

Research Article

Detection of High Impedance Faults in Electrical Distribution Systems Based on the RF-ICEEMDAN Approach

Mohammad Sadegh Attar, *Master's Degree*

Department of Electrical Engineering, Tafresh University, Tafresh, Iran, pwr400.Attar@tafreshu.ac.ir

Abstract:

One of the most critical faults in electrical distribution networks, often leading to hazardous incidents such as fires, is the high impedance fault (HIF). Due to the low magnitude of the fault current, this type of fault typically remains undetected by conventional protection devices, such as overcurrent relays. Moreover, the amplitude variations and waveform of HIF currents closely resemble those of other phenomena, such as variations in linear and nonlinear loads. In this paper, a Random Forest classification algorithm combined with an improved noise-assisted Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) is employed solely for the identification of high impedance faults. In the proposed RF-ICEEMDAN method, the dataset consisting of measured current signals from high impedance faults and load variations is first collected. The ICEEMDAN method is then used to decompose the signals and extract features from each sample. The extracted features are fed into the Random Forest classification algorithm. The proposed approach is implemented on the IEEE 34-bus distribution test system. The results demonstrate high accuracy and effectiveness in detecting high impedance faults. For simulation and evaluation purposes, EMTP-RV software is used for modeling the power system, and Python programming in the Google Colab environment is utilized for feature extraction and implementation of the Random Forest algorithm. The obtained results confirm the high performance and reliability of the proposed approach.

Keywords: High impedance fault, Random forest, Load switching, Fault current, ICEEMDAN method

Received: 15 March 2025

Revised: 19 May 2025

Accepted: 28 May 2025

*** Corresponding Author:** Mohammad Sadegh Attar

Citation: Mohammad Sadegh Attar, "Detection of High Impedance Faults in Electrical Distribution Systems Based on the RF-ICEEMDAN Approach", Journal of Novel Researches on Smart Power Systems, vol. 14, no. 1, pp. 83-103, June 2025 (in Persian)

مقاله پژوهشی

تشخیص خطای امپدانس بالا در شبکه توزیع انرژی الکتریکی با استفاده از روش RF-ICEEMDAN

محمدصادق عطار، کارشناسی ارشد

۱- دانشکده مهندسی برق، دانشگاه تفرش، تفرش، ایران، pwr400.Attar@tafreshu.ac.ir

چکیده: از مهم‌ترین خطاها که در شبکه‌های توزیع انرژی الکتریکی منجر به حوادث ناگواری هم‌چون آتش‌سوزی می‌شود، خطای امپدانس بالا می‌باشد. این نوع از خطاها به دلیل پایین بودن دامنه جریان، توسط تجهیزات حفاظتی مرسوم، هم‌چون رله‌های اضافه جریان، تشخیص داده نمی‌شوند. از طرفی دامنه تغییرات و شکل موج جریان خطای امپدانس بالا نزدیک به پدیده‌های دیگری هم‌چون تغییرات بارهای خطی و غیرخطی می‌باشد. در این مقاله از الگوریتم طبقه‌بندی جنگل تصادفی به همراه بهبود یافته تجزیه تجربی کامل مجموعه‌ای با نويز تطبیقی به منظور تنها شناسایی خطای امپدانس بالا استفاده شده است. در روش RF-ICEEMDAN، ابتدا مجموعه داده‌ها که شامل سیگنال‌های جریان‌های اندازه‌گیری شده خطای امپدانس بالا و تغییرات بارها می‌باشد، جمع‌آوری می‌شوند. سپس توسط روش ICEEMDAN تجزیه می‌شود تا ویژگی‌های هر نمونه از مجموعه داده‌ها، استخراج شود. ویژگی‌های استخراج شده از این بخش به عنوان ورودی به الگوریتم طبقه‌بندی جنگل تصادفی مورداستفاده قرار می‌گیرد. روش پیشنهادی بر روی شبکه توزیع تست ۳۴ شینه IEEE پیاده‌سازی شده است. در این روش، توانایی و دقت بالا در تشخیص خطای امپدانس بالا قابل مشاهده است. به منظور شبیه‌سازی و بررسی روش پیشنهادی، از نرم‌افزارهای شبیه‌سازی EMTP-rv برای مدل‌سازی شبکه قدرت و از زبان برنامه‌نویسی Python در محیط Google Colab برای استخراج ویژگی و پیاده‌سازی الگوریتم جنگل تصادفی استفاده شده است. نتایج به دست آمده، دقت بالای روش پیشنهادی را تأیید می‌کند.

کلمات کلیدی: خطای امپدانس بالا، جنگل تصادفی، تغییرات بار، جریان خطا، روش ICEEMDAN

تاریخ دریافت: ۱۴۰۴/۱۲/۲۵

تاریخ بازنگری: ۱۴۰۴/۲/۲۹

تاریخ پذیرش: ۱۴۰۴/۳/۷

* نام نویسنده مسئول: محمدصادق عطار

نشانی نویسنده مسئول: تفرش، دانشگاه تفرش، دانشکده فنی و مهندسی، گروه مهندسی برق

۱- مقدمه

انتقال انرژی الکتریکی با قابلیت اطمینان بالا از اولویت‌های اصلی در سیستم‌های قدرت می‌باشد. خطای امپدانس بالا^۱ (HIF) ناشی از اتصال هادی برق‌دار به زمین از طریق یک جسم با امپدانس بالا است که معمولاً با قوس الکتریکی همراه است. به دلیل جریان کم این خطا که نزدیک به جریان بار است، رله‌های اضافه جریان قادر به شناسایی این خطا نیستند. لذا این موضوع چالشی مهم در شبکه‌های توزیع انرژی الکتریکی محسوب می‌شود. خطاهای امپدانس بالا ویژگی‌های غیرخطی مشابهی با پدیده‌هایی مانند سوئیچینگ بار، بانک‌های خازنی، و جریان هجومی ترانسفورماتورها دارند. این شباهت، شناسایی آن‌ها را پیچیده‌تر کرده و ضرورت استفاده از تجهیزات حفاظتی دقیق را افزایش می‌دهد. عدم تشخیص صحیح این خطاها می‌تواند به خطرات جدی، از جمله برق‌گرفتگی، تلفات جانی و مالی، قطعی شبکه، و حتی آتش‌سوزی منجر شود. بنابراین، شناسایی سریع و دقیق خطاهای امپدانس بالا از اهمیت بالایی برخوردار است [۱، ۲].

برای شناسایی خطاهای امپدانس بالا، روش‌های متنوعی ارائه شده است که بسیاری از آن‌ها بر پایه یادگیری ماشین هستند. در [۳] مطالعه‌ای جامع بر روی انواع روش‌ها به منظور تشخیص خطای امپدانس بالا و مقایسه این خطا با سایر پدیده‌های مشابه در شبکه‌های توزیع انرژی الکتریکی اشاره کرده است. به عنوان مثال، روش‌هایی مانند استفاده از ماشین بردار پشتیبان در تحلیل پیش‌بینی و وضعیت سیستم‌ها [۴] و همچنین در تشخیص این نوع از خطاها [۵-۹]، کانولو شنال چگال^۲ [۱۰]، و تکنیک‌های یادگیری عمیق مانند شبکه حافظه طولانی کوتاه مدت^۳ (LSTM) [۱۱، ۱۲] توسعه یافته‌اند. برخی از پژوهش‌ها نیز از الگوریتم‌های مبتنی بر موجک [۱۳-۱۵] و مدل‌های شبکه عصبی [۱۶-۱۸] بهره برده‌اند. استفاده از تجزیه حالت تجربی و نسخه‌های پیشرفته آن در شناسایی خطاهای امپدانس بالا بسیار مورد توجه بوده است. در [۱۹]، از روش تجزیه حالت تجربی همراه با شبکه عصبی مصنوعی برای تحلیل سیگنال‌های قوس در زمان واقعی استفاده شده است. همچنین، در [۲۰] از روش تجزیه حالت تجربی تجمعی کامل با داده کمکی نويز^۴ (CEEMDAN)، نسخه‌ای بهبود یافته از تجزیه حالت تجربی^۵ (EMD)، برای تشخیص خطای امپدانس بالا استفاده شده است. در [۲۱] از روش‌های یادگیری ماشین به‌ویژه تکنیک‌های بدون نظارت مانند خود رمزگذار^۶ و تحلیل مؤلفه‌های اصلی، به دلیل توانایی در شناسایی ناهنجاری‌ها بدون نیاز به داده‌های برچسب‌دار استفاده کرده است. این روش‌ها قادرند تغییرات ظریف در شکل موج جریان را که ناشی از خطاهای امپدانس بالا هستند، تشخیص دهند. مطالعات اخیر نشان داده‌اند که ترکیب این تکنیک‌ها با داده‌های شبیه‌سازی شده و واقعی می‌تواند دقت و قابلیت اطمینان تشخیص خطاهای امپدانس بالا را به‌طور قابل توجهی بهبود بخشد. در [۲۲] به تشخیص خطای امپدانس بالا در یک شبکه توزیع واقعی با استفاده از روش ترکیبی تبدیل موجک گسسته به همراه ماشین بردار پشتیبان اشاره کرده است. مرجع [۲۳]، یک روش نوین دو مرحله‌ای مبتنی بر داده برای تشخیص خطاهای امپدانس بالا پیشنهاد می‌کند که از تخمین امپدانس خط استفاده کرده و نتایج تخمین زده شده را با اطلاعات محل خطا تطبیق می‌دهد. این روش هر خط را به‌طور مجزا هدف قرار داده و امکان تشخیص دقیق خطاهای امپدانس بالا بین گره‌های اندازه‌گیری خاص را فراهم می‌کند. نتایج آزمایش‌ها روی سیستم ۳۳ شینه‌ی IEEE نشان می‌دهد که این روش قادر است خطاهای امپدانس بالا را در شرایط بدون نویز و با دقت ۹۱٪/۷۲ تشخیص دهد، که اثربخشی آن روش را تأیید می‌کند. مرجع [۲۴]، با استفاده از تشخیص‌دهنده‌های انرژی، راه‌حلی برای شبکه توزیع موجود پیشنهاد می‌کند. تشخیص‌دهنده انرژی از دو بخش اصلی تشکیل شده است: (۱) تشخیص‌دهنده انرژی مبتنی بر «ولتاژ سمت منفی - جریان سمت منفی» (۲) تشخیص‌دهنده انرژی مبتنی بر «ولتاژ سمت منفی - جریان سمت منفی». این تشخیص‌دهنده‌ها در کنار رله‌های حفاظتی اضافه‌جریان موجود استفاده می‌شوند، به‌طوری‌که رله‌های اضافه‌جریان از خطای امپدانس بالا محافظت می‌کنند و تشخیص‌دهنده‌های انرژی، خطاهای امپدانس بالا را تشخیص می‌دهند. تشخیص‌دهنده‌ها بر اساس آستانه‌های انرژی تلف‌شده در طول خطاهای امپدانس بالا و مناطق همگرایی مربوطه عمل می‌کنند. عملکرد این سیستم روی یک سیستم ساده ۷ شینه در یک پلتفرم شبیه‌سازی آزمایش شده و با داده‌های واقعی از پست ۱۱/۳۳ کیلوولت اعتبارسنجی شده است. این سیستم راه‌حلی کم‌هزینه و قابل اعتماد برای مشکل خطاهای امپدانس بالا ارائه می‌دهد. در [۲۵]، یک روش طبقه‌بندی نیمه‌نظارت شده برای تشخیص خطای امپدانس بالا و سایر حالت‌های گذرا در یک

