوریتم دسته‌بندی گروهی داده‌های عددی که به اختصار GMDH خوانده می‌شود، سیستمی خود سازمانده بوده که توسط پرفسور ایواخنکو به‌عنوان روش آنالیز چندگانه، برای مدل‌سازی و شناسایی سیستم‌های پیچیده توسعه یافت. به عبارت دیگر می‌توان بدون داشتن اطلاعات تخصصی در مورد سیستم از GMDH در مدل‌سازی استفاده نمود [۱۹].

شبکه عصبی GMDH، شبکه‌ای یک سو به است که از چندین لایه و هر لایه نیز از چندین نرون تشکیل یافته است. تمامی

¹ Response Surface Methodology (RSM)

² Minitab16

³ Probability of Obtaining a Test Statistic Result (P-value)

⁴ Volterra Series

دارای ویژگی‌هایی هستند که آنها را برای پرسش‌های تحلیلی و مدیریتی داده‌های جهان واقعی، کاربردی و مفید ساخته است [۲۴].

مدل‌های بیزی از روش‌های آمار و احتمال برای استنتاج روی ساختارهای اطلاعاتی بهره می‌گیرد. یکی از احتمالات معروف که در آن احتمال شرطی به نحوی بارز به‌کار می‌رود فرمول بیز است. این فرمول توسط یک کشیش انگلیسی به نام توماس بیز ابداع شده است و توسط لاپلاس تعمیم داده شده است که نقطه آغازین پرداختن به مسائل استنباطی با استفاده از منطق تئوری احتمالات قرار گرفت. فرمول بیز در حقیقت نقطه شروع رشته‌ای از آمار به نام آمار بیز است. قضیه بیز روشی برای دسته‌بندی پدیده‌ها، بر پایه احتمال وقوع یا عدم وقوع یک پدیده است و از آن جهت مفید است که می‌توان از طریق آن احتمال یک پیشامد را با مشروط کردن نسبت به وقوع و یا عدم وقوع یک پیشامد دیگر محاسبه کرد [۲۵].

رابطه بیز به صورت زیر بیان می‌شود

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \quad (11)$$

که در آن

$P(A)$ احتمال وقوع پدیده A و $P(A|B)$ احتمال وقوع پدیده B از میان احتمال وقوع پدیده A نامیده می‌شود [۲۵].

مدل‌های خودباوری بیزین، گراف‌های جهت‌دار بدون دوری می‌باشند که شامل گره‌های والد^۵ و مولد^۶ هستند و روابط علی و معلولی آن‌ها با استفاده از یک سری پیکان موسوم به یال^۷ مشخص می‌شود. یال‌ها بیانگر وابستگی‌ها هستند، به طوری که یک یال که بین دو گره وجود دارد بیانگر وابستگی آن دو گره است [۲۶].

در این شبکه اگر X شامل n متغیر معلول $\{x_1 \dots x_n\}$ و $(parent(x_i))$ علت حادث شدن آن متغیر باشد، احتمال رخداد X به صورت رابطه ۱۲ بیان می‌شود. به این پدیده معلول اصطلاحاً مولد و به علت آن پدیده والد اطلاق می‌شود [۲۷ و ۲۸].

$$P(X) = \prod_{i=1}^n P(x_i | parent(x_i)) \quad (12)$$

۱- ایجاد یک مدل خودباوری بیزین به‌طور کلی شامل چهار گام:

⁴ Node
⁵ Parent
⁶ Child
⁷ Link

$p, q \in \{1, 2, \dots, m\}$ در حالی که $\{x_{ip}, x_{iq}, i = 1, 2, 3, \dots, N\}$ باشند، خروجی مطلوب $\{y_i, i = 1, 2, \dots, N\}$ برقرار شود. به همین دلیل تابع G بر اساس قاعده کمترین مربعات خطا^۱ پایه‌ریزی می‌شود.

$$\sum_{k=1}^N \left[(G(x_{ki}, x_{kj}) - y_k)^2 \right] \rightarrow Min \quad (5)$$

با شرایط حاکم بر مسئله، دستگاه معادله فوق را می‌توان به فرم ماتریسی نیز نمایش داد

$$Aa = Y \quad (6)$$

که در آن A ماتریس متغیرها، Y ماتریس خروجی‌ها و ماتریس a ترانزاده ماتریس ضرایب است

$$a = \{a_0, a_1, a_2, a_3, a_4, a_5\} \quad (7)$$

$$Y = \{y_1, y_2, y_3, \dots, y_N\}^T \quad (8)$$

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}^2 & x_{1q}^2 & x_{1p}^2 x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}^2 & x_{2q}^2 & x_{2p}^2 x_{2q}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{Np} & x_{Nq} & x_{Np}^2 & x_{Nq}^2 & x_{Np}^2 x_{Nq}^2 \end{bmatrix} \quad (9)$$

برای حل معادله لازم است که معکوس ماتریس غیر مربعی A محاسبه شود. به همین دلیل برای محاسبه معکوس ماتریس غیر مربعی A ، از روش حل معادلات نرمال^۲ استفاده می‌شود. بنابراین بردار ضرایب مجهول a توسط رابطه زیر به دست می‌آید [۲۰]

$$a = (A^T A)^{-1} A^T Y \quad (10)$$

۲-۵- شبکه خودباوری بیزین^۳

شبکه تصمیم‌گیری بیزین، روشی سودمند است که می‌تواند اطلاعات و داده‌های کمی را به خوبی و با دانش کیفی کارشناسی جمع نماید [۲۳]. از طرف دیگر این شبکه‌ها یکی از شاخه‌های هوش مصنوعی سمبولیک می‌باشند. روش‌های مدل‌سازی بیزین

¹ Least Squares Error

² Solving Normal Equations (SNE)

³ Bayesian Belief Network (BBN)

