



وزارة التعليم العالي والبحث العلمي  
جامعة الموصل  
كلية علوم الحاسوب والرياضيات  
قسم الإحصاء والمعلوماتية

# تصنيف السلاسل الزمنية بالطرائق الإحصائية والشبكات العصبية العميقة

رسالة مقدمة

الى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل  
كجزء من متطلبات نيل شهادة ماجستير علوم في الإحصاء

من قبل

عمر أكرم محمد سعيد بكر

بإشراف

أ.م.د. أسامة بشير شكر الحنون



# تصنيف السلاسل الزمنية بالطرائق الإحصائية والشبكات العصبية العميقة

رسالة تقدم بها

عمر أكرم محمد سعيد بكر

إلى

مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل

وهي جزء من متطلبات نيل شهادة الماجستير

في علوم الإحصاء

بإشراف

أ.م.د. أسامة بشير شكر الحنون

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وَقَدْ

رَبِّ زَيْنِ عَالِمِنَا

صدق الله العظيم

(طه ١١٤)

## إقرار المشرف

أشهد بأن إعداد الرسالة الموسومة بـ "تصنيف السلاسل الزمنية بالطرائق الإحصائية والشبكات العصبية العميقة" قد جرى بإشرافي في جامعة الموصل/كلية علوم الحاسوب والرياضيات وهي جزء من متطلبات نيل شهادة ماجستير علوم في الاحصاء.

التوقيع:

الاسم: أ.م.د. أسامة بشير شكر

التاريخ: / / ٢٠٢٣

## إقرار الخبير اللغوي

أشهد بأن الرسالة الموسومة بـ " تصنيف السلاسل الزمنية بالطرائق الإحصائية والشبكات العصبية العميقة" قد تمت مراجعتها من الناحية اللغوية وتصحيح ما ورد فيها من أخطاء لغوية وتعبيرية وبذلك أصبحت مؤهلة للمناقشة بقدر تعلق الأمر لسلامة الأسلوب وصحة التعبير.

التوقيع:

الاسم: أ.د. منتصر عبد القادر الغضنفر

الكلية: التربية للعلوم الانسانية

التاريخ: / / ٢٠٢٣

## إقرار رئيس لجنة الدراسات العليا

بناءً على توصيتي المشرف وإقرار الخبير اللغوي، أرشح هذه الرسالة للمناقشة.

التوقيع:

الاسم: أ.د. زكريا يحيى الجمال

التاريخ: / / ٢٠٢٣

## إقرار رئيس القسم

بناءً على توصيات المشرف وإقرار الخبير اللغوي وإقرار رئيس لجنة الدراسات العليا، أرشح هذه الرسالة للمناقشة.

التوقيع:

الاسم: أ.م.د. مثنى صبحي سليمان

التاريخ: / / ٢٠٢٣

## قرار لجنة المناقشة

نشهد أننا أعضاء لجنة المناقشة قد اطلعنا على الرسالة الموسومة " تصنيف السلاسل الزمنية بالطرائق الإحصائية والشبكات العصبية العميقة " وناقشنا الطالب " عمر أكرم محمد سعيد بكر " في محتوياتها وفيما له علاقة بها بتاريخ ٤ / ١ / ٢٠٢٤ ونعتقد أنها جديرة بالقبول لنيل شهادة ماجستير علوم في الاحصاء.

التوقيع:  
الاسم: د. هيام عبدالمجيد حياوي  
المرتبة العلمية: أستاذ مساعد  
التاريخ: / / ٢٠٢٤  
عضو اللجنة

التوقيع:  
الاسم: د. هيثم طه محمد علي  
المرتبة العلمية: أستاذ  
التاريخ: / / ٢٠٢٤  
رئيس اللجنة

التوقيع:  
الاسم: د. أسامة بشير شكر  
المرتبة العلمية: أستاذ مساعد  
التاريخ: / / ٢٠٢٤  
عضو اللجنة (المشرف)

التوقيع:  
الاسم: د. نجلاء سعد ابراهيم  
المرتبة العلمية: أستاذ مساعد  
التاريخ: / / ٢٠٢٤  
عضو اللجنة

## قرار مجلس الكلية

اجتمع مجلس كلية علوم الحاسوب والرياضيات بجلسته ( المنعقدة بتاريخ / / ٢٠٢٤ وقرر منحه شهادة ماجستير علوم في الاحصاء.

التوقيع:  
الاسم: أ. د. ضحى بشير عبد الله  
عميد كلية علوم الحاسوب والرياضيات  
التاريخ: / / ٢٠٢٤

التوقيع:  
الاسم: أ. د. صفوان عمر حسون  
مقرر مجلس كلية علوم الحاسوب والرياضيات  
التاريخ: / / ٢٠٢٤

## شكر وتقدير

الحمد لله رب العالمين، له الكمال وحده. والصلاة والسلام على سيدنا محمد رسوله الأمين، وعلى سائر الأنبياء والمرسلين.

أحمد الله تعالى الذي بارك لي في انهاء هذه الرسالة وإتمامها. واتقدم بشكري وامتناني الى استاذي الفاضل الأستاذ المساعد الدكتور اسامة بشير شكر الحنون الذي أسهم في اثناء موضوع رسالتي في جوانبها المختلفة. كما أتقدم بالشكر والتقدير الى السيدة عميد كلية علوم الحاسوب والرياضيات الأستاذ الدكتور ضحى بشير عبدالله على ما قدمته من دعم ومساندة لطلبة الدراسات العليا في الكلية فلها كل الاحترام والتقدير.

كذلك اقدم شكري وتقديري الى السيد رئيس قسم الاحصاء والمعلوماتية المحترم الاستاذ المساعد الدكتور مثنى صبحي سليمان على ما قدمه من توجيه ومساعدة لطلبة الدراسات العليا.

كما اتقدم بالشكر والتقدير الى اساتذتي في قسم الاحصاء والمعلوماتية الذين كان لهم الفضل الكبير في دراستي الاكاديمية.

وأنتقدم بالشكر الى امي الغالية، التي وقفت معي في مراحل الحياة غير ناسٍ دعمها وصبرها ورعايتها إياي، والى سندي في هذه الحياة ابي الغالي الذي اسأل الله ان يشفيه ويعافيه ويمد في عمره. والى زوج الغالية، التي تحملت معي مصاعب الدراسة والحياة فقد كانت خير سند، والى طفلي أحمد وعمران الذين كانا وسبباً في أمني المتجدد، والى اخوتي واخواتي حفظهم الله ورعاهم.

وأخيراً اقدم شكري وتقديري الى اصدقائي الأعزاء، الذين كانوا داعمين ومحفزين إياي في هذه الفترة، وارجو أن اكون قد وفقت في اعداد هذه الرسالة ..... والله ولي التوفيق.

الباحث

## المستخلص

في هذه الرسالة تم تصنيف السلاسل الزمنية باستخدام الانحدار اللوجستي Logistic Regression (LR) وشبكة المدرك العصبية Perceptron Neural Network (PNN) وآلة المتجه الداعم Support Vector Machine (SVM) و K الجار الأقرب K-nearest neighbor (KNN) وطريقة النواة الكاوسية Gaussian Kernel Method (GK). وتمت الاستفادة من متغيرات الانحدار الذاتي Auto Regressive (AR) وذلك من خلال رتبة أنموذج الانحدار الذاتي  $AR(p)$  في تحديد متغيرات الادخال للطرائق المستخدمة. ولغرض المقارنة وزيادة الدقة في التصنيف تم استخدام الشبكة العصبية الالتغافية Convolutional Neural Network (CNN) بوصفها احدى تقانات التعلم العميق الخاصة في التصنيف الصوري، ولغرض انشاء متغير الادخال لطريقة CNN حُوِّلت بيانات السلاسل الزمنية المستخدمة من صيغتها الرقمية وشُكِّلت صوراً ثنائية البعد. وقد استخدمت في هذه الدراسة بيانات السلسلة الزمنية لدودة الربداء الرشيقية Caenorhabditis Elegans (CE) بزوايا حركتها الموجية التي ستمثل حالة الدراسة. إن البيانات المستخدمة في هذه الرسالة هي من السلاسل الزمنية الطويلة وبعدها مشاهدات كبير جدا فضلا عن تعدد الخصائص والصفات التي تمر بها السلسلة الزمنية قد يكون سببا في عدم تجانس البيانات الذي قد يؤدي الى تصنيفات لا ترقى الى ان تكون دقيقة. ان استخدام طريقة CNN ستؤدي الى نتائج دقيقة؛ وذلك لانها ملائمة للتعامل مع بيانات الدراسة لحل مشكلات عدم التجانس من خلال التمثيل الصوري للبيانات الرقمية. ومن خلال مقارنة النتائج بين الطرائق المستخدمة تبين ان طريقة CNN يتفوق على بقية الطرائق الأخرى المستخدمة. ولذلك فمن الممكن استنتاج ان استخدام طريقة CNN التي تعتمد التصنيف الصوري الثنائي تؤدي الى نتائج تصنيفية دقيقة مقارنة بالطرائق الأخرى التي تعتمد التصنيف الرقمي الثنائي.

## قائمة المحتويات

رقم الصفحة	العنوان
i	المستخلص
ii-iv	قائمة المحتويات
v-vi	قائمة الجداول
vii-ix	قائمة الأشكال
<b>5-1</b>	<b>الفصل الأول: مفاهيم اساسية</b>
2-1	1.1 المقدمة
2	2.1 مشكلة البحث
2	3.1 هدف البحث
5-3	4.1 الاستعراض المرجعي
<b>28-6</b>	<b>الفصل الثاني: الجانب النظري</b>
6	1.2 المقدمة
7-6	2.2 أنموذج الانحدار الذاتي
13-7	3.2 الشبكة العصبية الالتفافية
16-13	4.2 شبكة المدرك العصبية
17-16	5.2 الانحدار اللوجستي
18-17	6.2 تقدير الإمكان الأعظم
21-18	7.2 آلة المتجه الداعم
24-21	8.2 K الجار الأقرب
27-25	9.2 طريقة النواة الكاوسية
28-27	10.2 مقاييس الدقة للتصنيف
<b>61-29</b>	<b>الفصل الثالث: الجانب التطبيقي</b>
29	1.3 المقدمة
31-29	2.3 البيانات المستخدمة في الدراسة

35-31	الانحدار الذاتي	3.3
42-35	الشبكة العصبية الالتفافية	4.3
44-42	شبكة المدرك العصبية	5.3
50-45	الانحدار اللوجستي	6.3
54-50	آلة المتجه الداعم	7.3
58-55	K الجار الأقرب	8.3
61-59	طريقة النواة الكاوسية	9.3
<b>64-63</b>	<b>الفصل الرابع: الاستنتاجات والتوصيات</b>	
63	الاستنتاجات	1.4
64	التوصيات	2.4
<b>67-65</b>	<b>المصادر</b>	
67-65	المصادر	
<b>الملاحق</b>		
الملحق A: تحويل درجات زوايا CEM الى اشكال رسومية		
الملحق B: تطبيق CNN		
الملحق C: تطبيق PNN		
الملحق D: تطبيق LR		
الملحق E: تطبيق SVM		
الملحق F: تطبيق KNN		
الملحق G: تطبيق GK		
الملحق H: البيانات		
الملحق I: مصفوفات الارتباك وسلوك ومعلومات عملية تدريب CNN بمعدل تعلم 0.01 للسلاطات الأولى والثانية والثالثة والرابعة		
الملحق J: مصفوفات الارتباك وسلوك ومعلومات عملية تدريب CNN بمعدل تعلم 0.05 للسلاطات الأولى والثانية والثالثة والرابعة		

الملحق K: مصفوفات الارتباك لانموذج PNN للسلاسل الثانية والثالثة والرابعة والخامسة
الملحق L: مصفوفات الارتباك لانموذج LR للسلاسل الثانية والثالثة والرابعة والخامسة
الملحق M: مصفوفات الارتباك لانموذج SVM للسلاسل الثانية والثالثة والرابعة والخامسة
الملحق N: مصفوفات الارتباك لانموذج GK للسلاسل الثانية والثالثة والرابعة والخامسة