ریز شبکه متصل به سیستم فتوولتائیک پیشنهاد می‌کند. روش پیشنهادی ترکیبی از خوشه‌بندی K-میانگین (بدون نظارت) و الگوریتم شبکه عصبی چندلایه با نظارت است. در مرحله اول، از تکنیک K-میانگین برای شناسایی و حذف نمونه‌های نامرتبط از مجموعه داده‌ها استفاده می‌شود. سپس، داده‌های خوشه‌بندی شده به‌عنوان ورودی به شبکه عصبی پرسپترون چندلایه^۷ (MLP) داده می‌شوند تا پیش‌بینی نهایی برچسب‌ها انجام شود. این روش ترکیبی نسبت به استفاده جداگانه از K-میانگین یا شبکه عصبی پرسپترون چندلایه، دقت و نرخ موفقیت بالاتری در تشخیص خطای امپدانس بالا تحت شرایط آزمایش استاندارد و تغییرات آب‌وهوایی فتوولتائیک^۸ (PV) نشان می‌دهد. هم‌چنین، نتایج مطالعه عملکردی نشان می‌دهد که این مدل در مقایسه با روش‌های مستقل، مقاوم‌تر بوده و عملکرد بهتری تحت شرایط عدم قطعیت فتوولتائیک در ریز شبکه‌ها ارائه می‌دهد. مرجع [۲۶]، یک الگوریتم جدید برای تشخیص خطای امپدانس بالا در سیستم‌های توزیع با استفاده از قابلیت‌های برتر تبدیل S ارائه می‌دهد. در این روش، تبدیل S بر روی یک سیگنال جریان اعمال شده و اطلاعات زمانی و فرکانسی سیگنال استخراج می‌شود. سپس، چهار شاخص عددی شامل اعوجاج هارمونیک کل، انرژی هارمونیک زوج، ضریب تغییرات و انحراف فاز از خروجی تبدیل S محاسبه می‌شوند. بر اساس مقادیر این شاخص‌ها و برخی آستانه‌های از پیش تعیین شده، خطاهای امپدانس بالا از شرایط عادی و سایر رویدادهای سیستم توزیع تشخیص داده می‌شوند. عملکرد این الگوریتم با استفاده از سیگنال‌های شبیه‌سازی شده توسط نرم‌افزار PSCAD/EMTDC و سیگنال‌های آزمایشی واقعی از یک شبکه توزیع ارزیابی شده است. مرجع [۲۷]، یک روش پیشرفته تشخیص خطاهای امپدانس بالا مبتنی بر یادگیری افزایشی تفسیرپذیری^۹ ارائه می‌دهد تا تفسیرپذیری مدل‌های داده‌محور در تشخیص خطا را بهبود بخشد. در این روش، از تبدیل موجک گسسته^{۱۰} (DWT) برای استخراج ضرایب موجک جریان سمت صفر استفاده شده و با محاسبه انحراف معیار، دید کمی از خطاهای امپدانس بالا ارائه می‌شود. سپس، یک شبکه عصبی مصنوعی با استفاده از یک اصل مبتنی بر رگولاریزاسیون بهبود می‌یابد. این رویکرد، پارامترهای مدل را به‌صورت پویا بهینه کرده و ضمن یادگیری دانش جدید از جریان‌های داده موج‌شکل، از فراموشی دانش قبلی جلوگیری می‌کند. عملکرد روش پیشنهادی با استفاده از داده‌های شبیه‌سازی شده توسط پلتفرم نرم‌افزاری PSCAD/EMTDC و داده‌های واقعی اعتبارسنجی شده است.

این روش‌ها به دلیل توانایی در استخراج ویژگی‌های غیرخطی و جزئیات فرکانسی سیگنال‌ها، عملکرد بهتری در مقایسه با تکنیک‌های سنتی ارائه می‌دهند. در پژوهش‌های دیگر نیز از ابزارهای مشابه مانند تکنیک‌های پیش‌پردازش پیشرفته هم‌چون تبدیل ویگنر-ویل^{۱۱} [۲۸] استفاده شده است. با این حال، روش‌های تجزیه حالت تجربی و تجزیه حالت تجربی تجمعی کامل با داده کمی‌نویز به دلیل دقت بالا و قابلیت تطبیق با سیگنال‌های غیرایستا، جایگاه ویژه‌ای در این حوزه دارند. این مطالعات نشان می‌دهند که استفاده از تکنیک‌های تجزیه سیگنال و یادگیری ماشین نقش کلیدی در شناسایی دقیق و سریع خطاهای امپدانس بالا ایفا می‌کند. روش‌های ارائه شده در مقالات گذشته با محدودیت‌هایی نظیر دقت پایین، پیچیدگی‌های ریاضیاتی، و استفاده از روش‌هایی ابتدایی در حوزه استخراج ویژگی مواجه بوده‌اند. این مشکلات ضرورت توسعه روشی با دقت بالا، معیارهای ارزش‌یابی قابل اعتماد را برجسته می‌کنند. چنین روشی باید با استفاده از زبان‌های برنامه‌نویسی و نرم‌افزارهای رایج صنعتی، راه‌حلی عملی و مؤثر برای این مسئله ارائه دهد.

در این مقاله، به‌منظور رفع محدودیت‌ها و بهبود دقت در شناسایی خطاهای امپدانس بالا، از روش تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی^{۱۲} (ICEEMDAN) برای استخراج ویژگی‌ها استفاده شده است. این روش، در مقایسه با روش‌های پیشین، توانایی بالایی در تحلیل سیگنال‌های غیرخطی و نویزی نشان می‌دهد. ویژگی‌های استخراج شده از روش پیشنهادی به‌عنوان ورودی الگوریتم جنگل تصادفی^{۱۳} (RF) به کار گرفته شده و داده‌ها به دو کلاس صفر و یک طبقه‌بندی شده‌اند. از طرفی به‌منظور افزایش کیفیت روش پیشنهادی در این مقاله، ویژگی داده‌ها با استفاده از تجزیه حالت تجربی تجمعی کامل با داده کمی‌نویز استخراج شده و مجدد به‌عنوان ورودی به الگوریتم جنگل تصادفی داده می‌شود. سپس نتایج به دست آمده با نتایج روش اصلی مقایسه می‌شود. نتایج نشان می‌دهند که روش پیشنهادی نسبت به روش تجزیه حالت تجربی تجمعی کامل با داده کمی‌نویز عملکرد بهتری در دقت شناسایی و تحلیل سیگنال‌های خطاهای امپدانس بالا دارد. این مقایسه در نهایت به انتخاب

روش تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی به عنوان روش بهینه برای این کاربرد منجر می شود. روش پیشنهادی تنها بر روی شبکه تست استاندارد ۳۴ شینه IEEE پیاده سازی شده است. در این پژوهش، از زبان برنامه نویسی Python به عنوان محیطی شی گرا و سطح بالا همراه با نرم افزار EMTP-RV استفاده شده است. داده های جمع آوری شده از شبیه سازی شبکه تست، پس از پردازش و استخراج ویژگی ها، به الگوریتم طبقه بندی داده شده و نتایج آن برای ارزیابی دقت و عملکرد روش پیشنهادی تحلیل شده اند.

۲- روش تحقیق

در این مقاله، یک روش جدید مبتنی بر تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی برای استخراج ویژگی های سیگنال های خطاهای امپدانس بالا در شبکه های توزیع انرژی الکتریکی پیشنهاد شده است. هدف استفاده از این روش، بهبود دقت شناسایی و کاهش نويز در سیگنال ها می باشد. در این بخش، مراحل مختلف پردازش داده ها، استخراج ویژگی ها، و طبقه بندی سیگنال ها به شرح زیر بیان شده است.

مرحله اول: پیش پردازش داده ها

در ابتدا، داده های ورودی به دست آمده از سیگنال های جریان شبکه های توزیع، برای پردازش آماده می شود. این داده ها شامل سیگنال های جریان خطاهای امپدانس بالا و سایر اختلالات شبکه هستند. پیش پردازش شامل فیلتر کردن سیگنال ها، حذف نويزهای اولیه و تنظیم داده ها برای مراحل بعدی است.

مرحله دوم: استخراج ویژگی با استفاده از روش تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی

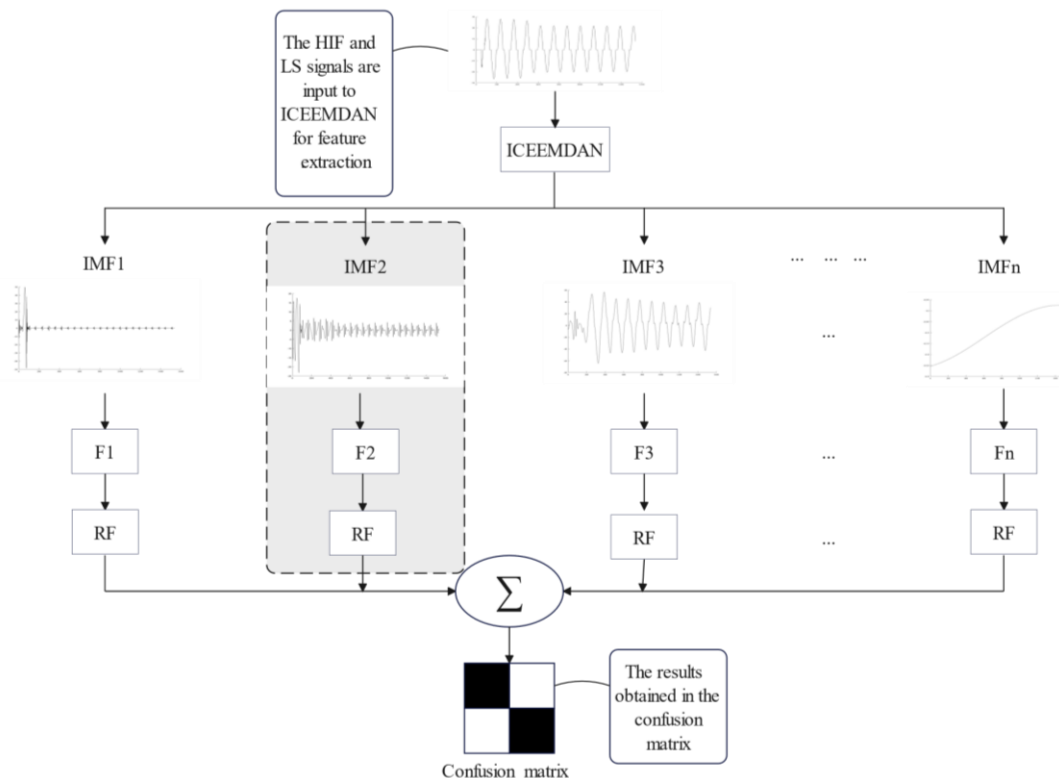
در این مرحله، از روش پیشنهادی برای تجزیه و تحلیل سیگنال ها استفاده شده است. این روش به عنوان یک ابزار مؤثر برای استخراج ویژگی از سیگنال های غیر خطی و پیچیده شناخته شده است. این روش سیگنال های ورودی را به مولفه های مختلف^{۱۴} (IMF) تجزیه کرده و ویژگی های مختلفی مانند انرژی، انحراف معیار و سایر ویژگی های آماری از هر مؤلفه استخراج می شود. در ادامه، ویژگی های استخراج شده برای مراحل طبقه بندی استفاده شده است.