جدول ۲- محدوده داده‌های جمع‌آوری شده از واحد بهره‌برداري

متغیرهای ورودی	واحد	حداقل	حداکثر	میانگین
دما	°C	۳/۹	۲۸/۸	۱۷/۱۱
pH	-	۷/۵	۸/۵	۸/۰۲
جامدات معلق	mg/L	۴	۶۸۰	۷۸
پلی‌کترولیت	ppm	۰/۰۷۲	۰/۲۶۳	۰/۱۴۳
سولفات آلومینیوم	ppm	۳/۹۸۹	۲۷/۱۲۸	۱۲/۱۳۴
کدورت باقیمانده	NTU	۰/۴	۱/۶	۰/۶۷

تصفیه‌خانه آب گیلان است، می‌باشد. این میزان در واحد بهره‌برداري با توجه به گزارش‌های روزانه واحد آزمایشگاه و نیز تجربه خود اپراتور در تعیین مقدار مصرف مواد منعقدکننده، به دست می‌آید. نمونه داده‌های جمع‌آوری شده به منظور مدل‌سازی در جدول ۳ آورده شده است. برای نمایش توانایی و استعداد شبکه‌های عصبی نوع GMDH و شبکه خودباوری بیزین در زمینه پیش‌بینی پارامتر خروجی در حالتی که پارامترهای ورودی تغییر می‌کنند، داده‌های آزمایشگاهی به دو دسته (۷۰ درصد برای آموزش و ۳۰ درصد برای آزمایش) تقسیم شدند تا عملکرد مدل مورد ارزیابی قرار گیرد.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_{(i,exp)} - Y_{(i,model)})^2 \quad (13)$$

$$NRMSE = \left[\sum_{i=1}^n (Y_{(i,exp)} - Y_{(i,model)})^2 / n \right]^{0.5} / (Y_{(exp,max)} - Y_{(exp,min)}) \quad (14)$$

$$AADP(\%) = \frac{100}{n} \sum_{i=1}^n \left| \left(\frac{Y_{(i,model)}}{Y_{(i,exp)}} \right) - 1 \right| \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{(i,exp)} - Y_{(i,model)})^2}{\sum_{i=1}^n (Y_{(i,ex)})^2} \quad (16)$$

که در این روابط

$Y_{i,model}$ کدورت آب محاسبه شده توسط مدل، $Y_{i,exp}$ کدورت آب تجربی، $Y_{exp,min}$ حداقل کدورت تجربی، $Y_{exp,max}$ حداکثر کدورت خروجی و n تعداد دسته داده‌ها است.

۳-۱- مدل‌سازی با شبکه GMDH

به منظور طراحی شبکه GMDH چند هدفه باید توابع هدف تعریف شود که این توابع عبارت‌اند از کمینه کردن خطای آموزش و

متغیرها ۳- تعریف حالات متغیرها ۴- محاسبه جدول احتمال شرطی^۱ برای هر گره است.

این مدل با استفاده از نرم‌افزار نتیکا^۲ پیاده‌سازی شده است. نرم‌افزار نتیکا از دو الگوریتم برای آموزش استفاده می‌کند: یکی الگوریتم امید- بیشینه‌سازی^۳ و دیگری الگوریتم گرادیان نزولی^۴.

الگوریتم امید- بیشینه‌سازی یک روش تکرارشونده است که به دنبال یافتن برآوردی با بیشترین درست‌نمایی برای پارامترهای یک توزیع پارامتری است. به بیان دیگر این الگوریتم یک تناوب را بین انجام دادن مرحله امید (E) و بیشینه‌سازی (M) ایجاد می‌کند، به این صورت که در مرحله امید، امید احتمال را ایجاد می‌کند و در مرحله بیشینه‌سازی، تخمین احتمال بیشینه پارامترها را توسط امید احتمال ایجاد شده در مرحله قبل (E)، محاسبه می‌کند. پارامترهای یافته شده در مرحله M برای آغاز مرحله دیگری از E، به کار گرفته می‌شوند و این فرایند تکرار می‌شود.

الگوریتم گرادیان نزولی، پایه بسیاری از الگوریتم‌های بهینه‌سازی است و از آن در یادگیری شبکه‌های عصبی و حداقل‌سازی مقدار خطای مدل برای تنظیم پارامترهای مدل استفاده می‌شود. الگوریتم گرادیان نزولی در فضای وزن‌ها به دنبال برداری می‌گردد که خطا را حداقل کند. با توجه به این‌که، سطح خطا به صورت یک سهمی است، این الگوریتم از یک مقدار دلخواه برای بردار وزن شروع کرده و در هر مرحله وزن‌ها را طوری تغییر می‌دهد که در جهت شیب کاهشی منحنی فوق خطا کاهش داده شود [۲۷].

۳-۲- مدل‌سازی و تحلیل یافته‌ها

هدف از مدل‌سازی، یافتن و شناسایی رابطه‌ای بین پارامترهای ورودی و خروجی و به کارگیری شبکه‌های GMDH و بیزین به منظور کاهش در پیچیدگی سیستم شبکه عصبی است؛ همچنین افزایش دقت در مدل‌سازی و پیش‌بینی اثر تغییرات متغیرهای ورودی بر خروجی فرایند در نظر گرفته می‌شود. متغیرهای ورودی شامل مشخصات آب خام (دما، pH، جامدات معلق) و نیز مقادیر مصرفی سولفات آلومینیوم و پلی‌کترولیت است. محدوده مورد مطالعه این متغیرها در جدول ۲ آورده شده است.