## قائمة الجداول

رقم الصفحة	العنوان	رقم الجدول
28	مصفوفة الارتباك	1.2
32	مثال لهيكلية نماذج الانحدار الذاتي	1.3
34	معاملات الانحدار الذاتي لسلاطات العينة الأولى	2.3
34	معاملات الانحدار الذاتي لسلاطات العينة الثانية	3.3
35	نماذج الارتباط الذاتي الأنسب	4.3
37	معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الأولى بمعدل تعلم 0.01	5.3
39	معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.01	6.3
40	نتائج قياس دقة التصنيف بأسلوب CNN بمعدل تعلم 0.01	7.3
40	معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الاولى بمعدل تعلم 0.05	8.3
41	معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الاولى بمعدل تعلم 0.05	9.3
42	نتائج قياس دقة التصنيف بأسلوب CNN بمعدل تعلم 0.05	10.3
44	نتائج قياس دقة التصنيف لأسلوب PNN	11.3
45	جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الأولى	12.3
46	جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الثانية	13.3
46	جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الثالثة	14.3
46	جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الرابعة	15.3
46	جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الخامسة	16.3
47	جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الأولى	17.3

47	جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثانية	18.3
47	جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثالثة	19.3
47	جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الرابعة	20.3
48	جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الخامسة	21.3
48	دقة التصنيف للسلالات الخمس لبيانات التدريب والاختبار باستخدام نموذج LR	22.3
51	نتائج أداء نموذج SVM لسلالات العينة الأولى	23.3
53	نتائج أداء نموذج SVM لسلالات العينة الثانية	24.3
54	الوقت المستغرق للخطوات والمراحل جميعا لعملية تدريب نموذج SVM	25.3
56	نتائج أداء نموذج KNN لسلالات العينة الأولى	26.3
57	نتائج أداء نموذج KNN لسلالات العينة الثانية	27.3
60	دقة التصنيف للسلالات الخمس لبيانات التدريب والاختبار باستخدام أسلوب GK	28.3
62	نتائج الدقة لجميع طرائق التصنيف على بيانات التدريب والاختبار ولجميع السلالات	29.3

## قائمة الأشكال

رقم الصفحة	العنوان	رقم الشكل
8	هيكلية الشبكة العصبية الالتقافية	1.2
10	الدالة اللوجستية	2.2
10	دالة tan	3.2
11	الدالة الخطية	4.2
11	دالة ReLU	5.2
13	خطوات عمل الشبكة العصبية الالتقافية	6.2
14	هيكلية PNN	7.2
15	دالة التحويل hardlim	8.2
15	دالة التحويل hardlims	9.2
16	خطوات عمل شبكة المدرك العصبية	10.2
18	فصل الفضاء بشكل خطي مع اظهار الهامش	11.2
21	خطوات عمل آلة المتجه الداعم	12.2
22	آلية عمل KNN	13.2
24	خطوات عمل KNN	14.2
25	فصل البيانات في فضاء أو مساحة الادخال الأصلية	15.2
26	فصل البيانات في فضاء عالي الابعاد	16.2
27	خطوات عمل GK	17.2
30	سباحة وحركة دودة اليربداء الرشيقية في زوايا مختلفة	1.3
30	نماذج من تحويل درجات زوايا CEM الى اشكال رسومية	2.3
33	الارتباط الذاتي لعينة السلالة الاولى (ACF)	3.3
33	الارتباط الذاتي الجزئي لعينة السلالة الاولى (PACF)	4.3
35	عملية التقاف المرشح على جزء من الصورة	5.3

36	عملية تدريب CNN لبيانات العينة الاولى من السلالة الرابعة	6.3
37	سلوك عملية تدريب CNN لبيانات العينة الأولى من السلالة الأولى بمعدل تعلم 0.01	7.3
38	مصفوفة الارتباك لأنموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى) بمعدل تعلم 0.01	8.3
39	سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.01	9.3
39	مصفوفة الارتباك لأنموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى) بمعدل تعلم 0.01	10.3
41	سلوك عملية تدريب CNN لبيانات العينة الأولى من السلالة الأولى بمعدل تعلم 0.05	11.3
41	سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.05	12.3
43	مصفوفة الارتباك لأنموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)	13.3
44	مصفوفة الارتباك لأنموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)	14.3
49	مصفوفة الارتباك لأنموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)	15.3
49	مصفوفة الارتباك لأنموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)	16.3
52	مصفوفة الارتباك لأنموذج SVM لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)	17.3
58	مصفوفة الارتباك لأنموذج KNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)	18.3

58	مصفوفة الارتباك لأنموذج KNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)	19.3
59	مصفوفة الارتباك لأنموذج GK لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)	20.3
60	مصفوفة الارتباك لأنموذج GK لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)	21.3

# الفصل الأول

## مقدمة عامة

### 1.1 المقدمة Introduction

إن أهمية التصنيف تكمن في عمل تنبؤات لبيانات العينة الجديدة وتحديد الفئات التي لم تظهر في عملية التدريب. وقد تطرقنا في هذه الرسالة الى دراسة تصنيف السلسلة الزمنية باستخدام طرائق التعلم تحت الاشراف supervised learning algorithms، التي تشترط توفر بيانات متغير الهدف Target variable للحصول على تصنيفات دقيقة. وقد استفدنا من الانحدار الذاتي (AR) Auto Regressive باعتماد رتبة نموذج الانحدار الذاتي (AR (p في تحديد هيكلية الأساليب المستخدمة. وعادة ما يستخدم في التصنيف الثنائي الانحدار اللوجستي Logistic Regression (LR) والذي يعد من أهم الطرائق الإحصائية المستخدمة في التصنيف وأكثرها شيوعاً. وهناك طرق أخرى وتستخدم في حالة اهمال توزيع البيانات لانها لا توفر نموذج بل توفر تنبؤات بدون نموذج لا معلمي. ومن ضمن هذه الطرائق طريقة آلة المتجه Support Vector Machine (SVM) الداعم، وطريقة K الجار الأقرب K-nearest neighbor (KNN)، وطريقة النواة الكاوسية Gaussian Kernel Method (GK)، التي تتعامل مع البيانات على نحو يتناسب مع هذا النوع من البيانات؛ من اجل تصنيفها بدقة عالية. وتبقى مشكلة عدم تجانس البيانات، ولغرض تحسين أداء التصنيف والحصول على دقة تصنيف أعلى تُستخدَم الشبكات العصبية العميقة، التي منها طرائق تتعامل مع البيانات على شكل صور، بدلاً من القيم مستمرة الحالة، وتحديداً استخدام الشبكة العصبية الالتفافية Convolutional Neural Network (CNN) وهي طريقة خاصة تستخدم على نطاق واسع في مجال معالجة الصور. تتميز شبكة CNN بقدرتها على تحليل الصور بشكل تدريجي حتى تتمكن في النهاية من تحديد المعالم الأكثر أهمية في الصورة، وتصنيف الصور الجديدة خارج عينة التدريب (Véstias, 2021).

تُعد دراسة الديدان الاسطوانية الشفافة عامة، من الدراسات المهمة في علم الأحياء المجهرية كون خلاياها ذات اشكال بسيطة وبفروع قليلة، فضلاً عن سرعة مراحل نموها (Alkema et al., 2005). وهناك مجموعة من الدراسات السابقة عُنيت بحركة الدودة وسرعتها والمسافات التي تقطعها خلال فترة زمنية محددة. ان حركة

الدودة الاسطوانية تكون بصورة متتابعة خلال زمن معين، أي على شكل سلسلة زمنية، لتمثل حالة الدراسة لمتغير سلسلة زمنية واحد لحركة الدودة، وخلال فترة زمنية تبلغ أجزاء من الثانية لكل حركة تقوم بها الدودة. تم في هذه الدراسة الحصول على بيانات عن الريداء الرشيقية (CE) *Caenorhabditis Elegans*، التي تُعد أحد أنواع الديدان الاسطوانية المهمة في فهم علم الوراثة السلوكية. وتضم البيانات مشاهدات تعبر كل منها عن زاوية خاصة بحركة الريداء الرشيقية<sup>1</sup> (CEM) *Caenorhabditis Elegans Motion*. ولعدم إمكان تحديد CEM تحديداً نقطياً، بل حصرها ضمن فترات، صار من الضروري توصيفها ضمن فئات، وعندها يُركّز الاهتمام على تصنيف المشاهدات الجديدة من خلال أنموذج تصنيفي يُبنى من خلال سلوك السلسلة الزمنية خلال فترة التدريب Training period.

## 2.1 مشكلة البحث Problem of Study

تعد مسألة التصنيف في السلاسل الزمنية مسألة بالغة الأهمية، وستستخدم في هذه الدراسة بيانات CE وهي تعد من السلاسل الزمنية الطويلة أي بعدد مشاهدات كبير جداً وبدليل زمني بواقع مشاهدة لكل نصف ثانية فضلاً عن تعدد الخصائص والصفات والمركبات التي تمر بها البيانات من بداية السلسلة الى نهايتها، الذي قد يكون سبباً رئيساً في عدم التجانس، وبالتالي صعوبة التعامل مع أمثال هذه البيانات علاوة على ان التصنيف النقطي حينها قد يكون ذا نتائج غير دقيقة.

## 3.1 هدف البحث Objective of Study

لتحسين نتائج التصنيف والتقليل من مشكلة عدم تجانس البيانات، فيمكن تمثيل البيانات بالصور واستخدام التصنيف الثنائي Binary Classification للأنماط الصورية، بديلاً عن التنبؤ النقطي؛ لتحسين دقة النتائج التصنيفية، مقارنة بنتائج التنبؤ بوجود عدم التجانس للبيانات المستخدمة، وذلك عبر استخدام طريقة CNN التي هي جديرة بالتعامل مع بيانات كهذه والحصول على افضل التصنيفات فضلاً عن طرائق أخرى تتعامل مع البيانات بصورتها الرقمية الاصلية والحصول على دقة تصنيف عالية باستخدام SVM و KNN وطرائق أخرى لأغراض المقارنة وهي GK و PNN و LR.

<sup>1</sup> أرشيف تصنيف السلاسل الزمنية UEA&UCR المتاح لجمهور الباحثين (UEA&UCR Time Series Classification Repository):  
<http://www.timeseriesclassification.com/description.php?Dataset=EigenWorms>

## 4.1 الاستعراض المرجعي Literature Review

تم في دراسات سابقة استخدام بيانات السلاسل الزمنية، مع الطرائق الإحصائية والشبكات العصبية العميقة؛ حيث استخدم (Labbé *et al.*, 2008) أساليب إحصائية عديدة مثل تحليل المكون الرئيسي Principal Component Analysis (PCA) والمربعات الصغرى الجزئية Partial Least Square (PLS) وطريقة النواة Kernel من اجل العثور على افضل طريقة للتعامل مع بيانات الاشعة تحت الحمراء لتصنيفها لغرض قياس محتوى الفحم والرماد، وقد كانت طريقة النواة باقل خطأ في التصنيف. واستخدم (Janghorbani *et al.*, 2011) سلاسل زمنية بناءً على معدل ضربات القلب وضغط الدم، وذلك لتصنيف المرضى الذين يتعرضون لنوبات انخفاض ضغط الدم، وفي فترة زمنية مدتها ساعة واحدة في المستقبل، والذين لا يتعرضون لهذه النوبات، وباستخدام نموذج الانحدار اللوجستي LR وآلة المتجه الداعم SVM ومن خلال المقارنة حققت SVM افضل دقة للتصنيف بمقدار 88%. وقدم (Fejfar *et al.*, 2012) في بحثه مقارنة بين نتائج ثلاث طرائق وهي طريقة KNN وخوارزمية PNN وخوارزمية تقدير متجهات التعلم Learning Vector Quantization (LVQ) في تصنيف السلاسل الزمنية المشتقة من المقطعات الموسيقية وكانت دقة التصنيف باستخدام طريقتي KNN و PNN افضل من LVQ. وكذلك قام (Wang *et al.*, 2013) باستخدام مهام تصنيف النواة Kernel لمنهجية التعلم عبر الانترنت التي يتم تطبيقها للكشف عن البريد العشوائي عبر الانترنت وكان أداء طريقة النواة في التصنيف فعالا وكفاءة عالية. وقام (Zhan *et al.*, 2015) باستخدام SVM للتصنيف الثنائي بسبب قوة هذا المصنف؛ اذ تم تحسين معاملات الأنموذج المستخدم وتم الحصول على أداء جيد لتحديد احد أجزاء هيكل CE ذي الأهمية وهو رأس الدودة. قام (Liu, H. *et al.*, 2016) باستخدام طريقة مقترحة وهي خوارزمية الانتشار Propagation algorithm ومقارنتها بطريقتي KNN و SVM لفحص مجموعة من الادوية التي تعمل على إطالة عمر دودة الربداء الرشيق؛ لغرض تصنيفها الى فعال وغير فعال، وتبين ان نتائج الطريقتين الكلاسيكيتين KNN و SVM كانت جيدة، ولكن الطريقة المقترحة كانت ذات نتائج افضل. قام الباحث (Javer *et al.*, 2018) باستخدام أسلوب CNN؛ لتمييز سلاسل CE المتنوعة وراثيا وتصنيفها، عن طريق تدريب الأنموذج على بيانات سلسلة زمنية لاوضاع الدودة، باستخدام عينات من صور حركة الدودة بوصفها متغير ادخال، وقد تم الحصول على أنموذج قادر على تصنيف

