

مدلسازی بهینه نرخ تهویه در ساختمان‌های مجهز به سرمایش غیر فعال با استفاده از یادگیری ماشین انباشته و الگوریتم ژنتیک: یک رویکرد نوین برای داده‌های محدود و نامتوازن

دانشجوی کارشناسی ارشد، گروه مهندسی سیستم‌های انرژی، دانشکده مهندسی مکانیک، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران. m.mohsenpour@email.kntu.ac.ir

مجید محسن پور

استادیار، گروه پژوهشی انرژی‌های تجدیدپذیر، پژوهشگاه نیرو، تهران، ایران. msalimi@nri.ac.ir

محسن سلیمی

استاد تمام، گروه مهندسی سیستم‌های انرژی، دانشکده مهندسی مکانیک، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران. amidpour@kntu.ac.ir

مجید عمیدپور

چکیده

با افزایش تقاضای سیستم‌های تهویه به دلیل تغییرات اقلیمی و گازهای گلخانه‌ای، مصرف انرژی رشد چشم‌گیری یافته است. توسعه سیستم‌های سرمایش غیر فعال به‌عنوان راهکاری ضروری مطرح می‌شود زیرا به تهویه مطبوع وابسته نیستند. مدل‌سازی نرخ تهویه در این سیستم‌ها به‌صورت مرسوم زمان‌بر است؛ لذا این پژوهش با ارائه مدل مبتنی بر داده، به مدل‌سازی نرخ تهویه در ساختمان‌های دارای سرمایش غیر فعال پرداخته و این مدل را از سایر مطالعات متمایز می‌سازد. از ده مدل یادگیری ماشین به‌عنوان مدل پایه در انباشته‌سازی گروهی استاندارد استفاده شد که نتایج در پارامترهای ارزیابی RMSE و MAE رضایت‌بخش نبود. برای بهبود مدل، از الگوریتم ژنتیک بهره گرفته شد. ارزیابی دقت مدل بهبود یافته با چهار مجموعه داده مختلف انجام شد و ویژگی‌های کلیدی از طریق روش حذف ویژگی‌های تکراری انتخاب شدند. نتایج نشان می‌دهند که مدل پیشنهادی با دقت بالاتر نسبت به الگوریتم‌های کلاسیک، در پیش‌بینی نرخ تهویه به‌ویژه در شرایط داده محدود و توزیع غیر یکنواخت، قابل استفاده در مناطق جغرافیایی متنوع است و به بررسی سریع نرخ تهویه در ساختمان‌ها کمک می‌کند. **واژه‌های کلیدی:** سرمایش غیر فعال، تغییرات اقلیمی، مدلسازی نرخ تهویه، یادگیری ماشین، الگوریتم ژنتیک، مدل بهبود یافته.

Optimized Modeling of Ventilation Rate in Buildings with Passive Cooling Systems Using Ensemble Learning and Genetic Algorithm: An Innovative Approach for Limited and Imbalanced Data

Majid Mohsenpour

Department of Energy System Engineering, Faculty of Mechanical Engineering, K.N. Toosi University of Technology, Tehran, Iran. m.mohsenpour@email.kntu.ac.ir

Mohsen Salimi

Renewable Energy Research Department, Niroo Research Institute (NRI), Tehran, Iran. msalimi@nri.ac.ir

Majid Amidpour

Department of Energy System Engineering, Faculty of Mechanical Engineering, K.N. Toosi University of Technology, Tehran, Iran. amidpour@kntu.ac.ir

Abstract

With the rising demand for ventilation systems due to climate change and greenhouse gas emissions, energy consumption has significantly increased. The development of passive cooling systems has emerged as an essential solution, as these systems are independent of conventional air conditioning. Modeling ventilation rates in such systems is traditionally time-consuming; therefore, this study presents a data-driven model to predict ventilation rates in buildings equipped with passive cooling, distinguishing it from other studies. Ten machine learning models were employed as base models in a standard ensemble stacking, but the results for RMSE and MAE evaluation parameters were unsatisfactory. To enhance the model, a genetic algorithm was applied. The accuracy of the improved ensemble model was assessed using four different datasets, with key features selected through the Recursive Feature Elimination (RFE) method. The results demonstrate that the proposed model, with higher accuracy than classic algorithms, is suitable for predicting ventilation rates, especially in conditions with limited and non-uniform data distribution. It is adaptable to various geographic regions and assists in a rapid assessment of ventilation rates in buildings.

Keywords: Passive Cooling, Climate Change, Data-Driven Modeling, Ventilation Rate Prediction, Genetics Algorithm, Enhanced Ensemble Model

سیستم‌های HVAC در کشورهای توسعه یافته و در حال توسعه به صورت روزافزون در حال افزایش است [۱].

صنعت ساختمان به‌عنوان یکی از بخش‌های اصلی انرژی بر و وابسته به سیستم‌های HVAC شناخته می‌شود؛ زیرا تقریباً نیمی از انرژی مصرفی ساختمان‌ها به سیستم‌های سرمایشی و گرمایشی اختصاص دارد [۲]. در صورت عدم اصلاح الگوی مصرف انرژی در

۱- مقدمه

در سال‌های اخیر، مصرف انرژی به دلیل استفاده گسترده و کنترل‌نشده از سیستم‌های تهویه مطبوع (HVAC) که عموماً به سوخت‌های فسیلی وابسته هستند، به‌طور چشمگیری افزایش یافته است. این در حالی است که با تشدید گرمایش جهانی، تقاضا برای

ساختمان‌ها، پیامدهای ناگواری را برای کشورهای در حال توسعه می‌توان متصور بود. ادامه این روند به مشکلات زیست محیطی جدی نیز منجر می‌شود. وابستگی شدید به منابع انرژی فسیلی برای تهویه مطبوع، باعث افزایش انتشار گازهای گلخانه‌ای، تشدید تغییرات آب و هوایی، و افزایش آلودگی هوا می‌گردد. بنابراین، پرداختن به این چالش‌ها و کاهش اثرات زیست‌محیطی آنها امری ضروری و غیرقابل انکار است. یکی از روش‌های مناسب استفاده از مواد تابشی در ساختمان است. در این روش مواد با ضریب گسیل مطلوب، تابش ساطع شده از خورشید را دفع کرده و موجب کاهش دمای داخلی در ساختمان می‌شود. با بکارگیری این روش، میزان نیاز به بار سرمایشی به طور قابل توجهی کاهش می‌یابد. به عنوان مثال، تانگ و همکاران [۳] کامپوزیت خنک‌کننده تابشی را برای شیشه پنجره‌ها پیشنهاد کرده است. این مطالعه نشان داد که ماده پیشنهادی به طور مؤثر توانسته دمای داخلی را کاهش دهد. گونگ و همکاران [۴] طراحی نوآورانه‌ای از خنک‌کننده تابشی خودتنظیم ارائه داده‌اند. در این پژوهش، پارافین با یک خنک‌کننده تابشی ترکیب شده تا امکان سازگاری دینامیکی با تغییرات دما از طریق انتقال مایع-جامد را فراهم سازد. نتایج نشان می‌دهد که با بکارگیری خنک‌کننده‌های ارائه شده امکان صرفه‌جویی انرژی در ساختمان‌های شهرها و اقلیم‌های مختلف وجود خواهد داشت. لو و همکاران [۵] روش نوآورانه‌ای برای خنک‌سازی تابشی غیرفعال در طول روز ارائه کرده است که با استفاده از پوششی متشکل از اکسید ایتیریم (Y_2O_3) و پلی‌دی‌متیل‌سیلوکسان (PDMS) ساخته شده است. این طراحی بازتابش خورشیدی و گسیلندگی در پنجره ساختمان را به طور مؤثری بهبود می‌بخشد و بازده خنک‌سازی تابشی بالایی را حتی در شرایط تابش شدید خورشیدی فراهم می‌کند. اما این روش‌ها به دلیل نیاز به استفاده از مواد جدید، مستلزم هزینه‌های قابل توجهی برای اجرا در ساختمان‌ها هستند. علاوه بر این، چالش‌های زیست محیطی مرتبط با کاربرد این مواد در ساختمان‌ها، از دیگر مسائلی است که کارایی و پذیرش این روش‌ها را تحت تأثیر قرار می‌دهد.