مدل‌سازی فرایند با استفاده از داده‌های فرایندی شامل ۷۰۰ سری داده، انجام شد. مدل شامل ۵ متغیر مؤثر به عنوان ورودی و یک خروجی که میزان کدورت باقیمانده در آب تصفیه شده خروجی از

¹ Conditional Probability Table (CPT)

² Netica 512

³ Expectation - Maximization (EM)

⁴ Gradient Descent (GD)

جدول ۳- نمونه داده‌های واحد بهره‌برداری شامل ورودی‌ها و خروجی مدل‌سازی

شماره	ورودی‌های مدل					خروجی
	جامدات معلق (mg/L)	سولفات آلومینیوم (ppm)	pH (-)	دما (°C)	پلی‌کترولیت (ppm)	
۱	۲۵	۱۹/۹۴۷	۸/۴	۲۱/۳	۰/۱۷۶	کدورت باقیمانده (NTU) ۰/۶
۲	۱۵	۱۱/۱۷	۸/۴	۲۰/۹	۰/۱۳۶	...
...
...
۶۹۹	۸	۱۰/۳۷۲	۸	۲۴/۹	۰/۱۰۴	...
۷۰۰	۲۹	۸/۷۷۷	۷/۹	۲۲/۶	۰/۱۱۲	...

جدول ۴- میزان خطای آموزش و آزمایش در منحنی پاره‌تو مدل GMDH چند هدفه

نقاط طراحی	خطای آموزش	خطای آزمایش
A	۰/۰۱۲۷۷۶	۰/۰۱۳۱۷۹
B	۰/۰۱۶۵۰۷	۰/۰۰۹۱۸۷
M	۰/۰۱۲۸۸۶	۰/۰۱۳۰۱۵

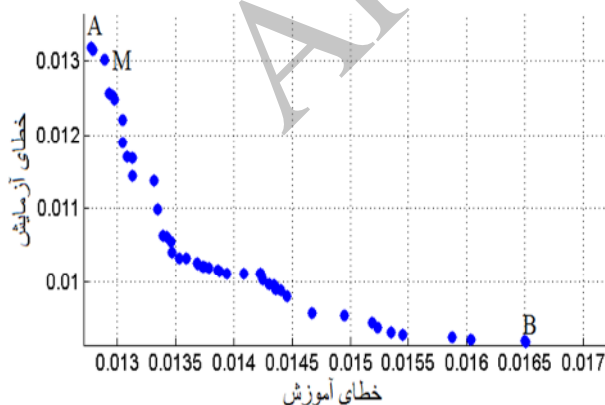
جدول ۵- کروموزوم برتر برای پیش‌بینی میزان کدورت باقیمانده در آب

خطای آموزش	خطای آزمایش	کروموزوم بهینه به دست آمده توسط مدل شبکه عصبی GMDH
۰/۰۱۲۸۸۶	۰/۰۱۳۰۱۵	۱۱۱۴۱۲۱۳۱۴۲۲۱۴۳۴۴۱۳۳۵۲۲۲۴۳۳۳۳۴۴۴

جدول ۶- مقداردهی پارامترهای الگوریتم ژنتیک در مدل GMDH چند هدفه

جمعیت اولیه	۴۰۰	احتمال جهش	۰/۰۹
تعداد نسل	۱۲۰۰	احتمال پیوند (تولید مثل)	۰/۹۷
تعداد لایه‌های مخفی	۴	تعداد توابع هدف	۲

باید توجه داشت که هر عدد متعلق به یک متغیر ورودی است که ترتیب آنها به شرح زیر است: ۱- جامدات معلق ۲- سولفات آلومینیوم ۳- pH ۴- دما ۵- پلی‌کترولیت.



شکل ۱- نقاط پاره‌تو خطای آموزش و آزمایش مدل GMDH چند هدفه

آزمایش شبکه که مربوط به اختلاف بین داده‌های تجربی و نتایج حاصل از مدل است. با طراحی بهینه توابع مورد نظر سیستم، میزان کدورت آب صاف خروجی به میزان قابل قبولی پیش‌بینی می‌شود. برای دستیابی به این مهم از طراحی تکاملی ساختار شبکه عصبی توسط الگوریتم ژنتیک استفاده شد. در الگوریتم ژنتیک ابتدا جمعیت اولیه ایجاد و کروموزوم‌ها تولید شدند، سپس کروموزومی که کمترین میزان خطای آموزش و خطای آزمایش را بر اساس رابطه ۱۳ دارا بود، به عنوان کروموزوم برتر انتخاب شد که در شکل ۱ منحنی پاره‌تو^۱ کروموزوم‌های تولید شده نمایش داده شده است. مقدار کمترین خطای آموزشی و آزمایشی با استفاده از این نمودار در جدول ۴ آمده است. نقطه مصالحه طراحی از بین تمامی نقاط پاره‌تو نقطه M می‌تواند باشد زیرا خطای مدل‌سازی نسبت به خطای آموزشی در این نقطه، دارای توازن بیشتری نسبت به نقاط دیگر است. به دست آوردن نقطه M آخرین مرحله از

¹ Pareto

جدول ۷ گزارش شده است. مدل‌سازی شرح داده شده با انجام کدنویسی در نرم‌افزار متلب^۱ انجام شد. پس از به دست آمدن مدل پیشنهادی، نتایج حاصل از مدل‌سازی با داده‌های تجربی مقایسه شد (شکل ۳). همانگونه که در شکل ۳ مشخص است، نتایج حاصل از مدل‌سازی تطبیق بسیار خوبی با نتایج تجربی دارند. مقادیر حداقل و حداکثر داده‌های تجربی با نتایج حاصل از مدل مقایسه و در جدول ۸ گزارش شده‌اند.