السلاطات. وقام (Newell Stamper *et al.*, 2018) بتحديد ما اذا كان العمر والنمط الجيني لدودة اليربوع الرشيقية يؤثران في حركتها في السلاطات جميعاً، وذلك باستخدام أنموذج الانحدار اللوجستي الثنائي؛ اذ تم تقييم فروق الحركة بين السلاطات من النوع البري والسلاطات الطافرة من خلال الانحدار اللوجستي. وكذلك تم الحصول على نتائج دقيقة في تصنيف السلاسل الزمنية باستخدام أحدث أساليب التعلم العميق مثل الشبكة العصبية الالتغافية CNN؛ اذ قام (Fawaz *et al.*, 2019) بتصنيف السلاسل الزمنية المتاحة للجمهور في أرشيف UEA&UCR باستخدام الشبكات العصبية العميقة. وقام (Fawaz, 2020) بإجراء مقارنة أساليب الشبكات العصبية العميقة لتصنيف بيانات السلاسل الزمنية المتاحة للجمهور في أرشيف UEA&UCR بأساليب حديثة أخرى قائمة على التعلم غير العميق مثل خوارزمية HIVE-COTE، التي تعد من الخوارزميات الخاصة بتصنيف السلاسل الزمنية، وقد تبين ان نتائج أساليب الشبكات العميقة وبيانات السلاسل الزمنية الطويلة متساوية في الدقة مع خوارزمية HIVE-COTE إلا ان خوارزمية HIVE-COTE لا يمكن تطبيقها على عدد من مجموعة البيانات الواقعية، نظرا لاستغراقها أياماً عديدة للتعلم من مجموعة بيانات صغيرة، بعكس الشبكات العصبية العميقة. كما تناولت مجموعة من الدراسات السابقة استخدام أساليب التصنيف والتنبؤ في مجال السلاسل الزمنية والأحياء المجهرية فيما يخص سلوك الديدان الاسطوانية. واستخدم (Uhlemann *et al.*, 2020) أسلوب CNN؛ لتصنيف الصور المجهرية لتحديد نوع معين من الديدان الخيطية، وكان الأنموذج المدرب يعمل على نحو جيد في التصنيف. وقام (Jiang, 2020) بإجراء مقارنة شاملة بين طريقة KNN وطريقة CNN وشبكة المدرك العصبية PNN لتصنيف 128 مجموعة بيانات من السلاسل الزمنية المتاحة للجمهور في أرشيف UEA&UCR، وكانت نتائج التصنيف متقاربة جدا بين الطرائق المستخدمة جميعاً. وقام (García Garvı *et al.*, 2021) كذلك باستخدام التعلم العميق من خلال اسلوبي CNN و Recurrent Neural Network (RNN) لحساب متوسط العمر المتوقع لبيانات CEM، عن طريق تصنيفها على انها على قيد الحياة او انها ميتة، عبر ملاحظة صور من حركة الدودة. وحقق الأسلوبان معدلات خطأ صغيرة مما يدل على جودتها. وقام (Maertens *et al.*, 2021) بتوقع الحركة الامامية لدودة اليربوع الرشيقية باستخدام الانحدار اللوجستي، من خلال التنبؤ بالخلايا العصبية المشاركة في سلوك الحركة وكان انموذج الانحدار اللوجستي مناسباً تماماً وحقق نتائج جيدة. وقام (Ikirigo, 2022) بتصنيف التركيبات الحينية لدودة اليربوع الرشيقية باستخدام الطرائق KNN و SVM والغابة العشوائية Random Forest (RF) وبالمقارنة بين كل

من الطرائق المذكورة كان KNN افضل أداء في التصنيف. وهذه البيانات عبارة عن مشاهدات دقيقة تم التقاطها على فترات زمنية صغيرة وقد طُبِقَ أنموذج الانحدار اللوجستي الإضافي المعمم Generalized Additive Logistic Regression Model للتنبؤ بالتأثيرات الزمانية في الاهتمام البصري والعمليات الأساسية التي تقوم بتوجيه العين. وقام (Chen *et al.*, 2023) بحساب طول العمر لدودة الريداء الرشيق، من خلال التمييز بين جزيئات الدهون، التي تمثل معدلاً مهماً للشيخوخة، باستخدام أسلوب شبكة المدرك العصبية عن طريق التصنيف الثنائي.

تضمنت الرسالة اربعة فصول في الفصل الأول مقدمة عن تصنيف السلسلة الزمنية ونوع البيانات والأساليب التي طُبِقَت في تصنيف البيانات المستخدمة في هذه الرسالة، وتضمن الفصل كذلك مشكلة الرسالة والهدف منها والاستعراض المرجعي لعدد من الدراسات السابقة. أما الفصل الثاني فقد ضم الجانب النظري الذي تضمن الأساليب المستخدمة في تصنيف السلاسل الزمنية. فيما إشتهل الفصل الثالث على الجانب التطبيقي الذي تضمن تطبيق أساليب التصنيف. وأخيراً جمع الفصل الرابع الاستنتاجات وما تم التوصل اليه من خلال تطبيق أساليب التصنيف.

## الفصل الثاني

### الجانب النظري

#### 1.2 المقدمة

يتناول هذا الفصل الجانب النظري من موضوع الرسالة، ويتضمن الأساليب المستخدمة في تصنيف السلاسل الزمنية، وهي أسلوب الانحدار اللوجستي LR وهو من أهم الأساليب الإحصائية في التصنيف الثنائي، ويُعد أساس عمل الأساليب المستخدمة في التصنيف. وأسلوب الشبكة العصبية الالتفافية CNN لتحسين نتائج التصنيف من خلال تمثيل البيانات بالصور ثم تصنيفها، فضلاً عن أساليب أخرى مثل أسلوب شبكة المدرك العصبية PNN، وآلة المتجه الداعم SVM، وأسلوب الجار الأقرب KNN، وطريقة النواة الكاوسية GK، بوصفها جميعاً أساليب تصنيف تتعامل مع السلاسل الزمنية بصورتها الرقمية، لأغراض المقارنة.

#### 2.2 نموذج الانحدار الذاتي (AR) Auto Regressive Model

إن السلاسل الزمنية Time Series هي مجموعة من المشاهدات التي تتولد في فترات زمنية متتالية وتتميز بعدم الاستقلالية؛ إذ أن المشاهدات فيها ترتبط بسابقتها زمنياً ويمكن من خلالها تصنيف السلاسل الزمنية المستقبلية بالاعتماد على مشاهدات لسلسلة زمنية وقعت في الماضي (Brockwell & Davis, 2006; Liu, L.-M., 2009).

إن السلسلة الزمنية الحالية يمكن أن يعبر عنها باستخدام دالة الانحدار الذاتي لقيم السلاسل الزمنية السابقة، وتمكن كتابة نموذج الانحدار الذاتي من الرتبة  $p$  كما في المعادلة (2.1) في أدناه.

$$x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t \quad (2.1)$$

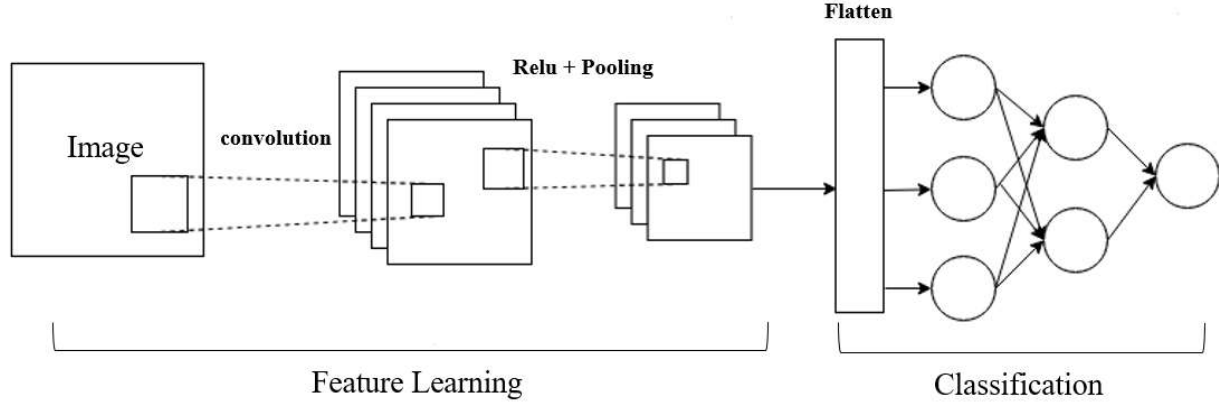
إذ أن  $\delta$  تمثل الحد المطلق للسلسلة الزمنية و  $\phi_p$  هي معاملات الانحدار الذاتي و  $p$  تمثل رتبة نموذج الانحدار الذاتي وان  $e_t$  تمثل الخطأ أو التغير العشوائي وهي عبارة عن عملية تشويش أبيض بمتوسط صفر وتباين ثابت  $\sigma_e^2$  أي أن  $e_t \sim i.i.d.N(0, \sigma_e^2)$  و  $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ .

ولغرض تحليل بيانات السلسلة الزمنية تُستخدم منهجية بوكس جنكنز Box-Jenkins بخطواتها الأربعة: التعرف Identification، وتقدير المعلمات parameter estimation، واجراء الفحوص التشخيصية Diagnostic checking، والتنبؤ Forecasting، أن الخطوة الأولى هي التعرف إلى أنموذج السلسلة الزمنية الملائمة ومنها الانحدار الذاتي AR، ويمكن من خلال الرسم تحديد الانموذج الملائم للسلسلة الزمنية باستخدام دالة الارتباط الذاتي Autocorrelation Function (ACF) اذا كانت الدالة متناقصة تدريجياً بشكل اسي، وتحديد رتبة الانموذج من خلال دالة الارتباط الذاتي الجزئي Partial Autocorrelation Function (PACF) اذا كانت الدالة تنقطع بعد إزاحة p فان الانموذج الملائم هو انموذج AR(p).

ان دالة الارتباط الذاتي ACF تستخدم لقياس الارتباط بين عناصر السلسلة الزمنية وهي مفيدة جدا في فهم هيكل السلسلة الزمنية وتحديد رتب نماذج الانحدار الذاتي للسلسلة الزمنية، اذ ان في هذه الدراسة تم اعتماد رتبة الانحدار الذاتي أنموذجاً احصائياً تقليدياً مختصاً بالسلاسل الزمنية الأحادية للاستفادة من الرتبة حسب في تحديد هيكلية المدخلات لطرائق التعلم تحت الاشراف والتعلم العميق.