یکی دیگر از روش‌ها ایجاد تهویه طبیعی در ساختمان است. تکنیک‌های تهویه مطبوع طبیعی^۱ از پدیده‌های طبیعی و اصول طراحی برای بهبود آسایش گرمایی بدون اتکال بیش از حد به سیستم‌های مکانیکی و مصرف انرژی استفاده می‌کنند. تهویه طبیعی ساختمان‌ها از اهمیت فزاینده‌ای برخوردار هستند. این امر نه تنها به ایجاد محیط‌های داخلی راحت و سالم با هدف ارتقای بهره‌وری ساکنین کمک می‌کند، بلکه نقش بسزایی در کاهش مصرف انرژی در انواع مختلف ساختمان‌ها ایفا می‌نماید. به عنوان مثال، فورتین و همکاران [۶] با به‌کارگیری آزمایش تجربی، تاثیر ترکیب تهویه طبیعی ناشی از اختلاف دما و استفاده از ماده سرمایش تابشی در سقف ساختمان را بررسی کرده‌اند. در این تحقیق، ورقه‌های آلومینیومی با پوشش پلی‌وینیلیدین فلوراید (PVDF) سفید به‌عنوان روکشی برای محفظه‌های آزمایشی استفاده شده‌اند که بر روی یک قاب آلومینیومی T شکل در بالای سقف نصب شده‌اند. همچنین، محفظه‌های پر شده از آب به‌عنوان

ذخیره گرمایی برای ایجاد اختلاف دما و نیروی بویانسی مورد استفاده قرار گرفته‌اند تا امکان سرمایش در طول روز فراهم شود. ساختمان تست فاقد پنجره بوده و لوله‌هایی در کف و سقف آن وظیفه تهویه طبیعی را بر عهده دارند. ساختمان مرجع، که در این پژوهش به‌عنوان مبنا استفاده شده است، دارای جرم گرمایی داخلی (بطری‌های پر شده از آب)، منبع گرمای داخلی و دو دودکش رو به بالا بوده است. تهویه تنها زمانی صورت می‌گرفت که دمای داخلی از دمای محیط بیشتر می‌شد، که این امر موجب جریان تهویه طبیعی از طریق کوتاه‌ترین دودکش و فراهم‌آوردن تهویه به سمت بالا می‌شد. در شب، زمانی که دمای داخلی بیشتر از دمای محیط می‌شد، طراحی این امکان را فراهم می‌کرد که جرم گرمایی گرما جمع‌شده را تخلیه کند. از سوی دیگر، ساختمان تست با داشتن همان جرم گرمایی و منبع گرمای داخلی ساختمان مرجع، سقفی پوشیده از ورق آلومینیومی غیرعایق داشت که سطح بالای آن با ماده سرمایشی تابشی ویژه‌ای برای روز پوشانده شده بود. هدف از این طراحی، نشان دادن تأثیر سرمایش تابشی و تهویه طبیعی بر دمای داخلی بود.

نویسندگان این پژوهش مدعی هستند، به دلیل انتخاب ماده ساده، ارزان، قابل دسترس و سازگار با محیط زیست امکان توسعه این ساختمان به خصوص در مناطق گرم و خشک وجود دارد. در این پژوهش، مدل نوآورانه در مقابل یک مدل مرجع محک زده شد و اثربخشی آن ساختمان ارائه شده در کاهش دمای داخلی اثبات شده است. قابل ذکر است، ساختمان جدید دمای داخلی را $3/9$ درجه سانتیگراد خنک‌تر از میانگین خارجی و $8/9$ درجه سانتیگراد پایین‌تر از دمای ساختمان مرجع حفظ کرده است

با توجه به مزایای متعدد تهویه طبیعی در زمینه صرفه‌جویی انرژی و بهبود کیفیت هوای داخلی، پژوهش‌های مختلفی اهمیت این روش را برجسته کرده و تلاش نموده‌اند با استفاده از راهبردهای کارآمد در مراحل طراحی و بهره‌برداری، نرخ تبادل هوای کافی را حفظ کنند. در مطالعات صورت گرفته، عموماً از طریق شبیه‌سازی توسط نرم افزارهای مرسوم همچو DesignBuilder و Energy Plus استفاده شده است. با وجود بهبود روش‌ها و کاربردهای عملی سیستم‌های تهویه طبیعی، چالش‌های مهندسی مربوط به عدم قطعیت این روش همچنان مانعی برای بهره‌برداری مؤثر از آن محسوب می‌شود. در مدل‌سازی نرخ تهویه طبیعی، عواملی نظیر اقلیم محلی، کیفیت هوای بیرونی و داخلی در این مسئله نقش دارند. به عنوان مثال خالد بتانه و همکاران [۶]، با استفاده از نرم افزار دیزاین بیلدر، اقدامات بهینه‌سازی ساختمان شامل ارتقاء سیستم تهویه، عایق‌کاری و نورپردازی بهینه را در مناطق آب و هوایی مختلف با دیدگاه اقتصادی و محیط زیستی بررسی کرده‌اند. اما مدل‌سازی چنین ساختمان‌هایی در نرم افزارهای مرسوم برای طراحان معمولاً زمان بر است و زمان زیادی را برای مدل‌سازی صرف می‌کنند [۷].

یکی از ابزارهای مناسب جهت مدل‌سازی روش مبتنی بر داده یا همان جعبه سیاه است. مدل‌های جعبه سیاه مزایای متمایز در مدل‌سازی ارائه می‌کنند. با اتخاذ روش‌های مبتنی بر داده، مانند تکنیک‌های یادگیری ماشین بر اساس داده‌های ثبت شده، مدل‌های جعبه سیاه قابلیت بیشتری در فهم داده‌های خطی و غیرخطی از خود نشان می‌دهند [۸]. علاوه بر آن، مدل‌های جعبه سیاه نیازمند زمان

1 Passive Cooling

عملکرد بهتری را ارائه دهد [۱۶].

نوآوری اصلی تحقیق حاضر در طراحی یک مدل انباشته گروهی بهبود یافته است که با استفاده از ۱۰ مدل یادگیری ماشین به عنوان مدل‌های پایه، یک چارچوب پیشرفته و قدرتمند برای مدل‌سازی نرخ تهویه در ساختمان با سرمایه‌گذاری غیر فعال ایجاد می‌کند.

در گام اول، برای افزایش دقت و کارایی، مدل‌های پایه مختلف (مانند ماشین بردار پشتیبان، جنگل تصادفی، و تقویت گرادیان شدید و...) به‌طور موازی با هم ترکیب شده‌اند تا از توانایی هر کدام در تحلیل داده‌های متنوع بهره‌برداری شود. سپس، به منظور بهینه‌سازی ترکیب و تعداد این مدل‌ها، از الگوریتم ژنتیک استفاده شده است که انتخاب بهینه‌ای از مدل‌ها و پارامترها را فراهم می‌سازد. این روش باعث می‌شود که مدل نهایی از انعطاف‌پذیری و توانایی یادگیری بهتری در مواجهه با داده‌های متغیر و نامتوازن برخوردار باشد.

این چارچوب مدل‌سازی نه تنها پاسخگوی نیاز به دقت بالا در پیش‌بینی میزان تهویه طبیعی است، بلکه امکان استفاده از داده‌های محدود را نیز فراهم می‌کند. علاوه بر این، این مدل با بهینه‌سازی ساختار و ترکیب مدل‌های پایه، به سوال مهمی که در مطالعات پیشین نادیده گرفته شده بود، پاسخ می‌دهد: آیا انتخاب مدل‌های پایه و تعداد آن‌ها تأثیر بسزایی در دقت پیش‌بینی دارند؟ در این تحقیق، با آزمایش ترکیب‌های مختلف و استفاده از الگوریتم ژنتیک، بهینه‌ترین ساختار برای مدل پایه بدست خواهد آمد.

به‌طور کلی، نوآوری این پژوهش در ترکیب مدل‌های یادگیری ماشین به شکل انباشته و بهینه‌سازی آن‌ها برای مدل‌سازی تهویه طبیعی با داده‌های محدود است. با پیش‌بینی دقیق تهویه طبیعی، می‌توان عملکرد سیستم‌های تهویه طبیعی را بهینه‌سازی کرده و اطمینان حاصل کنند که پنجره‌ها یا دریچه‌ها در زمان‌ها و مقادیر مناسب باز می‌شوند. این قابلیت پیش‌بینی، کنترل پیشگیرانه تهویه را ممکن می‌سازد. همچنین مدل ارائه شده می‌تواند برای مدل‌سازی و پیش‌بینی نرخ تهویه ساختمان برای هر داده جدیدی مورد استفاده قرار بگیرد. این پژوهش، بستری جدید برای استفاده از روش‌های داده‌محور در حوزه انرژی ساختمان فراهم کرده و زمینه‌ساز توسعه روش‌های پیشرفته در بهینه‌سازی مصرف انرژی است.

