KNO-1302-5201

ارزیابی شاخص سلامت و سن ظاهری ترانسفورماتورهای قدرت با استفاده از مدل یادگیری ماشین مبتنی بر رگرسیون گرادیان بوستینگ

محمدمهدی توکلی pezhman99t@gmail.com
سید سعید موسوی anchepoli@gmail.com
کارشناسی ارشد، گروه مهندسی برق، دانشگاه تخصصی فناوری های نوین آمل
استادیار، گروه انرژی های نو، دانشکده مهندسی، دانشگاه تخصصی فناوری های نوین آمل

چکیده: ترانسفورماتورهای قدرت از گران‌قیمت‌ترین و حیاتی‌ترین اجزای شبکه‌های انتقال و توزیع انرژی الکتریکی محسوب می شوند و پایش مستمر وضعیت عملکردی آن‌ها به‌منظور جلوگیری از خروج‌های ناگهانی و کاهش هزینه‌های سنگین تعمیرات، از اهمیت بالایی برخوردار است. ارزیابی دقیق وضعیت ناوگان ترانسفورماتورها نیازمند تحلیل جامع و یکپارچه داده‌های حاصل از آزمون‌های تشخیصی شیمیایی، آزمون‌های الکتریکی و همچنین پارامترهای حرارتی و بهره‌برداری است. در این پژوهش، یک سیستم هوشمند و داده‌محور مبتنی بر الگوریتم‌های یادگیری ماشین و یادگیری عمیق، با تمرکز بر الگوریتم پیشنهادی ***GBR***، به‌منظور محاسبه شاخص سلامت ***(Health Index - HI)*** و تخمین سن ظاهری یا امید به زندگی ***(Life Expectancy - LE)*** ترانسفورماتورهای قدرت توسعه داده شده است. مدل پیشنهادی با استفاده از مجموعه‌داده‌های واقعی آموزش داده شده و مراحل پیش‌پردازش، تنظیم پارامترها و ارزیابی عملکرد آن به‌صورت جامع انجام گرفته و در نهایت کارایی روش پیشنهادی در مقایسه با الگوریتم‌های متداول یادگیری ماشین و یادگیری عمیق با استفاده از معیارهای ارزیابی رگرسیون مقایسه شده است. نتایج شبیه سازی نشان می‌دهد که مدل پیشنهادی ***GBR*** به دلیل توانایی بالا در یادگیری روابط غیرخطی، با دستیابی به بیشترین مقدار ضریب تعیین ***(R^2)*** و کمترین میزان خطا عملکرد بهتری نسبت به سایر روش‌ها در تخمین شاخص سلامت و سن ظاهری ترانسفورماتور ارائه می‌دهد. این روش می‌تواند به‌عنوان ابزاری مؤثر در مدیریت بهینه دارایی‌ها، افزایش قابلیت اطمینان شبکه و برنامه‌ریزی تعمیرات پیشگیرانه در صنعت برق مورد استفاده قرار گیرد.

کلید واژه‌ها: *ترانسفورماتورهای قدرت، رگرسیون گرادیان بوستینگ، سن ظاهری، شاخص سلامت، یادگیری عمیق.*

۱. مقدمه

ترانسفورماتورهای قدرت از مهم‌ترین و گران‌قیمت‌ترین تجهیزات شبکه‌های انتقال و توزیع برق هستند و نقش اساسی در تبدیل سطوح ولتاژ و انتقال انرژی الکتریکی ایفا می‌کنند. هزینه بالای خرید، نصب، بهره‌برداری و جایگزینی این تجهیزات، همراه با پیامدهای گسترده خرابی آن‌ها بر عملکرد شبکه و مشترکان، ضرورت پایش وضعیت و مدیریت بهینه دارایی‌های ترانسفورماتور را دوچندان کرده است. اگرچه عمر نامی این تجهیزات معمولاً بین ۲۵ تا ۴۰ سال اعلام می‌شود، شرایط بهره‌برداری، بارگذاری، تنش‌های حرارتی و الکتریکی، عوامل محیطی و کیفیت نگهداری می‌توانند عمر واقعی آن‌ها را به میزان قابل توجهی کاهش دهند [۱-۸].

در این راستا، شاخص سلامت به عنوان ابزاری سریع و کارآمد برای ارزیابی، مقایسه و رتبه‌بندی وضعیت ترانسفورماتورهای یک ناوگان مورد توجه قرار گرفته است. این شاخص با ترکیب داده‌های حاصل از آزمون‌های تشخیصی، اندازه‌گیری‌های بهره‌برداری و سوابق عملکردی، وضعیت کلی تجهیز را در قالب یک امتیاز قابل تفسیر بیان می‌کند و امکان شناسایی ترانسفورماتورهای پرریسک، اولویت‌بندی

اقدامات نگهداری و پیشگیری از خرابی‌های ناگهانی را فراهم می‌سازد. با این حال، دقت و قابلیت اعتماد شاخص سلامت به کیفیت داده‌های ورودی، نحوه پردازش آن‌ها و توانایی مدل در برقراری ارتباط میان پارامترهای اندازه‌گیری‌شده و سازوکارهای خرابی و فرسودگی وابسته است [۹].

پیچیدگی رفتار ترانسفورماتورها، تأثیر هم‌زمان عوامل متعدد بر روند فرسودگی و وجود عدم قطعیت در داده‌های اندازه‌گیری و شرایط بهره‌برداری، از چالش‌های اصلی روش‌های سنتی ارزیابی سلامت به شمار می‌رود. ازاین‌رو، روش‌های مبتنی بر داده و الگوریتم‌های یادگیری ماشین و یادگیری عمیق به دلیل توانایی در استخراج الگوهای پنهان و مدل‌سازی روابط غیرخطی، به طور گسترده در این حوزه به کار گرفته شده‌اند. با وجود این، ارائه یک برآورد نقطه‌ای از شاخص سلامت یا سن ظاهری ترانسفورماتور، بدون در نظر گرفتن عدم قطعیت، ممکن است به تصمیم‌گیری‌های نگهداری و سرمایه‌گذاری کم‌اطمینان منجر شود [۱۰].

