

Fault Classification and Location on High Voltage Lines Using the Random Forest Algorithm

Dr. George Isber*
Dr. Ali Alsayed**
Rama Alkhtiar***

(Received 12 / 1 / 2023. Accepted 16 / 2 / 2023)

□ ABSTRACT □

Accurate identification of fault type and location in electrical power transmission lines contributes to improving the reliability of the electric power system.

In this research, the fault type and location in 230 kV high-voltage line were determined using the random forest algorithm, which is characterized by its high accuracy and its high ability to handle a large number of data, by using the amplitudes of the currents and voltages signals measured from one end of the transmission line, where they were used as input features for different faults on the line with different fault parameters (fault distance and resistance).

The results showed the ability of the proposed algorithm to classify the type of fault and determine its location in the transmission line with an accuracy of 100% in order to determine the type of fault and a maximum error in fault location of 0.402 km.

Keywords: electric power transmission system, high voltage lines, type of fault, fault location, random forest algorithm.

* Professor - Department of Electrical Power - Faculty of Mechanical and Electrical Engineering - Tishreen University - Lattakia – Syria. George.Isber@yahoo.com

**Associate Professor - Department of Electrical Power - Faculty of Mechanical and Electrical Engineering - Damascus University - Damascus – Syria. dr.alsayed@yahoo.com

*** Master Student - Department of Electrical Power - Faculty of Mechanical and Electrical Engineering - Tishreen University - Lattakia – Syria. rama.alkhtiar@tishreen.edu.sy

تصنيف العطل وتحديد موقعه على خطوط التوتر العالي باستخدام خوارزمية الغابات العشوائية

د. جورج اسبر*

د. علي السيد**

راما الختبار***

(تاريخ الإيداع 12 / 1 / 2023. قُبِلَ للنشر في 16 / 2 / 2023)

□ ملخّص □

يساهم التحديد الدقيق لنوع وموقع العطل على خطوط نقل الطاقة الكهربائية في تحسين موثوقية نظام القدرة الكهربائية. تم في هذا البحث تحديد نوع وموقع العطل على خط توتر عالي 230 kV باستخدام خوارزمية الغابة العشوائية التي تتميز بارتفاع دقتها وقدرتها العالية على التعامل مع عدد كبير من البيانات، باستخدام مطالات إشارات تيارات وجهود الأطوار المقاسة من نهاية واحدة لخط النقل، حيث أُستخدِمت كميزات دخل عند حدوث أعطال مختلفة على الخط ببارامترات عطل مختلفة (مسافة العطل ومقاومته). بينت النتائج قدرة الخوارزمية المقترحة في تصنيف نوع العطل وتحديد موقعه على خط النقل المستخدم بدقة وصلت إلى 100% من أجل تحديد نوع العطل وخطأ أعظمي في تحديد مكان العطل مقداره 0.402 km.

الكلمات المفتاحية: نظام نقل الطاقة الكهربائية، خطوط التوتر العالي، نوع العطل، مكان العطل، خوارزمية الغابات العشوائية.

*أستاذ- قسم الطاقة الكهربائية- كلية الهندسة الميكانيكية والكهربائية- جامعة تشرين- اللاذقية- سورية.

George.Isber@yahoo.com

**أستاذ مساعد- قسم الطاقة الكهربائية- كلية الهندسة الميكانيكية والكهربائية- جامعة دمشق-دمشق- سورية

dr.alsayed@yahoo.com

***طالبة ماجستير- قسم الطاقة الكهربائية-كلية الهندسة الميكانيكية والكهربائية-جامعة تشرين-اللاذقية- سورية

rama.alkhtiar@tishreen.edu.sy

مقدمة:

يتعرض نظام نقل الطاقة الكهربائية لأنواع مختلفة من الأعطال، تشكل هذه الأعطال 90-80 % من أعطال النظام ككل [1]، وتقسّم هذه الأعطال إلى قسمين:

- ❖ أعطال عابرة: وهي الأعطال التي تزول من تلقاء نفسها بعد فترة زمنية قصيرة.
 - ❖ أعطال دائمة: وهي غالباً أعطال القصر وتحتاج إلى إصلاح، تقلل هذه الأعطال من موثوقية تشغيل نظام القدرة الكهربائية، وتعد مصدراً يُهدد استمرارية التغذية الكهربائية، لذلك من الضروري تحديد نوع وموقع العطل على خطوط التوتر العالي بدقة عالية بهدف إزالته بسرعة واتخاذ التدابير المناسبة لتقليل حدوثه مستقبلاً عند تلك المواقع، وبالنتيجة تحسين جودة أنظمة الطاقة الكهربائية، وكذلك استقرار النظام [1].
- توجه الباحثون في الوقت الحالي إلى الاعتماد على تقنيات الذكاء الاصطناعي بدلا من الطرق التقليدية نظرا لدقتها العالية [2]، حيث حققت خوارزميات تعلم الآلة نجاحا كبيرا في مختلف مجالات الهندسة، وتميزت بالإضافة للدقة العالية بسرعة الاستجابة، مما يسهم في رفع سوية أمان نظام القدرة الكهربائية [3].

ذكر الباحثون في المرجع [4] أنه تم استخدام العديد من خوارزميات تعلم الآلة ومنها خوارزمية Support Vector Machine (SVM) وخوارزمية Extreme Learning Machine (ELM) لتحديد موقع العطل على خطوط النقل ولكن أظهرت النتائج أن أداء هذه الخوارزميات يتراجع بزيادة عدد ميزات الدخل، لذلك استخدمت خوارزمية الغابة العشوائية، التي تتميز بقدرة عالية على التعامل مع عدد كبير من ميزات الدخل في تصنيف العطل وتحديد موقعه على خط نقل 230 kV مغذى من جهتين، بينت النتائج الدقة العالية للخوارزمية لم يتجاوز متوسط أخطاء اختبار تحديد موقع العطل 0.11 % ولكن تم التحقق من أدائها فقط لأجل عطل أحادي طور مع أرض.

قام الباحثون أيضا في المرجع [5] بتصنيف نوع العطل على خط توتر متوسط 20 kV باستخدام خوارزمية الغابة العشوائية، حيث ثبت أن هذه الخوارزمية مُصنّف مثالي مقارنة بأنواع أخرى من المُصنّفات، ولكن البحث اكتفى بتحديد نوع العطل دون مكانه.

تم في هذا البحث تصنيف نوع العطل وتحديد موقعه على خط توتر عالي 230 kV مغذى من منبع واحد باستخدام خوارزمية الغابة العشوائية عند أنواع أعطال مختلفة (عطل طور مع أرض - عطل ثنائي طور دون أرض - عطل ثنائي طور مع أرض - عطل ثلاثي طور)، وذلك باستخدام بيئة "Python" البرمجية.

أهمية البحث وأهدافه:

تأتي أهمية البحث من استخدام إحدى الخوارزميات المتقدمة في مجال تعلم الآلة (خوارزمية الغابات العشوائية) في تحديد نوع وموقع العطل بدقة عالية على خطوط التوتر العالي بهدف زيادة موثوقية نظام الطاقة الكهربائية.

طرائق البحث ومواده:

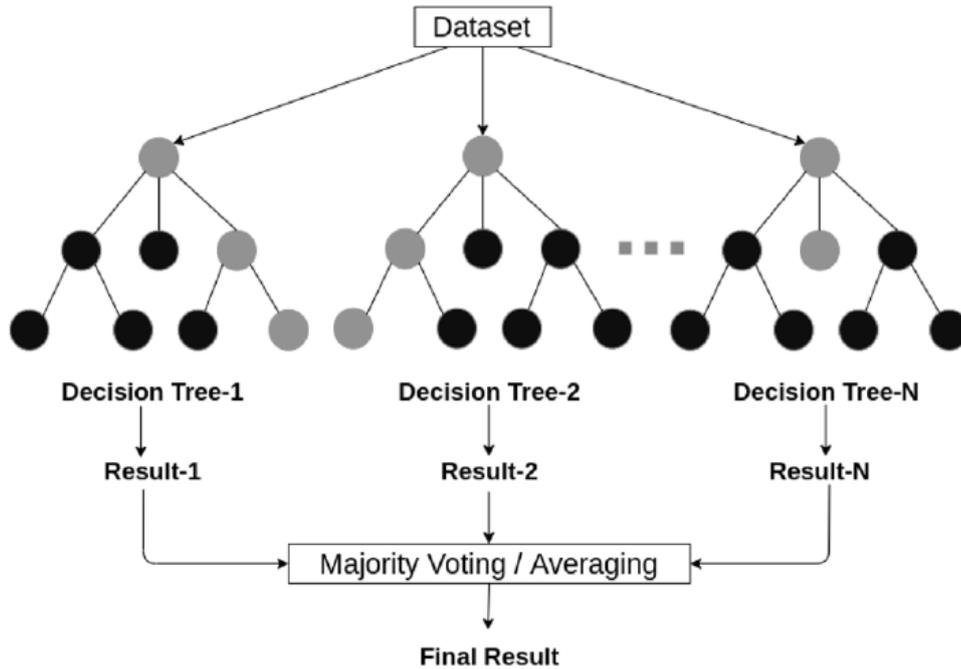
اعتمدنا في هذا البحث طريقة النمذجة والمحاكاة لخط نقل طاقة ثلاثي الطور 230 Kv في برنامج ماتلاب، وتم استخدام خوارزمية الغابة العشوائية لتحديد نوع العطل ومكانه على هذا الخط حيث تُستخدم مطالات جهود وتيارات الأطوار عند التردد الأساسي لشبكة التغذية 50Hz كميزات دخل للخوارزمية عند حدوث عطل على الخط ببارامترات مختلفة (مسافة العطل ومقاومته)، أما خرج الخوارزمية هو نوع العطل ومكانه.

1. خوارزمية الغابة العشوائية (Random Forest (RF))

تعرف خوارزمية الغابة العشوائية على أنها تقنية تجميع، تم تقديمها لأول مرة عام 2001 من قبل Breiman، تُستخدم في مجال تعلم الآلة لأجل مهام التصنيف والتنبؤ الترابطي.

تعتمد في عملها المعالجة على التوازي، إذ تقوم بتدريب عدة أشجار قرار (مجموعات بيانات فرعية مختلفة) مرتبة على التوازي بطريقة التعلم المراقب (حيث أن دقة التنبؤ عند استخدام شجرة قرار واحدة منخفضة نسبياً، أيضاً يمكن أن يسبب أي تغيير بسيط في ميزات التدريب تغييرات كبيرة في هيكلية الشجرة)، وأثناء عملية التصنيف يكون لكل شجرة خرج خاص، وبناء على قرار خرج الأغلبية تحدد النتيجة النهائية [5].

يبين الشكل (1) مبدأ عمل خوارزمية RF [6]:



الشكل (1) مبدأ عمل خوارزمية الغابة العشوائية

❖ مبدأ عمل شجرة القرار (Decision Tree):

تمثل أشجار القرار على شكل مجموعة من الأسئلة والتي تقسم عينات التدريب إلى قيم أكبر وأصغر، تسأل فقط أسئلة ذات جواب نعم أو لا، تبحث عن كل المتغيرات الممكنة وعن كل القيم الممكنة بهدف الوصول لأفضل تقسيم (السؤال الذي يقسم البيانات إلى جزأين مع أعلى دقة (فرع يحوي صنف عطل واحد))، ثم تكرر العملية لأجل كل جزء من البيانات [7].

- ميزات تطبيق خوارزمية الغابة العشوائية (استخدام مجموعة من أشجار القرار):
- ❖ تستخدم عدة أشجار قرار، لذلك هي فعالة جدا لتقليل مشكلة فرط التعلم عند استخدام شجرة قرار واحدة.
- ❖ تقلل من زمن تدريب البيانات.
- ❖ تعمل بكفاءة لأجل مجموعة كبيرة من البيانات وتعطي دقة تنبؤ عالية.