۲- مبانی و روش‌ها

دره توپانگا که در کوه‌های سانتا مونیکا و در شهرستان لس آنجلس، کالیفرنیا واقع شده است، بین مالبیو و پاسیفیک پالیسیدز قرار دارد. داده‌های این پژوهش از مطالعه اخیر فورتین و همکاران [۱۹] گردآوری شده و به دو دسته اصلی تقسیم شده‌اند: دسته مرجع (به عنوان مبنا) و دسته تست (به عنوان مدل پیشنهادی). هر مجموعه داده به صورت روزانه و با فواصل ۱۵ دقیقه‌ای طی ماه‌های سپتامبر و نوامبر ثبت گردیده و شامل اطلاعاتی از شرایط داخلی و خارجی ساختمان و میزان نرخ تهویه است. هدف این پژوهش، ارائه مدلی بهینه‌شده بر پایه الگوریتم ژنتیک (GA) است که به منظور دستیابی به این هدف، روش پیشنهادی با مدل‌های کلاسیک و مدل‌های انباشته مقایسه می‌شود. شایان ذکر است که این مدل قابلیت استفاده برای مدل‌سازی سرمایه‌گذاری غیرفعال در انواع ساختمان‌ها را دارد.

شکل ۱ فرآیند انجام پژوهش را به تصویر می‌کشد. در این

کمتر جهت محاسبه هستند. این امر باعث شده است تا به کارگیری این مدل‌ها از جذابیت زیادی برخوردار باشد. بنابراین توسعه مدل‌های یادگیری ماشین از اهمیت بالایی برخوردار است زیرا که دستیابی به مدل با دقت بالاتر، به کمک مهندسان جهت تصمیم‌گیری‌های دقیق خواهد شتافت [۹]. به عنوان مثال، هو و همکاران [۱۰] با بهبود الگوریتم شبکه عصبی پس انتشار^۱، شبکه عصبی تابع پایه شعاعی^۲ و RF به پیش‌بینی میزان سرمایه‌گذاری مورد نیاز در ساختمان پرداختند. سوی و همکاران [۱۱] به تحلیل مدل RNN^۳ جهت مدل‌سازی دمای هوای داخلی ساختمان پرداخت. نویسندگان این پژوهش معتقدند این مدل قابلیت به کارگیری در حوزه ساختمان را به قوت دارد. اما یکی از ضعف‌های این پژوهش، عدم مدل‌سازی میزان تهویه طبیعی در ساختمان است چرا که تأثیر بسزایی در میزان دمای ساختمان دارد [۱۲]. همچنین، در این پژوهش بیان نشده است در صورت کافی نبودن داده جهت مدل‌سازی، این مشکل چگونه مرتفع خواهد شد چرا که نرخ تهویه در ساختمان در هر بازه بندی زمانی ثبت شده از ۵ دقیقه تا ۳۰ دقیقه یکبار در طول شبانه روز در حال تغییر است. این امر باعث خواهد شد مشکل عدم توزیع یکنواخت داده شکل بگیرد و مدل‌سازی چنین سیستم‌هایی را با مشکل مواجه سازد و باعث می‌شود مدل با پدیده بیش برآزش مواجه شود [۱۳]. به عنوان مثال اودیم و همکاران [۱۴] به دلیل عدم یکنواختی داده از روش بیش نمونه‌گیری^۴ برای حل این مشکل استفاده شده است و نتایج حاصله نشان از بهبود یادگیری مدل پس از بکارگیری روش بیش نمونه‌گیری بوده است. اما وقتی توزیع غیر یکنواخت داده بسیار شدید باشد این روش هم پاسخگو نخواهد بود [۱۵].

این پژوهش با استفاده از داده‌های آزمایشی منتشر شده مقاله فورتین و همکاران به ارائه یک مدل بهبود یافته انباشته گروهی در مدل‌سازی میزان تهویه طبیعی در ساختمان با سرمایه‌گذاری غیرفعال با استفاده از یادگیری ماشین، به طور ویژه به چالش عدم یکنواختی و محدودیت داده‌ها در این زمینه پرداخته است. روش انباشته گروهی مزایای متعددی نسبت به مدل‌های کلاسیک دارد، از جمله:

۱. افزایش دقت پیش‌بینی: ترکیب نتایج چندین مدل پایه می‌تواند دقت پیش‌بینی را نسبت به استفاده از یک مدل منفرد به طور قابل توجهی افزایش دهد [۱۶].
۲. بهبود قابلیت تعمیم: این روش کمک می‌کند تا مدل در برابر داده‌های جدید و متغیر مقاوم‌تر باشد و قابلیت تعمیم بهتری داشته باشد [۱۷].
۳. کاهش احتمال بروز خطا: با استفاده از چندین مدل، خطر خطاهای پیش‌بینی ناشی از مدل‌های خاص کاهش می‌یابد و نتایج پایدارتر می‌شوند [۱۸].
۴. افزایش انعطاف‌پذیری در مقابل پارامترهای دارای عدم قطعیت: این روش می‌تواند به تنوع داده‌ها و تغییرات غیرقابل پیش‌بینی در داده‌ها بهتر پاسخ دهد و در نتیجه

¹ Back Propagation Neural Network

² Radial Basis Neural Network

³ Recurrent Neural Network

⁴ OverSampling

شماره [۱۹] استفاده گردید. جدول ۱ متغیرهای ورودی مدل را که برای پیش‌بینی نرخ تهویه به‌کار می‌روند، نشان می‌دهد. به طور خاص، داده‌های مرجع و تست در ماه‌های مختلف شامل مقادیر متفاوتی هستند. در ماه سپتامبر، ۶۷۳ داده مرجع با میانگین نرخ تهویه $5 \times 10^{-5} \times 6/87$ متر بر ثانیه و حداکثر نرخ تهویه 15×10^{-5} متر بر ثانیه به کار رفته‌اند؛ همچنین ۶۷۳ داده تست با میانگین نرخ تهویه $5 \times 10^{-5} \times 2/43$ متر بر ثانیه و حداکثر نرخ تهویه 13×10^{-5} متر بر ثانیه برای ارزیابی مدل استفاده شده‌اند. در ماه نوامبر نیز ۱۳۴۵ داده مرجع با میانگین نرخ تهویه $5 \times 10^{-5} \times 7/81$ متر بر ثانیه و حداکثر نرخ تهویه 16×10^{-5} متر بر ثانیه و ۱۳۴۵ داده تست با میانگین نرخ تهویه $5 \times 10^{-5} \times 5/08$ متر بر ثانیه و حداکثر نرخ تهویه 17×10^{-5} متر بر ثانیه استفاده شده است. متغیر "شار گرمایی ذخیره حرارتی" جریان گرما از طریق جرم گرمایی در ساختمان‌های مرجع و تست را اندازه‌گیری می‌کند که نشان‌دهنده ویژگی‌های نگهداری گرما است. دمای سطحی بیرونی تابش کننده، تبادل گرما بین تابش کننده و هوای اطراف را نشان می‌دهد، متغیر "منبع گرمایی مرجع/تست" بار حرارتی ساکنین که یکتفر در نظر گرفته شده است را نشان می‌دهد. متغیرهای شار گرمایی در بالا و پایین تابش کننده نیز بیانگر میزان تبادل گرما با بیرون و داخل ساختمان است. متغیر "دمای سطحی ذخیره ساز گرمایی مرجع/تست" بیانگر دمای سطحی محفظه‌های پر شده از آب در دو ساختمان است.

جدول ۱- ویژگی‌های ورودی مدل‌سازی

ویژگی	واحد
دمای هوا	(°C)
سرعت باد	(m/s)
تابش خالص	(W/m ²)
دمای موثر آسمان	(°C)
رطوبت نسبی هوا	(%)
جهت باد	(-)
تابش جوی	(W/m ²)
فشار بارومتريک	(mbar)
دمای نقطه شبنم	(°C)
سرعت باد لحظه‌ای	(m/s)
تابش خورشیدی	(W/m ²)
شار گرمایی در بالای تابش کننده	(W/m ²)
شار گرمایی در داخل تابش کننده	(W/m ²)
شار گرمایی در دیواره‌های داخلی	(W/m ²)
دمای سطحی ذخیره ساز گرمایی مرجع/تست	(°C)
دما در بالای محافظ همرفت	(°C)
سرعت جریان ورودی در ورودی مرجع/تست	(m/s)
شار گرمایی ذخیره گرمایی مرجع/تست	(W/m ²)
دمای سطحی بیرونی تابش کننده	(°C)
دمای سطحی دیواره‌های داخلی مرجع/تست	(°C)
دمای داخلی در ارتفاع ۰/۱	(°C)