## 3.2 الشبكة العصبية الالتفافية (CNN) Convolutional Neural Network

هي إحدى الأدوات الأساسية للتعلم العميق، وتتضوي تحت مظلة الشبكات العصبية العميقة Deep neural network (DNN)، التي تضم نوعاً آخر من DNN وهي الشبكات العصبية المعادة RNN. وتسمى CNN احياناً بالشبكة العصبية الالتفافية العميقة Deep-CNN، وذلك عندما تكون شبكة متعددة الطبقات تحتوي على اكثر من طبقتين. وبما ان CNN في الوقت الراهن يحتوي بهيكليته على طبقتين فاكثر، فان المصطلحين CNN و Deep-CNN أصبح لهما المفهوم العلمي نفسه (Theobald, 2017). إن هيكلية CNN تتألف من جزأين أساسيين وهما: أولاً طبقة تعلم الميزات (متغيرات الإنحدار الذاتي) التي تتم فيها عمليات الالتفاف والتجميع للتعرف إلى خصائص الصورة، مثل الحواف وتدرج اللون. وثانياً الطبقة المتصلة بالكامل والتي تستقبل مخرجات طبقة تعلم الميزات بوصفها مدخلات لتتم فيها عملية التصنيف كما يتضح في الشكل 2 . 1 أدناه.



الشكل 2 . 1: هيكلية الشبكة العصبية الالتفافية

بداية تُختار هيكلية CNN المناسبة عن طريق تحديد عدد الطبقات Layers. وتتكون الطبقات من طبقة الإدخال، والطبقات الخفية، وطبقة الإخراج. أما طبقة الإدخال فتتحدد بعدد المدخلات Inputs وتقابل مفهوم المتغيرات التفسيرية Explanatory Variables في نموذج الانحدار الخطي المتعدد Multiple Linear Regression وقد تسمى بمتغيرات الانحدار الذاتي للسلسلة الزمنية. أما حجم الطبقة الخفية فيتحدد بعدد العصبونات التي تحتويها الطبقة، وغالبا ما يكون أكثر من عدد المدخلات. أما طبقة الإخراج فعدد المخرجات فيها مطابق لعدد خيارات التصنيف المطلوب (Gamboa, 2017).

يُحدّد في الطبقة الأولى الخفية عدد المرشحات number Filters ويرمز لها بـ (m) ويُرمز لحجم كل مرشح filter Size بـ (f) ويحتوي كل مرشح على القيم الابتدائية للأوزان ويرمز لها بـ (w). وتقوم هذه المرشحات بالإنفاف حول نقاط الصورة جميعاً للتعرف إلى الحواف والخطوط الرأسية أو الأفقية أو زوايا أخرى للصورة والتفاصيل التي تشكل الصورة. أما الخطوة stride التي يتحرك فيها المرشح (يعني عدد الوحدات pixels التي يتم تجاوزها) فما لم تُحدّد فستكون افتراضية بمقدار خطوة واحدة للمرشح باتجاه يسار الصورة. وعند التقاف المرشح حول الصورة سوف يتقلص حجم الصورة ويتم فقدان جزء من البيانات التي تشكل الصورة، ولحل هذه المشكلة يُضاف الحشو Padding بصفوف وأعمدة على طول وعرض الصورة ومحيطها. إن حجم المصفوفة الناتجة من عملية التقاف كل مرشح حول الصورة يتحقق عن طريق عملية رياضية مبسطة وكما موضح في المعادلة (2.2) في أدناه (Kattenborn et al., 2021).

$$H_{i+1} \times W_{i+1} = \left[ \left( \frac{H + 2p - f}{s} \right) + 1 \right] \times \left[ \left( \frac{W + 2p - f}{s} \right) + 1 \right] \quad (2.2)$$

إذ أن  $(H_{i+1} \times W_{i+1})$ : أبعاد الصورة بعد عملية الالتفاف.

$(H)$ : طول الصورة.

$(W)$ : عرض الصورة.

$(P)$ : الحشو.

$(f)$ : أحد أبعاد المرشح.

$(s)$ : الخطوة.

بعد ذلك تُجمع قيمة التحيز (b) biased value لكل مرشح مع عناصر المصفوفة التي تقابلها كما في المعادلة (2.3) في أدناه، التي يمكن تسمية مخرجاتها بمخرجات الدالة الجمعية.

$$\text{SUM} = \sum_{i=1}^{f_2} \sum_{j=1}^{f_1} w_{ij} x + b_k \quad (2.3)$$

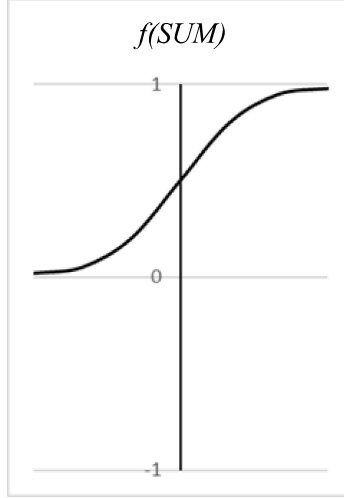
إذ أن  $i, j$  تمثلان أبعاد أو عدد الصفوف، وعدد الأعمدة في كل صورة على التوالي. وأن  $f_1, f_2$  تمثلان عدد الصفوف وعدد الأعمدة في كل مرشح على التوالي. وأن  $k=1,2,\dots,m$  ترمز الى تسلسل ناتج كل مرشح و  $x$  تمثل متغير الإدخال. وتُطبق إحدى دوال التحويل على المعادلة (2.3). حيث تعتمد دقة التصنيف للشبكة العصبية على عدد الطبقات المستخدمة ونوع دالة التحويل المستخدمة وان دوال التحويل الأكثر استخداما هي الدوال من النوع غير الخطي للتعامل مع الانماط غير الخطية للبيانات ومن أكثر دوال التحويل استخداما في الشبكة العصبية هي الآتي:

1. الدالة اللوجستية (Logistic Sigmoid function): هي دالة غير خطية تعمل على تسريع تدريب

الشبكة ومعادلتها كالآتي.

$$f(\text{SUM}) = \frac{1}{1 + e^{-(\text{SUM})}} \quad (2.4)$$

إذ أن SUM تمثل مدخلات دالة التحويل، التي تمثل مخرجات الدالة الجمعية. وتولد مخرجاتها ضمن الحدود (0,1) وكما موضح في الشكل 2. 2 في أدناه.

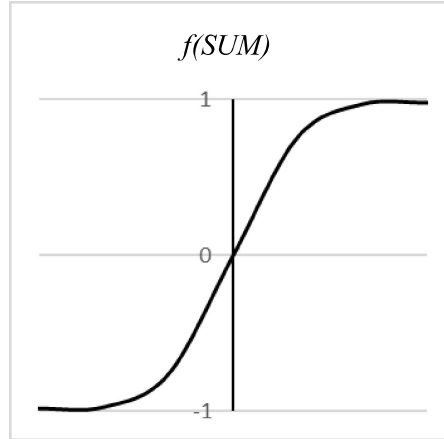


الشكل 2. 2 : الدالة اللوجستية

2. دالة (Tan sigmoid function) tan: دالة غير خطية تعمل على تنعيم ناتج الخلايا العصبية ومعادلتها كالآتي.

$$f(SUM) = \frac{2}{1 + e^{-2(SUM)}} - 1 \quad (2.5)$$

إذ أن SUM تمثل مدخلات دالة التحويل، التي تمثل مخرجات الدالة الجمعية. وتولد مخرجاتها ضمن الحدود (-1,1) وكما موضح في الشكل 2. 3 في أدناه.

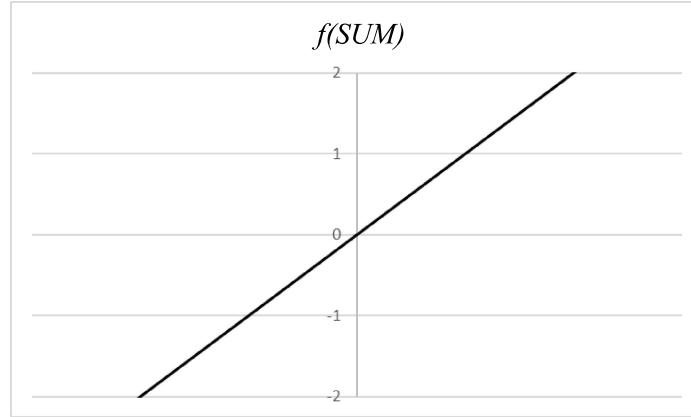


الشكل 3. 2 : دالة tan

3. الدالة الخطية (Pure Line function): دالة خطية حيث يكون الإخراج مشابهاً للادخال ومعادلتها كالآتي.

$$f(SUM) = SUM \quad (2.6)$$

إذ أن SUM تمثل مدخلات دالة التحويل، التي تمثل مخرجات الدالة الجمعية. والمخرجات مشابهة للمدخلات وكما موضح في الشكل 2. 4 في أدناه.

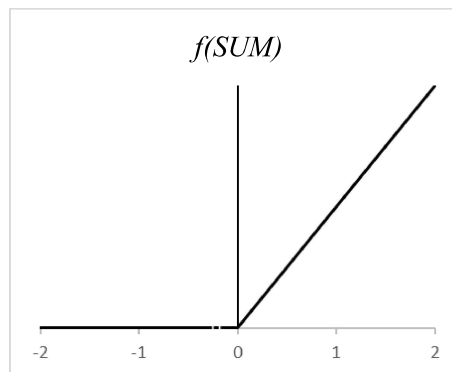


الشكل 2. 4 : الدالة الخطية

4. دالة الوحدة الخطية المصححة Rectified Linear Unit function (ReLU): دالة خطية على القسم الموجب تستخدم عادة في الشبكات ذات الطبقات المتعددة أو العميقة ومعادلتها كالآتي.

$$f(SUM) = \begin{cases} 0 & SUM \leq 0 \\ x & SUM > 0 \end{cases} \quad (2.7)$$

إذ أن SUM تمثل مدخلات دالة التحويل وان  $x=1,2,3,\dots$ ، التي تمثل مخرجات الدالة الجمعية. وتولد مخرجاتها بما هو أكبر أو يساوي (0) وكما موضح في الشكل 2. 5 في أدناه.



الشكل 2. 5 : دالة ReLU

## 5. دالة Softmax (Softmax function):

$$f(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^k \exp(z_j)} \quad (2.8)$$

اذ ان  $z_i$  تمثل مخرجات الطبقة المتصلة بالكامل، التي تُستخدم في التصنيف الثنائي احتمالي النجاح أو الفشل لكل مخرج، أما  $z_j$  فتتمثل مجموع الاحتمالات، تُطبق هذه الدالة على مخرجات الطبقة المتصلة بالكامل، ويكون الناتج احتمالات بعدد فئات متغير الاستجابة. وهذه الاحتمالات هي ارقام موجبة، كل رقم فيها يمثل احتمالية انتمائه الى فئة من الفئات ويساوي مجموعها واحد (Sharma et al., 2017).

بعد عملية ادخال دالة التحويل إلى مخرجات الدالة الجمعية تُطبّق عملية التجميع pooling عن طريق تقسيم المصفوفة على مناطق تجميع مربعة أو مستطيلة؛ إذ أن المصفوفة تقسم على عدد من المصفوفات الصغيرة ويتم ذلك بتحديد حجمها وعدد خطواتها. وتؤخذ أعلى قيمة في حال تطبيق max pooling أو متوسط القيم في حال تطبيق average pooling لكل منطقة. ثم تُختزل كل مصفوفة صغيرة الى قيمة واحدة حسب مما ينتج عنه تقليل حجم المصفوفة لتصبح المصفوفة في نهاية الطبقة الأولى الخفية بابعاد أقل. ثم تُحول الى متجه هو عبارة عن عمود واحد فيه الخصائص جميعاً وذلك عن طريق عملية تسمى التسوية Flatten. إن المتجه يمثل مدخلات الطبقة المتصلة بالكامل (Fully Connected Layer Neapolitan & Jiang, 2018).