بنابراین، توسعه مدل‌های هوشمند و قابل اعتماد که علاوه بر پیش‌بینی شاخص سلامت و سن ظاهری، عدم قطعیت پیش‌بینی‌های خود را نیز کمی‌سازی کنند، از اهمیت ویژه‌ای برخوردار است. چنین مدل‌هایی می‌توانند تصویری واقع‌بینانه‌تر از وضعیت تجهیزات ارائه داده و با پشتیبانی از تصمیم‌های فنی و اقتصادی، به افزایش قابلیت اطمینان شبکه، کاهش هزینه‌های ناشی از خرابی و بهره‌برداری مؤثرتر از دارایی‌های قدرت کمک کنند.

در سال‌های اخیر، مدل‌سازی شاخص سلامت به یکی از روش‌های اصلی ارزیابی وضعیت ترانسفورماتورهای قدرت تبدیل شده است. در این رویکرد، داده‌هایی مانند کیفیت روغن، تحلیل گازهای محلول، وضعیت عایق، شرایط کاغذ عایقی و تنش‌های بهره‌برداری در قالب یک شاخص کلی ترکیب می‌شوند. بدلیل وجود عدم‌قطعیت و ابهام در این داده‌ها، روش‌های فازی و سیستم‌های استنتاج مبتنی بر دانش خبرگان برای تلفیق اطلاعات و ارائه ارزیابی جامع‌تر مورد استفاده قرار گرفته‌اند. همچنین، مدل‌های جدید امکان ترکیب هم‌زمان داده‌های برخط، آزمون‌های دوره‌ای و آمار خرابی را فراهم کرده و زمینه ارتباط شاخص سلامت با مفهوم سن ظاهری را ایجاد کرده‌اند [۱۱].

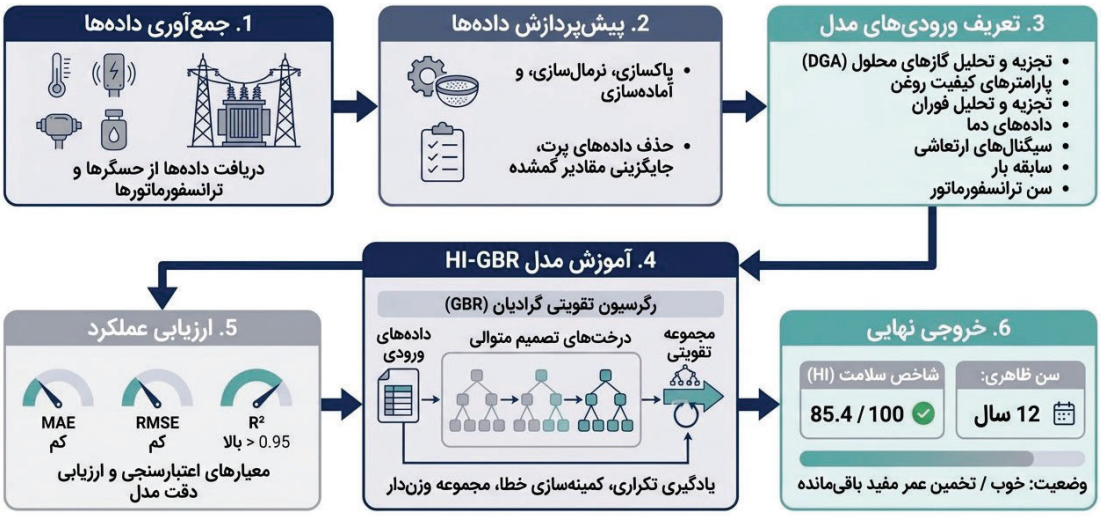
سن ظاهری، برخلاف سن تقویمی، میزان فرسودگی واقعی ترانسفورماتور را نشان می‌دهد و برای رتبه‌بندی تجهیزات و تصمیم‌گیری درباره بازرسی، تعمیر، نوسازی یا جایگزینی آن‌ها اهمیت دارد. در ادامه، مدل‌های سلسله‌مراتبی با هدف سازمان‌دهی عوامل مؤثر بر سلامت و ردیابی تأثیر عیوب موضعی بر وضعیت کلی تجهیز توسعه یافته‌اند. از سوی دیگر، روش‌های یادگیری ماشین با استفاده از داده‌های تاریخی و تشخیصی، توانایی مناسبی در بازتولید شاخص سلامت و پشتیبانی از برآورد سن ظاهری نشان داده‌اند.

در مجموع، روند پژوهش‌ها از روش‌های ساده امتیازدهی ثابت به سمت مدل‌های ترکیبی و هوشمند مبتنی بر منطق فازی، دانش خبرگان، مدل‌های احتمالاتی، ساختارهای سلسله‌مراتبی و یادگیری ماشین حرکت کرده است. هدف این مدل‌ها، ارائه ارزیابی دقیق‌تر و قابل‌تفسیرتر از وضعیت سلامت، امکان رتبه‌بندی تجهیزات و برآورد میزان فرسودگی و عمر واقعی ترانسفورماتورها است[۱۲].

۲. روش پیشنهادی

روش پیشنهادی Gradient Boosting Regression (GBR) یک چارچوب داده‌محور دو مرحله‌ای برای ارزیابی وضعیت ترانسفورماتورهای قدرت است. در این چارچوب، ابتدا مجموعه‌ای از داده‌های تشخیصی، آزمون‌های وضعیت و متغیرهای بهره‌برداری به‌عنوان ورودی مدل در نظر گرفته می‌شوند. سپس با استفاده از این الگوریتم، مقدار شاخص سلامت و سن ظاهری هر ترانسفورماتور برآورد می‌شود تا شدت فرسودگی تجهیز در قالب یک معیار قابل‌فهم‌تر و مناسب‌تر برای تصمیم‌گیری مدیریتی بیان گردد.

منطق اصلی این روش بر این فرض استوار است که وضعیت واقعی ترانسفورماتور تابعی پیچیده و غیرخطی از متغیرهای مختلف شیمیایی، الکتریکی، حرارتی و بهره‌برداری است. ازاین‌رو، استفاده از یک مدل یادگیری ماشین که توانایی استخراج این روابط را داشته باشد، می‌تواند نسبت به روش‌های مبتنی بر امتیازدهی ثابت عملکرد دقیق‌تری ارائه کند. به‌طور کلی، مراحل اجرای روش GBR در شکل ۱ نشان داده شده است.