پژوهش، ده مدل کلاسیک متنوع با ویژگی‌های متفاوت انتخاب شده‌اند که به بهبود فرآیند مدل‌سازی کمک می‌کنند. انتخاب این مدل‌ها به دلیل توانایی آنها در تحلیل بهتر داده‌های کم حجم و مقابله با مشکل عدم توزیع یکنواخت (Imbalance Data) است. این مدل‌های متنوع به این دلیل برگزیده شده‌اند که هر یک با تکنیک‌های مختلف، ابعاد گوناگون داده‌ها را تحلیل می‌کنند و به کاهش حساسیت مدل نسبت به داده‌های نامتوازن کمک می‌کنند، که می‌تواند در بهبود دقت نهایی و پایداری مدل موثر باشد. در این فرآیند، داده‌های خام ابتدا از طریق شناسایی داده‌های پرت و انتخاب ویژگی (RFE) پیش‌پردازش می‌شوند. سپس این مدل‌های پایه مورد استفاده: نزدیک‌ترین همسایه (K-NN)، پرسپترون چندلایه (MLP)، جنگل تصادفی (RF)، رگرسیون بردار پشتیبان (SVR)، رگرسیون خطی (LR)، آدا بوست (Adaboost)، لاسو (Lasso)، گرادیان تقویتی سریع (Xgboost)، درخت تصمیم (DT)، و گرادیان تقویتی نوری (LightGbm) به عنوان یادگیرنده‌های اولیه اعمال می‌شوند. در گام بعد، الگوریتم ژنتیک (GA) برای یافتن بهترین ترکیب از مدل‌های پایه به منظور بهینه‌سازی عملکرد استفاده می‌شود. مدل ترکیبی نهایی با استفاده از یک مدل متا (Meta Learner) که LR است توسعه یافته و در مرحله ارزیابی، بر اساس شاخص RMSE سنجیده می‌شود. در نهایت، ترکیبی از مدل‌ها که کمترین مقدار RMSE را داشته باشد به عنوان بهترین ترکیب معرفی خواهد شد، که نشان‌دهنده دقت بالاتر و عملکرد بهینه‌تر مدل نهایی است.



شکل ۱: نحوه انجام پژوهش

۱-۲- استخراج داده

در این پژوهش، با هدف توسعه و آزمون یک مدل جدید برای مدل‌سازی نرخ تهویه ساختمان با سرمایه‌ش غیر فعال اقدام شد. به منظور ارائه مدل یادگیری ماشین، از داده‌های منتشرشده در پژوهش

(m/s)	سرعت جریان ورودی در خروجی مرجع/تست
(°C)	دمای خروجی مرجع/تست
(°C)	دمای سطحی داخلی تابش کننده
(W/m ²)	منبع گرمایی مرجع/تست
(°C)	دمای داخلی در ارتفاع ۰/۳
(°C)	دمای ورودی مرجع/تست

است، ما را به کار گیری انتخاب متغیرهای مناسب واداشت. جهت دستیابی به این هدف، با اجرای روش حذف تکراری ویژگی‌ها (RFE)^۱ به دسته کوچک‌تری از مجموعه داده‌ها دست بیابیم. با انجام این امر، مدل قادر خواهد بود تا با دقت و سرعت بیشتری الگوهای پیچیده را در داده شناسایی کند [۲۲].

۲-۴- مدل یادگیری گروهی انباشته:

روش انباشت‌کننده با ترکیب یادگیری ماشین و مدل‌های آماری، از چندین مدل پایه و یک مدل فرامدل برای افزایش دقت پیش‌بینی استفاده می‌کند. در این روش، ابتدا ۸۰ درصد از کل داده‌ها برای یادگیری مدل انتخاب شد. سپس، از این ۸۰ درصد، ۲۰ درصد برای ارزیابی اولیه مدل‌های پایه جدا شد؛ به این ترتیب، مدل‌ها با ۶۰ درصد از داده‌ها آموزش داده شده و ارزیابی اولیه بر ۲۰ درصد باقی‌مانده از داده‌های آموزشی انجام شد. در ادامه، داده‌های آموزشی به پنج بخش تقسیم شده و از اعتبارسنجی کافولد پنج‌تایی برای آموزش مدل‌های پایه و کاهش احتمال بیش‌برازش استفاده گردید. خروجی مدل‌های پایه نیز به عنوان ورودی مدل فرامدل قرار گرفت.

برای ارزیابی عملکرد مدل، از پارامترهای R²، MAE و RMSE استفاده شده است؛ R² تطابق کلی با داده‌ها را می‌سنجد، MAE میانگین خطا را اندازه‌گیری می‌کند و RMSE با تأکید بر خطاهای بزرگ‌تر، به ارزیابی دقیق‌تر خطاها کمک می‌نماید. در نهایت، ۲۰ درصد از کل داده‌ها نیز به عنوان داده‌های خارج از نمونه (out-of-sample) کنار گذاشته شد تا دقت مدل بر روی داده‌های جدید ارزیابی شود [۲۴]، [۲۳].

۲-۵- مدل یادگیری گروهی انباشته بهبود یافته:

روش یادگیری ترکیبی انبوه‌سازی دارای دو بخش اصلی است: ساخت مدل‌های پایه و ترکیب آن‌ها [۱۷]. انتخاب مدل‌های پایه باید دارای دقت و تنوع مناسب باشد، در حالی که مدل‌های فرا یادگیری معمولاً الگوریتم‌هایی با پایداری مناسب هستند. از اینرو در این پژوهش، از رگرسیون خطی جهت مدل فرا یادگیری استفاده شده است. جهت ارائه مدل جدید، از الگوریتم ژنتیک برای تعیین بهینه ترکیب مدل‌ها در روش یادگیری ترکیبی انبوه‌سازی استفاده شده است. تابع هدف این الگوریتم RMSE در نظر گرفته شده است تا فاصله نقاط پیش‌بینی و واقعی به حداقل‌ترین میزان برسد.

۲-۶- پارامترهای ارزیابی:

عملکرد و دقت مدل‌های یادگیری ماشین در هر مطالعه موردی با سه معیار ارزیابی خواهد شد، که معادله (۵) ضریب تعیین (R-square)، معادله (۶) میانگین درصد مطلق خطا (MAPE) و معادله (۷) ریشه میانگین مربعات خطا (RMSE) ارزیابی می‌شود [۲۵].

۲-۲- پیش پردازش داده:

در این پژوهش، به منظور بهبود دقت مدل و کاهش خطاهای پیش‌بینی، داده‌های پرت با استفاده از نمودار جعبه‌ای (Box Plot) شناسایی و از مجموعه داده‌ها حذف شدند. نمودار جعبه‌ای به عنوان یک ابزار آماری مؤثر، برای تشخیص داده‌های خارج از محدوده (Outliers) مورد استفاده قرار می‌گیرد. در این روش، ابتدا چارک اول (Q₁) و چارک سوم (Q₃) محاسبه می‌شوند و فاصله بین چارک‌ها (IQR) به شرح زیر توسط معادله (۱) تعیین می‌گردد:

$$IQR = (Q_3) - (Q_1) \quad (1)$$

سپس، محدوده پذیرش داده‌ها توسط معادله (۲) و (۳) به صورت زیر تعریف می‌شود:

$$Upper\ Bound = 1/5 \times IQR + (Q_3) \quad (2)$$

$$Lower\ Bound = 1/5 \times IQR - (Q_1) \quad (3)$$

داده‌هایی که خارج از محدوده حد پایین و حد بالا (Lower and Upper Bound) قرار دارند، به عنوان داده‌های پرت شناسایی شده و از مجموعه داده‌ها کنار گذاشته می‌شوند. این فرآیند باعث می‌شود که تأثیر منفی داده‌های پرت بر دقت و عملکرد مدل کاهش یابد و انسجام مجموعه داده‌ها حفظ شود [۲۰].

سپس، داده‌های ورودی با استفاده از مقیاس‌بندی استاندارد بر اساس معادله مقیاس‌بندی تنظیم شدند. لازم به ذکر است که به منظور حفظ واریانس و میانگین واقعی داده‌ها، نرخ تهویه از این فرآیند مقیاس‌بندی مستثنی گردید [۲۱]. معادله (۴) نحوه مقیاس‌بندی پارامترهای ورودی را نشان می‌دهد. این عمل باعث خواهد شد توزیع پارامترهای ورودی تبدیل به توزیع نرمال شوند تا از مشکل بیش‌برازش در مدل‌سازی جلوگیری شود.

$$Z = \frac{(X - U)}{S} \quad (4)$$

داده واقعی: X

میانگین داده‌های واقعی: U

انحراف معیار داده‌های واقع: S

۲-۳- انتخاب متغیر مناسب:

با توجه به تعداد ویژگی‌ها در این پژوهش، که عددی بین ۲۰ تا ۳۰

¹ Recursive Elimination Feature

یکسانی با ماه نوامبر از نظر اهمیت تعداد ویژگی‌های شرایط بیرونی دارد.

نتایج نشان می‌دهد که تعداد ویژگی‌ها برای مدل‌سازی در این دو مجموعه داده از ۲۸ و ۲۶ به ۱۰ تقلیل پیدا کرده است. روش RFE بصورت تکراری ویژگی‌ها را اضافه کرده و آن‌هایی را که در پیش بینی تأثیر زیادی دارند حفظ می‌کند. انتخاب ۱۰ ویژگی به‌عنوان نهایی به این دلیل است که این تعداد بهینه‌ترین توازن را میان دقت پیش‌بینی و کاهش پیچیدگی مدل فراهم می‌کند. آزمایش‌های مختلف نشان داده‌اند که انتخاب ۱۰ ویژگی نسبت به ۹ یا ۱۱ ویژگی، به بهبود دقت و کارایی مدل کمک می‌کند، زیرا با کاهش تعداد ویژگی‌ها، مدل از سادگی و سرعت بیشتری برخوردار می‌شود و در عین حال اطلاعات کلیدی برای پیش‌بینی به‌خوبی حفظ می‌شود.