تُحدد مخرجات الطبقة المتصلة بالكامل بعدد فئات التصنيف بالاعتماد على عدد فئات المتغير المعتمد، وتُحدد الاوزان كذلك عشوائياً في كل عصبون. وغالباً ما تتبع هذه الطبقة في حالة التصنيف بدالة التحويل Softmax. وفي طبقة التصنيف Classification Layer تُحسب قيمة الخطأ؛ اذ ان الهدف من ادراج طبقة التصنيف هو تقليل الخطأ عن طريق تحديث قيم الوزن (weight value (w)، والتحيز (biased value (b). وهي عملية تكرارية إلى حين الوصول الى القيم المثلى؛ إذ أن كل عملية تكرار يتم فيها الحصول على قيم معينة من الاوزان والتحيز واختبار نسبة الخطأ. وتسمى هذه العملية بالانحدار المتدرج stepwise regression. وهذا الاجراء يسمى اصطلاحاً Gradient Descent (GD) مما يسمح بتحديد الاتجاه نحو تقليل الأخطاء وتكون معادلة GD كما يأتي.

$$p_{i+1} = p_i - \alpha \frac{1}{n} \sum_{j=1}^n (y_{ij} - \hat{y}_{ij}) \cdot x_{ij} \quad (2.9)$$

إذ أن  $(\alpha)$  تمثل معدل التعلم Learn Rate وهي قيمة صغيرة جدا محصورة بين (0 و 1)، وإن  $p_i$  تمثل  $w_i$  أو  $b_i$  في التكرار الحالي، وإن  $p_{i+1}$  تمثل  $w_{i+1}$  أو  $b_{i+1}$  في التكرار اللاحق الجديد،  $y_{ij}$  تمثل قيمة المشاهدة  $z$  لمتغير الهدف في التكرار الحالي  $i$ ، و  $\hat{y}_{ij}$  تمثل قيمة المشاهدة  $z$  لمتغير الإخراج في التكرار الحالي  $i$  لطبقة التصنيف، و  $x_{ij}$  تمثل المشاهدة  $z$  لمتغير الإدخال في التكرار الحالي  $i$  (Zhao et al., 2017).

وبعد الحصول على القيم المثلى للأوزان والتحيز من خلال طبقة التصنيف، فإن الاجراء النهائي الذي تقوم به CNN هو تصنيف السلسلة الزمنية بمقارنة المتغير  $\hat{y}_{final}$  بمتغير القيم الأصلية، وهو متغير الهدف. فيتم حساب دقة نموذج التصنيف إلى القيم الحقيقية للسلسلة الزمنية باستخدام مقياس دقة التصنيف. (O'Shea

& Nash, 2015) والشكل 6. 2 في أدناه يوضح خطوات عمل CNN.

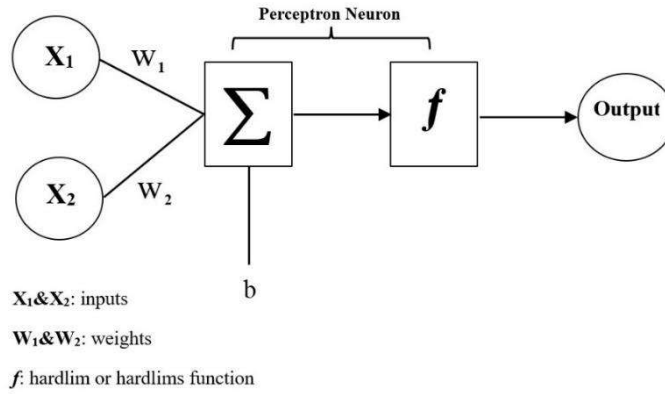


الشكل 6. 2 : خطوات عمل الشبكة العصبية الالتفافية

#### 4.1 شبكة المدرك العصبية (PNN) Perceptron Neural Network

إن عملية التعلم باستخدام الشبكات العصبية هي عبارة عن عملية تحاكي آلية عمل الخلايا العصبية في جسم الانسان. والهدف من تدريب أنواع مختلفة من الشبكات العصبية هو تحقيق أغراض التصنيف والانحدار مثل الشبكة العصبية الالتفافية CNN، والشبكة العصبية المتكررة RNN، وشبكة المدرك العصبية

PNN. ويمكن استخدام هذه الشبكات في التصنيف أو التنبؤ لأنواع البيانات جميعاً سواء كانت هذه البيانات صوراً أو أرقاماً أو نصوصاً أو أفلام فيديو. إن شبكة المدرك العصبية كما في الشكل 7. 2 في أدناه هي شبكة ذات طبقة واحدة وتعد أساس عمل الشبكات العصبية جميعاً؛ إذ إن فهم عمليات هذه الشبكة يعد أساساً جيداً لكي يتم فهم الشبكات الأخرى، التي تكون أكثر تعقيداً. ولغرض انشاء هذه الشبكة وتدريبها لأغراض التصنيف يجب أولاً تجهيز البيانات من خلال تقسيمها على مجموعات بيانات التدريب والاختبار، ثم تدريب الشبكة على بيانات التدريب، وأخيراً تُحسب دقة التصنيف نسبة إلى بيانات الاختبار (Gabrié et al., 2023).



### الشكل 7. 2 : هيكلية PNN من اقتراح الباحث بالاعتماد على (Demuth et al., 1992)

تتكون شبكة المدرك العصبية من طبقة الادخال التي يتم فيها ادخال المتغيرات  $x_i$  وعصبون واحد يسمى المدرك Perceptron، الذي يستقبل المدخلات باعطائها وزناً ابتدائياً مناسباً  $w_i$  ليتم ضرب المدخلات جميعاً في الاوزان الافتراضية، وجمعها مع بعض، واطافة قيمة التحيز  $bias$  التي يمكن تمثيلها بالدالة الجمعية في المعادلة (2.10) في أدناه.

$$SUM = \sum_{i=1}^n w_i x_i + b \quad (2.10)$$

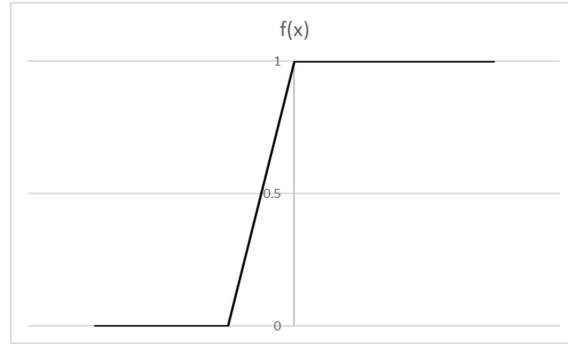
إذ أن  $x_i$  تمثل متغير الادخال، وان  $n$  تمثل عدد المتغيرات (Sagheer & Zidan, 2013).

ثم تُطبق إحدى دوال التحويل الخاصة بالشبكات أحادية الطبقة على الدالة الجمعية في المعادلة (2.10).

إن دوال التحويل الخاصة بأسلوب PNN هي كالآتي:

1. دالة التحويل hardlim (hard-limit transfer function) وتُستخدم في الشبكات ذوات الطبقة الأحادية، ومنها شبكة المدرك العصبية. وتعمل على تحويل الدالة الجمعية الى مخرجات ثنائية 0 أو 1 كما في المعادلة (2.11) والشكل 2.8.

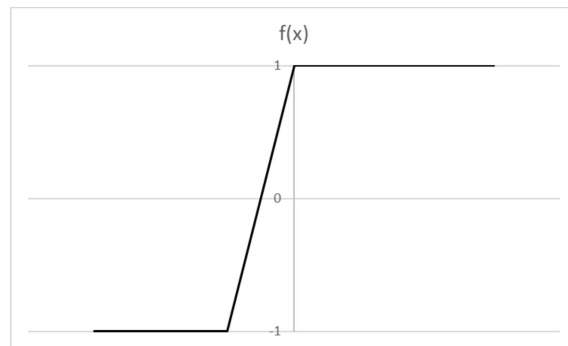
$$f(SUM) = \begin{cases} 0 & SUM < 0 \\ 1 & SUM \geq 0 \end{cases} \quad (2.11)$$



الشكل 2.8 : دالة التحويل hardlim

2. دالة التحويل hardlims (symmetric hard-limit transfer function) وتُستخدم في شبكة المدرك العصبية التي تعمل على تحويل الدالة الجمعية الى مخرجات -1 أو 1 كما في المعادلة (2.12) و الشكل 2.9 في أدناه.

$$f(SUM) = \begin{cases} -1 & SUM < 0 \\ 1 & SUM \geq 0 \end{cases} \quad (2.12)$$



الشكل 2.9 : دالة التحويل hardlims

ان دالة التحويل تقوم في الأساس بتحويل المعادلة (2.10) الى مخرجات ثنائية تسمى طبقة الإخراج، ويتم في هذه الطبقة حساب قيمة الخطأ، ثم تقليل الخطأ عن طريق تحديث قيم الوزن (w) weight value، والتحيز

(b) biased value، وهي عملية تكرارية إلى حين الوصول إلى القيم المثلى. وبعد الحصول على القيم المثلى للأوزان والتحيز من خلال طبقة التصنيف فإن الاجراء النهائي الذي تقوم به PNN هو تصنيف السلسلة الزمنية بمقارنة المتغير  $\hat{y}_{final}$  مع متغير القيم الأصلية وهو متغير الهدف. ويتم حساب دقة نموذج التصنيف نسبة إلى القيم الحقيقية للسلسلة الزمنية باستخدام مقياس تقييم دقة التصنيف (Glorot & Bengio, 2010). ان الشكل 2. 10 في أدناه يوضح خطوات عمل PNN.



الشكل 2. 10 : خطوات عمل شبكة المدرك العصبية

## 5.2 الانحدار اللوجستي Logistic Regression (LR)

يُعد الانحدار اللوجستي من أهم الطرائق المستخدمة في التصنيف، ويستخدم عندما يكون المتغير المعتمد فنوياً ثنائياً ويسمى بالأنموذج اللوجستي أو اللوجيت (Logit)، وهو أنموذج مفيد في تصنيف البيانات التي تتكون من متغير ثنائي الاستجابة كما في المعادلة (2.13) في أدناه.

$$Ln \left[ \frac{P(y=1 | x_1, x_2, \dots, x_p)}{1 - P(y=1 | x_1, x_2, \dots, x_p)} \right] = Ln \left[ \frac{\pi}{1 - \pi} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \beta_0 + \sum_{j=1}^p \beta_j x_j \quad (2.13)$$

إذ أن  $\beta_0$  تمثل معلمة الحد الثابت،  $\beta_j$  تمثل معلمات الانحدار،  $x_j$  تمثل المتغيرات المستقلة،  $\frac{\pi}{1 - \pi}$

يمثل نسبة الترجيح (Odds) وهو الاحتمال الشرطي للنجاح مقسوم على الاحتمال الشرطي للفشل. ولغرض

تبسيط المعادلة (2.13) نحصل على الآتي.

$$\frac{P(y=1|x_1, x_2, \dots, x_p)}{1 - P(y=1|x_1, x_2, \dots, x_p)} = \frac{\pi}{1 - \pi} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p} = e^{\beta_0 + \sum_{j=1}^p \beta_j x_j} \quad (2.14)$$

وعليه يكون الاحتمال الشرطي للنجاح كما في المعادلة (2.15) في أدناه.

$$\pi = P(y=1|x_1, x_2, \dots, x_p) = \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}} = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} \quad (2.15)$$

حيث ان z تمثل  $\beta_0 + \sum_{j=1}^p \beta_j x_j$  وإن احتمالية وقوع الحدث يكون 1 إذا كانت قيمة الاحتمال للمعادلة (2.15) أكبر أو تساوي 0.5. أما الاحتمال الشرطي للفشل فيكون كما في المعادلة (2.16) في أدناه.

$$1 - P(y=1|x_1, x_2, \dots, x_p) = 1 - \frac{e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}}{1 + e^{\beta_0 + \sum_{j=1}^p \beta_j x_j}} = 1 - \frac{e^z}{1 + e^z} = \frac{1}{1 + e^z} \quad (2.16)$$

حيث ان z تمثل  $\beta_0 + \sum_{j=1}^p \beta_j x_j$  وإن احتمالية عدم وقوع الحدث يكون 0 إذا كانت قيمة الاحتمال للمعادلة (2.16) أصغر من 0.5 (Hosmer Jr et al., 2013).

