

Introducing a fast and efficient machine learning model for lightning localization via Lightning-Induced Voltages on Transmission Lines

Mostafa Asadi¹, Mohammad Hassan Moradi², Hamid Reza Karami³, Farhad Rachidi⁴, Marcos Rubinstein⁵

¹ Ph.D. student, Department of Electrical Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran
m.asadi@eng.basu.ac.ir

² Professor, Department of Electrical Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran
mhmoradi@basu.ac.ir

³ Associate Professor, Department of Electrical Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran
hamidr.karami@basu.ac.ir

⁴ Professor, Electromagnetic Compatibility Laboratory, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland – farhad.rachidi@epfl.ch

⁵ Professor, Institute for Information and Communication Technologies, University of Applied Sciences of Western Switzerland (HES-SO), Yverdon-les-Bains, Switzerland
marcos.rubinstein@heig-vd.ch

Abstract :

Lightning localization is crucial for a wide range of applications. Conventional lightning localization methods typically use at least four antennas as sensors for detecting lightning electromagnetic radiation. One of the complications of these methods is the need for sensor synchronization. Most papers on lightning localization highlight the issues of complexity and synchronization requirements. In this paper, we propose a fast learning model using the XGBoost algorithm for lightning localization. This method utilizes two sensors to receive signals from the induced voltages on transmission lines, eliminating the need for synchronization. By employing the principal component analysis (PCA) algorithm, the input dimensions of the model are reduced, which decreases model complexity, speeds up calculations, reduces hardware resource usage, and enhances accuracy. By training the model with varying numbers of principal components, we can identify the smallest input dimensions that maintain model accuracy. The final model is evaluated using a noisy test dataset. The evaluation results show that the model achieves accuracy (R^2) of over 99%. Additionally, studies indicate that the model's accuracy depends on the configuration of the transmission lines and the position of the sensors.

Keywords: lightning localization, machine learning, principal component analysis (PCA), XGBoost..

Article Type: Research

Received: 23. 06. 2023

Revised: 15. 11. 2024

Accepted: 08. 01. 2025

Corresponding author: Mohammad Hassan Moradi

Corresponding author's address: Ahmadi Roshan Blvd, Department of Electrical Engineering, Faculty of Engineering, Bu-Ali Sina University. Hamedan , Iran



1. Motivation of the work

Knowing the exact geo-location of a lightning strike is important in a wide range of research and application domains, including geophysical research, lightning warning, aviation/air traffic, weather services, insurance claims, power transmission and distribution, etc. [1]. The location of the lightning strike is generally obtained using the so-called lightning location system (LLS). LLSs detect cloud-to-ground (CG) discharge signals using electromagnetic VLF/LF range sensors [1].

2. Contributions

Setting up location systems by using the effect of lightning on transmission lines can have a significant impact on reducing costs. Therefore, this article attempts to investigate lightning location through induced voltages on transmission lines.

3. Procedures

In this paper, problem geometry is defined in which two transmission lines are located as infinite single wires on perfectly conducting ground. Two voltage sensors are placed on each line. A Monte Carlo approach is used to determine the random locations of a number of lightning strikes. The lightning-induced voltages at the sensor locations are calculated using the Rusk coupling model [2]. A model based on the XGBoost algorithm is trained using data from only two sensors.

4. Findings

The trained model can estimate the location of lightning with more than 99% accuracy using lightning transient voltage data received from only two sensors.

5. Conclusion

In this paper, a machine learning-based method for lightning location through induced voltages on transmission lines was investigated. The efficiency of the XGBoost algorithm was also demonstrated as a relatively fast learning method for finding the lightning strike location using noisy induced voltage data on transmission lines. It is worth noting that the strong performance of the model is evident even in the presence of noisy data. Using only two independent sensors, reducing the input dimensions of the model from 4000 to 80 by the PCA algorithm, and data with a signal-to-noise ratio of 30 dB, the proposed model achieved an accuracy of 99%. The dimensionality reduction significantly increased the computational speed. By examining the accuracy of the model based on the number of PCA components, it was determined that the speed and accuracy of the model are improved with a lower number of components of about 21 components. It was also shown that the location of the sensors and the configuration of the transmission lines can affect the performance of the model.

معرفی یک مدل یادگیری ماشین سریع و کارآمد برای مکان‌یابی صاعقه از طریق ولتاژهای القایی صاعقه در خطوط انتقال

مصطفی اسدی^۱، محمد حسن مرادی^۲، حمیدرضا کرمی^۳، فرهاد رشیدی^۴، Marcos Rubinstein^۵

۱- دانشجوی دکتری-گروه مهندسی برق، دانشکده مهندسی، دانشگاه بوعلی سینا

m.asadi@eng.basu.ac.ir

۲- استاد- گروه مهندسی برق، دانشکده مهندسی، دانشگاه بوعلی سینا

mhmoradi@basu.ac.ir

۳- دانشیار- گروه مهندسی برق، دانشکده مهندسی، دانشگاه بوعلی سینا

hamidr.karami@basu.ac.ir

4- Professor- Electromagnetic Compatibility Laboratory, Swiss Federal Institute of Technology (EPFL),
Lausanne, Switzerland

farhad.rachidi@epfl.ch

5-Professor- Institute for Information and Communication Technologies, University of Applied Sciences of
Western Switzerland (HES-SO), Yverdon-les-Bains, Switzerland;

marcos.rubinstein@heig-vd.ch

چکیده: مکان‌یابی صاعقه در گستره وسیعی از کاربردها از اهمیت بسزایی برخوردار است. در روش‌های مرسوم مکان‌یابی صاعقه حداقل از ۴ آنتن به عنوان حس‌گر تشعشعات الکترومغناطیسی ناشی از صاعقه استفاده می‌شود. یکی از پیچیدگی‌های این روش‌ها لزوم سنکرون بودن حس‌گرها می‌باشد. در اکثر مقالات مرتبط با مکان‌یابی صاعقه پیچیدگی و نیاز به سنکرون بودن دیده می‌شود. در این مقاله، یک مدل یادگیری ماشینی سریع به کمک الگوریتم XGBoost برای مکان‌یابی صاعقه پیشنهاد شده است. در این روش دو حس‌گر برای دریافت سیگنال ولتاژهای القایی بر روی خطوط انتقال استفاده می‌شود که نیازی به سنکرون‌سازی هم ندارند. در روش پیشنهادی به کمک الگوریتم تجزیه و تحلیل مولفه اصلی (PCA) ابعاد ورودی مدل کاهش می‌یابد که منجر به کاهش پیچیدگی‌های مدل، افزایش سرعت محاسبات، تقلیل استفاده از منابع سخت افزاری و نیز افزایش دقت آن می‌شود. همچنین با آموزش مدل در یک بازه از تعداد مولفه‌های اصلی، کمترین ابعاد ورودی با حفظ دقت مدل بدست خواهد آمد. مدل نهایی با داده‌های آزمایشی نوپزی ارزیابی می‌شود. نتایج ارزیابی نشان می‌دهد که دقت R2 به دست آمده از مدل، بیش از ۹۹ درصد است. همچنین بررسی‌ها نشان می‌دهد که دقت مدل به پیکربندی خط انتقال و موقعیت سنسورها نیز بستگی دارد.