در مرحله بهینه‌سازی برای سیستم بررسی شده که به صورت چندهدفه انجام می‌گیرد، پارامترهای ژنتیکی مطابق جدول ۶ در نظر گرفته شدند. سپس مدل‌سازی انجام شد. ساختار شبکه عصبی GMDH به دست آمده برای پیش‌بینی میزان کدورت آب در شکل ۲ ارائه شده است. در نهایت برای پیدا کردن و شناسایی رابطه‌ای بین پارامترهای ورودی و خروجی، معادلات مربوطه برای ساختار GMDH به منظور پیش‌بینی میزان کدورت آب به دست آمد که در

¹ Matlab 7.11.0 (R2010b)

جدول ۷- روابط بازگشتی مدل GMDH چند هدفه به منظور پیش‌بینی میزان کدورت باقیمانده در آب شرب

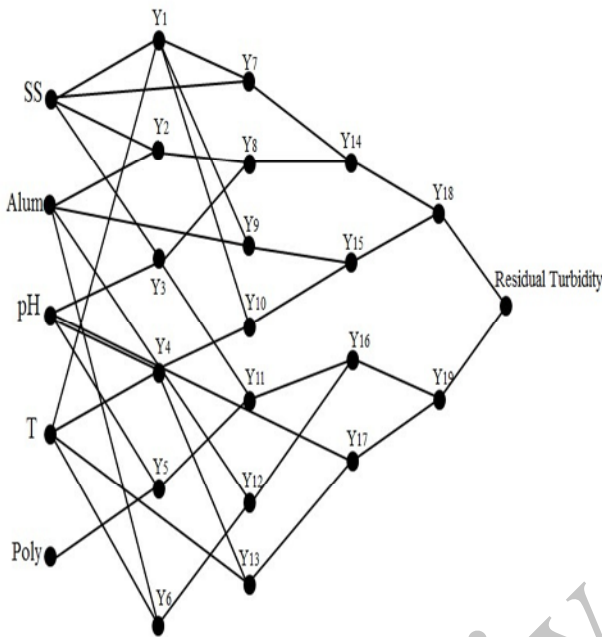
Y_1	$=0.760916446+0.001548329*SS-0.024782086*T-1.95E-0.6*SS^2-0.000728066*T^2-2.43E-0.5*SS*T$
Y_2	$=0.512418806-0.000607938*SS+0.015276741*A1-1.29E-0.6*SS^2-0.000548721*A1^2+9.35E-0.5*SS*A1$
Y_3	$=0.064929774+0.012164276*SS+0.260177081*pH-272E-0.6*SS^2-0.024407036*pH^2-0.001328418*ss*pH$
Y_4	$=28.8759121-7.129578094*pH+0.16566787*T+0.45292647*pH^2+0.000779599*T^2-0.023718746*pH*T$
Y_5	$=-24.11212108+5.645305829*pH+32.69628155*PE-0.317030446*pH^2+13.29804522*PE^2-4.432498938*pH*PE$
Y_6	$=0.224479916+0.041200908*A1+0.016282438*T-0.000231045*A1^2+0.000113744*T^2-0.001756626*A1*T$
Y_7	$=1.613343354+0.000641579*SS-3.89801005*Y_1-6.00E-0.7*SS^2+3.6744351*Y_1^2-0.000688731*SS*Y_1$
Y_8	$=2.437983375-3.365645426*Y_2-3.175321654*Y_3-0.373978624*Y_2^2-0.499900334*Y_3^2+6.677600013*Y_2*Y_3$
Y_9	$=2.153345668-3.097498219*Y_1-0.098660662*A1+0.254556918*Y_1^2-0.001690494*A1^2+0.233377382*Y_1*A1$
Y_{10}	$=0.071611192-2.288468407*Y_1+2.758643878*Y_4+2.362092064*Y_1^2-1.671389658*Y_4^2-0.066401672*Y_1*Y_4$
Y_{11}	$=2.536466374-0.798231174*Y_3-5.663781498*Y_5-1.280164621*Y_3^2+1.991408983*Y_5^2+4.727335823*Y_3*Y_5$
Y_{12}	$=2.214007964-0.015971069*A1-4.606528653*Y_6-0.002598266*A1^2+2.020608352*Y_6^2+0.14641095*A1*Y_6$
Y_{13}	$=-1.270582764+4.008023549*Y_4+0.029825297*T-1.760052814*Y_4^2-0.000165049*T^2-0.035792461*Y_4*T$
Y_{14}	$=1.004392255+2.539931531*Y_7-4.529525336*Y_8-7.080595162*Y_7^2-1.395782498*Y_8^2+10.68795438*Y_7*Y_8$
Y_{15}	$=-1.010558178+0.295468529*Y_9+3.587190852*Y_{10}+2.128864411*Y_9^2-0.815446094*Y_{10}^2-3.352883273*Y_9*Y_{10}$
Y_{16}	$=0.955979401-2.601282139*Y_{11}+0.5948808*Y_{12}-0.665505341*Y_{11}^2-2.934767734*Y_{12}^2+5.941836596*Y_{11}*Y_{12}$
Y_{17}	$=-313.7124706+8.55624753*pH+232.8856909*Y_{13}-2.733816002*pH^2-43.18878584*Y_{13}^2-21.59163165*pH*Y_{13}$
Y_{18}	$=1.145132634-3.863531725*Y_{14}+1.534999251*Y_{15}-2.903954063*Y_{14}^2-6.495420752*Y_{15}^2+11.81246067*Y_{14}*Y_{15}$
Y_{19}	$=-3.158136413+2.934146779*Y_{16}+7.340534098*Y_{17}+0.61634615*Y_{16}^2-3.612683826*Y_{17}^2-3.471518899*Y_{16}*Y_{17}$
Tu	$=-0.28491988-0.549955818*Y_{18}+1.621005165*Y_{19}+1.620308387*Y_{18}^2+0.446600078*Y_{19}^2-2.121046842*Y_{18}*Y_{19}$

جدول ۸- محدوده نتایج مدل‌سازی کدورت آب با شبکه‌های نوع GMDH چند هدفه و بیزین

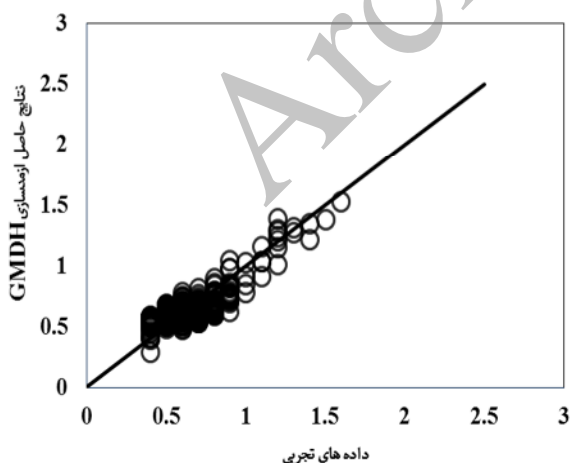
حداکثر خروجی (NTU)		حداقل خروجی (NTU)		نوع مدل‌سازی
مدل	تجربی	مدل	تجربی	
۱/۵۳	۱/۶	۰/۲۹	۰/۴	GMDH
۱/۶	۱/۶	۰/۳۵	۰/۴	EM
۱/۶	۱/۶	۰/۴	۰/۴	GD