[1] شکل ۱. الگوریتم چارتی روش پیشنهادی

۱،۲. ساختار الگوریتم رگرسیون گرادیان بوستینگ

الگوریتم رگرسیون گرادیان بوستینگ یکی از روش‌های قدرتمند یادگیری تجمعی در مسائل رگرسیونی است که به‌ویژه برای داده‌هایی با روابط پیچیده و غیرخطی کاربرد فراوانی دارد. در این الگوریتم، به‌جای استفاده از یک مدل منفرد، مجموعه‌ای از درخت‌های تصمیم رگرسیونی به‌صورت ترتیبی ساخته می‌شوند. هر درخت جدید به‌گونه‌ای آموزش داده می‌شود که خطاهای باقی‌مانده از درخت‌های قبلی را کاهش دهد.

فرض شود مجموعه داده آموزشی به‌صورت $\{(x_i, y_i)\}_{i=1}^{N}$ در اختیار باشد که در آن x_i بردار ویژگی‌های ورودی و y_i مقدار شاخص سلامت مرجع برای نمونه iام است. هدف الگوریتم، تقریب تابعی است که بتواند نگاشت بین ورودی‌ها و خروجی را با حداقل خطا مدل کند. مدل نهایی در روش گرادیان بوستینگ به‌صورت زیر تعریف می‌شود:

$$F_M(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \quad (۱)$$

که در آن:

$h_m(x)$ درخت تصمیم ساخته‌شده در مرحله mام است؛

γ_m وزن متناظر با درخت مرحله mام است؛

M تعداد کل درخت‌های مورد استفاده در مدل نهایی است.

در هر مرحله، الگوریتم سعی می‌کند با دنبال کردن گرادیان تابع زیان، مدلی جدید بیافزاید که خطای مدل قبلی را اصلاح کند. در مسائل رگرسیونی، تابع زیان معمولاً بر پایه خطای مربعی تعریف می‌شود. توانایی این الگوریتم در مدل‌سازی روابط غیرخطی، اندرکنش متغیرها و الگوهای پیچیده، آن را به گزینه‌ای مناسب برای برآورد شاخص سلامت ترانسفورماتور تبدیل کرده است.

۲،۲- آموزش مدل GBR

برای ارزیابی صحیح عملکرد مدل و جلوگیری از بیش‌برازش، داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمون تقسیم می‌شوند. مجموعه آموزش برای یادگیری مدل، مجموعه اعتبارسنجی برای تنظیم پارامترها و مجموعه آزمون برای ارزیابی نهایی عملکرد مدل مورد استفاده قرار می‌گیرد. در صورتی که حجم داده محدود باشد، می‌توان از روش اعتبارسنجی متقاطع نیز بهره گرفت.

عملکرد الگوریتم GBR به تنظیم مناسب پارامترهای آن وابسته است. مهم‌ترین پارامترهای مورد استفاده در این پژوهش شامل این موارد هستند: تعداد درخت‌ها[1] نرخ یادگیری[2] عمق بیشینه درخت‌ها[3] حداقل تعداد نمونه لازم برای تقسیم یک گره، حداقل تعداد نمونه در برگ‌ها و تابع هزینه. تنظیم این پارامترها با هدف دستیابی به توازنی مناسب میان دقت پیش‌بینی و توان تعمیم‌پذیری مدل انجام می‌شود. انتخاب نامناسب پارامترها می‌تواند منجر به بیش‌برازش یا کم‌برازش شود. پس از نهایی‌سازی پارامترها، مدل بر روی داده‌های آموزشی اجرا می‌شود. در هر مرحله، یک درخت جدید برای پیش‌بینی خطاهای باقی‌مانده ساخته شده و به مدل افزوده می‌شود. این روند تا زمان رسیدن به معیار توقف یا تکمیل تعداد از پیش تعیین‌شده درخت‌ها ادامه می‌یابد. خروجی نهایی مدل، مقدار پیش‌بینی‌شده شاخص سلامت برای هر ترانسفورماتور است.

روش GBR در مقایسه با روش‌های سنتی مبتنی بر امتیازدهی ثابت، دارای مزایای مهمی است. نخست آن‌که این روش قادر است روابط غیرخطی و پیچیده میان متغیرهای تشخیصی و وضعیت کلی تجهیز را به‌خوبی مدل‌سازی کند. دوم آن‌که وابستگی آن به وزن‌دهی ذهنی و قواعد از پیش تعریف‌شده کمتر است و در نتیجه با ساختار واقعی داده‌ها سازگاری بیشتری دارد. سوم آن‌که امکان تحلیل اهمیت نسبی ویژگی‌ها را فراهم می‌سازد و از این نظر در شناسایی عوامل مؤثر بر سلامت تجهیز مفید است.

با وجود این مزایا، این روش با محدودیت‌هایی نیز همراه است. مهم‌ترین محدودیت آن وابستگی به کیفیت داده‌های ورودی و شاخص سلامت مرجع است. چنانچه داده‌های آموزشی دارای سوگیری، خطا یا کمبود نمایندگی باشند، عملکرد مدل نیز تحت تأثیر قرار خواهد گرفت. افزون بر این، الگوریتم GBR به‌طور ذاتی یک پیش‌بینی نقطه‌ای ارائه می‌دهد و عدم‌قطعیت پیش‌بینی را مستقیماً مدل نمی‌کند. همچنین اگرچه این روش نسبت به برخی مدل‌های بسیار پیچیده شفاف‌تر است، اما همچنان از منظر تفسیرپذیری به سادگی مدل‌های کاملاً قاعده‌محور نیست.