کلمات کلیدی: مکان‌یابی صاعقه، یادگیری ماشین، تجزیه و تحلیل مولفه اصلی، XGBoost، PCA

نوع مقاله: پژوهشی

دریافت: ۱۴۰۳/۰۴/۰۳

بازنگری: ۱۴۰۳/۰۸/۲۵

پذیرش: ۱۴۰۳/۱۰/۱۹

نام نویسنده‌ی مسئول: دکتر محمد حسن مرادی

نشانی نویسنده‌ی مسئول: ایران - همدان - چهارباغ شهید مصطفی احمدی روشن - دانشگاه بوعلی سینا - دانشکده‌ی مهندسی - گروه

مهندسی برق

۱- مقدمه

دانستن موقعیت جغرافیایی دقیق برخورد صاعقه به زمین در بسیاری از حوزه‌های تحقیقاتی و کاربردی، از جمله ژئوفیزیک، هشدار صاعقه، حمل و نقل هوایی، خدمات مرتبط با هواشناسی، صنعت بیمه، شبکه انتقال و توزیع الکتریکی و... از اهمیت بالایی برخوردار است [۱]. تشعشعات الکترومغناطیسی ساطع شده از تخلیه صاعقه طیف وسیعی از فرکانس بسیار پایین (VLF) تا فرکانس بسیار بالا (VHF) را در بر می‌گیرد [۲، ۳، ۴، ۵]. در حال حاضر، به طور کلی محل برخورد صاعقه با استفاده از سیستم مکان‌یابی صاعقه (LLS) به دست می‌آید. LLS ها سیگنال‌های تخلیه الکتریکی ابر به زمین را با استفاده از حسگرهای الکترومغناطیسی شناسایی می‌کنند [۶]. یک LLS زمینی معمولاً از حداقل چهار حسگر تشخیص پالس الکترومغناطیسی، یک مرکز پردازش داده و یک لینک انتقال شبکه تشکیل شده است [۳]. همچنین این سیستم‌ها مجهز به یک مجموعه GPS بسیار دقیق با دقت نانو ثانیه جهت سنکرون نمودن حسگرها می‌باشند [۷].

روش‌های مختلف مکان‌یابی صاعقه را می‌توان مطابق با شکل (۱) به دو دسته اصلی تقسیم نمود؛ دسته اول روش‌های مبتنی بر داده‌های دریافتی از میدان‌های الکترومغناطیسی منتشر شده توسط صاعقه هستند. این دسته به دو زیر مجموعه روش‌های مرسوم و نوین تقسیم‌بندی می‌شود. روش‌های مرسوم مکان‌یابی صاعقه مبتنی بر میدان عبارتند از جهت‌یاب مغناطیسی (MDF)، زمان رسیدن (TOA)، اختلاف زمانی رسیدن (TDOA) و تداخل سنج (ITF) [۸]. روشی TOA روشی است که برای مکان‌یابی ۲ بعدی و ۳ بعدی استفاده می‌شود. این روش که متکی به زمان رسیدن موج ناشی از صاعقه می‌باشد به حداقل سه حسگر نیاز دارد تا مکان‌یابی صاعقه در درون مثلث حسگرها به درستی کار کند [۸، ۹]. لازم به ذکر است در این روش برای عملکرد کامل حداقل چهار حسگر نیاز است. تکنیک MDF نیز به حداقل سه حسگر نیاز دارد و برای مکان‌یابی دو بعدی استفاده می‌شود [۱۰]. این روش به تنهایی دقت کافی را ندارد. لازم به ذکر است در اغلب روش‌های مرسوم و حتی نوین لازم است حسگرها یا آنتن‌ها با یکدیگر سنکرون شوند و این موضوع خود بر پیچیدگی روش‌ها می‌افزاید.

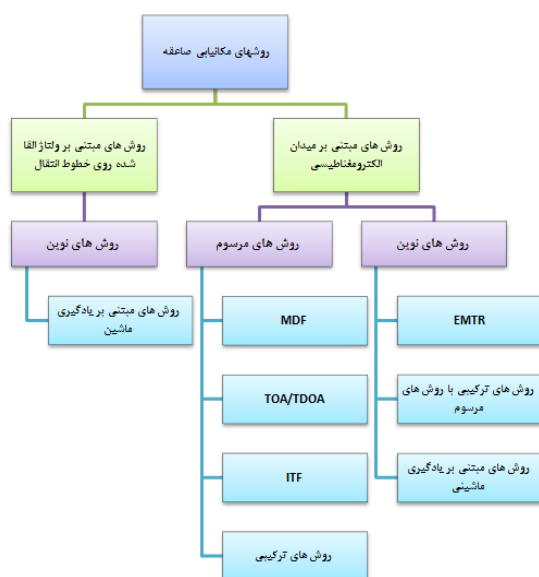
اخیراً، روش‌های نوین توسط محققان جهت مکان‌یابی صاعقه پیشنهاد شده است. در برخی از این روش‌ها با به‌کارگیری روش‌های مرسوم و ترکیب آن‌ها با روش‌های جدید به مدل‌های کارآمدتری دست یافته‌اند. در برخی دیگر نیز روش‌هایی کاملاً جدید پیشنهاد شده است. از جمله روش‌هایی ترکیبی در [۲] معرفی شده است. نویسندگان مقاله مذکور با استفاده از روش بهینه‌سازی PSO به منظور یافتن نقطه بهینه در مکان‌یابی ترکیبی MDF و TDOA استفاده نموده‌اند. در روش پیشنهادی آن‌ها سه آنتن VHF و سه آنتن متقاطع (در مجموع ۶ حسگر) و یک ارتباط GPS جهت سنکرون سازی سیگنال‌ها مورد استفاده قرار گرفته است. یکی دیگر از روش‌های نوین، استفاده از روش

معکوس زمانی الکترومغناطیسی (EMTR) است که به عنوان ابزاری برای مکان‌یابی صاعقه پیشنهاد شده است [۱۱]. ثابت شده است که EMTR در شناسایی نقطه برخورد صاعقه دقت بالایی دارد، اما برای دقت کافی به حداقل سه حسگر نیاز دارد [۱۲، ۱۳]. در [۷]، ترکیبی از تکنیک TDOA و شبکه‌های عصبی مصنوعی برای مکان‌یابی صاعقه مبتنی بر امواج منتشر شده ناشی از برخورد صاعقه پیشنهاد شده است. برخی از مطالعات از یادگیری عمیق برای یافتن منبع صاعقه استفاده کرده‌اند که همگی مبتنی بر میدان الکتریکی ناشی از برخورد صاعقه می‌باشند [۱۱، ۱۴].

در [۱۵]، داده‌های رادار سه بعدی و الگوریتم‌های یادگیری ماشینی مانند KNN، جنگل تصادفی (RF) و شبکه‌های عصبی کانولوشنال (CNN) برای شناسایی مکان‌های برخورد صاعقه استفاده می‌شوند.

دسته دوم از روش‌های مکان‌یابی براساس نمودار شکل (۱)، مبتنی بر سیگنال‌های دریافتی از ولتاژ القا شده بر روی خطوط انتقال می‌باشند. ایده استفاده از ولتاژهای القایی در خطوط انتقال برای مکان‌یابی محل برخورد صاعقه در [۱] مطرح گردیده است. در آن از یک روش مبتنی بر یادگیری ماشین برای یافتن موقعیت‌های جغرافیایی دو بعدی نقاط برخورد صاعقه استفاده می‌شود که به دو حسگر که با یکدیگر سنکرون شده‌اند نیاز دارد و از ولتاژ القایی ناشی از صاعقه در خطوط انتقال استفاده می‌کند اما به دلیل وجود تقارن در خط انتقال دقت بالایی ندارد و مدل پیشنهادی آن‌ها نمی‌تواند مکان صاعقه در سمت چپ و راست خط انتقال را از هم متمایز نماید همچنین همان‌گونه که ذکر گردید حسگرهای مورد استفاده نیز باهم سنکرون می‌باشند که خود سبب افزایش پیچیدگی و هزینه اجرایی طرح می‌شود.

در این مقاله، تشخیص برخورد صاعقه از طریق ولتاژهای القا شده



شکل (۱): روش‌های مکان‌یابی صاعقه

به سنکرون سازی آن ها نیست. همه پیاده سازی کدها در Google Colab اجرا می شوند.

ساختار مقاله به شرح زیر است. بخش ۲ روش تحقیق، از جمله ایجاد پایگاه داده، انتخاب ویژگی، و تولید مدل را ارائه می دهد. بخش ۳ به آموزش و آزمایش مدل تولید شده می پردازد و از نتایج شبیه سازی عددی برای تخمین نقطه برخورد صاعقه استفاده می کند. در بخش ۴ تاثیر تغییرات پیکربندی خطوط انتقال در داخل ناحیه محاسباتی و نیز تاثیر دو الگوریتم PCA و FA بر کارایی مدل مورد بحث قرار گرفته است. در پایان مقاله، نتیجه گیری نهایی ارائه شده است.