مرحله سوم: طبقه بندی با استفاده از الگوریتم جنگل تصادفی

ویژگی های استخراج شده از روش پیشنهادی به عنوان ورودی به الگوریتم جنگل تصادفی داده شده است. این الگوریتم یک روش قدرتمند در طبقه بندی داده ها می باشد که سیگنال های خطاهای امپدانس بالا را از سایر اختلالات شبکه تفکیک می کند. داده ها به دو کلاس صفر (بدون خطا) و یک (خطای امپدانس بالا) تقسیم بندی شده و مدل جنگل تصادفی به منظور شناسایی خطاها آموزش داده شده است.

مرحله چهارم: ارزیابی عملکرد

برای ارزیابی عملکرد روش پیشنهادی، نتایج طبقه بندی با استفاده از معیارهای مختلف ارزیابی مانند دقت، حساسیت، اختصاصیت و امتیاز F1 بررسی شده است. این مقایسه با روش های مشابه مانند تجزیه حالت تجربی تجمعی کامل با داده کمکی نويز انجام شده تا برتری روش پیشنهادی به طور دقیق تر مشخص شود. شکل (۱) شماتیک کلی روش پیشنهادی را نشان می دهد.



شکل (۱): شماتیک کلی روش پیشنهادی

Figure (1): General schematic of the proposed method

۲-۱- جمع‌آوری و آماده‌سازی داده‌های ورودی

در این پژوهش از سیگنال‌های جریان برای تشخیص خطاهای امپدانس بالا استفاده شده است. مطالعات صورت گرفته، بر روی شبکه توزیع تست ۳۴ شینه IEEE می‌باشد. مدل پیشنهادی خطای امپدانس بالا به شبکه تست ۳۴ شینه IEEE متصل شده است. سپس سیگنال‌های جریان خطاهای امپدانس بالا به‌عنوان داده به جهت استخراج ویژگی در مرحله بعد، جمع‌آوری شده است.

۲-۲- روش تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی

تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی یک روش تجزیه سیگنال است که هدف آن کاهش خطاهای تجزیه و تفکیک بهتر مؤلفه‌های سیگنال به اجزای مختلف است. این روش به‌طور خاص برای سیگنال‌های غیرخطی مناسب است. این روش به‌عنوان یک نسخه بهبود یافته از روش تجزیه حالت تجربی تجمعی توسعه یافته است. تجزیه حالت تجربی تجمعی خود نسخه‌ای اصلاح شده از روش تجزیه حالت تجربی می‌باشد. تجزیه حالت تجربی یک روش غیرخطی است که برای تجزیه سیگنال‌های پیچیده به اجزای ساده‌تر و قابل تحلیل استفاده می‌شود. هدف از تجزیه حالت تجربی این است که یک سیگنال پیچیده را به مجموعه‌ای از مؤلفه‌ها تجزیه کند. این اجزا به‌طور طبیعی ویژگی‌های فرکانسی مختلف سیگنال را نمایان می‌کنند. یکی از ویژگی‌های مهم این روش این است که نیازی به فرضیه‌های اولیه مثل خطی بودن یا ایستایی سیگنال ندارد. روش تجزیه حالت تجربی به‌طور کلی شامل مراحل هم‌چون استخراج نوسانات مقیاس‌های بالا، استخراج مؤلفه‌ها و باقی‌مانده‌ها تقسیم شده است. روش تجزیه حالت تجربی تجمعی کامل با داده کمکی نويز به‌عنوان نسخه‌ای پیشرفته‌تر از تجزیه حالت تجربی، با هدف کاهش مشکلات نويز و بهبود دقت تجزیه سیگنال‌های پیچیده توسعه یافته است. در این روش، به‌جای اضافه کردن تنها یک نويز به سیگنال، از نويز تصادفی و متعدد استفاده شده است. از ویژگی‌های این روش می‌توان به استفاده از مجموعه‌ای از نويزها و میانگین‌گیری از مؤلفه‌های به‌دست‌آمده اشاره نمود.

روش پیشنهادی یک بهبود دیگر از روش تجزیه حالت تجربی تجمعی کامل با داده کمکی نویز است که به طور خاص برای بهبود دقت تجزیه سیگنال‌ها و کاهش اثرات نویز طراحی شده است. تفاوت اصلی این دو روش در نحوه اضافه کردن نویز است. در روش پیشنهادی، نویز به صورت آدپتیو یا تطبیقی به سیگنال اضافه می‌شود. این روش شامل ویژگی‌های منحصر به فردی می‌باشد. برخلاف روش دیگر که نویز به صورت ثابت به سیگنال‌ها اضافه می‌شود، در روش پیشنهادی نویز به طور آدپتیو و با توجه به ویژگی‌های سیگنال اضافه می‌شود. این نویز به گونه‌ای طراحی شده است که بر اساس ویژگی‌های سیگنال اصلی تغییر می‌کند، که این امر باعث افزایش دقت در استخراج مؤلفه‌ها و کاهش اثرات نویز می‌شود. استفاده از نویز آدپتیو باعث شده است که این روش بتواند بهتر از دیگر روش‌ها، سیگنال‌های پیچیده و دارای نویز را تجزیه کند. این امر به ویژه در سیگنال‌هایی که ویژگی‌های متغیر و غیرخطی دارند، مهم است. همچنین اضافه کردن نویز آدپتیو سبب بهبود کیفیت مؤلفه‌ها و کاهش خطا در تحلیل سیگنال شده است. این روش می‌تواند مشکلاتی که در روش‌های قبلی وجود داشت (مانند نویز زیاد یا تجزیه نادرست) را به حداقل برساند. روش پیشنهادی به ویژه برای سیگنال‌هایی که رفتار غیرخطی دارند، بسیار مناسب است. این ویژگی روش بیان شده را از دیگر روش‌ها متمایز می‌کند.

پس باید بیان کرد که تجزیه حالت تجربی به دلیل سادگی و توانایی در تجزیه سیگنال‌های پیچیده، به طور گسترده‌ای استفاده می‌شود اما به دلیل حساسیت زیاد به نویز و مشکلات در تفکیک مؤلفه‌ها، محدودیت‌هایی دارد. از طرفی تجزیه حالت تجربی تجمعی کامل با داده کمکی نویز این مشکلات را با استفاده از نویزهای تصادفی متعدد حل کرده و دقت تجزیه را بهبود می‌بخشد، اما همچنان با مشکلات نویزهایی که به صورت ثابت به سیگنال‌ها اضافه می‌شود، مواجه است. در نهایت روش تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی به عنوان پیشرفته‌ترین نسخه، با استفاده از نویز آدپتیو و بهینه‌سازی فرآیند تجزیه، بهترین دقت را در تحلیل سیگنال‌های پیچیده و دارای نویز فراهم می‌کند. رابطه (۱) ترکیب سیگنال اصلی و نویز تصادفی را نشان می‌دهد.

$$\eta(t) + x(t) = y(t) \quad (1)$$

که در آن، $y(t)$ سیگنال با نویز، $x(t)$ سیگنال اصلی و $\eta(t)$ نویز تصادفی می‌باشد. با استفاده از تجزیه حالت تجربی، سیگنال $y(t)$ به مجموعه‌ای از مؤلفه‌ها تجزیه می‌شود. در این مرحله، سیگنال به مؤلفه‌هایی با فرکانس‌های مختلف تقسیم می‌شود. رابطه کلی (۲) برای تجزیه به مؤلفه‌ها به قرار زیر است.

$$r_n(t) + \text{IMF}_i(t) \sum_{i=1}^n = x(t) \quad (2)$$

که در آن، $x(t)$ سیگنال اصلی، $\text{IMF}_i(t)$ مؤلفه‌های درونی و $r_n(t)$ باقی مانده است که به عنوان روند نهایی سیگنال در نظر گرفته شده است.

در ادامه مجموعه‌ای از نویزها که به طور موازی چندین بار نویزهای مختلف را به سیگنال اصلی اضافه می‌کند و فرآیند تجزیه حالت تجربی را برای هر بار اضافه کردن نویز تکرار می‌کند، اعمال شده است. این فرآیند به طور کلی برای چندین مجموعه از نویزها انجام می‌شود. پس از انجام تجزیه سیگنال برای مجموعه‌های مختلف از نویزها میانگین مؤلفه‌های به دست آمده محاسبه می‌شود تا مؤلفه‌های نهایی با دقت بیشتری استخراج شود. رابطه (۳) این موضوع را نشان می‌دهد.

$$\text{IMF}_i^n(t) \sum_{n=1}^N \frac{1}{N} = \text{IMF}_i^{\text{final}}(t) \quad (3)$$

که در آن، $\text{IMF}_i^{\text{final}}(t)$ نیز مؤلفه‌های تجزیه حالت تجربی نهایی، $\text{IMF}_i^n(t)$ در n -امین نمونه از نویز و N تعداد نمونه‌های نویز است.

از ویژگی‌های روش پیشنهادی می‌توان به کاهش نویز (توانایی کاهش نویزهای موجود در سیگنال)، دقت بیشتر در استخراج مؤلفه‌ها (به دلیل استفاده از میانگین‌گیری بر روی چندین تجزیه، دقت استخراج مؤلفه‌ها بهبود می‌یابد) و مناسب برای سیگنال‌های غیرخطی و پیچیده (به ویژه برای سیگنال‌هایی که دارای رفتار غیرخطی هستند، مناسب است) اشاره نمود.