2. مراحل العمل:

جرت عملية المعالجة وفق المراحل التالية والمبينة في الشكل(2):



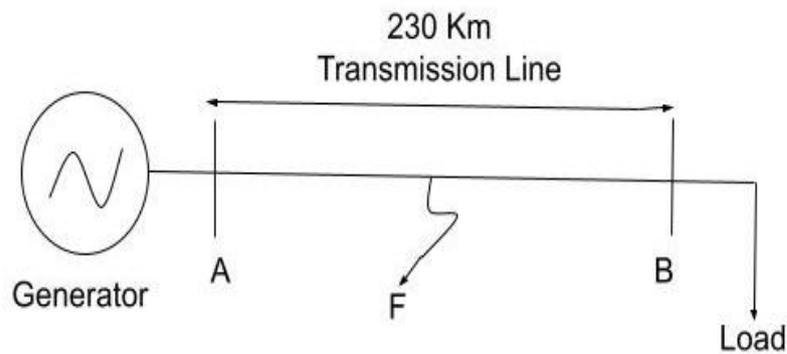
الشكل (2) المخطط التدفقي لآلية العمل

1- بناء نموذج خط نقل ثلاثي الطور بطول 50 Km، مغذى من منبع واحد بتوتر 230 kV، ويتردد 50 Hz في بيئة MATLAB –SIMULINK.
تعطى بارامترات النظام المدروس بالجدول (1):

جدول (1) بارامترات نظام نقل الطاقة المدروس [8]

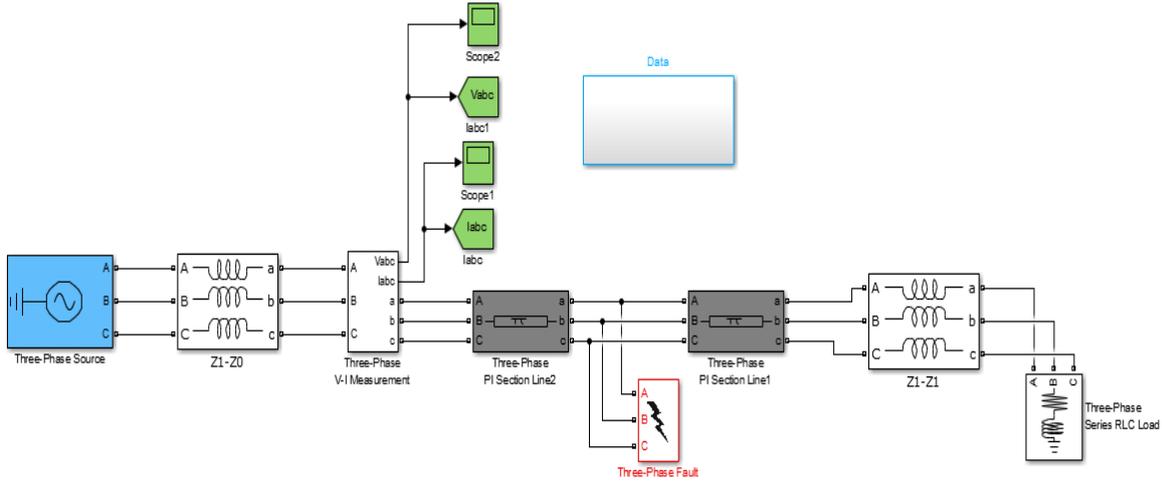
بارامترات منبع التغذية	
1.32+j15 [Ω]	ممانعة التتابع الموجب Z1
2.33+j26.8 [Ω]	ممانعة التتابع الصفري Z0
بارامترات خط التوتر الثلاثي الطور	
0.01273 [Ω/Km]	مقاومة التتابع الموجب Z1
0.1273 [Ω/Km]	مقاومة التتابع الصفري Z0
0.9337 [Ω/mH]	تحريضية التتابع الموجب L1
2.8 [Ω/mH]	تحريضية التتابع الصفري L0
12.74 [nF/Km]	سعة التتابع الموجب C1
7.75 [nF/Km]	سعة التتابع الصفري C0

يبين الشكل (3) مخطط الخط الواحد (One-Line diagram) لخط التوتر العالي المدروس:



الشكل (3) مخطط الخط الواحد لخط النقل المستخدم

وبين الشكل (4) نموذج نظام نقل الطاقة المدروس في برنامج MATLAB:



الشكل (4) نموذج نظام نقل الطاقة المدروس في بيئة SIMULINK

2- محاكاة الأعطال:

AG-BG-CG-AB-BC-AC-ABG-BCG-ACG-ABC

حيث:

A: رمز الطور الأول.