۳. مجموعه‌های داده

برای ارزیابی عملکرد، روش پیشنهادی بر روی دو مجموعه داده واقعی اعمال گردید. داده‌های ورودی از میان شاخص‌هایی انتخاب می‌شوند که در ادبیات موضوع به‌عنوان مهم‌ترین نشانگرهای وضعیت عایقی، شیمیایی و بهره‌برداری ترانسفورماتور شناخته شده‌اند. انتخاب این متغیرها با هدف پوشش ابعاد اصلی فرسودگی تجهیز انجام شده است تا مدل بتواند تصویر واقع‌بینانه‌تری از سلامت کلی ترانسفورماتور ارائه دهد.

۱،۳- مجموعه داده اول

در این پژوهش، یکی از مجموعه‌داده‌های اصلی مورد استفاده، مجموعه‌داده Sample Power Transformers Health Condition Dataset است که از پایگاه Kaggle دریافت شده است [۱۳]. این مجموعه‌داده شامل نمونه‌هایی از نتایج آزمون‌ها و شاخص‌های ارزیابی وضعیت ترانسفورماتورهای قدرت بوده و برای مدل‌سازی رابطه میان پارامترهای تشخیصی و وضعیت سلامت تجهیز به‌کار گرفته شده است. ساختار این مجموعه‌داده به‌گونه‌ای است که برای هر نمونه، مجموعه‌ای از ویژگی‌های مرتبط با گازهای محلول در روغن، ویژگی‌های فیزیکی و الکتریکی روغن عایقی، و نیز شاخص‌های سلامت و عمر تجهیز ثبت شده است.

مجموعه‌داده مذکور شامل ۴۷۰ سطر (نمونه) و ۱۶ ستون است که از این میان، ۱۴ ستون نخست به‌عنوان ویژگی‌های ورودی و ۲ ستون پایانی به‌عنوان خروجی‌های مرتبط با ارزیابی وضعیت تجهیز در نظر گرفته می‌شوند. همچنین ۸۰ درصد داده‌ها (۳۷۶ نمونه) بعنوان داده‌های آموزش و اعتبارسنجی[4] و ۲۰ درصد باقیمانده (۹۴ نمونه) بعنوان داده آزمون[5] در نظر گرفته شده است. ستون‌های ورودی عبارت‌اند از پارامترهای مربوط به گازهای محلول در روغن ترانسفورماتور، شاخص‌های شیمیایی و آلودگی روغن و شاخص‌های

[1] n_estimators
[2] learning_rate
[3] max_depth
[4] Validation
[5] Test

الکتریکی و فیزیکی روغن عایقی. دو ستون خروجی نیز عبارتند از دو شاخص اندیس سلامت Health Index (HI) و عمر تخمینی Life Expectation (LE) ترانسفورماتور. HI بیانگر وضعیت کلی سلامت ترانسفورماتور است و می‌تواند به‌عنوان یک شاخص ترکیبی برای بیان سطح سلامت تجهیز استفاده شود. LE نیز بیانگر تخمین عمر مورد انتظار یا عمر باقیمانده تجهیز است و از منظر برنامه‌ریزی نگهداری، تعمیرات و تصمیم‌گیری بهره‌برداری اهمیت دارد.

۲,۳- مجموعه داده دوم

مجموعه داده دوم مورد استفاده در این پژوهش از پایگاه Mendeley Data دریافت شده است. این مجموعه‌داده در سال ۲۰۲۲ توسط Diego Zaldivar Sanchez منتشر شده و با عنوان A comprehensive methodology for the optimization of condition-based maintenance in power transformer fleets در دسترس قرار گرفته است. این منبع داده در چارچوب یک روش جامع برای بهینه‌سازی نگهداری مبتنی بر وضعیت در ناوگان ترانسفورماتورهای قدرت توسعه یافته است [۱۴].

فایل داده‌ها شامل اطلاعات مربوط به ۵۸ واحد ترانسفورماتور قدرت است و برای محاسبه و تحلیل شاخص سلامت به‌کار گرفته شده است. بررسی ساختار این فایل نشان می‌دهد که داده‌های ثبت‌شده برای هر ترانسفورماتور، مجموعه‌ای از شاخص‌های مرتبط با کیفیت روغن عایقی، وضعیت عایق کاغذی و ارزیابی سلامت تجهیز را در بر می‌گیرد. همانند مجموعه داده نخست، ۸۰ درصد داده‌ها (۴۷ نمونه) بعنوان داده‌های آموزش و اعتبارسنجی و ۲۰ درصد باقیمانده (۱۱ نمونه) بعنوان داده آزمون در نظر گرفته شده است. ساختار این مجموعه‌داده شامل ۱۳ ستون ورودی و یک ستون (متغیر) خروجی است. مجموعه‌داده دوم نسبت به مجموعه‌داده اول، تمرکز بیشتری بر ارزیابی شرایط عایق مایع و جامد دارد و به‌ویژه از منظر تحلیل پیرشدگی کاغذ عایقی و ارتباط آن با سلامت کلی ترانسفورماتور قابل توجه است.

۴. معیارهای ارزیابی عملکرد مدل

برای ارزیابی دقت و کارایی مدل‌های پیش‌بینی، از چندین شاخص آماری استفاده می‌شود که هر یک جنبه‌ای متفاوت از خطای پیش‌بینی را اندازه‌گیری می‌کنند. فرض شود مقدار واقعی برابر با y_i، مقدار پیش‌بینی‌شده برابر با $\hat{y}_i$، و تعداد نمونه‌ها برابر با nباشد.

- میانگین قدرمطلق خطا (MAE)

شاخص میانگین قدرمطلق خطا، متوسط قدرمطلق اختلاف بین مقادیر واقعی و مقادیر پیش‌بینی‌شده را نشان می‌دهد و یکی از ساده‌ترین و قابل‌فهم‌ترین معیارهای ارزیابی خطاست.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i | \quad (۲)$$

این شاخص میانگین فاصله مطلق بین مقادیر واقعی و مقادیر پیش‌بینی‌شده را نشان می‌دهد. هرچه مقدار MAE کوچک‌تر باشد، دقت مدل بیشتر است.