۲-۲-۱- اعمال روش پیشنهادی برای سیگنال خطای امپدانس بالا

برای استخراج ویژگی‌های سیگنال خطای امپدانس بالا، ابتدا سیگنال‌های ولتاژ و جریان که حاوی خطاهای امپدانس بالا هستند، تحت تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی قرار می‌گیرند. در مرحله اول، سیگنال‌های ورودی از شبکه توزیع که شامل سیگنال‌های خطاهای امپدانس بالا هستند، وارد فرآیند تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی می‌شوند. این سیگنال‌ها به مجموعه‌ای از مؤلفه‌ها تجزیه می‌شوند. در مرحله دوم، انتخاب مؤلفه مناسب از میان مؤلفه‌های استخراج شده، برای استخراج ویژگی‌ها از مؤلفه‌های دوم استفاده می‌شود. دلیل انتخاب مؤلفه‌های دوم این است که این مؤلفه معمولاً شامل ویژگی‌های مهم و نوسانات متوسط سیگنال‌های خطاهای امپدانس بالا است که می‌تواند اطلاعات مفیدی در مورد خطا ارائه دهد.

۲-۲-۲- استخراج ویژگی‌ها از مؤلفه‌های دوم

پس از استخراج مؤلفه‌های دوم از سیگنال‌های خطاهای امپدانس بالا، ویژگی‌های مختلفی از آن استخراج می‌شود. ویژگی‌هایی که در این تحقیق استخراج می‌شوند، شامل موارد زیر هستند:

میانگین دامنه^{۱۵}

میانگین دامنه یک ویژگی ساده است که معمولاً برای سیگنال‌های با نوسانات کوچک یا بدون نوسانات بزرگ استفاده می‌شود. این ویژگی میزان "میانگین" سیگنال را با در نظر گرفتن مقادیر مطلق سیگنال اندازه‌گیری می‌کند. به عبارت دیگر، این ویژگی میزان حرکت سیگنال را بدون توجه به علامت آن (مثبت یا منفی بودن) اندازه‌گیری می‌کند. رابطه (۴) میانگین دامنه را نشان می‌دهد.

$$|x_i| \sum_{i=1}^N \frac{1}{N} = \text{Mean Amplitude} \quad (4)$$

که در آن x_i هر نقطه از سیگنال مؤلفه‌های دوم و N تعداد نقاط سیگنال است. به عبارتی $|x_i|$ مقدار مطلق هر نقطه از سیگنال است که به طور مستقیم نشان دهنده اندازه نوسانات در سیگنال می‌باشد. این ویژگی برای شناسایی سیگنال‌های با رفتار نوسانی استفاده می‌شود. سیگنال‌های خطاهای امپدانس بالا معمولاً دارای نوسانات بزرگ و شدید هستند که می‌توانند به طور ناگهانی تغییر کنند. میانگین دامنه به طور مستقیم می‌تواند شدت این نوسانات را اندازه‌گیری کند. سیگنال‌های خطاهای امپدانس بالا دارای دامنه بزرگ و تغییرات واضح هستند، و این ویژگی می‌تواند به شناسایی نوسانات خاص در این سیگنال‌ها کمک کند.

انرژی^{۱۶}

انرژی سیگنال به طور معمول به عنوان معیاری برای اندازه‌گیری قدرت سیگنال در یک بازه زمانی خاص است. این ویژگی به ویژه در تحلیل سیگنال‌هایی که در آن‌ها سیگنال‌های پر قدرت مهم هستند، مفید است. انرژی یک سیگنال معمولاً به عنوان مجموع مربعات سیگنال محاسبه می‌شود که در رابطه (۵) نشان داده شده است.

$$x_i^2 \sum_{i=1}^N = \text{Energy} \quad (5)$$

که در آن x_i^2 هر نقطه از سیگنال مؤلفه‌های دوم است. انرژی سیگنال در شناسایی خطاهای امپدانس بالا بسیار مهم است، زیرا این خطاها باعث افزایش ناگهانی قدرت سیگنال می‌شوند. اندازه‌گیری انرژی می‌تواند به راحتی سیگنال‌های خطاهای امپدانس بالا را از سیگنال‌های نرمال جدا کند، زیرا خطاهای امپدانس بالا معمولاً دارای انرژی بیش‌تری نسبت به سیگنال‌های معمولی هستند. این ویژگی می‌تواند تغییرات ناگهانی در قدرت سیگنال را شناسایی کرده و به تشخیص صحیح این خطا کمک کند.

انحراف معیار^{۱۷}

انحراف معیار یک معیار آماری است که میزان پراکندگی یا پراش سیگنال را از میانگین آن اندازه‌گیری می‌کند. این ویژگی به شدت برای تحلیل سیگنال‌های با نوسانات بزرگ یا کوچک و شناسایی تغییرات در سیگنال‌های خطاهای امپدانس بالا مفید

است. هرچه انحراف معیار بزرگتر باشد، یعنی سیگنال بیش‌تر از مقدار میانگین خود تغییر کرده است. رابطه (۶) انحراف معیار را نشان می‌دهد.

$$\sqrt{(x_i - \mu)^2 \sum_{i=1}^N \frac{1}{N}} = \text{StdDev} \quad (6)$$

که در آن μ میانگین سیگنال است. تغییرات ناگهانی و زیاد در سیگنال‌های خطاهای امپدانس بالا معمولاً باعث افزایش انحراف معیار می‌شوند. این ویژگی قادر است نوسانات زیاد سیگنال را که معمولاً در اثر وقوع این خطا ایجاد می‌شود، شناسایی کند. بنابراین، انحراف معیار یک شاخص کلیدی در تشخیص خطاهایی است که با نوسانات شدید همراه هستند.

پیک تا پیک^{۱۸}

ویژگی پیک تا پیک نشان‌دهنده تفاوت بین بزرگ‌ترین و کوچک‌ترین مقدار در سیگنال است. این ویژگی برای شناسایی نوسانات شدید در سیگنال خطای امپدانس بالا اهمیت دارد. اگر در یک سیگنال تغییرات ناگهانی و بزرگ رخ دهد، این ویژگی قادر به اندازه‌گیری آن است. رابطه (۷) مقدار پیک تا پیک را نشان می‌دهد.

$$\min(x) - \max(x) = \text{Peak_to_Peak} \quad (7)$$

که در آن به ترتیب $\max(x)$ بزرگ‌ترین و $\min(x)$ کوچک‌ترین مقدار سیگنال می‌باشد. این ویژگی می‌تواند تفاوت بین بزرگ‌ترین و کوچک‌ترین مقادیر سیگنال را اندازه‌گیری کند. در سیگنال‌های خطاهای امپدانس بالا، تغییرات ناگهانی و بزرگ معمولاً باعث افزایش مقدار پیک تا پیک می‌شوند. این ویژگی به شناسایی نوسانات شدید و غیرمنتظره که ممکن است نشان‌دهنده خطاهای امپدانس بالا باشند، کمک می‌کند.

چولگی^{۱۹}

چولگی یکی از ویژگی‌های آماری است که میزان عدم تقارن توزیع یک سیگنال حول میانگین آن را اندازه‌گیری می‌کند. اگر چولگی مثبت باشد، به این معنی است که توزیع سیگنال به سمت راست کشیده شده است، و اگر چولگی منفی باشد، توزیع به سمت چپ کشیده شده است. رابطه (۸) چولگی را نشان می‌دهد.

$$\left(\frac{x_i - \mu}{\sigma}\right)^3 \sum_{i=1}^N \frac{N}{(N-2)(N-1)} = \text{Skewness} \quad (8)$$

که در آن، μ میانگین سیگنال، σ انحراف معیار سیگنال، x_i هر نقطه از سیگنال مولفه‌های دوم و N تعداد نقاط سیگنال است. در سیگنال‌های خطای امپدانس بالا، توزیع سیگنال‌ها به‌طور معمول از حالت متقارن خارج می‌شود. چولگی می‌تواند به خوبی عدم تقارن توزیع سیگنال‌های خطا را نشان دهد و به شناسایی سریع سیگنال‌های غیرطبیعی و خطاهای احتمالی کمک کند. سیگنال‌های خطای امپدانس بالا معمولاً دارای چولگی مثبت یا منفی هستند که با نوسانات غیرقابل پیش‌بینی ایجاد می‌شوند.

کشیدگی^{۲۰}

کشیدگی یک ویژگی آماری است که میزان "نوک‌داری" یا "پهنای" توزیع سیگنال را اندازه‌گیری می‌کند. کشیدگی بالا نشان‌دهنده توزیع‌های با پیک‌های تیز است، در حالی که کشیدگی پایین نشان‌دهنده توزیع‌های پهن‌تر و صاف‌تر است. رابطه (۹) کشیدگی را نشان می‌دهد.

$$\frac{(N-1)^2 3}{(N-3)(N-2)} - \left(\frac{x_i - \mu}{\sigma}\right)^4 \sum_{i=1}^N \frac{N(N+1)}{(N-3)(N-2)(N-1)} = \text{Kurtosis} \quad (9)$$

کشیدگی می‌تواند پیک‌های تیز و نوک‌دار در سیگنال‌ها را شناسایی کند که در سیگنال‌های خطاهای امپدانس بالا به‌طور خاص مشاهده می‌شود. این سیگنال‌ها معمولاً ویژگی‌های کشیدگی بالاتری دارند که نشان‌دهنده تغییرات ناگهانی و پر قدرت هستند. این ویژگی می‌تواند به شناسایی دقیق‌تر خطاها و بررسی رفتار غیرعادی سیگنال‌ها کمک کند.

تمامی ویژگی‌های ذکر شده از سیگنال مؤلفه‌های دوم، که حاصل از پردازش روش تجزیه حالت تجربی مجموعه کامل با نویز تطبیقی هستند، به‌طور کلی برای شناسایی خصوصیات خاص سیگنال‌های خطاهای امیدانس بالا استخراج می‌شوند. به‌ویژه، ویژگی‌هایی مانند انرژی، انحراف معیار، و پیک تا پیک می‌توانند نوسانات بزرگ و تغییرات ناگهانی موجود در این نوع از خطاها را شناسایی کنند. همچنین ویژگی‌های چولگی و کشیدگی می‌توانند به شناسایی توزیع غیرطبیعی و ناهنجاری‌های سیگنال‌های خطاهای امیدانس بالا کمک کنند.

این ویژگی‌ها به‌عنوان ورودی برای الگوریتم‌های یادگیری ماشین استفاده شده است تا سیگنال‌های خطاهای امیدانس بالا از سیگنال‌های نرمال جدا شود. به این ترتیب، استخراج ویژگی‌ها از سیگنال‌های مؤلفه‌های دوم و استفاده از آن‌ها در فرآیند یادگیری ماشین، به‌طور مؤثر می‌تواند به شناسایی و تشخیص خطا کمک کند. این ویژگی‌ها برای شناسایی دقیق‌تر و کارآمدتر خطا در شبکه‌های توزیع انرژی الکتریکی استفاده می‌شوند و به‌عنوان ابزارهای مهم در تحلیل سیگنال‌های خطاهای امیدانس بالا عمل می‌کنند.