## 6.2 تقدير الإمكان الأعظم (MLE) Maximum Likelihood Estimation

تُقدَّر معاملات الانحدار اللوجستي باستخدام أحد طرائق التقدير المهمة، وهي طريقة تقدير الإمكان الأعظم؛ وذلك للحصول على الاحتمال الشرطي للملاحظات جميعاً عن طريق المعادلة (2.17) في أدناه.

$$P(y_i | x_i) = \prod_{i=1}^n (\pi(x_i))^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (2.17)$$

وتكون دالة الإمكان الأعظم لأنموذج الانحدار اللوجستي كما في المعادلة (2.18) في أدناه.

$$L(\beta_0, \beta_j) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (2.18)$$

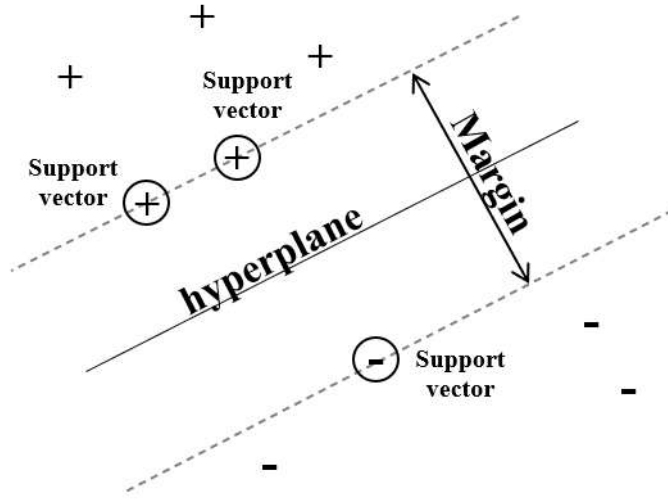
ولغرض تبسيط معادلة دالة الإمكان الأعظم يُؤخذ اللوغارتم للطرفين كما في المعادلة (2.19) في أدناه.

$$\log(\beta_0, \beta_j) = \sum_{i=1}^n (y_i \ln \pi(x_i) + (1 - y_i) \ln(1 - \pi(x_i))) \quad (2.19)$$

ولغرض ايجاد مقدر الإمكان الأعظم تُشتق معادلة الإمكان الأعظم نسبة الى معلمات الانحدار ومساواتها بالصفر (Gould *et al.*, 2006).

## 7.2 آلة المتجه الداعم (SVM) Support Vector Machine

إن فكرة آلة المتجه الداعم SVM تبحث عن المستوى الأمثل الذي يفصل البيانات الى فئتين، بمعنى أن نقاط المشاهدات على جانبي الخط تمثل فئتين مختلفتين يفصل بينهما خط فاصل يسمى المستوى الفائق (hyperplane). وغالبا ما تكون إحدى الفئتين إيجابية والأخرى سلبية، إذ يكون للمستوى الأمثل هامشان متوازيان عند أقرب النقاط، التي تسمى المتجهات الداعمة (support vectors) وكما موضح في الشكل 2.11. في أدناه. ويعد SVM من طرائق تعلم الآلة تحت الاشراف، ومن تطبيقاته المتنوعة التعرف إلى الوجوه، وتصنيف البيانات، والتعرف إلى خط الأيدي وغيرها كثير (Du & Swamy, 2013).



الشكل 2.11 : فصل الفضاء بشكل خطي مع اظهار الهامش، من اقتراح الباحث بالاعتماد على (Jade *et al.*, 2006)

يمكن تمثيل المستوى الأمثل كما يأتي:

$$w^T x + b = 0 \quad (2.20)$$

ويتمثل المتجه الداعم للفتة الإيجابية والسلبية كما في المعادلتين وعلى التوالي :

$$w^T x_2 + b = +1 \quad , y_i = +1 \quad (2.21)$$

$$w^T x_1 + b = -1 \quad , y_i = -1 \quad (2.22)$$

إذ أن  $w = \{w_1, w_2, \dots, w_d\}$  متجه يتحكم في اتجاه المستوى الأمثل ويكون متعامداً مع المستوى الأمثل و  $d = \{(x_1, y_1), (x_2, y_2), \dots, (x_p, y_p)\}$  أي أن  $d$  تمثل عدد تخلفات السلسلة الزمنية المقابلة لمتغيرات الانحدار الذاتي، و  $b$  يمثل التحيز الذي يتحكم في المسافة بين المستوى الأمثل ونقطة الأصل، أي أنه يمثل ثابت التقاطع intercept constant مع محور  $y$ ، أما  $x$  فتمثل التخلفات الزمنية، أي متغيرات الانحدار الذاتي المؤثرة بشكل معنوي (Schölkopf et al., 2003).

إن المسافة التي تمثل الإزاحة بين أقرب متجه داعم بشكل عامودي عن المستوى الأمثل hyperplane تسمى الهامش Margin. والهدف من هذه الخوارزمية هو الحصول على أكبر هامش لتحقيق أمثل مستوى بأقل أخطاء تصنيفية وبحسب المعادلة (2.23) في أدناه.

$$\max_{w,b} = \frac{2}{\|w\|} \quad (2.23)$$

$$y_i(w^T x_i + b) \geq 1 \quad \forall \quad x_i, i = 1, 2, \dots, n$$

إذ أن  $\|w\|$  تمثل حجم المتجه  $w$  ويسمى كذلك (Norm). وإن الغرض من المعادلة (2.23) هو الحصول على أكبر هامش وإيجاد المعاملات  $w$  و  $b$  ويمكن تحسين الهامش عن طريق إعادة كتابة المعادلة بالشكل الآتي.

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (2.24)$$

$$y_i(w^T x_i + b) \geq 1 \quad \forall \quad x_i, i = 1, 2, \dots, n$$

وحل هذه المشكلة يكون بإدخال معاملات لاگرانج Lagrangian's Multipliers  $\lambda_i$  وكما يأتي.

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^L \lambda_i y_i (x_i \cdot w + b) + \sum_{i=1}^L \lambda_i \quad (2.25)$$

هنالك نوعان من أنواع آلة المتجه الداعم وهي:

أولاً: آلة المتجه الداعم الخطية: عبارة عن مصنف يستخدم لفصل البيانات القابلة للفصل خطياً بواسطة مستوى فائق وهو عبارة عن مستقيم يفصل مجموعة البيانات الى مجموعتين.

ثانياً: آلة المتجه الداعم غير الخطية: عبارة عن مصنف يستخدم لفصل البيانات الغير قابلة للفصل خطياً بواسطة مستقيم حيث يعمل على تحويل البيانات من فضاءها الحالي الى فضاء اعلى حيث يمكن من خلالها فصل البيانات. فمثلا اذا كانت البيانات في فضاء ثنائي البعد يتم إضافة بعد ثالث وبعدها يتم اختبار فصل البيانات في الفضاء ثلاثي الابعاد وفي حال تحقق الشرط يتم فصل البيانات بالمستوى الأمثل المناسب واذا لم يتحقق الشرط يتم إضافة بعد رابع وخامس وهكذا الى ان تكون البيانات قابلة للفصل وتسمى هذه الطريقة خدعة النواة Kernel trick. وهناك عدة أنواع من دوال النواة وهي (Ng, 2000).

1. دالة النواة الشعاعية (RBF) Radial Basis Function (Gaussian) : وتعرف أيضا بدالة نواة غاوسي وهي من اكثر الدوال المستخدمة ومعادلتها كالتالي.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) = \exp\left(-\sum_{i=1}^n \sum_{j=1}^n \frac{(x_i - x_j)^2}{2\sigma^2}\right) \quad (2.26)$$

إذ أن  $x_i$  و  $x_j$  يمثلان متجهين في مصفوفة  $K$  للملاحظات  $i$  و  $j$  في المتغير  $x$ ، وان المقام في معادلة دالة النواة هو عبارة عن معلمة ضعف مربع  $\sigma$ .

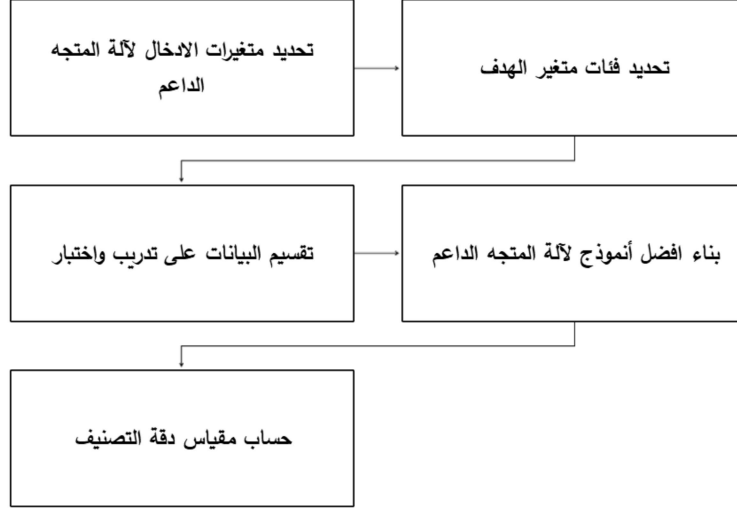
2. دالة النواة الخطية (LK) Linear Kernel function : ويتم عن طريق ضرب متجهين وليكن مثلا  $x_i$  و  $x_j$ .

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (2.27)$$

3. دالة النواة كثيرة الحدود (PK) Polynomial Kernels function : تعرف دالة النواة كثيرة الحدود عن طريق ضرب متجهين مرفوع الى  $d$  وهي درجة كثيرة الحدود كما في المعادلة التالية.

$$K(x_i, x_j) = (x_i^T \cdot x_j)^d \quad (2.28)$$

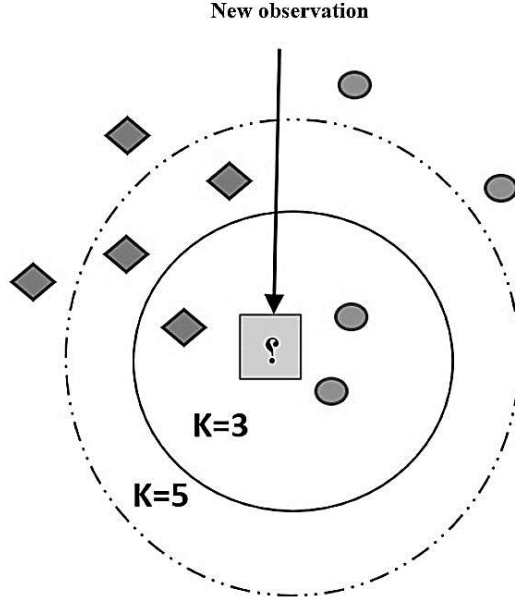
حيث ان  $d$  تمثل مرتبة النواة (Shalev-Shwartz & Ben-David, 2014). والشكل 2. 12 في أدناه يوضح خطوات عمل SVM.



الشكل 2. 12 : خطوات عمل آلة المتجه الداعم

## 8.2 K الجار الأقرب (KNN) K-nearest neighbor

تعد KNN إحدى طرائق التقدير الالعملية المستخدمة في التصنيف الثنائي، وتقوم فكرتها على أنه لو كان هناك عدد من المشاهدات لصنفين مختلفين من البيانات، ولغرض تصنيف كل مشاهدة بحسب الفئة التي تنتمي إليها، فيكون ذلك عن طريق قياس المسافة بين المشاهدة وعدد من النقاط القريبة منها، بواسطة مقياس مسافة محدد، وتسمى هذه النقاط بالجار الأقرب  $K$ ؛ إذ يجب تحديد عدد الجيران قبل تطبيق KNN، ويتم بعد ذلك ترتيب المسافات من الأصغر إلى الأكبر، إذ إن المسافات ذات القيمة الأصغر هي التي تمثل اقرب الجيران. وبعد ذلك تُحدد الفئة التي تنتمي إليها اغلب النقاط المجاورة للمشاهدة المراد تصنيفها، وهكذا يتم التصنيف بتحديد الفئة التي تنتمي إليها المشاهدة. إن الشكل 2. 13 في أدناه يحتوي على عينات تدريب من فئتين (دائرة ومعين). وتقوم طريقة KNN بتحديد المسافة بين المشاهدة الجديدة -وهي عينة الاختبار- مع كل عينة من عينات التدريب، بواسطة أحد مقاييس المسافة. وعندما  $k=3$  تُصنّف المشاهدة الجديدة إلى فئة الدائرة. أما عندما  $k=5$  تُصنّف المشاهدة الجديدة إلى فئة المعين (Steinbach & Tan, 2009).