جدول ۹- پارامترهای آماری و مقادیر خطای مدل‌سازی کدورت باقیمانده در آب شرب

مدل GMDH چند هدفی				مدل بیزین					عملکرد مدل
کل	آموزش	آزمایش	کل	آموزش	آزمایش	کل	آموزش	آزمایش	
۰/۰۱۲۹۳	۰/۰۱۲۸۹	۰/۰۱۳۰۱۵	۰/۰۰۹۲۸۸	۰/۰۰۱۴۸۷	۰/۰۲۷۴۹	۰/۰۱۲۵۳۷	۰/۰۰۲۴۰۶	۰/۰۳۶۱۲۹	MSE
۰/۰۹۴۷۴	۰/۰۹۴۵۹	۰/۰۲۲۸۱۶۳	۰/۰۰۸۰۳۱۱	۰/۰۰۳۲۱۳۱	۰/۰۳۳۱۶۰۵	۰/۰۰۹۳۳۰۸	۰/۰۰۴۰۸۷۸	۰/۰۱۵۸۳۹۶	NRMSE
۱۴/۳۰۱۵	۱۴/۸۰۸۹	۱۳/۱۱۷۳۳	۷/۶۶۷۰۲۴	۱/۹۲۶۵۵	۲۱/۰۶۱۴۵	۸/۲۹۶۲۳	۲/۲۷۱۲۳۳	۲۲/۳۵۴۵۵	AADP (%)
۰/۹۷۱۸۲	۰/۹۷۲۱۸	۰/۹۷۰۹۵	۰/۹۷۹۷	۰/۹۹۶۷	۰/۹۳۸۸	۰/۹۷۲۷	۰/۹۹۴۸	۰/۹۱۹۶	R ²



شکل ۲- ساختار پیشنهادی شبکه عصبی نوع GMDH چند هدفه برای پیش‌بینی میزان کدورت باقیمانده در آب تصفیه شده



شکل ۳- مقایسه خروجی واقعی و خروجی مدل شده در مدل‌سازی GMDH چند هدفه

همچنین دقت و مقادیر خطا در جدول ۹ ارائه شده است. برای بهینه‌سازی میزان مصرف مواد شیمیایی از الگوریتم ژنتیک استفاده شد. مقادیر بهینه متغیرهای مؤثر بر میزان کدورت باقیمانده در جدول ۱۰ آورده شده است. در این شرایط میزان مصرف منعقدکننده‌ها برای دستیابی به حد مطلوب کدورت آب شرب، حداقل است. اثر میزان مصرف مواد منعقدکننده روی کدورت باقیمانده در شرایط بهینه در شکل ۴ نشان داده شده است. در این شکل سایر پارامترها مقدار بهینه خود را خواهند داشت. شکل ۴ نشان می‌دهد که با افزایش میزان مصرف سولفات آلومینیوم تا ۵ppm، میزان کدورت تا مرز حد مطلوب آن یعنی ۱ NTU کاهش می‌یابد که با افزایش بیشتر سولفات آلومینیوم، این میزان به کمتر از ۱ NTU می‌رسد. همچنین این شکل نشان می‌دهد سولفات آلومینیوم به نسبت پلی‌الکترولیت تأثیر بیشتری در کاهش کدورت دارد. برای بهینه‌سازی از ابزار بهینه‌سازی نرم‌افزار متلب استفاده شد.

۳-۲- مدل‌سازی با شبکه بیزین

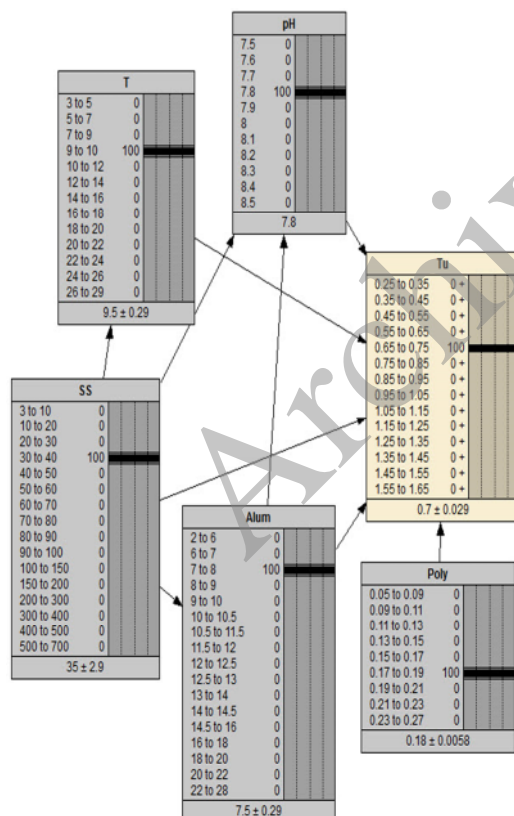
در این پژوهش ابتدا سعی شد ارتباط بین گره‌ها (نقاط والد و مولد) به وسیله روش سطح پاسخ مشخص شود. در روش سطح پاسخ ابتدا متغیرهای تأثیرگذار بر مدل و ارتباط بین متغیرهای مستقل (فاکتورها) و متغیر وابسته (پاسخ)، توسط معیار مقدار P مشخص شد. در ادامه پس از مشخص شدن ساختار مدل، جداول احتمال شرطی برای متغیرها تخمین زده شد. مدل ساخته شده در نرم‌افزار نتیکا در شکل ۵ نشان داده شده است.