B: رمز الطور الثاني.

C: رمز الطور الثالث.

G: رمز الأرضي.

تم تجهيز بيانات التدريب والاختبار بتغيير كل من موقع العطل (L) ومقاومته (R_f) عند كل نوع من أنواع الأعطال السابقة الذكر كما هو مبين في الجدول (2).

الجدول (2) بيانات العطل

AG-BG-CG-AB-BC-AC-ABG-BCG-ACG-ABC	نوع العطل N
5, 15	مقاومة العطل R_f [Ω]
0.5-50 Step: 0.5	مكان حدوث العطل L [Km]

3- معالجة إشارات تيارات وجهود الأطوار المقاسة عند قضيبة التجميع A عند حدوث العطل، وذلك بتمريرها عبر مرشح anti-aliasing ثم تقطيعها بتردد تقطيع (Sampling Frequency) مساوٍ ل 1 KHz .

4- الحصول على مطال التردد الأساسي (fundamental frequency phasor) لإشارات تيارات وجهود الأطوار عند حدوث الأعطال لكل حالة من الحالات المختلفة السابقة باستخدام تحويل فورييه المتقطع . Discrete Fourier Transform (DFT)

5- تخزين بيانات العطل بصيغة EXCEL حيث يشكل كل سطر من مجموعة البيانات عينة تتألف من مجموعة سمات وهي على التوالي: نوع العطل (N)، مقاومة العطل (Rf)، مسافة العطل (S)، مطالات الجهود (V1-V2-V3) والتيارات (I1-I2-I3) عند التردد الأساسي 50 Hz، كما يبين الجدول (3)، تستخدم هذه القيم في تدريب واختبار الخوارزمية المقترحة حيث تمثل مطالات الجهود والتيارات دخل الخوارزمية ونوع العطل وموقعه خرجها.

N	Rf	S	V1	V1	V3	I1	I2	I3
BCG	5	21.5	187108.357	75306.6386	52986.046	1872.108	7867.624	7547.456
AG	5	28.5	95623.957	176330.416	189732.4933	5926.291	1676.262	1955.644
BG	15	27.5	181854.693	116543.260	174527.282	1876.199	5474.329	1732.625
ACG	5	49.0	89030.941	183898.626	101950.074	5369.733	1938.569	5959.000
ACG	15	24.0	98453.662	200075.998	102120.232	6538.758	1867.396	6413.594
ABC	5	6.0	51004.229	62170.217	56952.127	10375.848	10373.522	10374.284

الجدول (3) بنية بيانات التدريب المستخدمة

7 - تقسيم البيانات بعد معالجتها إلى 80 % من أجل تدريب الخوارزمية و 20 % من أجل اختبارها. تم تنفيذ الخوارزمية باستخدام لغة البرمجة Python ضمن بيئة Google Colab، حيث تم استخدام المصنف Random Forest Classifier من مكتبة تعلم الآلة سايكيت ليرن (Scikit Learn Library) لبناء الغابة العشوائية لمهمة التصنيف، وتقنية Random Forest Regressor لبناء الغابة العشوائية لأجل مهمة التنبؤ الترابطي.

3. معايير تقييم الأداء:

تم في هذا البحث تقييم أداء خوارزمية تعلم الآلة المستخدمة من أجل التنبؤ بنوع العطل على خط التوتر العالي المدروس باستخدام مصفوفة الارتباك (Confusion Matrix) التي تعتبر من أفضل المقاييس في مجال التحليل التنبؤي [9]، كما أظهرنا النتائج باستخدام تقرير التصنيف (Classification Report) والذي يعتمد على المقاييس الثلاثة التالية التي يتم حسابها من خلال هذه المصفوفة [10]:

1- مقياس الدقة ويعرف أيضا باسم التنبؤية الإيجابية: يعتبر أهم المقاييس، كلما زادت دقة المصنف كلما كان أفضل. ويعطى بالمعادلة:

$$precision = \frac{TP}{TP+FP} \dots\dots\dots(1)$$

2- مقياس Recall:

يعطى بالمعادلة:

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(2)$$

3- مقياس F1-Score

يعطى بالمعادلة:

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \dots\dots\dots(3)$$

حيث كلما اقتربت قيمة هذه المقاييس من الواحد كلما كان الحساب أدق.