۲- تئوری تحقیق

همانند مقالات مرتبط با یادگیری ماشین، در این مقاله نیز روند مشابهی جهت دستیابی به مدل نهایی مکان یابی صاعقه بر اساس هدف مد نظر یعنی استفاده از ولتاژهای القا شده ناشی از صاعقه پیش گرفته شده است. اقدامات کلی انجام شده در نمودار شکل (۲) نشان داده شده است. این اقدامات به ترتیب عبارتند از ۱. تولید مجموعه داده ۲. پیش پردازش داده ها ۳. انتخاب مدل یادگیری ماشین ۴. ارزیابی مدل و انتخاب مدل نهایی. در ادامه توضیحات هر بخش ارائه خواهد شد.

۲-۱- هندسه مسئله و تولید داده آموزشی

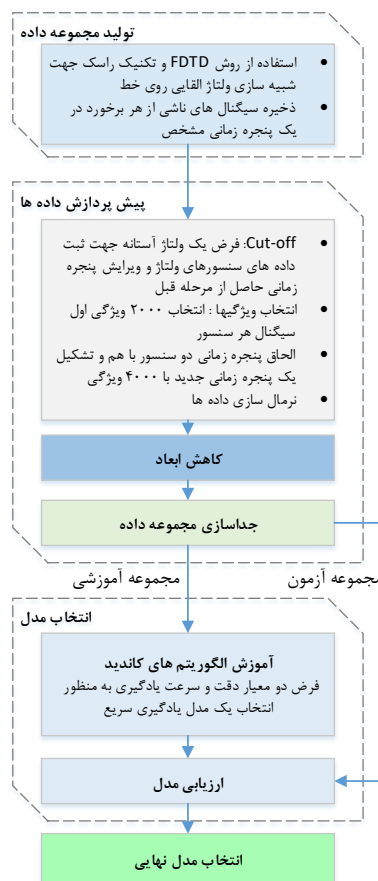
هندسه مسئله در شکل (۳) نشان داده شده است. ناحیه محاسباتی به صورت یک ناحیه مستطیل شکل 50×50 کیلومتر با دو خط انتقال موازی و بی نهایت بصورت تک سیم واقع بر بالای یک زمین کاملاً رسانا در نظر گرفته شده است. به منظور یافتن بهترین نقطه عملکرد حس گر ها، در هر خط انتقال دو حس گر ولتاژ در نظر گرفته شده است. ۱۰۰۰۰ موقعیت تصادفی یکنواخت برای منبع صاعقه در ناحیه مورد نظر در شکل (۳) توزیع شده است. شبیه سازی هریک از ولتاژهای القایی ناشی از صاعقه روی خط انتقال فرضی در نقطه مورد نظر برای سنسور ولتاژ با استفاده از فرمول راسک در نرم افزار متلب انجام شده است. طبق تعریف راسک، ولتاژ القا شده ناشی از صاعقه در یک نقطه روی خط انتقال بدون تلفات بصورت تک سیم بینهایت در بالای زمین رسانای کامل بصورت رابطه ۱ است [۱۶]:

$$v(x, t) = \frac{Z_0 I_0 \beta}{4\pi} \left[\frac{\gamma_-}{(d^2 + \beta^2 \gamma_-^2)} \left(1 + \left(r + \frac{\beta^2 \gamma_-^2}{\sqrt{(\beta c t)^2 + \rho^2 / \delta}} \right) \right) + \frac{\gamma_+}{(d^2 + \beta^2 \gamma_+^2)} \left(1 + \left(r + \frac{\beta^2 \gamma_+^2}{\sqrt{(\beta c t)^2 + \rho^2 / \delta}} \right) \right) \right] \quad (1)$$

که در آن :

$$\gamma_- = (ct - x) \quad \gamma_+ = (ct + x) \quad \delta = \frac{1}{1 - \beta^2} \quad \rho = x^2 + d^2$$

دامنه جریان بازگشتی صاعقه I_0 و سرعت آن در امتداد کانال صاعقه برابر βc است که در آن c سرعت نور در فضای آزاد و β که در این مقاله روی 0.4 تنظیم شده است که همان نسبت سرعت ضربه بازگشتی به سرعت نور است. t نشان دهنده زمان است و باید بزرگتر از d/c باشد. d فاصله افقی بین کانال صاعقه (نقطه برخورد)

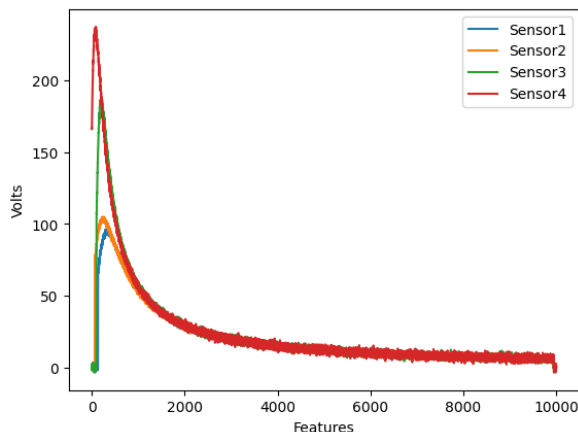


شکل (۲): روش انجام عملیات یادگیری ماشین و استخراج مدل نهایی

ناشی از آن بر روی خطوط انتقال صورت می گیرد و بر خلاف اکثر روش های رایج که حداقل به چهار سنسور برای مکان یابی نیاز دارند، در این روش با استفاده از ترکیب داده های سری زمانی به دست آمده از تنها دو سنسور مستقر روی خطوط بعنوان ورودی مدل مکان یابی صاعقه مبتنی بر یادگیری ماشین استفاده می شود. مدل مورد استفاده نیز یک مدل یادگیری ماشین XGBoost است که یک الگوریتم سریع و کارآمد می باشد. دلیل استفاده از XGBoost به عنوان الگوریتم اصلی، برتری مدل بدست آمده از آن نسبت به دو مدل مبتنی بر الگوریتم شبکه عصبی عمیق (DNN) است. این برتری بر اساس اعمال دو معیار دقت و زمان یادگیری حاصل شده است. مدل بدست آمده از الگوریتم XGBoost در مدت زمانی بسیار کوتاه به دقتی بالاتر از دو مدل دیگر دست یافت. همچنین برای بهبود کارایی محاسباتی ابعاد ورودی مدل کاهش داده شد. برای این منظور ابتدا چند الگوریتم کاهش ابعاد از جمله تجزیه و تحلیل مؤلفه اصلی (PCA) و تحلیل عاملی (FA) ارزیابی شد، در نهایت از PCA برای کاهش ابعاد ورودی به مدل استفاده گردید که منجر به کاهش قابل توجه در زمان محاسبه و در عین حال حفظ دقت مدل در سطح قابل قبولی شده است. علاوه بر این، بهینه سازی مدل با اصلاح پیکربندی هندسی مسئله به دست می آید که منجر به انتخاب بهترین آرایش سنسورها نیز می شود. لازم به ذکر است در این مطالعه سنسورها مستقل از یکدیگر عمل نموده و نیازی

۲-۲- پیش پردازش داده‌ها

در این بخش مطابق با شکل (۲) مراحل زیر انجام می‌گردد:



شکل (۴): نمونه‌ای از سیگنال‌های ولتاژ شبیه‌سازی شده نویزی ناشی از برخورد صاعقه در نقطه‌ای داخل ناحیه محاسباتی در محل

سنسورهای شکل (۳)