۳-۲- مدل گروهی انباشته

در این پژوهش، مدل یادگیری انباشته برای پیش‌بینی دقیق نرخ تهویه بر اساس پارامترهای داخلی و خارجی ساختمان به‌کار گرفته شده است. ده مدل کلاسیک، شامل نزدیکترین همسایه‌ها، جنگل تصادفی، شبکه عصبی، و گرادیان بوستینگ، به‌عنوان مدل‌های پایه انتخاب شدند و از یک رگرسیون خطی به‌عنوان فرا مدل برای کاهش خطا استفاده شده است. نتایج نشان دادند که مدل انباشته در سپتامبر با R^2 معادل ۰٫۹۶ عملکرد بهتری نسبت به بهترین مدل کلاسیک (لایت گرادیان بوستینگ با R^2 معادل ۰٫۹۵۵) (ارائه داده است. این بهبود اندک، با توجه به نرخ تهویه کم، در توسعه ساختمان‌های پایدار اهمیت دارد. همچنین با توجه به پارامترهای MAE و RMSE، مشخص شده است که مدل انباشته ضعیف‌تری دارد؛ چرا که میزان خطای آن نسبت به مدل گرادیان بوستینگ بالاتر است، و این مدل کمترین خطا در MAE و RMSE را نشان داده است. برای مثال، MAE و RMSE به ترتیب برابر $۴/۲۲ \times (۱۰)^{-۶}$ و $۱/۱۵ \times (۱۰)^{-۵}$ برای مدل گرادیان بوستینگ محاسبه شده‌اند، در حالی که برای مدل انباشته این مقادیر به ترتیب $۵/۲ \times (۱۰)^{-۶}$ و $۱/۱۸ \times (۱۰)^{-۵}$ گزارش شده‌اند. همچنین، در میان مدل‌های پایه، مدل‌های شبکه عصبی، رگرسیون بردار پشتیبان و لاسو ضعیف‌ترین عملکرد را نشان داده‌اند. به منظور بهبود مدل انباشته در مدل‌سازی نرخ تهویه، استفاده از الگوریتم ژنتیک در ترکیب مدل‌های پایه مورد بررسی قرار گرفته است.

در ماه نوامبر، با بررسی دو مجموعه داده مرجع و تست، مشاهده شده که ترکیب ده مدل پایه به مدل نهایی امکان تخمین دقیق نرخ تهویه را می‌دهد. نتایج به دست آمده از مجموعه داده مرجع نشان داده است که مدل انباشته توانایی پیش‌بینی نرخ تهویه با دقت R^2 برابر با ۰٫۹۶۳ را داراست، اما همچنان ضعف‌هایی در کاهش خطاهای MAE و RMSE مشاهده شده است، که این ضعف‌ها در مقایسه با مدل گرادیان بوستینگ به چشم می‌خورند. علاوه بر این، در داده تست نیز مدل شبکه‌های عصبی، رگرسیون بردار پشتیبان و لاسو نتوانسته‌اند عملکرد مناسبی ارائه دهند، در حالی که مدل‌های مبتنی بر درخت تصمیم عملکرد مناسبی نشان داده‌اند.

در نهایت، نتایج مدل انباشته در داده تست نشان داده‌اند که این مدل در هر سه پارامتر R^2 ، MAE و RMSE بهترین عملکرد را داشته

که در آن $(Y_{i,a})$ نشان‌دهنده مقادیر واقعی، $(Y_{i,p})$ مقادیر پیش‌بینی شده، و (\bar{Y}_a) میانگین داده‌های واقعی است. مقدار R^2 بین صفر و یک است که مقادیر بالاتر به معنای تطابق بهتر مدل با داده‌های واقعی است. این امر نشان می‌دهد، هرچه مدل نزدیک به داده واقعی پیش‌بینی کند و مقدار خطا به صفر میل کند میزان R^2 یک خواهد شد. بنابراین R^2 یکی از پارامترهای مناسب جهت ارزیابی مدل خواهد بود. در معادله (۶)، MAE به صورت درصد بیان می‌شود و مقادیر کمتر آن نشان‌دهنده دقت بیشتر مدل است. این معیار، برخلاف RMSE که بر خطاهای بزرگ تمرکز دارد، همه خطاها را یکسان وزن دهی می‌کند و دیدگاه ساده‌تری از انحراف میانگین مدل از داده‌های واقعی ارائه می‌دهد. در معادله (۷) RMSE میزان خطاهای بزرگ را بیشتر از خطاهای کوچک نشان می‌دهد، زیرا تفاوت‌ها را به صورت مربعی محاسبه می‌کند. مقادیر کمتر RMSE نشان‌دهنده دقت بالاتر مدل است. این معیار برخلاف MAE، که همه خطاها را به طور یکنواخت در نظر می‌گیرد، به خطاهای بزرگ وزن بیشتری می‌دهد. بنابراین ارزیابی این ۳ پارامتر کمک خواهد کرد تا میزان دقت مدل، میانگین انحراف از داده واقعی و میزان خطا از داده واقعی محاسبه گردد [۲۵].

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_{i,a} - Y_{i,p})^2}{\sum_{i=1}^N (Y_{i,a} - \bar{Y}_a)^2} \quad (5)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{i,a} - Y_{i,p}| \times 100 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{i,a} - Y_{i,p})^2} \quad (7)$$

۳- نتایج

۳-۱- انتخاب متغیرهای مناسب (RFE)

در این پژوهش، با توجه به تعداد داده‌ها، از روش انتخاب ویژگی RFE جهت کاهش پیچیدگی مسئله استفاده گردید. این روش توانسته داده‌های ورودی را به حداقل‌ترین میزان کاهش دهد و ویژگی‌های داده مرجع و تست تا ۴۰ درصد کاهش یافته است. با توجه به ویژگی‌های حفظ شده در داده مرجع، مشخص می‌شود که میزان تهویه در ساختمان مرجع بیشتر به عوامل داخلی وابسته است؛ زیرا ۶۰ درصد از ویژگی‌های حفظ شده شامل دما محیط، سرعت باد، تابش خورشید، فشار بارومتريک، شار گرمایی داخلی، شار گرمایی ذخیره‌ساز گرمایی، دماهای داخلی در فاصله‌های یک‌دهم و دو دهم متر، دمای هوای ورودی و خروجی است که تأثیر زیادی روی نرخ تهویه دارند. این پارامترها بسیار مهم هستند و نقش تعیین‌کننده‌ای در تنظیم و بهینه‌سازی تهویه ساختمان ایفا می‌کنند.

ویژگی‌های حفظ شده در داده تست (از جمله دمای محیط، دمای نقطه شبنم، سرعت باد، جهت باد، تابش خورشید، شار گرمایی داخلی و خارجی، شار گرمایی ذخیره‌ساز گرمایی و دمای هوای خروجی) در ماه نوامبر به طور یکسان به شرایط داخلی و خارجی وابسته است. همچنین، ویژگی‌های ثابت شده در داده ارائه شده در پژوهش فورتین و همکاران در ماه سپتامبر افزایش یافته‌اند، اما ماه سپتامبر شرایط

است و به ترتیب مقادیر 0.973 ، $4/74 \times (10)^{-6}$ و $9/63 \times (10)^{-6}$ برای این مدل ثبت شده است. این نتایج نشان‌دهنده کارایی مدل انباشته در ارائه نتایج قابل قبول است. با این حال، ضعف عملکرد مدل‌های شبکه عصبی، لاسو و رگرسیون بردار پشتیبان در داده‌های کم مشهود بوده و از این رو در پژوهش‌های مبتنی بر داده‌های اندک در زمینه مدل‌سازی نرخ تهویه، این مدل‌ها گزینه مناسبی نیستند. با این حال، حضور این مدل‌ها در مدل انباشته به دلیل ذخیره‌سازی الگوهای جزئی داده‌ها کمک‌کننده بوده و در بهبود مدل انباشته در گام‌های بعدی با استفاده از انتخاب بهینه ترکیب مدل‌های پایه توسط الگوریتم ژنتیک اقدام خواهد شد.