- ومن أجل تقييم الأداء عند تحديد موقع العطل اعتمدنا على مقياس الخطأ المطلق (Absolute Error) والذي يعطى بالمعادلة:

$$\text{Error(km)} = | \text{RF output} - \text{Target} | \dots\dots\dots(4)$$

حيث :

RF output: مسافة العطل المقدره من قبل خوارزمية الغابات العشوائية ب km.

Target: مسافة العطل الحقيقية ب km.

كما اعتمدت أيضاً في عملية التقييم قيمة متوسط القيمة المطلقة للخطأ Mean Absolute Error (MAE) والتي تعطى بالمعادلة:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - x| \dots\dots\dots(5)$$

حيث :

X: القيمة الحقيقية، و x_i : القيمة المتنبأ بها.

كلما اقتربت قيمة متوسط القيمة المطلقة للخطأ من الصفر كلما كانت الخوارزمية أفضل.

النتائج والمناقشة:

تم ضبط بارامترات الغابة العشوائية لكل من مهمة التصنيف والتنبؤ الترابطي بحيث نحصل على الأداء الأفضل، وكان ذلك عند احتواء الغابة على 500 شجرة قرار (n-estimators= 500)، وعملية التقريع غير محددة العمق (max-depth= None).

أعطت عملية التدريب واختبار الغابة العشوائية النتائج التالية:

1- نتائج عملية تصنيف نوع العطل:

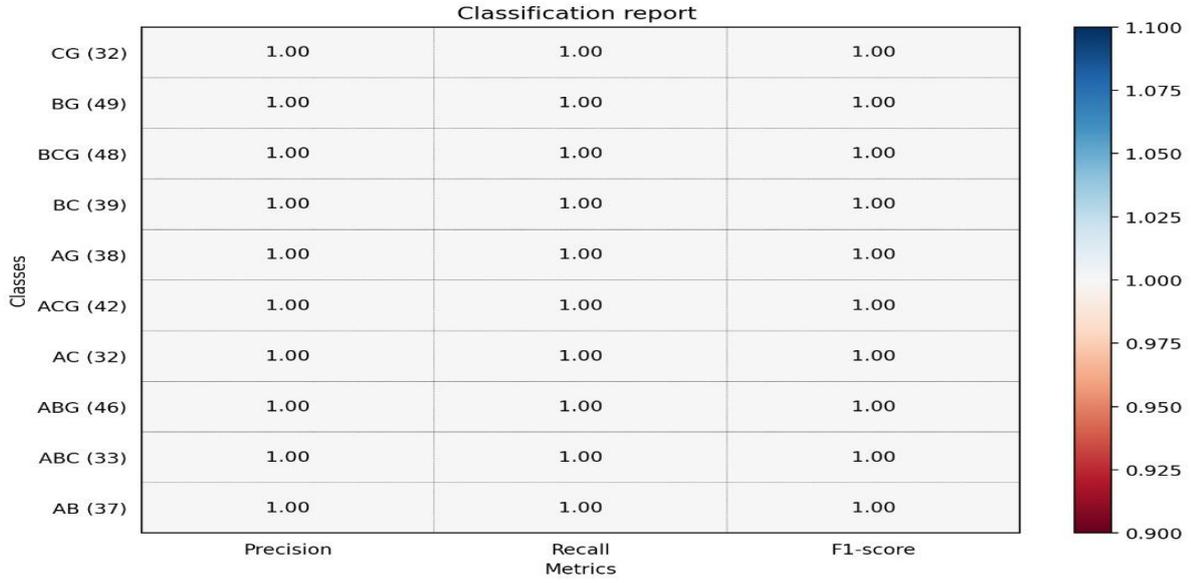
يبين الشكل (5) مصفوفة الارتباك:



الشكل (5) مصفوفة الارتباك

تظهر مصفوفة الارتباك السابقة عدد العينات التي تم توقعها بشكل صحيح وعدد العينات التي تم توقعها بشكل خاطيء لأجل كل صنف من الأعطال، على سبيل المثال نلاحظ في العمود الأول من الجهة اليسرى من المصفوفة أنه لأجل عينات الصنف الأول AG والتي تشكل 9.34 % من مجموعة عينات الاختبار الكلية كانت التوقعات صحيحة 100 % ويمكن قراءة هذا بسبب وجود قيم صفرية مقابلة لكل الأصناف باستثناء الصنف AG، وبشكل مماثل نستنتج ذلك لبقية أنواع الأعطال.

ويبين الشكل (6) تقرير التصنيف الذي حصلنا عليه:



الشكل (6) تقرير تصنيف نوع العطل

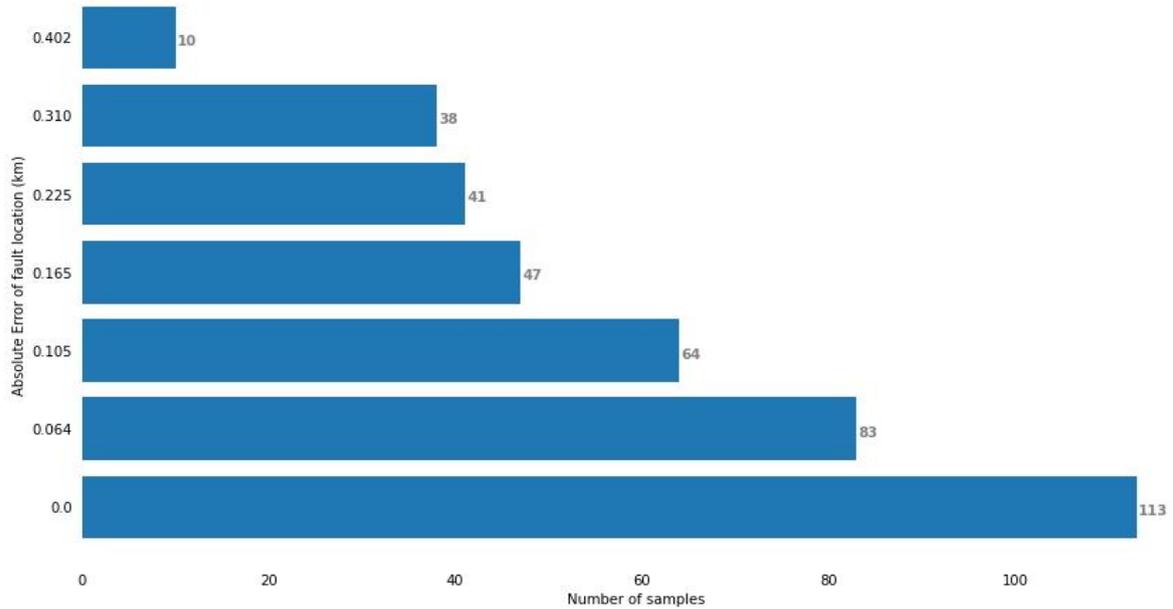
نلاحظ من خلال تقرير التصنيف السابق أننا حصلنا على دقة 100% لجميع أنواع الأعطال من أجل جميع المقاييس. نتائج تحديد موقع العطل:

2- يبين الجدول (4) نتائج تحديد مكان العطل لعينات عشوائية من بيانات الاختبار:

جدول (4) تحديد مسافة العطل والخطأ المطلق عند قيم مختلفة لمقاومات العطل المختلفة

رقم العينة	مكان العطل [km]	مقاومة العطل [Ω]	مكان العطل المقدر [km]	الخطأ المطلق [km]
1	25	15	25.001	0.001
2	43.5	5	43.105	0.105
3	23.0	15	23.000	0
4	14	5	14.000	0
5	18	5	18.000	0
6	4	5	4.170	0.170
7	21.5	15	21.500	0
8	27.0	15	27.009	0.009
9	40.0	5	4.011	0.011
10	37.5	5	37.723	0.223

ويبين الشكل (7) الخطأ المطلق في تحديد مكان العطل الموافق لحجم عينة الاختبار عند قيمتي مقاومة العطل.