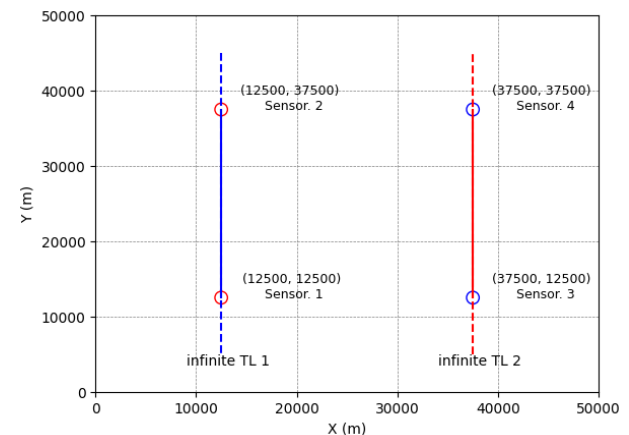
Cut-off: برای اینکه بتوان داده‌های شبیه‌سازی شده را به داده‌های ثبت شده توسط ضبط کننده‌های داده واقعی تقریب داد، لازم است ولتاژهای القایی محاسبه شده را تا لحظه اول رسیدن به آستانه معینی از ولتاژ، که معمولاً شرط لازم برای تحریک و ذخیره‌سازی در تجهیزات ثبت کننده سیگنال می‌باشد، حذف کرده و سپس تمام داده‌ها به ابتدای آرایه داده انتقال داد. به این ترتیب داده‌ها بسیار شبیه به داده‌های ثبت کننده داده‌های واقعی خواهند بود. مقدار ولتاژ آستانه ۱۰ ولت در نظر گرفته شده است.

انتخاب ویژگی‌ها: هر پنجره زمانی سیگنال ولتاژ القا شده که یک نمونه‌ی آموزشی در فرآیند یادگیری ماشین محسوب می‌گردد ۱۰۰۰۰ ویژگی دارد. از طرف دیگر یکی از مراحل عملیات یادگیری ماشین انتخاب ویژگی به منظور کاستن از ورود ویژگی‌های دارای اهمیت کم است. با توجه به شکل موج ولتاژ القا شده (شکل (۴)) می‌توان گفت که با ارزش‌ترین ویژگی‌ها در شکل موج ناشی از صاعقه در زمان اولیه آن قرار دارند. به همین دلیل، تنها ۲۰۰۰ ویژگی اول از پنجره زمانی انتخاب گردید. این امر به منظور حفظ منابع سیستم صورت گرفت.

الحاق سیگنال‌های دو سنسور: یکی از اهداف این مقاله استفاده از حداقل تعداد سنسور است. بنابراین تعداد دو عدد سنسور برای مکان‌یابی در نظر گرفته شد و سیگنال‌های سنسورها دوبه‌دو مطابق با شکل (۵) هم ادغام شدند. همان‌گونه که در شکل مشاهده می‌شود، داده‌های دو حسگر که هرکدام دارای ۲۰۰۰ ویژگی می‌باشند پس از الحاق با یکدیگر یک پنجره زمانی را تشکیل داده‌اند که دارای ۴۰۰۰ ویژگی است.

نرمال‌سازی داده‌ها: نرمال‌سازی داده‌ها یکی از مهم‌ترین مراحل پیش‌پردازش داده‌ها است. در این مرحله تمام داده‌های مورد نظر بر

و خط انتقال و x فاصله در امتداد خط انتقال است. Z_0 امپدانس مشخصه فضای آزاد است.



شکل (۳): هندسه مسئله و موقعیت (پیکربندی) خطوط انتقال واقع در ناحیه محاسباتی و محل قرار گیری سنسورهای ولتاژ

ارتفاع سیم از سطح زمین ۱۰ متر فرض شده است. منبع تحریک (در اینجا صاعقه) یک جریان پله ای است که در امتداد کانال برخورد صاعقه صعود می‌کند مقدار پیک جریان ۱۲ کیلوآمپر در نظر گرفته شده است. برای محاسبه ولتاژهای القایی در نزدیکترین نقطه خط انتقال به صاعقه ($x = 0$)، از رابطه ۲ استفاده می‌شود:

$$v(0, t) = \frac{Z_0 I_0 h}{4\pi d^2} \cdot \frac{2\beta ct}{1 + \left(\frac{\beta ct}{d}\right)^2} \cdot \left(1 + \beta^2 \left(\frac{ct/d}{\sqrt{1 + \beta^2((ct/d)^2 - 1)}}\right)\right) \quad (2)$$

این رابطه فقط برای $t \geq d/c$ معتبر است. لازم به ذکر است که سایر روش‌های محاسبه ولتاژ القایی دقیق‌تر، چه بر اساس تئوری خطوط انتقال یا شبیه‌سازی‌های موج کامل می‌توانند به راحتی در روش پیشنهادی جایگزین شوند که در کارهای آینده به آن نیز پرداخته خواهد شد. با استفاده از روابط ۱ و ۲، ولتاژهای القایی ناشی از صاعقه در حوزه زمان در مکان‌های سنسورها محاسبه شده‌است.

ذکر این نکته ضروری است که دلیل قرار دادن چهار سنسور در شکل (۳) فقط پیدا نمودن یک ترکیب داده‌ای مناسب از دو سنسور نهایی است.

به منظور نزدیک‌تر کردن داده‌های جمع‌آوری شده به داده‌های واقعی، پس از شبیه‌سازی سیگنال‌های ولتاژ، با اعمال یک نویز با نسبت سیگنال به نویز 30dB سیگنال‌ها نویزی شدند. نمونه‌هایی از سیگنال‌های شبیه‌سازی شده به روش فوق مربوط به سنسورهای ۱ تا ۴ از شکل (۳) در شکل (۴) نشان داده شده است.

در شبیه‌سازی ولتاژهای القا شده ناشی از صاعقه از گام‌های زمانی $1\mu s$ استفاده شده است. پنجره زمانی ثبت هر سیگنال 10ms است که با احتساب گام‌های زمانی، دارای ۱۰۰۰۰ داده (سمپل) می‌باشد (مطابق با شکل (۴)) که هرکدام یک ویژگی برای عملیات یادگیری ماشین محسوب می‌گردد.

جدول (۳) مقایسه مدل‌ها با توجه به معیارهای در نظر گرفته شده نشان داده شده است.

مطابق با نتایج جدول (۳) و بر اساس معیارهای انتخاب مدل، می‌توان نتیجه گرفت که مدل مبتنی بر الگوریتم XGBoost در مدت زمانی به مراتب کمتر از دو مدل دیگر به دقتی بالاتر از آن‌ها دست یافته است لذا با توجه به برتری نتایج XGBoost از نظر رسیدن به دقت مطلوب در زمان کوتاه‌تر، نتایج مفصلاً و به طور خاص برای روش XGBoost ارائه شده است. XGBoost یک سیستم تقویت درخت سرتاسر مقیاس پذیر است و به عنوان یک الگوریتم یادگیری ماشینی قوی است که می‌تواند به درک داده‌ها و تصمیم‌گیری بهتر کمک کند [۱۹]. این الگوریتم توسعه‌ای از الگوریتم تقویت گرادینت^{۱۱} است. تقویت گرادینت یک روش یادگیری گروهی است که با ترکیب پیش‌بینی‌های چندین مدل ضعیف‌تر (معمولاً درخت‌های تصمیم) یک مدل پیش‌بینی قوی می‌سازد. این کار را با آموزش متوالی مدل‌های جدید برای تصحیح خطاهای مدل‌های قبلی انجام می‌دهد

جدول (۱): مشخصات دو مدل شبکه عصبی مورد بررسی

DNN2	DNN1	
۲	۳	تعداد لایه مخفی در شبکه عصبی
۱۰۰	۶۴	تعداد نوروں در هر لایه مخفی
۷۲	۳۲	مقدار batch در شبکه عصبی

جدول (۲): مشخصات پارامترهای مدل XGBoost

estimators	max_depth	eta	subsample	colsample_bytree
۴۰۰	۶۰	۰/۰۵۰	۰/۶	۰/۷

جدول (۳): مقایسه نتایج اولیه آموزش مدل‌های کاندید جهت

انتخاب مدل یادگیری سریع

XGBoost	DNN2	DNN1	
۸۶	۵۶۲	۵۹۵	زمان یادگیری (ثانیه)
۰/۹۹	۰/۸۵۴	۰/۹۸۱	دقت R2

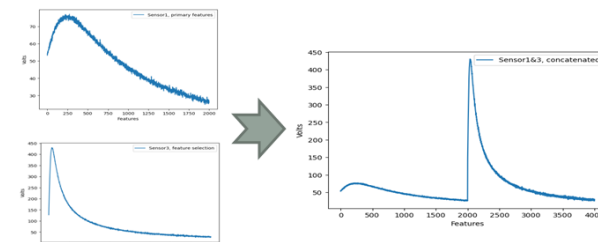
مرحله بعد انتخاب جفت حسگری است که اطلاعات آنها منجر به تولید یک مدل با دقت بالا می‌شود.