جدول ۲- نتایج مدل‌های کلاسیک و مدل گروهی انباشته در مرجع، سپتامبر

ارزیابی مدل‌سازی مرجع			الگوریتم
R ²	RMSE(m/s)	MAE	
0/91	$1/17 \times (10)^{-5}$	$4/37 \times (10)^{-6}$	K-NN
0/89	$1/24 \times (10)^{-5}$	$4/45 \times (10)^{-6}$	RF
0/41	$1/24 \times (10)^{-5}$	$6/62 \times (10)^{-5}$	SVR
0/165	0/077	0/118	MLP
0/87	$1/41 \times (10)^{-5}$	$5/07 \times (10)^{-6}$	DT
0/06	$6/72 \times (10)^{-5}$	$6/6 \times (10)^{-5}$	Lasso
0/955	$1/15 \times (10)^{-5}$	$4/22 \times (10)^{-6}$	GBM
0/931	$1/38 \times (10)^{-5}$	$8/15 \times (10)^{-6}$	LR
0/956	$1/16 \times (10)^{-5}$	$5/54 \times (10)^{-6}$	LightGBM
0/92	$1/46 \times (10)^{-5}$	$6/68 \times (10)^{-6}$	AdaBoost
0/96	$1/18 \times (10)^{-5}$	$5/2 \times (10)^{-6}$	StackedEN

جدول ۳- نتایج مدل‌های کلاسیک و مدل گروهی انباشته در تست، سپتامبر

ارزیابی مدل‌سازی تست			الگوریتم
R ²	RMSE(m/s)	MAE	
0/83	$1/36 \times (10)^{-5}$	$7/15 \times (10)^{-6}$	K-NN
0/84	$1/34 \times (10)^{-5}$	$6/99 \times (10)^{-6}$	RF
0/012	$5/2 \times (10)^{-5}$	0/043	SVR
0/005	0/057	0/0256	MLP
0/66	$1/97 \times (10)^{-5}$	$9/82 \times (10)^{-6}$	DT
0/39	$3/46 \times (10)^{-5}$	$3/33 \times (10)^{-5}$	Lasso
0/81	$3/46 \times (10)^{-5}$	$3/19 \times (10)^{-6}$	GBM
0/78	$1/59 \times (10)^{-5}$	$8/06 \times (10)^{-6}$	LR
0/81	$1/46 \times (10)^{-5}$	$8/13 \times (10)^{-6}$	LightGBM
0/73	$1/76 \times (10)^{-5}$	$1/3 \times (10)^{-5}$	AdaBoost
0/85	$1/29 \times (10)^{-5}$	$7/14 \times (10)^{-6}$	StackedEN

جدول ۴- نتایج مدل‌های کلاسیک و مدل گروهی انباشته در مرجع، نوامبر

ارزیابی مدل‌سازی تست			الگوریتم
R ²	RMSE(m/s)	MAE	
0/92	$1/38 \times (10)^{-6}$	$5/54 \times (10)^{-6}$	K-NN
0/94	$9/29 \times (10)^{-6}$	$4/06 \times (10)^{-6}$	RF
0/31	$7/1 \times (10)^{-5}$	$6/96 \times (10)^{-5}$	SVR
0/11	0/0455	0/032	MLP
0/925	$1/33 \times (10)^{-5}$	$4/6 \times (10)^{-6}$	DT
0/04	$7/1 \times (10)^{-5}$	$7/01 \times (10)^{-5}$	Lasso
0/9578	$8/78 \times (10)^{-6}$	$3/93 \times (10)^{-6}$	GBM
0/95	$1/57 \times (10)^{-5}$	$9/75 \times (10)^{-6}$	LR
0/94	$1/12 \times (10)^{-5}$	$5/42 \times (10)^{-6}$	LightGBM
0/9448	$1/43 \times (10)^{-5}$	$7/87 \times (10)^{-6}$	AdaBoost
0/963	$1/1 \times (10)^{-5}$	$5/9 \times (10)^{-6}$	StackedEN

جدول ۵- نتایج مدل‌های کلاسیک و مدل گروهی انباشته در تست، نوامبر

ارزیابی مدل‌سازی تست			الگوریتم
R ²	RMSE(m/s)	MAE	
0/95	$1/3 \times (10)^{-5}$	$6/95 \times (10)^{-6}$	K-NN
0/97	$9/75 \times (10)^{-6}$	$4/8 \times (10)^{-6}$	RF
0/3	$7/01 \times (10)^{-5}$	$6/39 \times (10)^{-5}$	SVR
0/11	0/1146	0/0863	MLP
0/95	$1/25 \times (10)^{-5}$	$5/85 \times (10)^{-6}$	DT
0/007	$6/14 \times (10)^{-5}$	$5/81 \times (10)^{-5}$	Lasso
0/959	$1/2 \times (10)^{-5}$	$5/52 \times (10)^{-6}$	GMB
0/956	$1/05 \times (10)^{-5}$	$5/52 \times (10)^{-6}$	LR
0/96	$1/01 \times (10)^{-5}$	$5/52 \times (10)^{-6}$	LightGBM
40/9	$1/35 \times (10)^{-5}$	$1/14 \times (10)^{-5}$	AdaBoost
0/973	$9/63 \times (10)^{-6}$	$4/74 \times (10)^{-6}$	StackedEN

۳-۳- مدل گروهی انباشته بهبود یافته

با توجه به نتایج ارائه‌شده در بخش پیشین، مشخص گردید که مدل گروهی انباشته در بهبود عملکرد پارامترهای کلیدی مانند MAE و RMSE با محدودیت‌هایی مواجه است. این مسئله می‌تواند ناشی از وجود خطا در مدل‌های پایه و نیز عدم توازن (imbalance) داده‌ها باشد که به عدم بهبود این دو پارامتر منجر شده است. از این رو، اهمیت به‌کارگیری الگوریتم‌های بهینه‌سازی در این زمینه نمایان می‌شود. به منظور دستیابی به اهداف پژوهش و بهبود عملکرد مدل، از الگوریتم ژنتیک برای انتخاب بهینه‌ترین ترکیب مدل‌های پایه استفاده شده

است. انتظار می‌رود که الگوریتم ژنتیک نه تنها باعث ارتقای پارامترهای ارزیابی گردد، بلکه پیچیدگی عملکردی مدل گروهی انباشته را نیز کاهش دهد. در پیاده‌سازی الگوریتم ژنتیک، از بهینه‌سازی تک‌هدفه بهره گرفته شده و پارامتر RMSE به عنوان تابع هدف در نظر گرفته شده است، به گونه‌ای که با کاهش تعداد مدل‌های پایه، کمینه‌سازی مقدار RMSE حاصل گردد. پیش‌فرض‌های لازم برای پیاده‌سازی الگوریتم ژنتیک در جدول ۶ ارائه شده است.

جدول ۶- متغیرهای الگوریتم ژنتیک

متغیرها	داده‌ها
اندازه جمعیت	۵۰
تعداد نسل‌ها	۱۰
استراتژی انتخاب	انتخاب مسابقه‌ای
نرخ عبور	۰/۵
استراتژی عبور	عبور دونقطه‌ای
نرخ جهش	۰/۲
استراتژی جهش	بیت فلیپ (۰/۰۵)

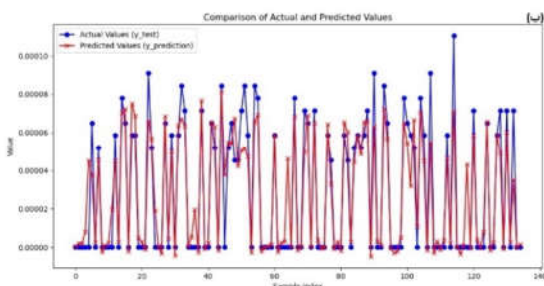
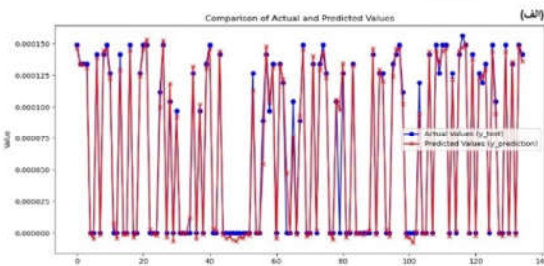
به کارگیری الگوریتم ژنتیک در مجموعه داده مرجع، منجر به شناسایی ترکیب صحیح مدل‌های پایه و کاهش چشمگیر تعداد مدل‌های پایه شد. این امر نشان‌دهنده توانمندی روش یادشده در کار با داده‌های مرجع است که شامل مدل‌های K-NN، DT، MLP، GBM و LightGBM می‌باشد. با وجود عملکرد ضعیف مدل MLP در تمامی مجموعه داده‌ها، نتایج نشان می‌دهد که این مدل همچنان به‌طور قابل توجهی به مدلسازی میزان تهویه در ساختمان کمک می‌کند. در عوض، در مجموعه داده تست، غیاب مدل RF مشهود است. با توجه به نتایج قبلی، می‌توان نتیجه گرفت که مجموعه داده تست از پیچیدگی‌های بیشتری برخوردار است، بنابراین حتی با وجود الگوریتم ژنتیک، ۹ مدل از ۱۰ مدل ابتدایی برای مدلسازی حفظ شده‌اند.