۲-۳- الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی یکی از معروف‌ترین و پرکاربردترین الگوریتم‌های یادگیری ماشین در مسائل طبقه‌بندی و رگرسیون است. این الگوریتم به‌ویژه در مسائل پیچیده‌ای که داده‌ها دارای ویژگی‌های غیرخطی یا روابط پیچیده هستند، عملکرد بسیار خوبی دارد.

۲-۳-۱- تعریف الگوریتم جنگل تصادفی

جنگل تصادفی یک روش مبتنی بر تجمع است که از مجموعه‌ای از درخت‌های تصمیم‌گیری استفاده می‌کند. در این الگوریتم، به‌جای استفاده از یک درخت تصمیم‌گیری واحد، چندین درخت تصمیم‌گیری به‌طور مستقل ساخته می‌شوند و پیش‌بینی‌ها از طریق رأی‌گیری اکثریت یا میانگین‌گیری از پیش‌بینی‌های درخت‌ها انجام می‌شود. این فرآیند باعث کاهش خطر بیش‌برازش^{۲۱} و بهبود دقت مدل می‌شود.

۲-۳-۲- دلایل انتخاب الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی برای مسائل طبقه‌بندی سیگنال‌ها در این تحقیق انتخاب شده است. از دلایل انتخاب این الگوریتم می‌توان به دقت بالا، قابلیت تحمل داده‌های نویزی، ساده بودن و قابل فهم بودن مدل و قابلیت پردازش ویژگی‌های زیاد اشاره نمود.

به‌دلیل محدودیت‌ها و پایداری ضعیف طبقه‌بندی‌های تکی، افزایش دقت طبقه‌بندی در مواجهه با شرایط خاص دشوار است. الگوریتم جنگل تصادفی از مجموعه‌ای از درخت‌های تصمیم‌گیری تشکیل شده است. جنگل‌های تصادفی، با توانایی تعمیم‌دهی و دقت بالا، یک الگوریتم ترکیبی است که توسط لئو بریمن در سال ۲۰۰۱ ارائه شد [۲۹]. این الگوریتم، ایده "جمع‌آوری بوت-استرپ" بریمن را با روش "زیرفضای تصادفی" هو ترکیب کرده و با ادغام چندین مدل درخت تصمیم^{۲۲} دقت پیش‌بینی را بهبود می‌بخشد. با استفاده از نتایج تصمیم‌گیری چندین مدل درخت تصمیم، خروجی نهایی با رأی‌گیری کلیه درخت‌های تصمیم تعیین می‌شود به‌عبارت دیگر، هدف الگوریتم این است که یک مدل ترکیبی از درخت‌های تصمیم‌گیری بسازد و پیش‌بینی نهایی را از طریق رأی‌گیری یا میانگین‌گیری انجام دهد. برای هر درخت تصمیم‌گیری، یک نمونه تصادفی از داده‌های آموزشی انتخاب می‌شود. برای هر درخت تصمیم‌گیری در هر گره، ویژگی‌ها به‌طور تصادفی انتخاب می‌شوند تا تصمیم‌گیری براساس آن‌ها انجام شود. درخت تصمیم‌گیری به‌طور تکراری گره‌ها را تقسیم می‌کند تا در نهایت به برگ‌هایی برسد که پیش‌بینی نهایی را ارائه دهد.

درخت تصمیم تصادفی، کوچک‌ترین واحد تصمیم‌گیری است که یک جنگل تصادفی را تشکیل می‌دهد. تولید آن بر اساس دو ویژگی تصادفی خلاصه می‌شود: نمونه‌های آموزشی تصادفی هر درخت تصمیم با روش بگینگ تولید می‌شوند و ویژگی‌های نمونه‌های آموزشی به صورت تصادفی برای تقسیم گره‌های درخت تصمیم انتخاب می‌شوند [۳۰]. مراحل ساخت جنگل‌های تصادفی به شرح زیر است:

الف) ایجاد مجموعه آموزشی با روش بوت استرپ:

از روش بوت استرپ برای نمونه‌گیری مجدد و ایجاد یک مجموعه آموزشی برای هر درخت تصمیم استفاده می‌شود. فرض کنید مجموعه‌ای شامل نمونه مختلف $\{C_1, C_2, \dots, C_N\}$ باشد. اگر یک نمونه از مجموعه با جای‌گذاری انتخاب شود و این فرآیند بارها تکرار شود تا مجموعه جدید W^* تشکیل شود، احتمال این که مجموعه جدید شامل یک نمونه خاص نباشد c_r برابر است با روابط (۱۰) و (۱۱).

$$p = \left(1 - \frac{1}{n}\right)^n \quad (10)$$

$$\lim_{n \rightarrow \infty} p = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \approx 0.368 \quad (11)$$

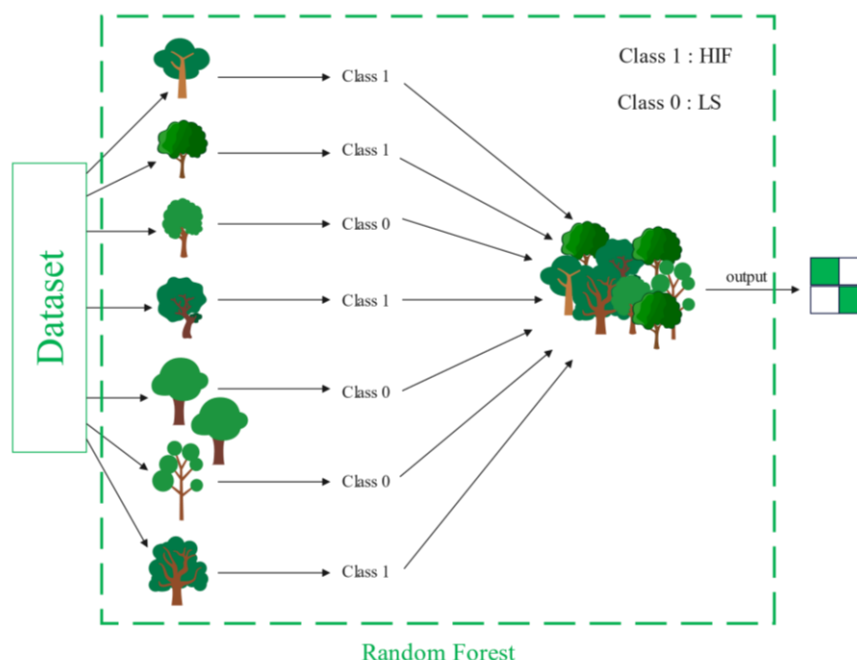
بنابراین، ۳۶/۸ درصد از نمونه‌ها انتخاب نخواهند شد. این امر موجب می‌شود که درخت تصمیم در جنگل تصادفی به یک حل بهینه محلی نرسد و داده‌های غیرعادی به‌طور مؤثر از ورود به مجموعه نمونه جلوگیری شوند.

ب) ساخت هر درخت تصمیم:

قبل از انتخاب ویژگی‌ها در هر گره غیر برگ، q ($q < Q$) ویژگی به‌صورت تصادفی از میان ویژگی‌های Q استخراج شده و به‌عنوان مجموعه ویژگی‌های طبقه‌بندی گره فعلی انتخاب می‌شود. گره در بهترین حالت ممکن میان این q ویژگی تقسیم شود. با تقسیم گره‌ها، یک درخت تصمیم کامل ایجاد می‌شود. هر درخت تصمیم بدون هرس رشد می‌کند تا به گره‌های برگ برسد. با استفاده از هر مجموعه آموزشی، یک درخت تصمیم متناظر تولید می‌شود.

ج) تشکیل جنگل تصادفی:

از درخت‌های تصمیم ساخته‌شده برای آزمایش نمونه‌ها استفاده می‌شود. خروجی دسته‌بندی هر درخت تصمیم جمع‌بندی شده و دسته‌ای که بیش‌ترین تعداد خروجی را دارد به‌عنوان دسته نمونه آزمایشی تعیین می‌شود. ساختار جنگل تصادفی در شکل (۲) نشان داده شده است.



شکل (۲): ساختار جنگل تصادفی

Figure (2): Random forest structure

۲-۳-۳- نحوه پیاده‌سازی الگوریتم جنگل تصادفی در حل مسئله

برای طبقه‌بندی سیگنال‌های استخراج‌شده و تشخیص خطای امیدانس بالا از تغییرات بار، در این مقاله از الگوریتم جنگل تصادفی به‌عنوان یک طبقه‌بند قدرتمند استفاده گردید. در ادامه، نحوه پیاده‌سازی این الگوریتم تشریح می‌شود.

الف) آماده‌سازی داده‌ها

ابتدا داده‌های سیگنال با استفاده از روش‌های استفاده شده به‌منظور استخراج ویژگی و از هر سیگنال مجموعه‌ای از شش ویژگی (F_1 تا F_6) استخراج شده است. ویژگی‌ها به‌همراه برچسب در نظر گرفته شده به‌عنوان ورودی طبقه‌بندی مورد استفاده قرار گرفته شده است. پیش از اعمال الگوریتم، داده‌ها نرمال‌سازی و به دو بخش داده‌های آموزش و آزمون تقسیم شده است. همان‌طور که اشاره شد، الگوریتم جنگل تصادفی مبتنی بر ایجاد تعداد زیادی درخت تصمیم است. برای هر درخت یک زیرمجموعه تصادفی از داده‌ها با روش Bootstrap انتخاب شده است. در هر گره، تنها تعدادی از ویژگی‌ها به‌صورت تصادفی در نظر گرفته شده است تا تقسیم‌بندی انجام شود. معیار انتخاب تقسیم، شاخص جینی در نظر گرفته شده است. این فرآیند تا رسیدن به عدم بهبود در دقت ادامه یافته است. پس از آموزش هر درخت، طبقه‌بندی نهایی با روش رأی‌گیری اکثریت میان خروجی‌های درخت‌ها تعیین شده است. بدین ترتیب، برای هر نمونه ورودی، طبقه‌ای انتخاب شد که بیش‌ترین رأی را از درخت‌های مختلف دریافت کرده است. در انتها، مدل آموزش‌دیده با استفاده از داده‌های آزمون مورد ارزیابی قرار گرفت. معیارهای دقت، حساسیت، دقت پیش‌بینی، امتیاز F_1 و ... محاسبه گردید تا کارایی الگوریتم در تشخیص خطای امیدانس بالا سنجیده شود.