جدول (۴) دقت مدل آموزش دیده را با استفاده از داده‌های شش جفت حسگر مختلف نشان می‌دهد. با توجه به این جدول، بهترین عملکرد مدل با استفاده از داده‌های بدست آمده از زوج سنسورهای ۱ و ۴ و همچنین ۲ و ۳ به دست می‌آید. همچنین جدول مذکور مقایسه عملکرد مدل یادگیری ماشینی با و بدون PCA را نیز نشان می‌دهد. مشاهده می‌شود که زمان محاسبه در هنگام استفاده از روش PCA بسیار کاهش می‌یابد، از بیش از نیم ساعت بدون استفاده از روش PCA به کمتر از ۹۰ ثانیه با PCA. علاوه بر این استفاده از PCA تا حدودی دقت مدل را نیز افزایش می‌دهد.

مقدار حداکثر تقسیم شدند به طوری که همه داده‌ها در محدوده ۰ تا ۱ قرار گرفتند.

کاهش ابعاد: تعداد ویژگی‌های انتخاب شده جهت آموزش مدل با توجه به الحاق داده‌های دو حسگر، ۴۰۰۰ است. این تعداد ویژگی ورودی هم پیچیدگی‌های مدل و هم زمان آموزش و میزان محاسبات سیستم را افزایش خواهد داد. کاهش ابعاد ورودی مدل می‌تواند تاثیر زیادی بر ساده‌سازی مدل و نیز کاهش زمان و میزان محاسبات سیستم داشته‌باشد. روش‌های متعددی برای کاهش ابعاد ورودی یک مدل در یادگیری ماشینی وجود دارد از جمله آن‌ها می‌توان به تجزیه و تحلیل مؤلفه اصلی (PCA) و تحلیل عاملی (FA) اشاره نمود. ابتدا به دلیل محبوبیت استفاده، الگوریتم PCA انتخاب گردید. ایده اصلی روش PCA یافتن مهم‌ترین مؤلفه‌ها در داده‌ها و جایگزینی داده‌های اصلی با مهم‌ترین مؤلفه‌ها برای دستیابی به هدف کاهش ابعاد داده است [۱۷]. [۱۸]. بطور کلی، کاهش خطر برازش بیش از حد (Overfitting) یک مدل با ویژگی‌های نویز دار، ساده‌سازی تجسم داده‌ها، افزایش سرعت آموزش الگوریتم یادگیری ماشینی و افزایش دقت مدل از مزایای استفاده از PCA در یک مسئله یادگیری ماشینی است.

در این مرحله، با استفاده از PCA، تعداد ویژگی‌های ورودی مدل از ۴۰۰۰ به ۸۰ عدد کاهش داده شد اما تعداد مولفه‌ی نهایی PCA در طی یک فرآیند تنظیم پارامترها^{۱۰} که در بخش ۴ توضیح داده شده‌است، انتخاب شد.



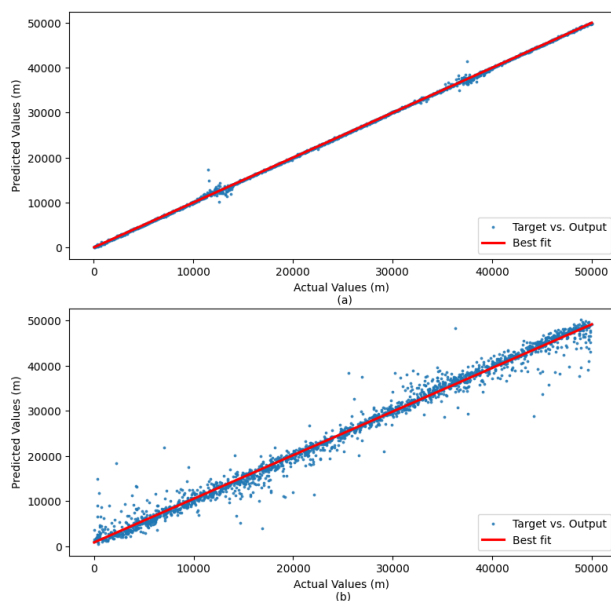
شکل (۵): الحاق داده‌های دو سنسور نمونه جهت آموزش مدل

تقسیم‌بندی داده‌های آموزشی و آزمون: برای ارزیابی مدل آموزش‌دیده، لازم است مجموعه داده‌ها به دو بخش داده‌های آموزشی و داده‌های آزمایشی (آزمون) تقسیم شود. با توجه به اینکه مختصات داده‌های آزمون مشخص بوده و در طی فرآیند یادگیری مدل کنار گذاشته می‌شوند لذا می‌توانند بعنوان معیاری برای ارزیابی دقت مدل محسوب شوند. نسبت تقسیم‌بندی به ترتیب ۲۰ و ۸۰ درصد برای داده‌های آزمایشی و آموزشی انتخاب شد.

۲-۳- استخراج مدل

سه مدل یادگیری ماشینی شامل XGBoost و دو مدل شبکه عصبی عمیق (DNN) با مشخصات ذکر شده در جداول ۱ و ۲ پیاده‌سازی شد. به منظور انتخاب مدل با یادگیری سریع، دو معیار زمان یادگیری و دقت مدل به عنوان معیار انتخاب مدل نهایی در نظر گرفته شد. در

دقت مدل مربوط به شکل (۳) را برای داده‌های زوج سنسور نمونه (۳و۲) نشان می‌دهد. همچنین نتایج ارزیابی مدل برای همه داده‌های زوج حسگرها در جدول (۵) آمده است. در این جدول علاوه بر خطای کلی مکان‌یابی مدل، خطا در راستای محور Xها و Yها نیز بطور جداگانه درج شده است. با دقت به نتایج جدول (۵) و با توجه به نوع استقرار خطوط (در داخل ناحیه محاسباتی) و حسگرهای انتخاب شده، دقت مدل تولید شده در ستون‌های مرتبط با زوج حسگرهای ۴و۱ و همچنین ۳و۲ به ۹۹ درصد رسیده است. ضمن آن‌که در هر دو ستون جدول، با توجه به نوع استقرار خطوط، دقت مدل در راستای محور Xها به مراتب کمتر از راستای محور Yها است.



شکل (۶): نتایج ارزیابی مدل برای (a) محور X و (b) محور Y در مورد پیکربندی هندسه شکل ۳ (داده‌های حسگرهای ۲ و ۳)

۴- بحث

در بخش‌های قبل مدل با یادگیری سریع انتخاب شد. اکنون به بررسی تاثیر پیکربندی‌های مختلف خط انتقال در دقت مدل، بررسی چند روش کاهش ابعاد ورودی مدل و در نهایت پیدا نمودن کمترین تعداد مولفه اصلی در روش PCA پرداخته خواهد شد.