الشكل (7) الخطأ المطلق في تحديد مكان العطل

كما هو واضح من الرسم البياني أن قيمة الخطأ المطلق في تحديد مسافة العطل ب km لعينات الاختبار تتراوح بين 0 km عند 113 عينة، و 0.402 km ل 10 عينات.

❖ كما حصلنا على قيمة صغيرة لمتوسط القيمة المطلقة للخطأ MAE مقدارها: 0.08 km

الاستنتاجات والتوصيات:

أظهرت الدراسة قدرة خوارزمية الغابة العشوائية على تحديد نوع ومكان العطل بدقة عالية عند شروط النظام المختلفة مما يزيد من موثوقية نظام القدرة الكهربائية، حيث:

- 1- من أجل تصنيف نوع العطل:
 - ✓ بلغت نسبة التقييم الصحيح لكافة أنواع الأعطال 100%.
- 2- من أجل تحديد موقع العطل:
 - ✓ بلغت أعظم قيمة للخطأ المطلق 0.402 km وكان ذلك عند قليل من العينات (10 عينات).
 - ✓ قيمة متوسط القيمة المطلقة للخطأ لجميع عينات الاختبار هي 0.08 km وهي قيمة صغيرة جدا.

نوصي بتطوير هذا البحث من خلال:

❖ تطبيق خوارزميات أخرى من خوارزميات الذكاء الاصطناعي ومقارنة أدائها مع خوارزمية الغابة العشوائية عند شبكات نقل أكثر تعقيدا كخطوط الدارات المزوجة أو تلك الحاوية على تجهيزات FACTS لضمان أعلى موثوقية لنظام نقل الطاقة الكهربائية.

❖ دراسة تأثير إدخال بارامترات إضافية على عملية تحديد نوع وموقع العطل مثل زاوية بدء العطل ومقارنة النتائج مع التي حصلنا عليها.

❖ إدخال لغة البرمجة Python إلى الخطة الدراسية لطلاب هندسة الطاقة الكهربائية في جامعة تشرين نظرا لبساطتها وسهولة وشيوع استخدامها في مجال الذكاء الاصطناعي.

References:

1. Tuba A. *Using Machine Learning Algorithms For Classifying Transmission Line Faults*. Dicle Universitesi Mühendislik Fakültesi Mühendislik Dergisi. 2022;13(2):227–34.
2. Ray P, Mishra DP. *Support vector machine based fault classification and location of a long transmission line*. 2016;
3. Hasan AN, Eboule PP, Twala B. *The use of machine learning techniques to classify power transmission line fault types and locations*. In: 2017 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) & 2017 Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP). IEEE; 2017. pp. 221–6.
4. Farshad M, Sadeh J. *Fault locating in high voltage transmission lines based on harmonic components of one-end voltage using random forests*. *Iranian Journal of Electrical and Electronic Engineering*. 2013;9(3):158–66.
5. Mehdi A, Hussain A, Haider W, Hassan SJ, Saad M, Kim CH. *Classification of Power System Faults Using Random Forests*. . 2020;957–8.
6. Sarker IH. *Machine learning: Algorithms, real-world applications and research directions*. *SN Computer Science*. 2021;2(3):1–21.
7. Upendar J, Gupta CP, Singh GK. *Statistical decision-tree based fault classification scheme for protection of power transmission lines*. *International Journal of Electrical Power & Energy Systems*. 2012;36(1):1–12.
8. Ray P, Panigrahi BK, Senroy N. *Hybrid methodology for fault distance estimation in series compensated transmission line*. *IET Generation, Transmission & Distribution*. 2013;7(5):431–9.
9. Da'ada J. *Improvement and classification of epileptic seizures from brain signaling* [Master]. [Syria]: Tishreen University; 2021
10. Abdualgalil B, Abraham S. *Applications of machine learning algorithms and performance comparison: a review*. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE; 2020. pp. 1–6