نتایج جدول ۷ نشان‌دهنده این است که با انتخاب ترکیب صحیح از مدل‌های پایه، مدل گروهی انباشته می‌تواند تأثیر بسزایی در مدلسازی ایفا کند. بر اساس جدول ۷، مدل بهبود یافته در هر سه پارامتر ارزیابی، از مدل انباشته و GBM که از بهترین مدل‌ها در قسمت قبلی بوده‌اند، عملکرد بهتری داشته است. به‌طور خاص، مدل بهبود یافته در مجموع داده مرجع، با بیش از یک درصد بهبود در پارامتر R^2 به عدد ۰/۹۷۱۷ رسید. همچنین در پارامترهای RMSE و MAE بهبود شگفت‌انگیزی حاصل شده است، به طوری که نه تنها مدل بهبود یافته رشد داشته، بلکه از بهترین مدل‌های پیشین، یعنی مدل انباشته و GBM نیز بهتر عمل کرده است. به‌طور مشخص، مقادیر MAE و RMSE به ترتیب $6 \times (10)^{-6}$ و $5 \times (10)^{-5}$ و $13 \times (10)^{-5}$ به دست آمده است. شکل ۳ الف نشان‌دهنده دقت مناسب مدل ارائه شده در مدلسازی است. روش در نظر گرفته شده، عملکرد مطلوبی در داده تست نیز از خود نشان داده است، به‌خصوص با توجه به پیچیدگی‌های موجود در رفتار تهویه. جدول ۸ که نتایج داده تست در ماه سپتامبر را ارائه می‌دهد، نشان می‌دهد که پارامتر R^2 با ۴ درصد بهبود به عدد ۰/۸۹ رسیده است، که این امر بیانگر دقت و عملکرد مدل مذکور است. همچنین در پارامترهای RMSE و MAE بهبود مشهودی مشاهده شده و اعداد $5 \times (10)^{-5}$ و $12 \times (10)^{-6}$ و $27 \times (10)^{-6}$ به ترتیب محاسبه شده است.

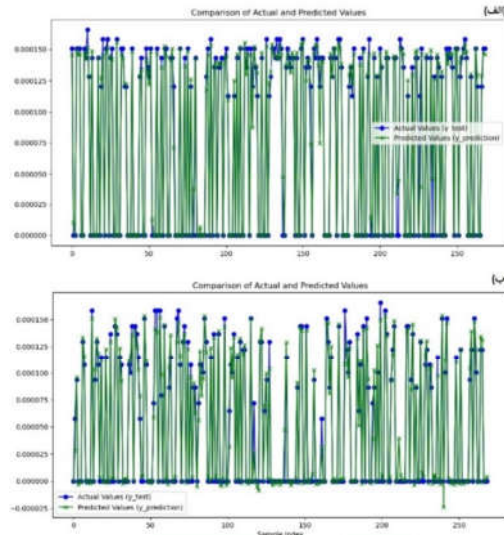
شایان ذکر است، روش ارائه شده به دلیل بهبود چشمگیر قابلیت به‌کارگیری در هر نوع داده‌ای را دارد. شکل ۳ ب نتایج مدل بهبود یافته و داده واقعی در مجموعه داده مرجع و واقعی در ماه سپتامبر را نشان می‌دهد.

جهت اعتبارسنجی روش ارائه شده، داده مرجع و تست در ماه نوامبر نیز مورد بررسی قرار گرفت تا اهمیت این روش در مدلسازی بیشتر نمایان گردد. نتایج جدول ۹ نشان می‌دهد که پس از به‌کارگیری الگوریتم ژنتیک برای انتخاب مناسب‌ترین الگوریتم پایه، در داده مرجع تعداد مدل‌ها به عدد ۵ تقلیل یافت. برای مدلسازی در ماه نوامبر، وجود مدل‌های K-NN، RF، LR، SVR و DT در پیش‌بینی اولیه نتایج قابل قبولی را ارائه می‌دهد. انتخاب مدل SVR توسط الگوریتم ژنتیک قابل توجه است، زیرا این مدل جزء مدل‌های با مقدار خطای قابل توجه بوده است. بنابراین، به دلیل ضعف در پارامترهای ارزیابی نمی‌توان این مدل را به‌عنوان مدل پایه کنار گذاشت، چراکه یادگیری بخشی از رفتارهای پیچیده داده بر عهده این مدل بوده است. بدین ترتیب، اهمیت روش ارائه شده در این پژوهش نمایان می‌شود.

با توجه به نتایج داده تست در جدول ۱۰، می‌توان نتیجه گرفت که تقلیل اندک مدل پایه پس از به‌کارگیری بهینه‌سازی، نشان‌دهنده وجود پیچیدگی و نوسان در نرخ تهویه است. این موضوع در پژوهش فورتین و همکاران نیز اشاره شده است. لذا می‌توان نتیجه گرفت مدل ارائه شده با حفظ ۸ مدل پایه از ۱۰ مدل اولیه، به خوبی توانسته است پیچیدگی‌های مدل را درک کند و بهترین نتیجه را در میان سایر مدل‌ها ارائه دهد که نشانگر دقت و قدرت مدل ارائه شده است. شکل ۴ الف و ب نشان‌دهنده دقت مدل ارائه شده در مجموعه داده مرجع و تست در ماه نوامبر است.



شکل ۳- نتایج مدل بهبود یافته، پیش‌بینی و داده واقعی در ماه سپتامبر، الف مجموعه داده مرجع، ب مجموعه داده تست



شکل ۴- نتایج مدل بهبود یافته، پیش‌بینی و داده واقعی در ماه نوامبر، الف مجموعه داده مرجع، ب مجموعه داده تست

جدول ۸- نتایج مدل بهبود یافته تست در ماه سپتامبر

الگوریتم	ارزیابی مدل‌سازی تست		
	R ²	RMSE(m/s)	MAE
GBM	۰/۸۱	۳/۴۶×(۱۰) ^{-۵}	۳/۱۹×(۱۰) ^{-۶}
Stacked EN	۰/۸۵	۱/۲۹×(۱۰) ^{-۵}	۷/۱۴×(۱۰) ^{-۶}
GA-Stacked	۰/۸۹	۱/۱۲×(۱۰) ^{-۵}	۲/۶۷×(۱۰) ^{-۶}

جدول ۹- نتایج مدل بهبود یافته داده مرجع در ماه نوامبر

الگوریتم	ارزیابی مدل‌سازی مرجع		
	R ²	RMSE(m/s)	MAE
GBM	۰/۹۵۵	۱/۱۵×(۱۰) ^{-۵}	۴/۲۲×(۱۰) ^{-۶}
Stacked EN	۰/۹۶۳	۱/۳۵×(۱۰) ^{-۵}	۵/۹×(۱۰) ^{-۶}
GA-Stacked	۰/۹۸۳	۹/۱۱×(۱۰) ^{-۶}	۳/۹۷×(۱۰) ^{-۶}

جدول ۱۰- نتایج مدل بهبود یافته تست در ماه نوامبر

الگوریتم	ارزیابی مدل‌سازی تست		
	R ²	RMSE(m/s)	MAE
GBM	۰/۹۵۷	۸/۷۸×(۱۰) ^{-۶}	۳/۹۳×(۱۰) ^{-۶}
Stacked EN	۰/۹۷۳	۹/۹۱×(۱۰) ^{-۶}	۴/۷۴×(۱۰) ^{-۶}
GA-Stacked	۰/۹۸۸	۸/۶۷×(۱۰) ^{-۶}	۳/۷۹×(۱۰) ^{-۶}

۴- نتیجه‌گیری

نتایج این پژوهش نشان‌دهنده کارایی بالای مدل‌سازی جعبه‌سیاه در مدل‌سازی تهویه طبیعی در ساختمان‌های دارای سرمایش غیرفعال است. با استفاده از دو مجموعه داده در ماه‌های سپتامبر و نوامبر، عملکرد دقیق مدل در پیش‌بینی تهویه طبیعی ارزیابی شد. این مدل با بهره‌گیری از تکنیک انباشته گروهی و انتخاب مدل‌های پایه مناسب توسط الگوریتم ژنتیک، بهینه‌سازی شده و برای داده‌های محدود و نامتوازن عملکرد مؤثری دارد.

نتایج پژوهش به طور خلاصه به شرح زیر است:

۱- استفاده از روش RFE به حذف پارامترهای کم‌اهمیت منجر شد. در ماه سپتامبر تعداد ویژگی‌ها به میزان ۴۰ درصد کاهش یافت و نشان داد که میزان اهمیت پارامترهای داخلی و بیرونی در این ماه مشابه است، در حالی که برای ماه نوامبر تفاوت‌های معناداری وجود داشت و پارامترهای داخلی در ساختمان مرجع و پارامترهای بیرونی در ساختمان تست اهمیت بیشتری داشتند.