۳- معیار ارزیابی (ماتریس درهم ریختگی):

معیار ارزیابی در این مقاله ماتریس درهم ریختگی می‌باشد. این ماتریس، کیفیت عملکرد طبقه‌بندی و کلاس بندی‌های مجموعه داده‌های ورودی را نمایش می‌دهد. این معیار معمولاً در الگوریتم‌های طبقه‌بندی بسیار کاربرد است و به‌عنوان یک معیار اصلی در ارزیابی نتایج استفاده می‌شود. شکل (۳) پارامترهای ماتریس درهم ریختگی را نمایش می‌دهد.

		مقادیر واقعی	
		مثبت	منفی
مقادیر پیش‌بینی	مثبت	مثبت درست TP	مثبت نادرست FP
	منفی	منفی نادرست FN	منفی درست TN

شکل (۳): پارامترهای ماتریس درهم ریختگی

Figure (3): Confusion matrix parameters

همان‌طور که مشاهده می‌شود، ماتریس درهم ریختگی از چهار سلول تشکیل شده است. هدف اصلی ما در این پژوهش شناسایی خطای امیدانس بالا می‌باشد. برای شناسایی این خطا با توجه به ماتریس درهم ریختگی، اگر تشخیص مثبت باشد، به این معنی است که خطای امیدانس بالا رخ داده است و اگر خلاف آن باشد و به عبارتی تشخیص منفی باشد، به این معنی است که خطای امیدانس بالا رخ نداده است. هر کدام از سلول‌های این ماتریس دارای یکسری مفاهیم می‌باشد که در زیر به ترتیب به آن‌ها پرداخته شده است:

- مثبت درست^{۲۳}: در این حالت خطای امپدانس بالا رخ داده است و الگوریتم هم پیش‌بینی درستی داشته است.
 - منفی درست^{۲۴}: در این حالت خطای امپدانس بالا رخ نداده است و الگوریتم هم به‌همین موضوع اشاره کند و عدم ایجاد خطا را اعلام کرده است.
 - مثبت نادرست^{۲۵}: در این حالت خطای امپدانس بالا رخ نداده است اما الگوریتم وقوع خطا را تشخیص داده است.
 - منفی نادرست^{۲۶}: در این حالت خطای امپدانس بالا رخ داده است اما الگوریتم عدم وقوع خطا را بیان کرده است.
- بر اساس تعاریفی که از سلول‌های ماتریس درهم‌ریختگی داریم، می‌توانیم روابط حاکم بر معیارهای ارزش‌یابی در مسائل طبقه‌بندی را بیان کنیم. در این پژوهش از شش معیار حساسیت^{۲۷}، دقت (نسبت مقادیر اندازه‌گیری شده نسبت به یک مقدار خاص)^{۲۸}، اختصاصیت^{۲۹}، دقت (نزدیکی مقادیر اندازه‌گیری شده نسبت به یکدیگر)^{۳۰}، امتیاز F_1 ^{۳۱} و ضریب تشابه^{۳۲} استفاده شده است. که روابط مربوطه این معیارها در (۱۲) تا (۱۷) نشان داده شده است.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (15)$$

$$F1 - Score = \frac{2 * (Precision * Sensitivity)}{Precision + Sensitivity} \times 100\% \quad (16)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \times 100\% \quad (17)$$

شبکه توزیع انرژی الکتریکی مورد مطالعه

در این مقاله از شبکه توزیع برق استاندارد ۳۴ شینه IEEE استفاده شده است. شبکه تست دارای ولتاژ نامی در سطح ۲۴/۹ کیلوولت و هم‌چنین دارای فرکانس نامی ۶۰ هرتز می‌باشد. نزدیکی سطح ولتاژ به شبکه واقعی و هم‌چنین تنوع در تجهیزاتی هم‌چون بانک‌های خازنی و از طرفی بارهای مختلف از دلایل اصلی انتخاب این شبکه می‌باشد. این موارد شبکه تست ۳۴ شینه IEEE را به یکی از رایج‌ترین شبکه‌های قدرت آزمایشی برای مطالعات در حوزه تشخیص خطای امپدانس بالا تبدیل کرده است.

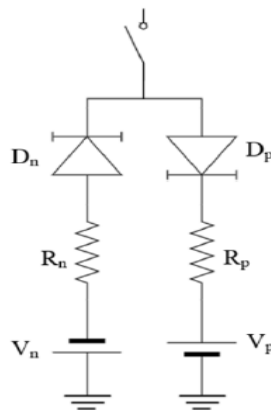
۴- نتایج و تحلیل شبیه‌سازی

در این مقاله به‌منظور ارزیابی نتایج به‌دست آمده، از شش معیار حساسیت، دقت، شفافیت (وضوح)، ظرافت، امتیاز F_1 و ضریب تشابه استفاده شده است.

۴-۱- جمع‌آوری داده

یکی از مهم‌ترین قسمت‌ها در جمع‌آوری داده‌ها، مدل‌سازی دقیق خطای امپدانس بالا می‌باشد. مدل خطا باید دارای یک سری ویژگی‌های کلی اعم از غیرخطی بودن، نامتقارن بودن، غیر ایستا بودن و ... باشد. به‌همین منظور، در این مقاله از مدل خطای امپدانس بالا امانوئل در مرجع [۳۱] نیز استفاده شده است. شکل (۵) مدل خطای امپدانس بالا استفاده شده در این پژوهش را

نشان می‌دهد. این مدل از جفت‌های دیودی، مقاومت و منابع ولتاژ جریان مستقیم تشکیل شده است که در مدار در دو مسیر و به شکل معکوس نسبت به هم قرار گرفته است. برای نیم سیکل مثبت در این مدل، زمانی که ولتاژ خط از ولتاژ جریان مستقیم بیش‌تر می‌شود، دیود شروع به هدایت می‌کند. کنترل جریان از طریق مقاومتی که در مسیر وجود دارد انجام می‌شود. همین روند برای نیم سیکل منفی معکوس می‌شود. جدول (۱) پارامترهای مدل خطاهای امپدانس بالا و محدوده تغییرات مقادیر پارامترهای بیان شده در مدل خطای امپدانس بالا انتخاب شده را نشان می‌دهد. شکل (۴) نمونه‌ای از سیگنال جریان خطاهای امپدانس بالا به‌دست آمده در نرم‌افزار EMTP-rv را نشان می‌دهد. محل خطاهای امپدانس بالا و تغییرات بار در شکل (۷) نمایش داده شده است. همچنین جریان خطا توسط آمپر متر در ابتدای فیدر اندازه‌گیری شده است. فرکانس نمونه‌برداری در شبکه تست، ۷۶۸۰ هرتز می‌باشد.



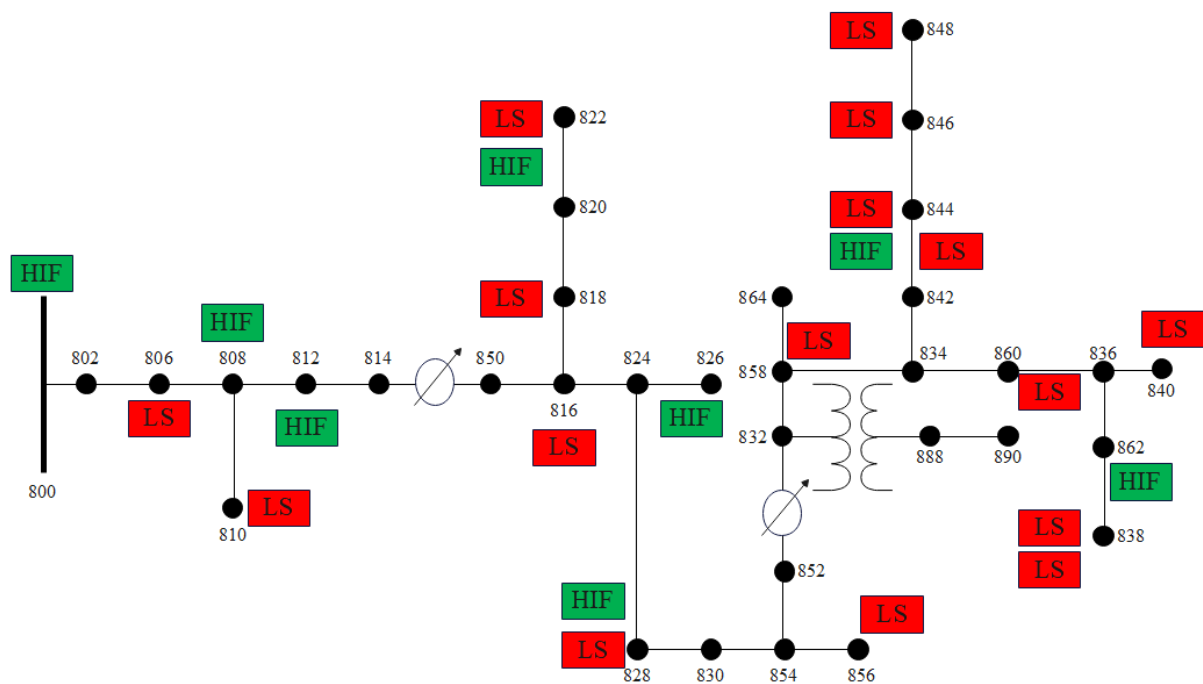
شکل (۴): مدل خطای امپدانس بالا استفاده شده در این پژوهش

Figure (4): HIF model used in this research

Title (1): Parameters of high impedance fault model

جدول (۱): پارامترهای مدل خطای امپدانس بالا

پارامترهای مدل خطای امپدانس بالا	محدوده تغییرات مدل خطای امپدانس بالا
V_p	۴۰۰-۸۰۰ ولت
V_n	۵۰۰-۱۴۰۰ ولت
R_p	۹۰-۸۰۰۰ اهم
R_n	۹۰-۸۰۰۰ اهم
D_p	۰/۷ ولت
D_n	۰/۷ ولت