۴-۱- تاثیر پیکربندی خطوط بر کارایی مدل

به منظور بررسی تاثیر پیکربندی‌های مختلف خط و همچنین حسگرها بر دقت و کارایی مدل چندین حالت مختلف قرارگیری خطوط انتقال در داخل ناحیه محاسباتی در نظر گرفته شد که به دلیل جلوگیری از طولانی شدن متن تنها به دو مورد خاص اشاره می‌شود. یکی از پیکربندی‌های مورد بررسی در شکل (۷) نشان داده شده است در این شکل، راستای قرارگیری خطوط انتقال بر خلاف شکل (۳)، در راستای محور Xها است. خطوط در داخل ناحیه قرار گرفته و

جدول (۴): دقت مدل آموزش دیده بر روی داده‌های تست از زوج سنسورهای ولتاژ انتخاب شده و همچنین مقایسه دو حالت با و بدون PCA

حسگرهای انتخاب شده	۲و۱	۳و۱	۴و۱	۳و۲	۴و۲	۴و۳
دقت (R2) با PCA	۰/۹۳	۰/۹۲	۰/۹۹	۰/۹۹	۰/۹۲	۰/۹۳
زمان آموزش با PCA (ثانیه)	۸۸	۹۱	۸۵	۸۶	۸۹	۸۵
دقت (R2) بدون PCA	۰/۹۳	۰/۹۱	۰/۹۸	۰/۹۷	۰/۹۱	۰/۹۲
زمان آموزش بدون PCA (ثانیه)	۱۹۳۰	۱۸۷۰	۱۸۵۷	۱۸۵۶	۱۸۶۳	۱۸۹۸

۳- ارزیابی مدل

در این مقاله، ارزیابی برخی الگوریتم‌های مختلف یادگیری ماشین، از جمله DNN و XGBoost، برای اطمینان از انتخاب یک مدل قوی، با یادگیری سریع و مناسب برای مکان‌یابی صاعقه انجام شد. ملاک انتخاب مدل نهایی در درجه اول دقت مدل و در درجه دوم مدت زمان آموزش بعنوان یک معیار کاهش هزینه در نظر گرفته شد. پس از انجام عملیات آموزش مدل‌ها برای یک نمونه مجموعه داده آموزشی (در اینجا مجموعه داده‌های حاصل از دو سنسور ۳و۲)، الگوریتم XGBoost به دلیل دقت قابل توجه و کارایی بالا به عنوان الگوریتم نهایی یادگیری ماشین انتخاب شد. فعالیت‌های بعدی در مقاله منحصراً با استفاده از مدل مبتنی بر XGBoost انجام شد.

جدول (۵): نتایج ارزیابی مدل برای داده‌های همه زوج سنسورها

حسگرهای انتخاب شده	۲و۱	۳و۱	۴و۱	۳و۲	۴و۲	۴و۳
R2 Score	۰/۹۲۶	۰/۹۲۱	۰/۹۸۹	۰/۹۹۰	۰/۹۲۲	۰/۹۲۷
RMSE [m]	۳۷۹۱	۴۰۹۹	۱۳۲۴	۱۳۹۲	۳۹۷۸	۳۹۵۱
RMSE(x) [m]	۵۲۰۰	۲۲۷	۲۶۴	۲۸۹	۲۴۳	۵۴۵۵
RMSE(y) [m]	۱۳۰۷	۵۷۹۲	۱۸۵۳	۱۹۴۷	۵۶۱۹	۱۲۰۹

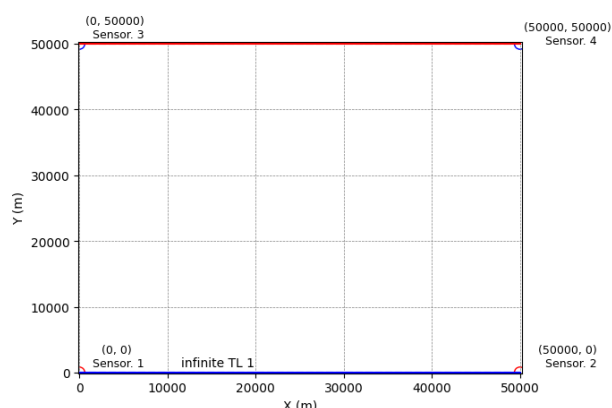
پس از انتخاب مدل، این بار مدل برای همه مجموعه داده‌های زوج حسگرها آموزش داده شد. در این مقاله از دو معیار R2 Score و ریشه میانگین مربع خطا (RMSE) برای ارزیابی دقت مدل استفاده شده است. امتیاز R2 نشان می‌دهد خط رگرسیون چقدر به داده‌های واقعی نزدیک است. خطای RMSE میانگین تفاوت بین مقادیر پیش‌بینی شده توسط یک مدل و مقادیر واقعی را اندازه‌گیری می‌کند و تخمینی از قدرت پیش‌بینی مدل را ارائه می‌دهد. شکل (۶) نمودار

است. با توجه به نتایج جدول می‌توان گفت که دقت مدل و خطای مکان‌یابی در همه داده‌های حسگرها افزایش یافته و در زوج حسگرهای روی یک خط و یا حسگرهای دارای مختصات X برابر بیشترین دقت را دارد.

بررسی‌ها نشان می‌دهد اولاً پیکربندی خطوط در میزان دقت مدل تاثیر داشته و ثانیاً حالتی که خطوط در لبه‌های ناحیه محاسباتی قرار می‌گیرد این دقت بیشترین مقدار خواهد بود.

جدول (۷): نتایج ارزیابی مدل آموزش دیده در مورد پیکربندی مربوط به شکل (۸)

زوج سنسور	۲و۱	۳و۱	۴و۱	۳و۲	۴و۲	۴و۳
دقت R2	۰/۹۹۸	۰/۹۹۶	۰/۹۹۳	۰/۹۹۳	۰/۹۹۷	۰/۹۹۸
RMSE [m]	۵۳۲	۶۹۸	۱۰۶۸	۱۱۲۵	۵۹۴	۶۳۸
RMSE(x) [m]	۷۴۴	۹۵۶	۱۴۷۹	۱۵۷۹	۸۱۳	۸۹۳
RMSE(y) [m]	۱۱۱	۲۴۷	۳۰۸	۲۰۰	۲۱۳	۱۳۶

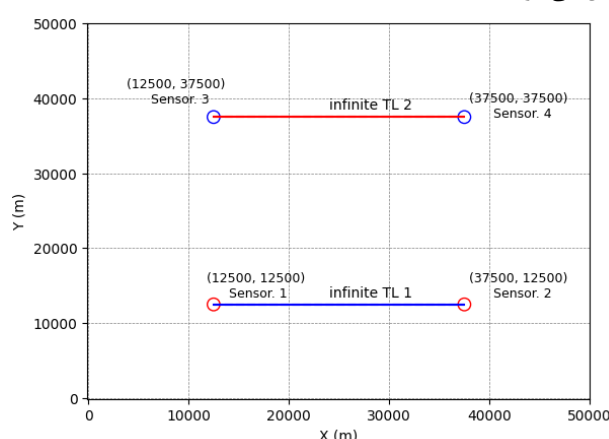


شکل (۸): پیکربندی نمونه دو خط انتقال موازی و در راستاری محور X واقع بر روی مرز ناحیه

۴-۲- تاثیر روش‌های کاهش ابعاد بر کارایی مدل

تکنیک‌های مختلفی برای کاهش ابعاد ویژگی‌های ورودی به مدل‌های یادگیری ماشین وجود دارد؛ از جمله آنها می‌توان به تجزیه و تحلیل مولفه اصلی (PCA) [۲۰]، تحلیل عاملی (FA) [۲۱]، حذف ویژگی روبه‌جلو و حذف ویژگی معکوس اشاره نمود. استفاده از الگوریتم‌های حذف ویژگی روبه‌جلو و حذف ویژگی معکوس به دلیل تعداد بالای ویژگی‌های موجود در این مسئله (۴۰۰۰ ویژگی) و افزایش زمان محاسبات منطقی نیست. همان‌گونه که در بخش قبلی مقاله توضیح داده شد در ابتدا و به منظور انتخاب الگوریتم یادگیری ماشین از الگوریتم PCA برای کاهش ابعاد استفاده شد. به منظور اطمینان از

حسگرها با حفظ نسبت موقعیت آنها نسبت به طول خط بر روی آن قرار گرفته‌اند. کلیه مراحل توضیح داده‌شده یادگیری مدل، برای این پیکربندی نیز اجرا شد. نتایج ارزیابی مدل مربوطه در جدول (۶) آمده است. با مقایسه بین نتایج مدل شکل (۳) با نتایج حاصل از مدل شکل (۷)، می‌توان گفت که با تغییر راستای قرارگیری خطوط از محور Y به محور X ، دقت مدل نیز در راستای محور Y بیشتر از محور X خواهد شد. ولی میزان دقت کلی مدل با مدل هندسه شکل (۳) تقریباً یکسان است زیرا بدون درنظر گرفتن راستای استقرار خطوط دو شکل باهم یکسانند. این امر نشان می‌دهد تغییر در راستای استقرار خطوط انتقال (بعنوان مثال X) سبب تغییر دقت در راستای مقابل (بعنوان مثال Y) خواهد شد. در این پیکربندی داده‌های حاصل از زوج حسگرهای ۴و۱ و همچنین ۳و۲ منجر به بیشترین دقت در مکان‌یابی صاعقه توسط مدل می‌شوند.