۲- برای مدل‌سازی ساختمان با سرمایش طبیعی، ۱۰ مدل کلاسیک یادگیری ماشین به‌عنوان مدل‌های پایه در یک مدل انباشته گروهی استفاده شد. ارزیابی‌ها بر اساس پارامترهای RMSE، MAE و R² نشان داد که مدل گروهی انباشته نسبت به مدل‌های پایه در بهبود پارامتر R² موفق‌تر عمل کرده است. در ماه‌های سپتامبر و نوامبر، مدل انباشته گروهی به ترتیب بهبود ۰/۳۸٪، ۳/۲۴٪، ۰/۷۲٪ و ۰/۴۹٪ در ساختمان‌های مرجع و تست داشته است، هرچند که در دو پارامتر دیگر (RMSE و MAE) بهبود چشمگیری دیده نشد.

۳- برای ارتقاء مدل انباشته گروهی استاندارد، از الگوریتم ژنتیک برای انتخاب بهترین ترکیب از مدل‌های پایه استفاده شد. نتایج حاکی از آن است که مدل بهبود یافته نه تنها در پارامتر R² بلکه در دو پارامتر ارزیابی دیگر نیز عملکرد بهتری داشته و بر مدل‌های کلاسیک برتری یافته است.

مدل پیشنهادی، علاوه بر برخورداری از توانایی قابل توجه در مدل‌سازی نرخ تهویه در ساختمان‌های با سیستم سرمایش غیرفعال، می‌تواند به عنوان ابزاری کارآمد برای مدل‌سازی و ارائه تصمیمات دقیق در طراحی ساختمان‌های پایدار و بهینه انرژی توسط مهندسان و پژوهشگران به کار گرفته شود. همچنین، این مدل قادر است نرخ تهویه طبیعی را در تمامی ساختمان‌ها با شرط وجود داده، صرف نظر از نوع سیستم سرمایشی، به‌طور مؤثر مدل‌سازی کند.

۵- مراجع

- [۱] S. Vashisht and D. Rakshit, "Recent advances and sustainable solutions in automobile air conditioning systems," J. Clean Prod., vol. 329, p. 129754, Dec. 2021, doi: 10.1016/j.jclepro.2021.129754.
- [۲] A. R. Hao Wang, Xiwen Chen, Natan Vital, Edward Duffy, "Energy Optimization for HVAC Systems in Multi-VAV Open Offices: A Deep Reinforcement Learning Approach", [Online]. Available: <https://arxiv.org/abs/2306.13333>
- [۳] Y. Tang, Q. Tao, Y. Chen, J. Zheng, and Y. Min, "Building envelopes with radiative cooling materials: A model for indoor

- 2025, doi: 10.1016/j.patcog.2024.110886.
- [17] T. Song, W. Zhu, B. Pan, H. Song, Z. Chen, and M. Yue, "Development of ensemble learning techniques and sequential model-based optimization for enhancing the generalizability of shale wettability predictions," *Mar. Pet. Geol.*, vol. 168, p. 107010, Oct. 2024, doi: 10.1016/j.marpetgeo.2024.107010.
- [18] Z. Ma, G. Jiang, and J. Chen, "Physics-informed ensemble learning with residual modeling for enhanced building energy prediction," *Energy Build.*, vol. 323, p. 114853, Nov. 2024, doi: 10.1016/j.enbuild.2024.114853.
- [19] R. Fortin, J. Mandal, A. P. Raman, and S. Craig, "Passive radiative cooling to sub-ambient temperatures inside naturally ventilated buildings," *Cell Reports Phys. Sci.*, vol. 4, no. 9, p. 101570, Sep. 2023, doi: 10.1016/j.xcrp.2023.101570.
- [20] M. Krzywinski and N. Altman, "Visualizing samples with box plots," *Nat. Methods*, vol. 11, no. 2, pp. 119–120, Feb. 2014, doi: 10.1038/nmeth.2813.
- [21] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," *Technologies*, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.
- [22] A. Sharma and M. Singh, "Batch reinforcement learning approach using recursive feature elimination for network intrusion detection," *Eng. Appl. Artif. Intell.*, vol. 136, p. 109013, Oct. 2024, doi: 10.1016/j.engappai.2024.109013.
- [23] K. Li, J. Zhang, X. Chen, and W. Xue, "Building's hourly electrical load prediction based on data clustering and ensemble learning strategy," *Energy Build.*, vol. 261, p. 111943, Apr. 2022, doi: 10.1016/j.enbuild.2022.111943.
- [24] Z. Wang, Z. Liang, R. Zeng, H. Yuan, and R. S. Srinivasan, "Identifying the optimal heterogeneous ensemble learning model for building energy prediction using the exhaustive search method," *Energy Build.*, vol. 281, p. 112763, Feb. 2023, doi: 10.1016/j.enbuild.2022.112763.
- [25] J. Fan et al., "Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China," *Renew. Sustain. Energy Rev.*, vol. 100, pp. 186–212, Feb. 2019, doi: 10.1016/j.rser.2018.10.018.
- thermal environment assessment based on climate adaptation," *J. Build. Eng.*, vol. 74, p. 106869, Sep. 2023, doi: 10.1016/j.jobe.2023.106869.
- [4] Q. Gong, L. Lu, and J. Chen, "Design and performance investigation of a novel self-adaptive radiative cooling module for thermal regulation in buildings," *Appl. Energy*, vol. 352, p. 121928, Dec. 2023, doi: 10.1016/j.apenergy.2023.121928.
- [5] C.-L. Luo et al., "Enhanced passive radiative cooling coating with Y2O3 for thermal management of building," *Opt. Mater. (Amst.)*, vol. 138, p. 113710, Apr. 2023, doi: 10.1016/j.optmat.2023.113710.
- [6] K. Bataineh and A. Al Rabee, "A cost effective approach to design of energy efficient residential buildings," *Front. Archit. Res.*, vol. 11, no. 2, pp. 297–307, Apr. 2022, doi: 10.1016/j.foar.2021.10.004.
- [7] M. Klimczak, J. Bojarski, P. Ziembicki, and P. Kęskiewicz, "Analysis of the impact of simulation model simplifications on the quality of low-energy buildings simulation results," *Energy Build.*, vol. 169, pp. 141–147, Jun. 2018, doi: 10.1016/j.enbuild.2018.03.046.
- [8] S. Royer, S. Thil, T. Talbert, and M. Polit, "Black-box modeling of buildings thermal behavior using system identification," *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 10850–10855, 2014, doi: 10.3182/20140824-6-ZA-1003.01519.
- [9] Y. Shen and Y. Pan, "BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization," *Appl. Energy*, vol. 333, p. 120575, Mar. 2023, doi: 10.1016/j.apenergy.2022.120575.
- [10] Y. Hu et al., "Adaptive corrected parameters algorithm applied in cooling load prediction based on black-box model: A case study for subway station," *Energy Build.*, vol. 297, p. 113429, Oct. 2023, doi: 10.1016/j.enbuild.2023.113429.
- [11] B. Cui, P. Im, M. Bhandari, and S. Lee, "Performance analysis and comparison of data-driven models for predicting indoor temperature in multi-zone commercial buildings," *Energy Build.*, vol. 298, p. 113499, Nov. 2023, doi: 10.1016/j.enbuild.2023.113499.
- [12] Q. Al-Yasiri, M. Alktrane, M. Szabó, and M. Arıcı, "Building envelope-enhanced phase change material and night ventilation: Effect of window orientation and window-to-wall ratio on indoor temperature," *Renew. Energy*, vol. 218, p. 119263, Dec. 2023, doi: 10.1016/j.renene.2023.119263.
- [13] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Learning from Class-imbalanced Data with a Model-Agnostic Framework for Machine Intelligent Diagnosis," *Reliab. Eng. Syst. Saf.*, vol. 216, p. 107934, Dec. 2021, doi: 10.1016/j.rss.2021.107934.
- [14] M. N. Uddin et al., "Thermal and visual comforts of occupants for a naturally ventilated educational building in low-income economies: A machine learning approach," *J. Build. Eng.*, vol. 94, p. 110015, Oct. 2024, doi: 10.1016/j.jobe.2024.110015.
- [15] J. Ahmed and R. C. Green II, "Predicting severely imbalanced data disk drive failures with machine learning models," *Mach. Learn. with Appl.*, vol. 9, p. 100361, Sep. 2022, doi: 10.1016/j.mlwa.2022.100361.
- [16] Y. Pang, L. Peng, H. Zhang, Z. Chen, and B. Yang, "Imbalanced ensemble learning leveraging a novel data-level diversity metric," *Pattern Recognit.*, vol. 157, p. 110886, Jan.