- میانه اندازه خطای نسبی (MDMRE)

معیار MDMRE معمولاً به صورت Median Magnitude of Relative Error برای سنجش خطای نسبی به‌ویژه در مطالعات تخمین و پیش‌بینی کاربرد دارد. تعریف رایج آن بصورت میانه اندازه خطای نسبی است. ابتدا اندازه خطای نسبی برای هر نمونه محاسبه می‌شود:

$$MRE_i = | \frac{y_i - \hat{y}_i}{y_i} | \quad (۳)$$

سپس مقدار MDMRE به صورت زیر به دست می‌آید:

$$MDMRE = \text{median}(MRE_i) \quad (۴)$$

این معیار نسبت به داده‌های پرت حساسیت کمتری نسبت به میانگین دارد.

- ریشه میانگین مربعات خطا (RMSE)

شاخص ریشه میانگین مربعات خطا، مجذور خطاها را میانگین گرفته و سپس ریشه دوم آن را محاسبه می‌کند. این معیار نسبت به خطاهای بزرگ حساس‌تر است و در سنجش پایداری و دقت کلی مدل اهمیت دارد. مقادیر کمتر RMSE نشان‌دهنده عملکرد بهتر مدل است.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (۵)$$

- ضریب تعیین ($\boldsymbol{R^2}$)

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (۶)$$

این شاخص بیان می‌کند که چه سهمی از تغییرات شاخص سلامت توسط مدل توضیح داده می‌شود. هرچه مقدار R^2به یک نزدیک‌تر باشد، عملکرد مدل مطلوب‌تر است. علاوه بر معیارهای فوق، برای تحلیل عمیق‌تر می‌توان از نمودار مقایسه مقادیر واقعی و پیش‌بینی‌شده، نمودار باقیمانده‌ها و تحلیل اهمیت ویژگی‌ها نیز استفاده کرد.

- میانگین درصد قدرمطلق خطا (□□□□)

شاخص میانگین درصد قدرمطلق خطا، خطای پیش‌بینی را به صورت درصدی نسبت به مقدار واقعی بیان می‌کند و برای مقایسه خطا در مقیاس‌های مختلف مفید است.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (۷)$$

هرچه MAPE کمتر باشد، مدل از دقت بیشتری برخوردار است. لازم به ذکر است که این معیار در حالتی که $y_i = 0$باشد، قابل استفاده نیست یا نیازمند اصلاح است.

- میانگین قدرمطلق خطای مقیاس‌شده (□□□□)

شاخص میانگین قدرمطلق خطای مقیاس‌شده، MAE مدل را نسبت به خطای یک روش مبنا (Naive Forecast) مقیاس‌بندی می‌کند. این معیار برای مقایسه عملکرد مدل‌ها در مجموعه‌داده‌های مختلف بسیار مناسب است.

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\frac{1}{n-m}\sum_{t=m+1}^{n}|y_t - y_{t-m}|} \quad (۸)$$

که در آن، mنشان‌دهنده فاصله زمانی یا دوره فصلی است. در داده‌های غیرفصلی معمولاً $m = 1$ در نظر گرفته می‌شود. هرچه مقدار $MASE < 1$ باشد، مدل عملکرد بهتری دارد.

- شاخص میانگین (□□)

شاخص MF یا Mean Factor میانگین نسبت مقادیر پیش‌بینی‌شده به مقادیر واقعی را محاسبه می‌کند. این معیار میزان تمایل کلی مدل به بیش‌برآوردی یا کم‌برآوردی را نشان می‌دهد. مقدار $MF = 1$ بیانگر آن است که پیش‌بینی‌های مدل به طور متوسط با مقادیر واقعی منطبق هستند. اگر $MF > 1$ باشد، مدل به طور کلی تمایل به بیش‌برآوردی دارد، در حالی که $MF < 1$ نشان‌دهنده تمایل مدل به کم‌برآوردی است.

$$MF = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_i}{y_i} \quad (۹)$$

۵. نتایج شبیه‌سازی

برای آن‌که نتایج ارائه‌شده از اعتبار علمی و تحلیلی کافی برخوردار باشند، عملکرد روش پیشنهادی تنها به‌صورت مستقل گزارش نمی‌شود، بلکه با چهار الگوریتم دیگر نیز مقایسه خواهد شد. این الگوریتم‌ها به‌گونه‌ای انتخاب شده‌اند که هم شامل روش‌های کلاسیک و شناخته‌شده یادگیری ماشین باشند و هم یک روش مبتنی بر یادگیری عمیق را در بر گیرند تا امکان تحلیل جامع‌تری از نقاط قوت و ضعف مدل پیشنهادی فراهم شود. الگوریتم‌های انتخاب‌شده برای مقایسه عبارت‌اند از:

- رگرسیون بردار پشتیبان (□□□)

الگوریتم رگرسیون بردار پشتیبان (SVR) به عنوان یکی از روش‌های شناخته‌شده و پرکاربرد در مسائل رگرسیونی، در این پژوهش به عنوان یکی از مدل‌های مبنا برای مقایسه با روش پیشنهادی مورد استفاده قرار گرفته است. این الگوریتم در واقع نسخه رگرسیونی ماشین بردار پشتیبان است و هدف آن یافتن تابعی است که بتواند با حفظ ساختار ساده و قابلیت تعمیم مناسب، رابطه بین متغیرهای ورودی و خروجی را با کمترین خطای ممکن مدل‌سازی کند. برخلاف بسیاری از روش‌های رگرسیونی که مستقیماً به دنبال کمینه‌سازی مجموع خطاها هستند، SVR بر این مبنا عمل می‌کند که خطاهای کوچک در یک بازه مشخص قابل اغماض باشند و تنها انحراف‌هایی که از این بازه فراتر می‌روند در تابع هزینه وارد شوند. این ویژگی سبب می‌شود که مدل در برابر نوسانات جزئی و نویزهای محدود از پایداری بیشتری برخوردار باشد.