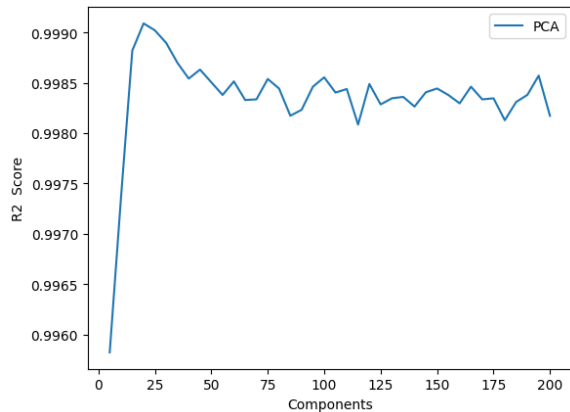
شکل (۷): پیکربندی نمونه خطوط انتقال موازی با محور X ها و در داخل ناحیه

جدول (۶): نتایج ارزیابی مدل آموزش دیده در مورد پیکربندی شکل (۷)

زوج سنسور	۲و۱	۳و۱	۴و۱	۳و۲	۴و۲	۴و۳
دقت R2	۰/۹۲۷	۰/۹۲۲	۰/۹۹۰	۰/۹۸۹	۰/۹۲۲	۰/۹۲۸
RMSE [m]	۳۸۷۳	۳۹۷۳	۱۲۳۳	۱۳۱۶	۴۰۵۶	۳۹۷۹
RMSE(x) [m]	۱۲۰۶	۵۶۱۵	۱۷۲۳	۱۹۳۱	۵۷۳۲	۱۱۶۵
RMSE(y) [m]	۵۳۴۳	۲۱۷	۲۷۳	۲۴۰	۱۹۲	۵۵۱۶

در پیکربندی دیگری که در شکل (۸) نشان داده شده است، راستای خطوط انتقال نسبت به شکل (۷) تغییر نکرده اما موقعیت خطوط به جای داخل ناحیه به کناره‌های آن و درست منطبق با اضلاع ناحیه و در راستای محور X ها منتقل شده است. هدف از بررسی این پیکربندی ارزیابی تاثیر نقاط قرینه نسبت خطوط در میزان دقت مدل می‌باشد. نتایج ارزیابی مدل مرتبط با این پیکربندی در جدول (۷) آمده

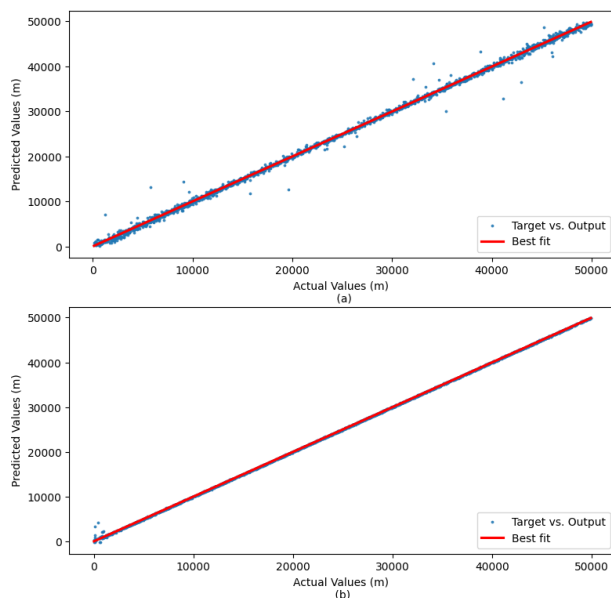
بالاترین دقت را در مدل مربوط به خود داشتند، مدل مجدداً آموزش داده شد. نتایج حاصل از ارزیابی مدل مربوط به آن در شکل (۱۰) و نیز جدول (۹) آمده است. در این حالت خطای کلی مدل در مکان‌یابی صاعقه ۴۳۲ متر، در راستای محور y ۱۳۲ متر و در راستای محور x ها به ۵۹۶ متر رسید.



شکل (۹): نمودار دقت مدل بر اساس تعداد مولفه‌های PCA

جدول (۹): نتایج ارزیابی مدل نهایی مربوط به داده‌های زوج حس‌گر ۲۱ از پیکربندی شکل با تعداد مولفه اصلی ۲۰ عدد در روش PCA

تعداد PC	دقت R2	RMSE [m]	RMSE(x) [m]	RMSE(y) [m]
۲۱	۰/۹۹۹۱	۴۳۲	۵۹۶	۱۳۲



شکل (۱۰): نمودار ارزیابی مدل نهایی با تعداد مولفه اصلی برابر با ۲۱

۵- نتیجه‌گیری

در این مقاله، روش مبتنی بر یادگیری ماشین برای مکان‌یابی صاعقه از طریق ولتاژهای القا شده روی خط انتقال بررسی شد.

درستی انتخاب روش کاهش ابعاد، الگوریتم تحلیل عاملی (FA) نیز به کار گرفته شد. در جدول (۸) نتایج حاصل از به کارگیری الگوریتم FA و PCA در مدل نهایی مربوط به هندسه شکل (۳) برای همه مجموعه داده‌ها نشان داده شده است. لازم به ذکر است در این مقایسه تعداد مولفه‌های هر دو روش یکسان و برابر ۸۰ در نظر گرفته شده است.

همان‌گونه که از نتایج جدول پیداست، روش FA، بجز در مواردی که کمی دقت را افزایش می‌دهد اختلاف چندانی با روش PCA ندارد. اما از طرف دیگر زمان محاسبات در روش FA بیشتر از PCA بوده و بنابراین با توجه به اهداف مقاله که یافتن مدل با یادگیری سریع است، می‌توان گفت در کل روش PCA انتخاب بهتری است.

جدول (۸): مقایسه نتایج ارزیابی کاهش ابعاد ورودی مدل با استفاده از تکنیک PCA و FA

زوج سنسور	۲۱	۳۱	۴۱	۳۲	۴۲	۴۳
دقت R2 با PCA	۰/۹۲۶	۰/۹۲۱	۰/۹۸۹	۰/۹۹۰	۰/۹۲۲	۰/۹۲۷
دقت R2 با FA	۰/۹۲۶	۰/۹۲۲	۰/۹۹۳	۰/۹۹۲	۰/۹۲۲	۰/۹۲۷
زمان آموزش PCA با	۶:۵۳	۶:۴۲	۶:۴۷	۶:۴۶	۶:۵۰	۶:۴۴
زمان آموزش FA با	۷:۴۲	۷:۳۰	۷:۳۸	۷:۳۵	۷:۴۳	۷:۳۶

۴-۳- کمترین تعداد مولفه اصلی با حفظ دقت مدل

به منظور انتخاب بهترین تعداد مولفه اصلی (PC) در روش PCA، مدل با تعداد مولفه‌های مختلف PCA آموزش داده شد. در شکل (۹)، نمودار دقت R^2 مدل به ازای تعداد مولفه‌های اصلی مختلف نشان داده شده است. همان‌گونه که در شکل مشاهده می‌شود با افزایش تعداد مولفه اصلی (PC) از عدد ۲ دقت مدل افزایش یافته تا حدود ۲۰ (دقیقتر ۲۱) به ۹۹/۹ درصد می‌رسد و پس از آن با افزایش بیشتر تعداد مولفه اصلی نه تنها افزایشی در دقت مدل مشاهده نمی‌شود که مقدار بسیار کمی کاهش نیز در دقت وجود دارد که یکی از دلایل آن را می‌توان تاثیر نویز در اطلاعات مربوط به مولفه‌های بالاتر دانست. ضمن آن‌که واضح است که هر چه تعداد ویژگی‌های ورودی یک مدل کمتر باشد سرعت آن مدل نیز در مقایسه با تعداد ویژگی‌های بیشتر، افزایش خواهد یافت و این یک امتیاز برای مدل نهایی محسوب می‌گردد.