جدول ۱۰- مقادیر بهینه متغیرهای مستقل مؤثر بر میزان کدورت باقیمانده در آب شرب شامل ورودی‌ها و خروجی

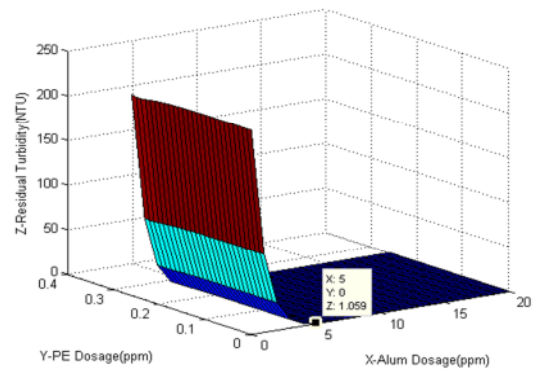
خروجی مدل	ورودی‌های مدل					نوع مدل
کدورت باقیمانده (NTU)	دما (C)	pH (-)	جامدات معلق (mg/L)	سولفات آلومینیوم (ppm)	پلی‌الکترولیت (ppm)	GMDH
۱	۱۲/۷۶	۷/۵۴	۱۷۰/۳۹	۵	۰/۱۹	

توزیع‌های احتمال نشان داده شده در این شکل‌ها مربوط به یک نمونه از مدل هستند که در آن ورودی‌های مدل از طریق دما، pH، جامدات معلق، مقدار پلی‌الکترولیت و سولفات آلومینیوم مشخص شده‌اند. مقدار عددی که توسط فرایند استنتاج به گره کدورت آب اختصاص داده شده است، همان‌طور که در این دو شکل نیز نشان داده شده است، برای الگوریتم امید-بیشینه‌سازی، 0.29 ± 0.07 و برای الگوریتم گرادیان نزولی، 0.3 ± 0.07 محاسبه شده است. مقدار تجربی کدورت آب برای این حالت 0.65 بود که نشان‌دهنده دقت خوب هر دو الگوریتم بوده است.

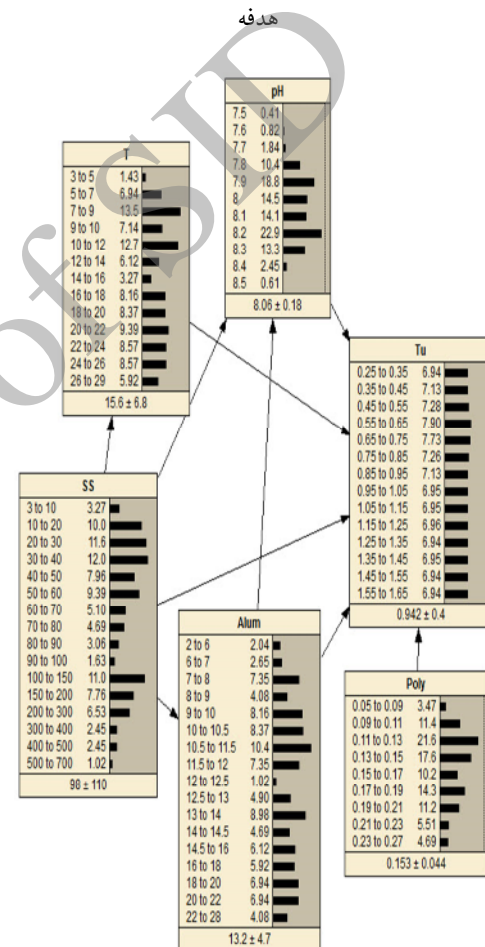
مجدداً پس از مدل‌سازی، برای حصول اطمینان از قابل اطمینان بودن مدل، نتایج حاصل از مدل‌سازی با هر دو الگوریتم آموزشی با داده‌های تجربی مقایسه شد (شکل‌های ۸ و ۹). همانگونه که در شکل‌ها مشخص است، نتایج مدل‌سازی با هر دو الگوریتم تطبیق بسیار خوبی با نتایج تجربی دارند. دقت و مقادیر خطای مدل در جدول ۹ ارائه شده است. همچنین مقادیر حداقل و حداکثر داده‌های تجربی با نتایج حاصل از مدل مقایسه و در جدول ۱۰ گزارش شده‌اند.



شکل ۶- نمونه‌ای از استنتاج مدل بیزی بر اساس الگوریتم امید بیشینه‌سازی در نتیکا



شکل ۴- اثر میزان مصرف سولفات آلومینیوم و پلی‌الکترولیت بر روی میزان کاهش کدورت آب در شرایط بهینه برای مدل GMDH چند



شکل ۵- مدل بیزی پیاده شده در نرم‌افزار نتیکا

مدل بیزی پس از پیاده‌سازی در نرم‌افزار نتیکا، به وسیله داده‌های تجربی مورد آموزش واقع شد. در این مدل‌سازی عمل آموزش توسط هر دو الگوریتم امید-بیشینه‌سازی و گرادیان نزولی صورت گرفت. یک نمونه از اجرای مدل بیزی در نتیکا توسط دو الگوریتم فوق در شکل‌های ۶ و ۷ نشان داده شده است.

۴- نتیجه‌گیری

در این پژوهش سعی شد با استفاده از شبکه‌های عصبی گامی در جهت مدل‌سازی میزان کدورت باقیمانده در آب تصفیه شده، با توجه به متغیرهای ورودی برداشته شود. هدف عمده این پژوهش ارائه مدلی بود که به ازای مقادیر تجربی از دقت بالایی برخوردار باشد، پیش‌بینی معقولی را به همراه داشته باشد، و میزان دخالت اپراتور را کاهش دهد و در نتیجه، میزان خطای انسانی در مقدار مصرف مواد شیمیایی را کمتر کند.