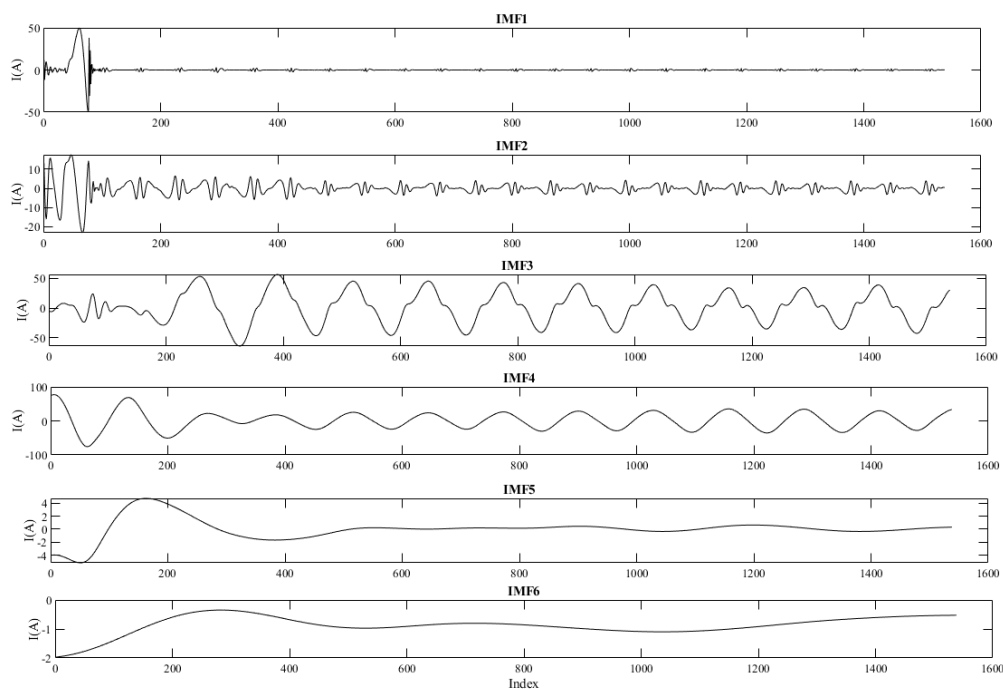
شکل (۵): نمودار تک خطی شبکه استاندارد ۳۴ شینه IEEE

Figure (5): Single-line diagram of the IEEE standard 34-bus network

۴-۲- استخراج ویژگی

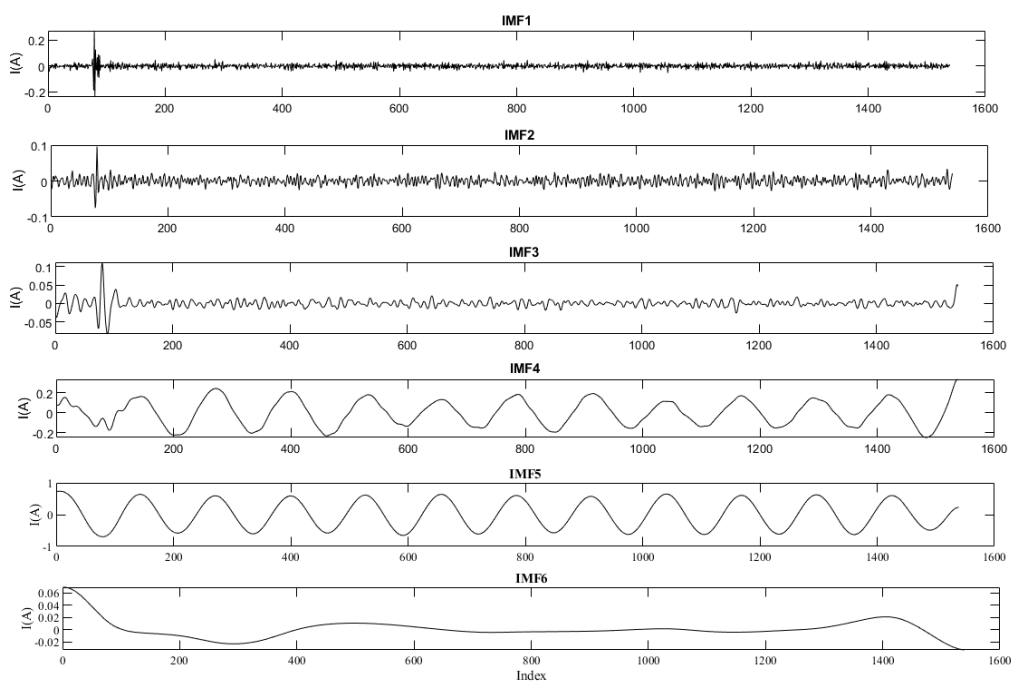
استخراج ویژگی در محیط برنامه نویسی Google Colab و با استفاده از زبان برنامه نویسی پایتون انجام شده است. در این پژوهش، سیگنال‌های خطای امپدانس بالا و تغییرات بار با استفاده از روش پیشنهادی تجزیه و تحلیل شده است. این روش به دلیل توانایی بالای آن در تفکیک سیگنال‌های غیرخطی و غیرایستا، به عنوان ابزار اصلی انتخاب شد. پس از اعمال روش، سیگنال‌ها به چندین مؤلفه ذاتی تفکیک شده است. مؤلفه‌های ذاتی به دست آمده پس از اعمال روش برای سیگنال خطاهای امپدانس بالا و تغییرات بار در شکل‌های (۶) و (۷) نمایش داده شده است.

از میان این مؤلفه‌ها، مؤلفه‌های دوم به عنوان سیگنال هدف برای استخراج ویژگی انتخاب شد. مؤلفه‌های دوم شامل اطلاعات فرکانسی میانی است که نه تنها نویزهای فرکانس بالا (مؤلفه‌های اولیه) را حذف کرده، بلکه از تلفیق بیش از حد سیگنال‌ها در مؤلفه‌های پایین‌تر جلوگیری کرده است. حساسیت به تغییرات خطا از دلایل انتخاب این مؤلفه ذاتی می‌باشد. تحلیل‌ها نشان داده است که مؤلفه‌های دوم توانایی بالایی در تمایز رفتار دینامیکی سیگنال‌های خطاهای امپدانس بالا نسبت به تغییرات بار دارد. ویژگی‌های استخراج شده از مؤلفه‌های دوم نسبت به سایر مؤلفه‌ها دارای واریانس کمتر و تکرارپذیری بیشتری بودند. از سیگنال مؤلفه‌های دوم، شش ویژگی کلیدی (F1 تا F6) استخراج شده است. این ویژگی‌ها شامل معیارهای آماری و حوزه فرکانسی بودند که به طور خاص برای تمایز سیگنال‌های خطا و تغییرات بار طراحی شدند. به عنوان نمونه، F1 (میانگین انرژی) نشان‌دهنده سطح انرژی کلی سیگنال و تفاوت چشم‌گیر آن در دو کلاس بود. F2 (انحراف معیار) نشان‌دهنده نوسانات سیگنال و حساسیت به وقوع خطا است. F3 تا F6 ترکیبی از شاخص‌های فرکانسی و زمانی بودند که دقت شناسایی را بهبود بخشیدند. نتایج نشان داد که ویژگی‌های استخراج شده از مؤلفه‌های دوم، به ویژه در تمایز سیگنال‌های خطا از تغییرات بار، عملکرد بسیار مناسبی داشته و زمینه‌ساز مدل‌سازی موفق در مراحل بعدی بودند.



شکل (۶): مؤلفه‌ها پس از اعمال روش پیشنهادی بر روی سیگنال خطاهای امپدانس بالا

Figure (6): IMFs after applying proposed method to the HIFs signal

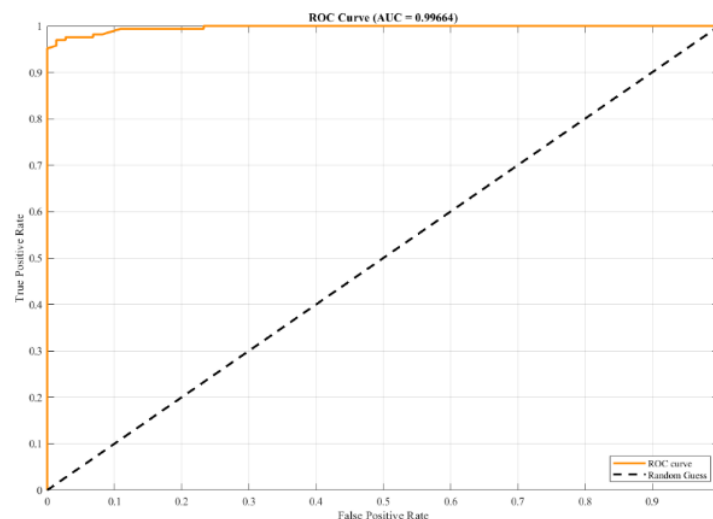


شکل (۷): مؤلفه‌ها پس از اعمال روش پیشنهادی بر روی سیگنال تغییرات بار

Figure (7): IMFs after applying proposed method to the load switching signal

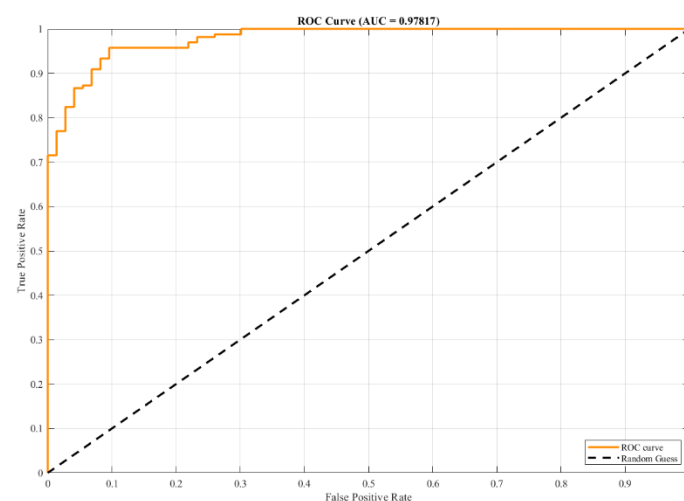
۴-۳- تصمیم‌گیری و طبقه‌بندی

در ابتدا کتابخانه‌ها و ابزارهای موردنیاز را در زبان برنامه نویسی پایتون فراخوانی می‌کنیم. از آنجایی که بخش زیادی از پردازش با استفاده از Google colab صورت گرفته است پس رابطی بین Google Colab و Google drive در نظر گرفته شده است. مجموعه داده‌ها از یک فایل Excel در Google Drive فراخوانی شده است. داده‌ها به ویژگی‌ها و برچسب‌ها و هم‌چنین به مجموعه داده‌های آموزشی و تست تقسیم شده است. یک مقیاس‌بندی بر روی آن‌ها اعمال شده است. مدل RF ایجاد شده است. مدل‌ها با استفاده از داده‌های آموزش، آموزش داده شده است. پیش‌بینی برچسب‌ها با استفاده از داده‌های تست به دست می‌آید. در مراحل پایانی ارزیابی دقت مدل‌ها بیان شده است و معیار اعتبارسنجی که ماتریس درهم‌ریختگی می‌باشد نیز نمایش داده شده است. به منظور ارزیابی عملکرد روش پیشنهادی، مقایسه‌ای با روش تجزیه حالت تجربی تجمعی کامل با داده کمکی نوین انجام شده است. به عبارتی دیگر از این روش در قسمت استخراج ویژگی استفاده شده است و مجدد روش الگوریتم جنگل تصادفی اعمال شده است و نتایج، نشان دهنده بهبود روش پیشنهادی می‌باشد. شکل‌های (۸) و (۹) نمودار منحنی ROC و AUC به دست آمده برای هر کدام از روش‌های تجزیه حالت تجربی مجموعه کامل با نوین تطبیقی و تجزیه حالت تجربی تجمعی کامل با داده کمکی نوین نشان داده شده است.