- جنگل تصادفی

الگوریتم جنگل تصادفی نیز در این پژوهش به عنوان یکی از روش‌های مقایسه‌ای اصلی برای ارزیابی عملکرد مدل پیشنهادی مورد استفاده قرار گرفته است. این الگوریتم یکی از روش‌های یادگیری جمعی است که بر مبنای ترکیب تعداد زیادی درخت تصمیم عمل می‌کند و تلاش دارد با تجمیع خروجی چندین مدل ساده‌تر، دقت پیش‌بینی و پایداری نهایی مدل را افزایش دهد. ایده اصلی در جنگل تصادفی آن است که به جای اتکا به یک درخت تصمیم منفرد، مجموعه‌ای از درخت‌ها بر روی نمونه‌ها و زیرمجموعه‌های متفاوتی از ویژگی‌ها آموزش داده شوند و سپس خروجی نهایی از طریق تجمیع نتایج این درخت‌ها به دست آید. این سازوکار موجب می‌شود که مدل نسبت به نوسانات داده و بیش‌برازش حساسیت کمتری داشته باشد و در عین حال بتواند روابط غیرخطی میان متغیرهای ورودی و خروجی را نیز به خوبی یاد بگیرد.

- شبکه عصبی پرسپترون چندلایه (□□□)

الگوریتم پرسپترون چندلایه یا MLP یکی دیگر از روش‌های مقایسه‌ای مورد استفاده در این پژوهش است که در رده شبکه‌های عصبی مصنوعی پیش‌خور قرار می‌گیرد. این مدل به دلیل توانایی در تقریب توابع غیرخطی، در بسیاری از مسائل رگرسیونی و طبقه‌بندی کاربرد گسترده‌ای یافته است. ایده اصلی در MLP آن است که با استفاده از چند لایه پردازشی متوالی، رابطه‌ای میان متغیرهای ورودی و خروجی ایجاد شود که بتواند الگوهای پنهان موجود در داده‌ها را استخراج کند. برخلاف مدل‌های خطی که معمولاً به یک تابع نگاشت ساده محدود می‌شوند، MLP قادر است برهم‌کنش‌های پیچیده‌تر میان متغیرها را بیاموزد و از این رو در مسائل مهندسی با روابط غیرخطی، گزینه‌ای قابل توجه به شمار می‌رود.

- مدل □□□□

الگوریتم LSTM یا حافظه بلندمدت-کوتاه‌مدت نیز در این پژوهش به عنوان یکی از روش‌های مقایسه‌ای مورد استفاده قرار گرفته است. این الگوریتم نوعی شبکه عصبی بازگشتی است که با هدف رفع محدودیت شبکه‌های بازگشتی متداول در یادگیری وابستگی‌های بلندمدت طراحی شده است. در مسائل مهندسی، به‌ویژه در شرایطی که داده‌ها دارای الگوهای وابسته، روندهای تدریجی یا ارتباطات غیرخطی پیچیده باشند، LSTM می‌تواند ابزار مناسبی برای مدل‌سازی رفتار سیستم باشد. اگرچه این مدل بیشتر در داده‌های زمانی و ترتیبی شناخته می‌شود، در برخی مسائل پیش‌بینی نیز به عنوان یک روش مقایسه‌ای قدرتمند به کار می‌رود تا توانایی مدل پیشنهادی در برابر ساختارهای عمیق‌تر ارزیابی شود.

در جدول ۱، عملکرد چهار الگوریتم مطرح شده و روش پیشنهادی مبتنی بر GBR بر روی مجموعه‌داده‌های اول و دوم مورد ارزیابی و مقایسه قرار گرفته است. هدف از این مقایسه، بررسی توانایی هر الگوریتم در تخمین شاخص سلامت ترانسفورماتور و سن ظاهری یا عمر مورد انتظار تجهیز بر اساس مجموعه‌ای از داده‌های وضعیت، روغن و گازهای محلول است.

جدول ۱. نتایج شبیه سازی

شاخص	SVR		Random Forest		MLP		LSTM		الگوریتم پیشنهادی GBR	
	مجموعه ۱ HI LE	مجموعه ۲ HI	مجموعه ۱ HI LE	مجموعه ۲ HI	مجموعه ۱ HI LE	مجموعه ۲ HI	مجموعه ۱ HI LE	مجموعه ۲ HI	مجموعه ۱ HI LE	مجموعه ۲ HI
MAE	۱۵.۶۶ ۶.۴۵	۰,۰۷	۶,۰۲ ۴,۹۹	۰,۰۶۳	۹,۵۴ ۷,۱۳	۰,۱۲	۷,۷۷ ۶,۴۸	۰,۰۵۴	۵,۵۱ ۴,۹۸	۰,۰۳
MDMRE	۱.۰۶ ۰.۰۶۶	۰,۱۷	۰,۱۴ ۰,۰۷	۰,۱۵	۰,۲۹ ۰,۰۹	۰,۲۳	۰,۱۹ ۰,۱۰	۰,۰۹۶	۰,۱۵ ۰,۰۸	۰,۰۵
RMSE	۱۸.۰۱ ۱۰.۲۷	۰,۰۹	۹,۶۷ ۸,۶۰	۰,۰۷	۱۲,۷۰ ۱۰,۶۴	۰,۱۸	۱۱,۶۶ ۹,۶۱	۰,۰۷	۸,۹۷ ۸,۵۲	۰,۰۴
R2	-۰.۰۱۲۱ ۰.۶۶	۰,۸۳	۰,۷۱ ۰,۷۶	۰,۹۰	۰,۵۰ ۰,۶۴	۰,۳۵	۰,۵۸ ۰,۷۰	۰,۹۰	۰,۷۵ ۰,۷۷	۰,۹۷
MAPE (%)	۷۹.۸۰ ۴۴.۴۱	۲۳,۹۶	۲۴,۶۴ ۳۷,۵۳	۱۷,۴۱	۴۳,۷۶ ۵۲,۵۵	۳۱,۲۰	۳۳,۴۴ ۴۶,۱۳	۱۳,۵۶	۲۲,۴۴ ۳۳,۰۱	۸,۶۰
MASE	۸۱.۴۳ ۰.۸۱	۰,۳۷	۳۱,۳۱ ۰,۶۳	۰,۳۴	۴۹,۶۴ ۰,۹۰	۰,۶۳	۴۰,۴۳ ۰,۸۱	۰,۲۹	۲۸,۶۳ ۰,۶۲۵	۰,۱۶
MF	۱.۵۵ ۱.۲۷	۱,۱	۱,۰۹ ۱,۲۶	۱,۰۲	۱,۲۴ ۱,۴۱	۰,۹۹	۱,۱۲ ۱,۳۱	۱,۰۵	۱,۰۷ ۱,۲۰	۰,۹۵
Time (sec)	۳.۳۳ ۳.۴۱	۴,۰۱	۰,۶۱ ۰,۵۰	۱۰۲,۴۹	۳,۹۱ ۴,۲۰	۰,۱	۱۰,۸۲ ۱۰,۱۸	۱,۷	۸۵,۰۵ ۱۲۳,۹۱	۱۴۴.۴۶