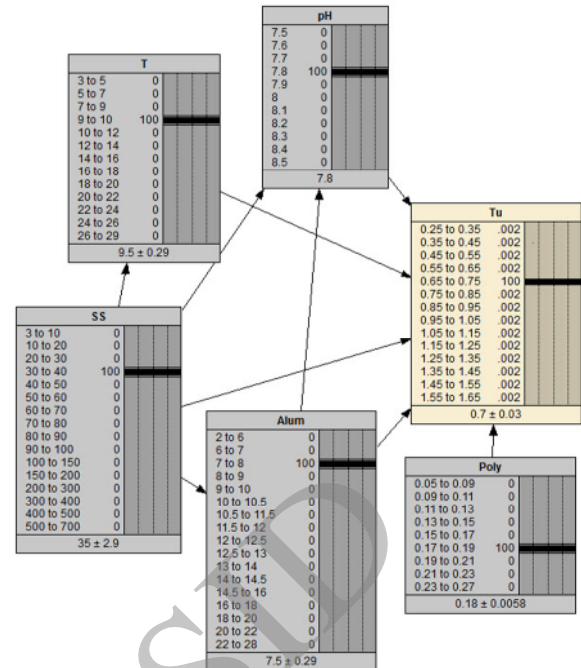
به منظور ارزیابی مدل از مقادیر خطا و ضریب تعیین (R^2) استفاده شد که با توجه به نتایج ارائه شده در شبکه‌های بیزی نسبت به شبکه GMDH، آموزش دقیق‌تر، و در میان الگوریتم‌های بیزین، الگوریتم امید-بیشینه‌سازی، دقت آموزش بالاتری را نشان داده است. در بخش آزمایش شبکه که به منظور اعتبارسنجی مدل انجام شد، اختلاف معنی‌داری بین مدل‌ها مشاهده شد. در این بخش شبکه GMDH دارای آزمایش دقیق‌تری است که این حاکی از قابل اطمینان بودن مدل به دست آمده است. در میان الگوریتم‌های بیزی، امید-بیشینه‌سازی دقت آموزش بالاتری را نشان داد.

با توجه به نتایج به دست آمده و مقایسه آن با داده‌های تجربی مشخص شد که هر دو نوع شبکه (GMDH و بیزین) نتایج قابل قبولی را ارائه داده‌اند. با وجود این به نظر می‌آید شبکه‌های عصبی نوع GMDH می‌تواند جایگزین مناسب‌تری برای تجربه اپراتور باشد؛ زیرا به نسبت هر دو الگوریتم بیزی، قابلیت پاسخگویی و تعیین کدورت آب در زمان بسیار کم و با دقت بسیار بالا را دارد. یکی از دلایل این امر را می‌توان در محدودیت‌های شبکه بیزی در آموزش و آزمون متغیرهای پیوسته عنوان کرد، زیرا در فرایند تصفیه آب بیشتر متغیرها از نوع پیوسته است.

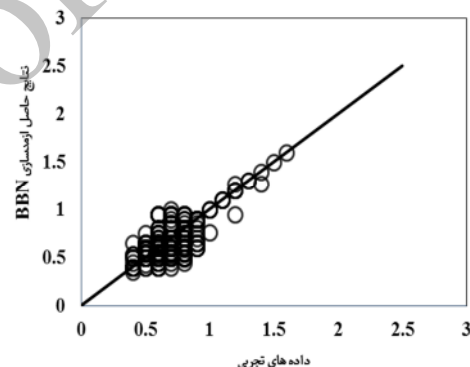
از مزایای این مدل‌ها این است که هر سیستم پیچیده و غیرخطی را می‌توان مدل کرد. هر قدر دقت بهره‌برداران در میزان مصرف مواد بیشتر باشد، خطای مدل ایجاد شده نیز کمتر خواهد بود و دقت مدل به طور محسوسی به دقت کار بهره‌بردار و داده‌های تجربی وابسته است.

با توجه به جوابگو بودن مدل، می‌توان در سایر قسمت‌های تصفیه‌خانه که نیاز به بهینه‌سازی مواد مصرفی است مانند مرحله گندزدایی، از شبکه‌های عصبی نوع GMDH استفاده نمود تا از این طریق هم میزان دخالت نیروی انسانی کاهش یابد و هم در هزینه و زمان صرفه‌جویی شود.

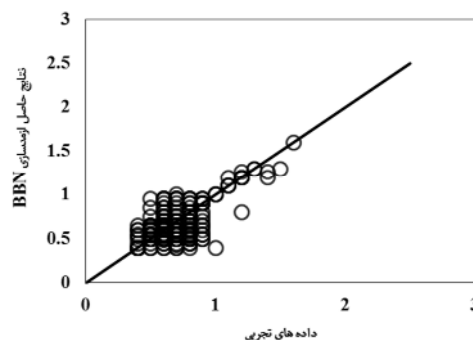
از محدودیت‌های این مدل‌سازی می‌توان به عدم وجود مقدار مواد آلی، هدایت الکتریکی، رنگ و زمان لخته‌سازی اشاره کرد که



شکل ۷- نمونه‌ای از استنتاج مدل بیزی بر اساس الگوریتم گرادیان نزولی در نتیکا



شکل ۸- مقایسه خروجی واقعی و خروجی مدل شده در مدل‌سازی بیزی بر اساس الگوریتم امید-بیشینه‌سازی



شکل ۹- مقایسه خروجی واقعی و خروجی مدل شده در مدل‌سازی بیزی بر اساس الگوریتم گرادیان نزولی

تمامی کارکنان محترم واحد بهره‌برداری تصفیه‌خانه بزرگ آب گیلان که نهایت همکاری را در جمع‌آوری داده‌های لازم داشته‌اند، تشکر و قدردانی نمایند.

این موارد در واحد بهره‌برداری تصفیه‌خانه مورد مطالعه، اندازه‌گیری و ثبت نمی‌شود.