با یافتن بهترین تعداد مولفه اصلی بعنوان ورودی مدل نهایی (PC=21) برای داده‌های حاصل از زوج حس‌گر ۲۱ از شکل (۸) که

- Kontani, "Broadband VHF sources locating system using arrival-time differences for mapping of lightning discharge process", *J. Atmos. Solar-Terrestrial Phys.*, vol. 67, no. 1031-1039, 2005.
- [10] E. P. Krider, R. C. Noggle, and M. A. Uman, "A Gated, Wideband Magnetic Direction Finder for Lightning Return Strokes", *J. Appl. Meteorol. Climatol.*, vol. 15, no. 3, pp. 301-306, 1976.
- [11] A. Mostajabi, H. Karami, M. Azadifar, A. Ghasemi, M. Rubinstein, and F. Rachidi, "Single-Sensor Source Localization Using Electromagnetic Time Reversal and Deep Transfer Learning: Application to Lightning", *Sci. Reports* 2019 91, vol. 9, no. 1, pp. 1-14, Nov. 2019, doi: 10.1038/s41598-019-53934-4.
- [12] G. Lugrin, N. M. Parra, F. Rachidi, M. Rubinstein, and G. Diendorfer, "On the location of lightning discharges using time reversal of electromagnetic fields", *IEEE Trans. Electromagn. Compat.*, vol. 56, no. 1, pp. 149-158, 2014, doi: 10.1109/TEMC.2013.2266932.
- [13] T. Wang, S. Qiu, L. H. Shi, and Y. Li, "Broadband VHF Localization of Lightning Radiation Sources by EMTR", *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 6, pp. 1949-1957, Dec. 2017, doi: 10.1109/TEMC.2017.2651142.
- [14] X. Wang, K. Hu, Y. Wu, and W. Zhou, "A Survey of Deep Learning-Based Lightning Prediction", *Atmos.* 2023, Vol. 14, Page 1698, vol. 14, no. 11, p. 1698, Nov. 2023, doi: 10.3390/ATMOS14111698.
- [15] M. Lu et al., "Lightning Strike Location Identification Based on 3D Weather Radar Data", *Front. Environ. Sci.*, vol. 9, p. 714067, Aug. 2021, doi: 10.3389/FENVS.2021.714067/BIBTEX.
- [16] S. Rusck, *Induced Lightning Over-Voltages on Power-Transmission Lines With Special Reference to the Over-Voltage Protection of Low-Voltage Networks*. Stockholm, Sweden: KTH, 1958.
- [17] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments", *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, Apr. 2016, doi: 10.1098/RSTA.2015.0202.
- [18] M. Ringnér, "What is principal component analysis?", *Nat. Biotechnol.* 2008 263, vol. 26, no. 3, pp. 303-304, Mar. 2008, doi: 10.1038/nbt0308-303.
- [19] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785-794, Mar. 2016, doi: 10.1145/2939672.2939785.
- [20] Jasmin Praful Bharadiya, "A Tutorial on Principal Component Analysis for Dimensionality Reduction in Machine Learning", *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 5, pp. 2028-2032, Jun. 2023, doi: 10.5281/ZENODO.8020810.
- [21] S. A. Mulaik, "Foundations of factor analysis", p. 524, 2010.

زیر نویس ها

¹ Lightning locating systems

² Magnetic direction finder

³ Time of arrival

⁴ Time difference of arrival

⁵ Interferometry

⁶ Electromagnetic time reversal

همچنین کارایی الگوریتم XGBoost بعنوان یک روش یادگیری نسبتاً سریع برای یافتن مکان برخورد صاعقه به زمین با استفاده از داده‌های ولتاژ القایی نویزدار در خطوط انتقال نشان داده شد. قابل ذکر است، عملکرد قوی مدل حتی در حضور داده‌های نویزدار مشخص است. با استفاده از تنها دو حسگر مستقل از هم، کاهش ابعاد ورودی مدل از ۴۰۰۰ به ۸۰ توسط الگوریتم PCA و داده‌ها با نسبت سیگنال به نویز ۳۰ دسی بل، مدل پیشنهادی به دقت ۹۹ درصد دست یافت. کاهش ابعاد به طور قابل توجهی سرعت محاسبات را افزایش داد. با بررسی دقت مدل بر اساس تعداد مولفه‌های PCA مشخص گردید که سرعت و دقت مدل با تعداد مولفه‌های کمتر در حدود ۲۱ مولفه بهبود می‌یابد. همچنین نشان داده شد که محل سنسورها و پیکربندی خطوط انتقال می‌تواند بر عملکرد مدل تأثیر بگذارد.

مراجع

- [1] H. Karami, A. Mostajabi, M. Azadifar, M. Rubinstein, C. Zhuang, and F. Rachidi, "Machine Learning-Based Lightning Localization Algorithm Using Lightning-Induced Voltages on Transmission Lines", *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 6, pp. 2512-2519, Dec. 2020, doi: 10.1109/TEMC.2020.2978429.
- [2] K. Mehranzamir, A. B. Pour, Z. Abdul-Malek, H. N. Afrouzi, S. M. Alizadeh, and M. Hashim, "Implementation of Ground-Based Lightning Locating System Using Particle Swarm Optimization Algorithm for Lightning Mapping and Monitoring", *Remote Sens.* 2023, Vol. 15, Page 2306, vol. 15, no. 9, p. 2306, Apr. 2023, doi: 10.3390/RS15092306.
- [3] T. Zhang, J. Wang, Q. Ma, and L. Fu, "Improving the Detection Effect of Long-Baseline Lightning Location Networks Using PCA and Waveform Cross-Correlation Methods", *Remote Sens.* 2024, Vol. 16, Page 885, vol. 16, no. 5, p. 885, Mar. 2024, doi: 10.3390/RS16050885.
- [4] H. Abolghasempour and A. A. Razi-Kazemi, "Transient modeling of photovoltaic system (PV) to investigating the peak of transient voltage generated by direct and indirect lightning strikes", *J. Iran. Assoc. Electr. Electron. Eng.*, vol. 20, no. 2, pp. 159-169, 2023.
- [5] R. Shariatinasab and P. Tadayyon, "Estimation of Lightning Performance and Failure Risk of Overhead Lines Caused by Direct Strokes Based on Monte Carlo Method", *Journal of Iranian Association of Electrical and Electronics Engineers*, vol. 13, no. 1, pp. 85-94, 2016.
- [6] G. Diendorfer et al., "Review of CIGRE Report 'Cloud-to-Ground Lightning Parameters Derived from Lightning Location Systems – The Effects of System Performance'", 2009.
- [7] K. Mehranzamir, Z. Abdul-Malek, H. Nabipour Afrouzi, S. Vahabi Mashak, C. leong Wooi, and R. Zarei, "Artificial neural network application in an implemented lightning locating system", *J. Atmos. Solar-Terrestrial Phys.*, vol. 210, p. 105437, Nov. 2020, doi: 10.1016/J.JASTP.2020.105437.
- [8] A. Alammari et al., "Lightning mapping: Techniques, challenges, and opportunities", *IEEE Access*, vol. 8, pp. 190064-190082, 2020, doi: 10.1109/ACCESS.2020.3031810.
- [9] T. Tantisattayakul, K. Masugata, I. Kitamura, and K.

⁷ k-nearest neighbors

⁸ Principal component analysis

⁹ Factor analysis

¹⁰ Hyper parameter tuning

¹¹ Gradient boosting algorithm

¹² Root mean square error