الشكل 2. 13 : آلية عمل KNN من اقتراح الباحث بالاعتماد على (Rithesh, 2017).

تُحسب المسافة بين المشاهدة المراد تصنيفها والجيران الأقرب، عن طريق أحد مقاييس المسافة التي تعمل على وفق حساب التشابه بين نقاط البيانات؛ إذ إن النقاط المتشابهة ستكون أقرب إلى بعضها بعضاً في المسافة. 1. مقاييس مسافة منكوفسكي Minkowski distance ويتضمن ثلاثة مقاييس للمسافة، تتوافق مع عدة قيم من  $p$  وتعتمد بصورة عامة على الفرق المطلق بين قيم متجهين. ومعادلتها كالآتي.

$$d(X, Y) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p} \quad (2.29)$$

إذ أن  $p$  هي قيمة موجبة وإن  $x_i$  تمثل قيمة  $i$  في المتجه  $X$  و  $y_i$  تمثل قيمة  $i$  في المتجه  $Y$ ، وعندما  $p=1$  فإنه يمثل مقياس Manhattan، وعندما  $p=2$  فيمثل مقياس المسافة الإقليدية، وعندما  $p=\infty$  فيمثل مقياس Chebychev وكما يأتي.

أ. مقياس المسافة الإقليدية Euclidean distance: تمثل المسافة الإقليدية أقصر مسافة بين متجهين، وهي الجذر التربيعي لمجموع مربعات الفروقات بين العناصر المتناظرة ومعادلتها كالآتي.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.30)$$

إذ أن  $x_i$  تمثل المشاهدات للمتغير المستقل المراد تصنيفها،  $y_i$  تمثل المشاهدات للمتغير المستقل الجيران الأقرب،  $n$  تمثل عدد المتغيرات المستقلة (Imandoust & Bolandraftar, 2013).

ب. مقياس مسافة كتلة المدينة City block distance ويسمى كذلك مقياس مانهاتن Manhattan وهي مجموع الاختلافات المطلقة بين القيمة  $i$  في المتجه  $X$  والقيمة  $i$  في المتجه  $Y$  ومعادلتها كالآتي.

$$d(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (2.31)$$

ج. مقياس مسافة شيبشيف Chebychev distance ويسمى كذلك أقصى قيمة للمسافة أو مسافة رقعة الشطرنج، وهو أكبر مسافة للفرق المطلق بين القيمة  $i$  في المتجه  $X$  والقيمة  $i$  في المتجه  $Y$  ومعادلته.

$$d(X, Y) = \max_i \{|x_i - y_i|\} \quad (2.32)$$

2. مقياس مسافة المطرقة Hamming distance وهو قياس عدم التطابق بين متجهين، سواء كان المتجه نصياً أو رقمياً بالطول نفسه، ومقارنة مدى تشابهها مع بعضها عن طريق حساب عدد الحروف أو الأرقام، التي تختلف عن بعضها بعضاً ومعادلته.

$$d(X, Y) = \sum_{i=1}^n 1_{x_i \neq y_i} \quad (2.33)$$

3. مقياس مسافة الجاكارد Jaccard distance ويستخدم لقياس التشابه والاختلاف بين مجموعتين من العينات ومعادلته (Prasatha et al., 2017).

$$d(X, Y) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i y_i} \quad (2.34)$$

إذ أن  $x_i$  تمثل قيمة  $i$  في المتجه  $X$  و  $y_i$  تمثل قيمة  $i$  في المتجه  $Y$ .

4. مقياس مسافة جيب التمام Cosine distance وتسمى كذلك المسافة الزاوية لقياس التشابه بين متجهين عن طريق حساب جيب تمام الزاوية بينهما، ومعادلته بطرح تشابه جيب التمام من واحد.

$$d(X, Y) = 1 - \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (2.35)$$

5. مقياس مسافة الارتباط Correlation distance يستخدم لقياس العلاقة الخطية، أو الارتباط بين نقطتين. ويكون قياس المسافة في المدى بين الصفر والواحد ومعادلته.

$$d(X, Y) = \frac{1}{2} \left( 1 - \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right) \quad (2.36)$$

اذ أن  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  وان  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  (Prasatha et al., 2017). والشكل 2. 14 في ادناه يوضح

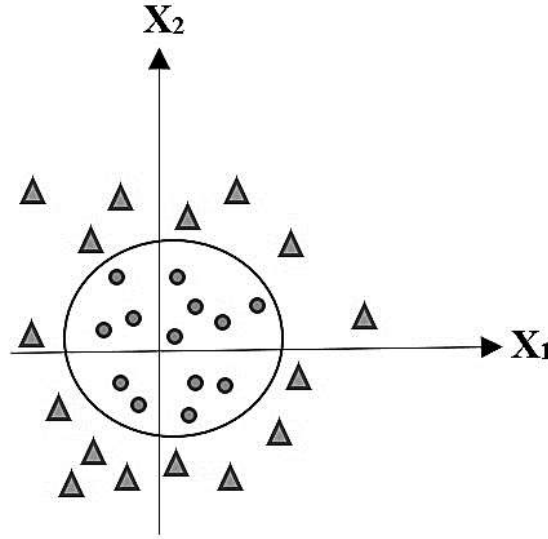
خطوات عمل KNN.



الشكل 2. 14 : خطوات عمل KNN

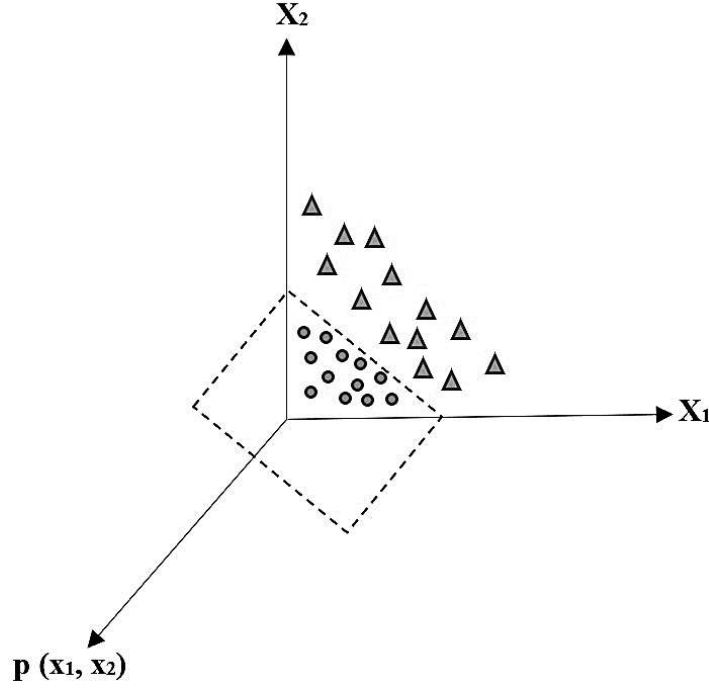
## 9.2 طريقة النواة الغاوسية Gaussian Kernel Method

تعد طريقة النواة الغاوسية من طرائق التعلم الآلي، التي تستخدم في التصنيف الثنائي ولاسيما في حالة البيانات، التي لا يمكن فصلها خطيا. ان طريقة النواة الغاوسية هي أكثر استخداما في البيانات الضخمة، التي تحتوي على مجموعات تدريب كبيرة. وكذلك يمكن تطبيقها على مجموعات البيانات الأصغر. وتستخدم كذلك في الطرائق القائمة على النواة مثل الانحدار الخطي وآلة المتجه الداعم (SVM). إن أساس عمل تصنيف النواة الغاوسية هو البحث عن افضل طريقة لفصل البيانات الى فئتين؛ اذ تقوم بتحويل البيانات التي لا يمكن فصلها خطيا، الموجودة في مساحة منخفضة الابعاد، الى فضاء عالي الابعاد. ثم تعمل على ملائمة أنموذج خطي للبيانات في الفضاء عالي الابعاد. إن الشكل 2. 15 في أدناه يحتوي على عينات تدريب من فئتين (دائرة ومثلث) لمتجهين ( $X_1$  و  $X_2$ ) في فضاء ثنائي الابعاد، وذلك للفصل بين فئتي البيانات وذلك من خلال ملائمة شكل النواة مع البيانات.



الشكل 2. 15 : فصل البيانات في فضاء أو مساحة الإدخال الأصلية

في الشكل 2. 15 في أعلاه نقاط لا تمكن رؤيتها بشكل واضح أو الوصول إليها، ولذلك فان عملية فصل البيانات في مساحة الإدخال الأصلية قد تمت بصورة معقدة وغير خطية. وفي هذه الحالة تجب إعادة صياغة الزوايا والاطوال والمسافات الى فضاء، أو أبعاد أعلى لكي تُفصل البيانات خطيا وبسهولة اكبر بواسطة النواة كما في الشكل 2. 16 في أدناه.



الشكل 2. 16 : فصل البيانات في فضاء عالي الابعاد

في الشكل 2. 16 في أعلاه فُصِّلت البيانات بشكل خطي وبسهولة بعد تحويل البيانات من فضاء ثنائي البعد الى فضاء ثلاثي الابعاد عن طريق دالة النواة الكاوسية (Gaussian Kernel) على وفق المعادلة أدناه.

$$K(x, x') = \exp(-\|x - x'\|^2 / (2\sigma^2)) \quad (2.37)$$

اذ ان  $x$  و  $x'$  متجهان. والمعلمة هي  $\sigma$ .

ان دالة النواة الكاوسية تقوم بتحويل المتجهين ( $x$  و  $x'$ ) الى فضاء متجه جديد. وتعمل هذه الدالة بوصفها مقياس تشابه؛ اذ تقوم بتحديد التشابه او الاختلاف بين كل نقطتين في الفضاء عالي الابعاد. وتعطي دالة النواة نتيجة 1 في حال التشابه الكبير بين النقاط، وتعطي النتيجة صفر في حال الاختلاف الكبير بين النقاط؛ اذ ان المقام في معادلة دالة النواة هو عبارة عن ضعف مربع  $\sigma$  وتؤدي زيادتها الى تقليل ما بداخل الدالة الاسية، وهو ما يجعل النتيجة تقترب من 1. واذا كانت  $\sigma$  قليلة فإن ذلك يؤدي الى زيادة ما بداخل الدالة الاسية، وهو ما يجعل النتيجة تقترب من صفر. وتعمل دالة النواة على تحديد مقدار التشابه والاختلاف بين المدخلات، بمعنى أنه يعمل على تقسيم البيانات على قسمين لغرض تصنيفها. أي أن قيم  $x$  و  $x'$  عندما تكون قريبة من بعضها فإن الفارق يكون قليلاً ويقترب من الصفر. وبهذا يقترب كل ما في داخل الدالة الاسية من

الصفر فيصبح الناتج 1 أما عندما تكون هذه القيم بعيدة عن بعضها فإن الفارق يكون كبيراً، وبذلك يكون داخل الدالة الاسية سالب رقم كبير والناتج يكون صفرًا أو ما يقرب من الصفر (Smola et al., 2005). ان دالة النواة الكاوسية هي في الأصل مشتقة من دالة التوزيع الاحتمالي الطبيعي Normal distribution في أدناه.

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad x \in R \quad (2.38)$$

اذ ان  $\mu$  تمثل المتوسط،  $\sigma$  تمثل الانحراف المعياري (Le et al., 2013). إن الشكل 2. 17 في ادناه يوضح خطوات عمل GK.