۵- قدردانی

نویسندگان مقاله بر خود لازم می‌دانند از زحمات مدیریت محترم و

۶- مراجع

1. Chalkosh Amiri, M. (2010). *Principes of water treatment*, Arkan Danesh Publications, Isfahan. (In Persian)
2. Davis, M. L. (2010). *Water and wastewater engineering, design principles and practice*, The Mc Graw-Hill Companies. Inc., Michigan State University, USA.
3. Water, C.O.D. (2012). *Turbidity in drinking water*, Federal, Provincial-Territorial Committee, Canada.
4. Gagnon, C., Grandjean, B., and Thibault, J. (1997). "Modelling of coagulant dosage in a water treatment plant." *Artificial Intelligence in Engineering*, 11 (4), 401-404.
5. Montgomery, J. M. (1985). *Water treatment: Principles and design*, Published by John Wiley and Sons Ltd., New York, USA.
6. Reckhow, K. H. (1999). "Water quality prediction and probability network models." *Canadian Journal of Fisheries and Aquatic Sciences*, 56 (7), 1150-1158.
7. Joo, D.-S., Choi, D.-J., and Park, H. (2000). "The effects of data preprocessing in the determination of coagulant dosing rate." *Water Research*, 34 (13), 3295-3302.
8. Pike, W. A. (2004). "Modeling drinking water quality violations with bayesian networks." *Journal of the American Water Resources Association (JAWRA)*, 40 (6), 1563-1578.
9. Benardos, P., and Vosniakos, G.-C. (2007). "Optimizing feedforward artificial neural network architecture." *Engineering Applications of Artificial Intelligence*, 20 (3), 365-382.
10. Wu, G. D., and Lo, S. L. (2008). "Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system." *Engineering Applications of Artificial Intelligence*, 21 (8), 1189-1195.
11. Olyaie, E., Banejad, H., Samadi, M. T., Rahmani, A., and Saghi, M. H. (2010). "Performance evaluation of artificial neural networks for predicting rivers water quality indices (BOD and DO) in Hamadan Morad Beik River." *Water and Soil Science*, 20 (3), 199-210. (In Persian)
12. Juntunen, P., Liukkonen, M., Pelo, M., Lehtola, M. J., and Hiltunen, Y. (2012). "Modelling of water quality: An application to a water treatment process." *Applied Computational Intelligence and Soft Computing*, 2012, (4), doi : 10.1155/2012/846321.
13. Nam, S. W., Jo, B. I., Kim, M.-K., Kim, W.-K., and Zoh, K.-D. (2013). "Streaming current titration for coagulation of high turbidity water." *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 419, 133-139.
14. Daghbandan, A., Akbarizadeh, M., and Yaghoobi, M. (2013). "Modeling and optimization of poly electrolyte dosage in water treatment process by GMDH Type-NN and MOGA." *International Journal of Chemoinformatics and Chemical Engineering (IJCCE)*, 3 (2), 94-106.
15. Daghbandan, A., Ali Taleshi, F., and Akbari Zadeh, M. (2014). "Modeling of coagulant dosage in water treatment process by artificial neural network." *The 8th International Chemical Engineering Congress and Exhibition (IChEC 2014)*, Kish, Iran. (In Persian)

16. Box, G. E., and Wilson, K. (1951). "On the experimental attainment of optimum conditions." *J. of the Royal Statistical Society Series B (Methodological)*, 13 (1), 1-45.
17. Khayet, M., Zahrim, A., and Hilal, N. (2011). "Modelling and optimization of coagulation of highly concentrated industrial grade leather dye by response surface methodology." *Chemical Engineering Journal*, 167 (1), 77-83.
18. Ghorbani, F., Younesi, H., Ghasempouri, S. M., Zinatizadeh, A. A., Amini, M., and Daneshi, A. (2008). "Application of response surface methodology for optimization of cadmium biosorption in an aqueous solution by *Saccharomyces cerevisiae*." *Chemical Engineering Journal*, 145 (2), 267-275.
19. Ivakhnenko, A.G. (1971). "Polynomial theory of complex systems." *Systems, Man and Cybernetics, IEEE Transactions on, SMC-I*, 4, 364-378.
20. Darvizeh, A., Nariman-Zadeh, N., and Gharababaei, H. (2003). "GMDH-type neural network modelling of explosive cutting process of plates using singular value decomposition." *Systems Analysis Modelling Simulation*, 43 (10), 1383-1397.
21. Nariman-Zadeh, N., and Jamali, A. (2007). "Pareto design of GMDH-type neural networks for nonlinear systems." *Proceedings of the International Workshop on Inductive Modelling*, Czech Technical University, Prague, Czech Republic, 96-103.
22. Dolenko, S., Orlov, Y. V., and Persiantsev, I.G. (1996). Practical implementation and use of Group Method of data Handling (GMDH): Prospects and problems." *Proceedings of ACEDC*, 96, PEDC, University of Plymouth, England.
23. Baran, E., and Jantunen, T. (2004). "Stakeholder consultation for bayesian decision support systems in eEnvironmental management." *Regional Conference on Ecological and Enviromental Modeling (ECOMOD 2004)*, Penang, Malaysia, 27 (35.6), 31-37.
24. Uusitalo, L.(2007). "Advantages and challenges of bayesian networks in environmental modelling." *Ecological Modelling*, 203 (3), 312-318.
25. Jensen, F. V., and Kjarulff, U. (2005). "Bayesian networks and decision graphs." *A 3-week Course at Reykjavik University*, Group of Machine Intelligence, Department of Computer Science, Aalborg University.
26. Pearl, J. (2000). *Models, reasoning and inference*, Cambridge University Press Cambridge, UK.
27. Grossman, D., and Domingos, P. (2004). "Learning bayesian network classifiers by Maximizing Conditional Likelihood." *Proc., Proceedings of the twenty-first international conference on Machine learning*, ACM, New York, USA. 46.
28. Khanteymooori, A., and Sameni, M. (2011). "Precipitation modeling using bayesian networks." *The Fifth Iran Data Mining Conference / IDMC 3122*, Amirkabir University of Technology Tehran, Iran. (In Persian)