شکل (۸): نمودار منحنی ROC و AUC به دست آمده از روش پیشنهادی

Figure (8): ROC and AUC curve obtained from the proposed method

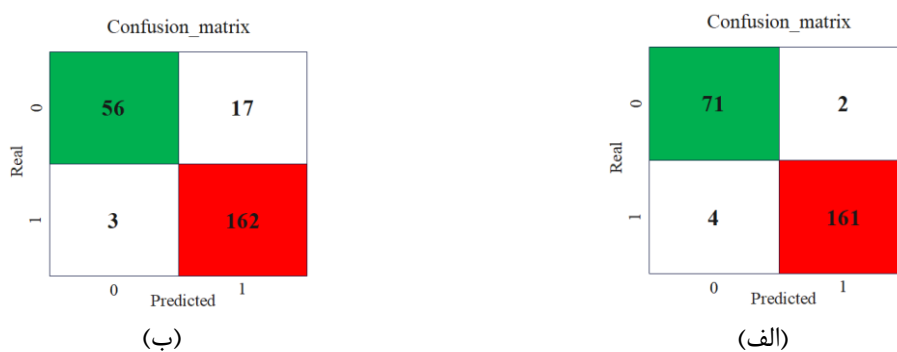


شکل (۹): نمودار منحنی ROC و AUC به دست آمده از روش پیشنهادی

Figure (9): ROC and AUC curve diagram obtained from the proposed method

همان طور که مشاهده می شود منحنی ROC روش پی شنهادی نتایج قابل قبولی نسبت به روش تجزیه حالت تجربی جمععی کامل با داده کمکی نويز از خود نشان داده است. هم چنین مقدار AUC همان طور که مشاهده می شود در روش پی شنهادی مقداری اثرگذارتر دارد.

ماتریس درهم ریختگی به دست آمده از طبقه بندی RF_CEEMDAN و RF_ICEEMDAN برای شبکه تست ۳۴ شینه IEEE در شکل (۱۰) نشان داده شده است. روابط معیارهای ارزیابی برای این ماتریس محاسبه شده است. هم چنین جدول (۲) نتایج به دست آمده از روابط معیارهای ارزیابی برای شبکه تست ۳۴ شینه IEEE با استفاده از روش های RF_ICEEMDAN و RF_CEEMDAN را نشان می دهد.



شکل (۱۰): ماتریس درهم ریختگی با استفاده از روش RF_ICEEMDAN (الف) و RF_CEEMDAN (ب)

Figure (10): Confusion matrix using RF ICEEMDAN (a) and RF_CEEMDAN (b) methods

Table (2): Evaluation results of proposed method with another method

جدول (۲): نتایج ارزیابی روش پیشنهادی با روش دیگر

روش	دقت (%)	حساسیت (%)	اختصاصیت (%)	دقت (%)	ضریب تشابه (%)	امتیاز F1 (%)
RF_CEEMDAN	۹۱/۵۹۶	۹۴/۹۱	۹۰/۵۰۲	۷۶/۷۱	۸۴/۸۴۸	۸۴/۸۴۴
RF_ICEEMDAN	۹۷/۴۷	۹۴/۶۶	۹۸/۷۷۳	۹۷/۲۶	۹۵/۹۴۵	۹۵/۹۴۲

همان طور که مشخص است، نتایج قابل قبول و با کیفیتی را در مقایسه با روش RF_CEEMAN به دست آمده است.

۵- نتیجه گیری

روش ارائه شده در این پژوهش ترکیبی از تجزیه حالت تجربی مجموعه کامل با نويز تطبیقی و یادگیری ماشین می باشد. ابتدا مجموعه داده ها جمع آوری شده است. از تجزیه حالت تجربی با نويز تطبیقی به منظور استخراج ویژگی بر روی مجموعه نمونه برداری های انجام شده از داده ها اعمال شده است. ویژگی های استخراج شده برای طبقه بندی به عنوان ورودی به الگوریتم جنگل تصادفی وارد شده است.

این روش بر روی شبکه تست ۳۴ شینه IEEE پیاده سازی شده است. همان طور که در نتایج قابل مشاهده است، این روش توانست وقوع و یا عدم وقوع تشخیص خطای امپدانس بالا در شبکه تست را با دقت قابل قبول و بالایی بیان کند. دقت به دست آمده از روش پیشنهادی بر روی شبکه تست ۳۴ شینه IEEE، ۹۷/۴۷ درصد می باشد. با توجه به نتایج به دست آمده، این روش حالت بهره برداری عادی شبکه قدرت را از حالتی که خطای امپدانس بالا رخ داده است، تفکیک کرده است.

مراجع

- ۱۰۱

- [15] M. Bhatnagar, A. Yadav, and A. Swetapadma, "Fuzzy based relaying scheme for high impedance faults in DG integrated distribution system," *Electric Power Systems Research*, vol. 202, p. 107602, 2022, (doi: [10.1016/j.epsr.2021.107602](https://doi.org/10.1016/j.epsr.2021.107602)).
- [16] S. Lavanya, S. Prabakaran, and N. A. Kumar, "A deep learning technique for detecting high impedance faults in medium voltage distribution networks," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9477-9482, 2022, (doi: [10.48084/etasr.5288](https://doi.org/10.48084/etasr.5288)).
- [17] M. Eissa, M. Awadalla, and A. Sharaf, "A new high impedance fault detection technique using deep learning neural network," in *Artificial Intelligence Applications in Electrical Transmission and Distribution Systems Protection*: CRC Press, 2021, pp. 179-195.
- [18] K. Rai, F. Hojatpanah, F. B. Ajaei, J. M. Guerrero, and K. Grolinger, "Deep learning for high-impedance fault detection and classification: transformer-CNN," *Neural Computing and Applications*, vol. 34, no. 16, pp. 14067-14084, 2022, (doi: [10.1007/s00521-022-07219-z](https://doi.org/10.1007/s00521-022-07219-z)).
- [19] H. Lala and S. Karmakar, "Detection and experimental validation of high impedance arc fault in distribution system using empirical mode decomposition," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3494-3505, 2020, (doi: [10.1109/JSYST.2020.2969966](https://doi.org/10.1109/JSYST.2020.2969966)).
- [20] X. Wang *et al.*, "High impedance fault detection method based on improved complete ensemble empirical mode decomposition for DC distribution network," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 538-556, 2019, (doi: [10.1016/j.ijepes.2018.12.021](https://doi.org/10.1016/j.ijepes.2018.12.021)).
- [21] Y. Liu, M. Razeghi-Jahromi, and J. Stoupis, "Unsupervised High Impedance Fault Detection in Distribution Networks With Autoencoder," *IEEE Transactions on Industry Applications*, 2025, (doi: [10.1109/TIA.2025.3544169](https://doi.org/10.1109/TIA.2025.3544169)).
- [22] M. S. Attar and M. R. Miveh, "High- Impedance Fault Detection in Distribution Networks Based on Support Vector Machine and Wavelet Transform Approach (Case Study: Markazi Province of Iran)," *Energy Science & Engineering*, vol. 13, no. 3, pp. 1171-1183, 2025, (doi: <https://doi.org/10.1002/ese3.2056>).
- [23] Z. Zhang, Y. Li, Z. Wang, J. Liu, A. Chen, and Y. Cao, "A data-driven impedance estimation and matching method for high impedance fault detection and location of distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 165, p. 110499, 2025, (doi: [10.1016/j.ijepes.2025.110499](https://doi.org/10.1016/j.ijepes.2025.110499)).
- [24] P. V. Dhawas, P. P. Bedekar, S. Dambhare, and P. V. Nandankar, "Advancement of High-Impedance Fault Detection in Primary Distribution Systems Using a Sequence Components Energy Detector," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2025, (doi: [10.1109/TCSI.2025.3544894](https://doi.org/10.1109/TCSI.2025.3544894)).
- [25] A. Vinayagam *et al.*, "Discrimination of high impedance fault in microgrid power network using semi-supervised machine learning algorithm," *Ain Shams Engineering Journal*, vol. 16, no. 1, p. 103187, 2025, (doi: [10.1016/j.asej.2024.103187](https://doi.org/10.1016/j.asej.2024.103187)).
- [26] A. A. Janiabadi and S. Hasheminejad, "A new algorithm for the identification of high impedance faults in distribution systems utilizing S transform," *Ain Shams Engineering Journal*, vol. 16, no. 4, p. 103334, 2025, (doi: [10.1016/j.asej.2025.103334](https://doi.org/10.1016/j.asej.2025.103334)).
- [27] H. Bai, J.-H. Gao, T. Liu, Z.-Y. Guo, and M.-F. Guo, "Explainable incremental learning for high-impedance fault detection in distribution networks," *Computers and Electrical Engineering*, vol. 122, p. 110006, 2025, (doi: [10.1016/j.compeleceng.2024.110006](https://doi.org/10.1016/j.compeleceng.2024.110006)).
- [28] A. Mohammadi, M. Jannati, and M. Shams, "Using deep transfer learning technique to protect electrical distribution systems against high-impedance faults," *IEEE Systems Journal*, vol. 17, no. 2, pp. 3160-3171, 2023, (doi: [10.1109/JSYST.2023.3234655](https://doi.org/10.1109/JSYST.2023.3234655)).
- [29] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001, (doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)).
- [30] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, pp. 123-140, 1996, (doi: [10.1007/BF00058655](https://doi.org/10.1007/BF00058655)).
- [31] T. Lai, L. Snider, E. Lo, C. Cheung, and K. Chan, "High impedance faults detection using artificial neural network," 2003.

زیر نویس ها

1. High Impedance Fault (HIF)
2. Dense Convolutional Neural Network
3. Long Short-Term Memory (LSTM)
4. Complete Ensemble Empirical Mode Decomposition with Noise-Assisted Data (CEEMDAN)
5. Empirical Mode Decomposition (EMD)
6. Autoencoder
7. Multi Layer Perceptron (MLP)
8. Photovoltaics (PV)
9. Explainable Incremental Learning
10. Discrete Wavelet Transform (DWT)
11. Wigner-Ville Transform
12. Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN)
13. Random Forest (RF)
14. Intrinsic Mode Function (IMF)
15. Mean Amplitude
16. Energy
17. Standard Deviation
18. Peak-to-Peak
19. Skewness
20. Kurtosis
21. overfitting
22. Decision Tree
23. True Positive
24. True Negative
25. False Positive
26. False Negative
27. Sensitivity
28. Accuracy
29. Specificity
30. Precision
31. F-Score
32. Dice