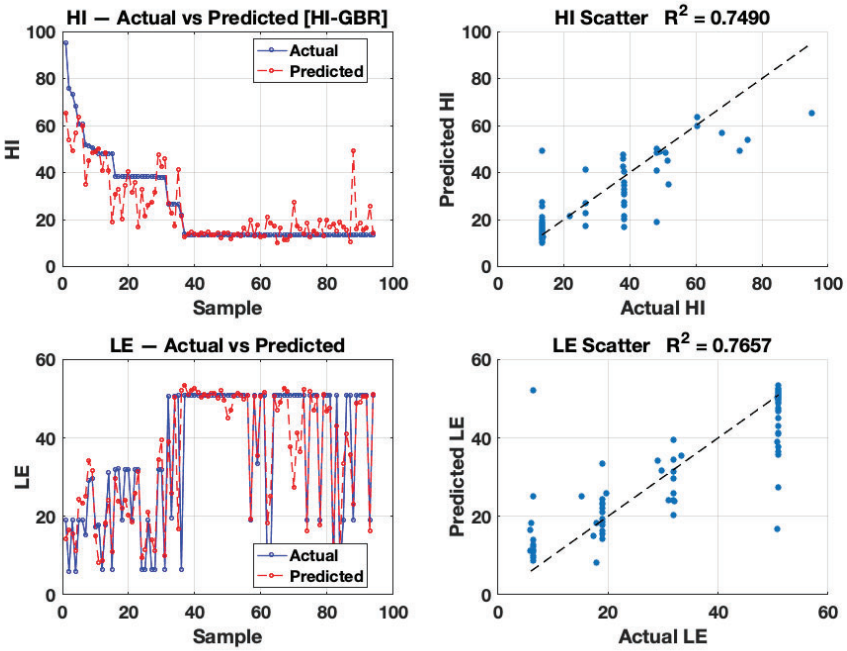
برآیند نتایج به‌دست‌آمده از مجموعه‌داده اول نشان می‌دهد که روش پیشنهادی □□□ در هر دو مسئله تخمین شاخص سلامت و سن ظاهری، بهترین عملکرد را در میان پنج الگوریتم مورد بررسی ارائه کرده است. این روش در اکثر معیارهای کلیدی شامل □□□ ، □□□□ ، $□^2$، □□□□ و □□ برتر از سایر مدل‌ها بوده و توانسته است تخمین دقیق‌تر، پایدارتر و قابل‌اعتمادتری از وضعیت ترانسفورماتور

به دست دهد. پس از روش پیشنهادی، جنگل تصادفی به‌عنوان نزدیک‌ترین رقیب ظاهر شده است و نتایج آن نشان می‌دهد که مدل‌های درختی برای داده‌های این حوزه از کارایی بالایی برخوردارند. در مقابل، □□□□ با وجود قدرت تئوریک بالا، تنها عملکردی متوسط ارائه کرده و □□□ و به‌ویژه □□□ نتوانسته‌اند به سطح دقت دو روش برتر برسند.

در مجموع، بر اساس نتایج تخمین شاخص سلامت (□□) و سن ظاهری (□□) ترانسفورماتور، رتبه‌بندی کلی الگوریتم‌ها را می‌توان به‌صورت زیر در نظر گرفت:

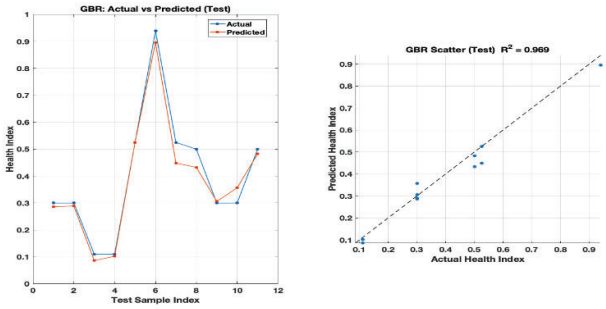
□□□ ◻ □□□ ◻ □□□□ ◻ □□□□□□ □□□□□□ ◻ □□□

نمودارهای بدست آمده از اعمال الگوریتم □□□ بر روی داده‌های تست مجموعه داده نخست در شکل ۲ نشان داده شده است.



شکل ۲. نتایج حاصل از اعمال الگوریتم پیشنهادی بر روی داده‌های تست مجموعه اول [2]

در مجموعه‌داده دوم، برآیند نتایج نشان می‌دهد که روش پیشنهادی □□□ باز هم بهترین عملکرد را در تخمین شاخص سلامت ترانسفورماتور ارائه کرده است. پس از روش پیشنهادی، □□□□ بهترین عملکرد را داشته و سپس جنگل تصادفی، □□□ و □□□ در رتبه‌های بعدی قرار گرفته‌اند. نتایج نشان می‌دهد که مدل‌های توانمند در یادگیری روابط غیرخطی، به‌ویژه مدل‌های مبتنی بر تقویت گرادیانی، برای تخمین شاخص سلامت ترانسفورماتور مناسب‌تر هستند. همچنین ضعف □□□ در داده‌های آزمون نشان می‌دهد که استفاده از شبکه‌های عصبی ساده در مجموعه‌داده‌های کوچک، بدون راهکارهای کنترلی کافی، می‌تواند منجر به بیش‌برازش و کاهش قابلیت تعمیم شود. نمودارهای بدست آمده از اعمال GBR بر روی داده‌های تست مجموعه داده دوم در شکل ۳ نشان داده شده است.



شکل ۳. نتایج حاصل از اعمال الگوریتم پیشنهادی بر روی داده‌های تست مجموعه دوم [3]

در مجموع، نتایج مجموعه‌داده دوم تأیید می‌کند که روش پیشنهادی □□□ قادر است با بهره‌گیری از ساختار تقویت گرادیانی و بهینه‌سازی فراپارامترها، تخمین دقیق‌تر و پایدارتری از شاخص سلامت ترانسفورماتور ارائه دهد. این یافته، در کنار نتایج مجموعه‌داده اول، نشان‌دهنده قابلیت اعتماد و کارایی روش پیشنهادی برای کاربردهای پایش وضعیت، برنامه‌ریزی تعمیرات و مدیریت دارایی ترانسفورماتورهای قدرت است.

۶. نتیجه‌گیری

در این پژوهش، یک چارچوب مبتنی بر ion□□□□□□□□ □□□□□□□□ □□□□□□□□ □□□□□□□□ برای تخمین شاخص سلامت و سن ظاهری ترانسفورماتورهای قدرت ارائه شد. هدف اصلی این تحقیق، توسعه مدلی بود که بتواند با بهره‌گیری از داده‌های وضعیت ترانسفورماتور، رابطه میان ویژگی‌های مؤثر بر سلامت تجهیز و خروجی‌های موردنظر را با دقت مناسب شناسایی کند. برای این منظور، دو مجموعه‌داده مستقل مورد استفاده قرار گرفت که یکی شامل شاخص‌های مبتنی بر آزمون‌های روغن و گازهای محلول در روغن بود و دیگری بر داده‌های مرتبط با محصولات فرعی تخریب عایق و شاخص سلامت متمرکز بود. استفاده از این دو مجموعه‌داده امکان ارزیابی مدل پیشنهادی را در دو سناریوی متفاوت فراهم ساخت و موجب شد عملکرد آن از جنبه‌های مختلف بررسی شود.

نتایج به‌دست‌آمده نشان داد که مدل پیشنهادی □□□ در مقایسه با روش‌های مبنا شامل □□□، □□□□□□ □□□□□□□، □□□ و □□□□ عملکرد برتری داشته است. این برتری هم در معیارهای دقت نظیر □□□، □□□□، □□□□، $□^2$ و هم در شاخص‌های خطای نسبی و پایداری پیش‌بینی مشاهده شد. در هر دو مجموعه‌داده، مدل پیشنهادی توانست خطای پیش‌بینی را کاهش دهد و هم‌زمان مقدار ضریب تعیین را در سطح بالاتری حفظ کند. این موضوع نشان می‌دهد که الگوریتم پیشنهادی قادر است روابط غیرخطی و پیچیده میان ورودی‌ها و شاخص سلامت ترانسفورماتور را به‌خوبی مدل‌سازی کند.

منابع

[1] Aeggegn, D. B., Akumu, A. O., & Nnachi, A. F., “AI-driven predictive maintenance for enhanced reliability of power transformers: Systematic review”. Next Energy, 13, 100822, 2026.

[2] M. Badawi, S. A. Ibrahim, D. A. Mansour, A. A. El-Faraskoury, S. A. Ward, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, ‘‘Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques,’’ IEEE Access, vol. 10, pp. 25954–25972, 2022.

[3] آرمان آذرخش، سید محمدتقی بطحایی، محاسبه تلفات ترانسفورماتور قدرت با استفاده از روش اجزای محدود، مجله کهربا، شماره ۲٤، تابستان ۹۸، ص ۳۸-۳۱.

[4] R. A. Prasojo, H. Gumilang, Suwarno, N. U. Maulidevi, and B. A. Soedjarno, ‘‘A fuzzy logic model for power transformer faults’ severity determination based on gas level, gas rate, and dissolved gas analysis interpretation,’’ Energies, vol. 13, no. 4, p. 1009, Feb. 2020.

[5] S. Padmanaban, M. Khalili, M. A. Nasab, M. Zand, A. G. Shamim, and B. Khan, ‘‘Determination of power transformers health index using parameters affecting the transformer’s life,’’ IETE J. Res., vol. 69, no. 11, pp. 8467–8488, Nov. 2023.

[6] Bazi, S., Nhaila, H., & El Khaili, M., “Data-driven machine learning for enhancing predictive maintenance of distribution transformers”, E-Prime – Nexus of Electrical, Electronic, and Intelligent Engineering, 17, 201164, 2026.

[7] D. Paul and A. K. Goswami, ‘‘Determination of health index of insulating oil of in-service transformer & reactors based on IFT predicted by multigene symbolic regression,’’ IEEE Trans. Ind. Appl., vol. 58, no. 5, pp. 5935–5943, Sep. 2022.

[8] Syamsiana, I. N., Febriani, N. A., Sumari, A. D. W., & Sutjipto, R., “Revolutionizing predictive maintenance: Remaining useful life forecasting based on cognitive artificial intelligence using knowledge growing system”, Ain Shams Engineering Journal, 17(1), 103880, 2026.

[9] Zhang, S., Cheng, W., Wang, W., Ahmad, H., Zhang, L., Liu, H., Nie, Z., & Chen, X., “Deep learning-based health monitoring and prognostics for nuclear power plants: a systematic review”, Applied Energy, 421, 128252, 2026.

[10] I. B. M. Taha, ‘‘Power transformers health index enhancement based on convolutional neural network after applying imbalanced-data oversampling,’’ Electronics, vol. 12, no. 11, p. 2405, May 2023.

[11] H. Zeinoddini-Meymand, S. Kamel, and B. Khan, ‘‘An efficient approach with application of linear and nonlinear models for evaluation of power transformer health index,’’ IEEE Access, vol. 9, pp. 150172–150186, 2021.

[12] A. M. Abdullah, R. Ali, S. B. Yaacob, K. Ananda-Rao, and N. A. Uloom, ‘‘Transformer health index by prediction artificial neural networks diagnostic techniques,’’ J. Phys., Conf. Ser., vol. 2312, no. 1, Aug. 2022, Art. no. 012002.

[13] https://www.kaggle.com/datasets/shashwatwork/failure-analysis-in-power-transformers-dataset

[14] https://data.mendeley.com/datasets/hcyg643sym/1