الشكل 2. 17 : خطوات عمل GK

## 10.2 مقاييس الدقة للتصنيف Classification Accuracy Measurement

تستخدم هذه المقاييس لقياس دقة أداء التصنيف. وللتعرف إلى هذه المقاييس تجب معرفة مصفوفة الارتباك (Confusion Matrix (CM) التي هي عبارة عن ملخص جدولي لعدد التنبؤات الصحيحة وغير الصحيحة التي قام بها المصنف، حيث تقارن المصفوفة القيم المستهدفة الفعلية مع القيم المتنبأ بها (Luque et al., 2019) كما موضح في المعادلة (2.39) والجدول 2. 1 في أدناه.

$$CM = \frac{-1 \left[ \begin{array}{l} \text{True Positive (TP)} \\ \text{False Positive (FP)} \end{array} \right] + 1 \left[ \begin{array}{l} \text{False Negative (FN)} \\ \text{True Negative (TN)} \end{array} \right]}{2} \quad (2.39)$$

إذ أن (TP) تمثل عدد المشاهدات التي صُنفت بشكل صحيح على أنها إيجابية. وأن (TN) تمثل عدد المشاهدات التي صُنفت بشكل صحيح على أنها سلبية. وأن (FP) تمثل عدد المشاهدات التي صُنفت بشكل غير صحيح على أنها إيجابية وقد كانت في الواقع سلبية. وأن (FN) تمثل عدد المشاهدات التي صُنفت بشكل غير صحيح على أنها سلبية وقد كانت في الواقع إيجابية.

### الجدول 2 . 1 : مصفوفة الارتباك

		القيم الاصلية	
		Yes	No
القيم التنبؤية	Yes	True Positive TP	False Positive TP
	No	False Negative FN	True Negative TN

ومن ابسط المقاييس المستخدمة في التصنيف مقياس الدقة التصنيفية Classification Accuracy الذي تُحسب فيه نسبة الحالات المتوقعة المطابقة للحالات الفعلية الى العدد الكلي للحالات المتوقعة والفعلية المطابقة وغير المطابقة جميعاً وكما في المعادلة (2.40).

$$Accuracy = \frac{\text{Number of correctly classified}}{\text{Total Number}} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2.40)$$

ويُحسب عدم الدقة لمقياس Accuracy عن طريق طرح ناتج المعادلة (2.40) من العدد 1 وكما يأتي.

$$inaccuracy = 1 - Accuracy = \frac{FP+FN}{TP+TN+FP+FN} \quad (2.41)$$

## الفصل الثالث

### الجانب التطبيقي

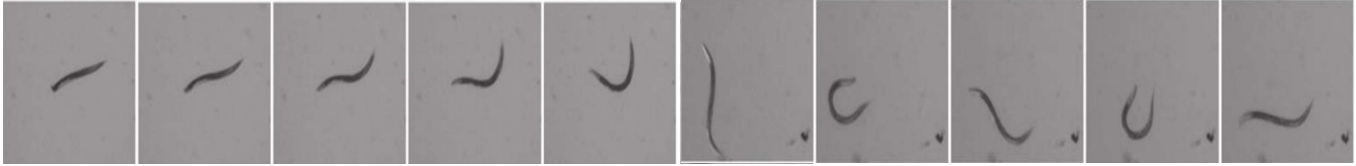
#### 1.3 المقدمة

يتناول هذا الفصل الجانب التطبيقي من الرسالة، ويتضمن تطبيق طرائق تصنيف السلاسل الزمنية لتحسين النتائج، من خلال تحويل مشاهدات السلسلة الزمنية من حالتها الرقمية، وتشكيلها بوصفها صوراً ثنائية البعد، باستخدام الشبكة العصبية الالتغافية CNN وذلك بالاعتماد على رتبة أنموذج الانحدار الذاتي  $AR(p)$  في تحديد عدد عصبونات أو مصفوفة الاوزان في الطبقة الأولى الخفية. وكذلك تطبيق طرائق تصنيف تتعامل مع السلاسل الزمنية بصورتها الرقمية، وهي الانحدار اللوجستي LR في التصنيف الثنائي. وكذلك آلة المتجه الداعم SVM و K الجار الأقرب KNN وطريقة النواة الغاوسية GK وشبكة المدرك العصبية PNN. وفي هذه الطرائق - وفيما يخص عدد متغيرات الادخال - تم الاعتماد على رتبة أنموذج الانحدار الذاتي  $AR(p)$ ، وذلك اعتماداً على شكل دالتي الارتباط الذاتي Autocorrelation Function (ACF) والارتباط الذاتي الجزئي Partial Autocorrelation Function (PACF) للسلاسل الخمس في فترات التدريب. أما بيانات الاختبار المقابلة فقد تم الاعتماد على العدد نفسه الذي تم تحديده. وبعد تحديد الرتب تم اخذ التخلفات الزمنية بوصفها متغيرات ادخال للطرائق (LR و SVM و KNN و GK و PNN) لكل سلسلة.

#### 2.3 البيانات المستخدمة في الدراسة

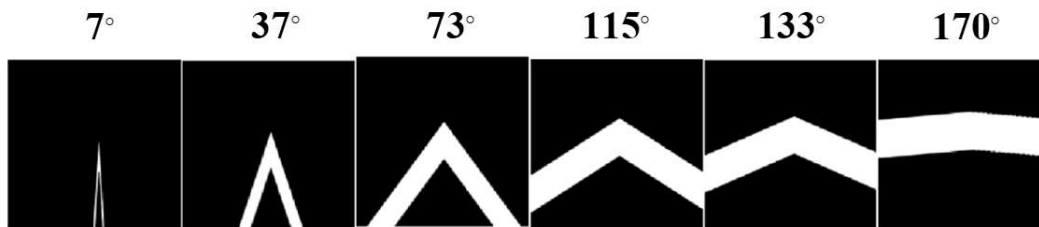
وهي عبارة عن زوايا CEM، التي تتحرك بحرية على مدى فترات طويلة من الزمن وعلى طول لوحة اجار Agar plate، التي تمثل زوايا جسم الدودة في اثناء حركتها على شكل سلاسل زمنية طويلة وبعدد مشاهدات كبير جداً. اذ ان كل مشاهدة في السلسلة تمثل اطاراً Frame من إطارات فيديو يصور الدودة وزاوية انحناء جسمها في اثناء حركتها عند ذلك الاطار وبمعدل لا يقل عن 2 إطار في الثانية، وبواقع 120 اطاراً في الدقيقة. أي ان المشاهدات هي مشاهدات جُمعت عبر الزمن، واعتماداً على الزمن. وقد جُمعت كل سلسلة زمنية من فيديو بطول ساعتين ونصف تقريباً مقسم على ما يقرب من 18000 اطار. كل اطار منها يمثل نصف ثانية؛ ذلك أن البيانات تم الحصول عليها بشكل رقمي بوصفها مشاهدات، ولم يتم الحصول على الفيديو الأصيل. ان زاوية الانحناء تحدد سرعة الحركة؛ لذلك تستخدم زاوية انحناء جسم الدودة للإشارة الى سرعتها،

فكلما كانت الزاوية حادة ازدادت السرعة فتقطع الدودة مسافة أكبر في زمن أقل. وتتراوح درجة زوايا جسم CE بين 1° عندما تنحني على نفسها تقريباً وأقل من 180° بوصفها أكبر زاوية ممكنة عندما تصل الى الاستقامة تقريباً. (Yemini et al., 2014). إن الشكل 3 . 1 في أدناه يوضح انحناءات CE في عدد من الزوايا المختلفة التي صُوّرت من ابعاد مختلفة.



الشكل 3 . 1 : سباحة وحركة دودة الربداء الرشيقية في زوايا مختلفة (Bilbao et al., 2018).

عُنيت دراسات معينة بدراسة متى تقطع الدودة مسافات أكبر في وقت أقل، وتبين أن ذلك يتحقق عندما تتحرك بسرعة، أي عندما تكون حركتها بزوايا حادة (Stephens et al., 2010). وهناك دراسات أخرى عُنيت بتوقيفات الدودة لوجود مشكلة ما أو بطء حركتها، أي عندما تكون حركتها بزوايا منفرجة نسبياً (Stephens et al., 2008). وهكذا فإن تحديد الصفة الإيجابية والسلبية يتم بحسب طبيعة الدراسة. ولأن من الصعب التنبؤ بهذا العدد من القيم الرقمية لزوايا CEM لكون القيم كثيرة ومتقاربة من بعضها فإن من الممكن تصنيف هذه الزوايا فنوياً بحسب سرعة الحركة، الى حركة سريعة بزوايا حادة تمثل الصفة الإيجابية (+1) وحركة بطيئة بزوايا منفرجة تمثل الصفة السلبية (-1) أي تصنيف ثنائي binary classification. وقد تم في هذه الدراسة الاعتماد على تحويل درجات زوايا CEM الى اشكال رسومية من خلال صور ثنائية البعد بتدرج الرمادي (Grayscale) وذلك باجتهاد من الباحث، بعد التفكير بعدة حلول كان انسبها تمثيل زاوية انحناء جسم الدودة باقرب شكل ممكن، وهو الذي يمثله الخط الأبيض في كل شكل من الاشكال المختلفة. ومثالاً على ذلك ما يتضح في الشكل 3 . 2 في أدناه.



الشكل 3 . 2 : نماذج من تحويل درجات زوايا CEM الى اشكال رسومية.

بلغ عدد مشاهدات السلسلة الزمنية CEM (17984) مشاهدة، ولخمس سلالات: هي ( N2 reference strain, goa-1 strain, unc-1 strain, unc-38 strain, unc-63 strain). وقد تم اختيار سلسلتين زمنيتين عشوائياً من كل سلالة. كل سلسلة زمنية منهما تمثل CEM لدودة CE واحدة. وتمثل المصفوفة رباعية البعد المتغير المستقل (X) وصُنِفَت الصور داخل المصفوفة الى فئتين (1 و -1) هما: الحادة والمنفرجة؛ لغرض تمثيل المتغير المعتمد (Y).

يعد أسلوب CNN مناسباً على نحو خاص لتحليل بيانات صورية لأشكال الزاوية التي تشكلها CE عند حركتها على شكل مشاهدة لكل وحدة زمنية ضمن المدى مما يقرب من ( $1^\circ$ ) الى ما يقرب من ( $177^\circ$ ). وإن حد العتبة بين الزوايا الحادة والمنفرجة هي الزاوية ( $90^\circ$ ) للتصنيف الثنائي. ولانشاء متغير الادخال لاسلوب CNN حُوِّلت كل سلسلة زمنية من صيغتها الرقمية numerical وشُكِّلت صوراً ثنائية البعد، باستخدام برنامج MATLAB؛ وذلك عبر رسم الزاوية التي تأخذ المدى (1-177). ثم حفظ كل متغير سلسلة زمنية بوصفه مصفوفة رباعية الأبعاد. يمثل البعدان الأول والثاني ابعاد كل صورة للزاوية المقابلة لكل مشاهدة. اما البعد الرابع للصورة فيمثل تسلسل المشاهدة التي تم التعبير عنها بوصفها صورة. وظهرت صور CEM في هذه الدراسة على نحو تلقائي بحجم (246×251) بكسل، وبعدها مشاهدات 17984 تعادل تقريبا 80% من المشاهدات الكلية وعددها 14400 مشاهدة تقابل فترة التدريب، و3584 مشاهدة أي 20% تقريبا من المشاهدات الكلية لفترة الاختبار. ولذلك فان الحجم النهائي لصورة متغير الادخال هو (246×251×1×17984) بكسل.

### 3.3 الانحدار الذاتي (AR) Auto Regressive

يمكن عموماً استخدام أي أسلوب يعتمد مبدأ الانحدار أو الانحدار اللوجستي مع بيانات متغيرات السلاسل الزمنية بعد هيكلتها باستخدام نموذج AR. فعلى سبيل المثال لو كانت رتبة نموذج الانحدار الذاتي هي AR(6) فإن من الممكن تكوين ستة متغيرات انحدار ذاتي، يمكن تمثيلها، من خلال التخلفات الزمنية، على شكل متغيرات تفسيرية. فيما يكون متغير السلسلة الزمنية الأصلية متغير استجابة (معتمداً) كما في الجدول 3 . 1.

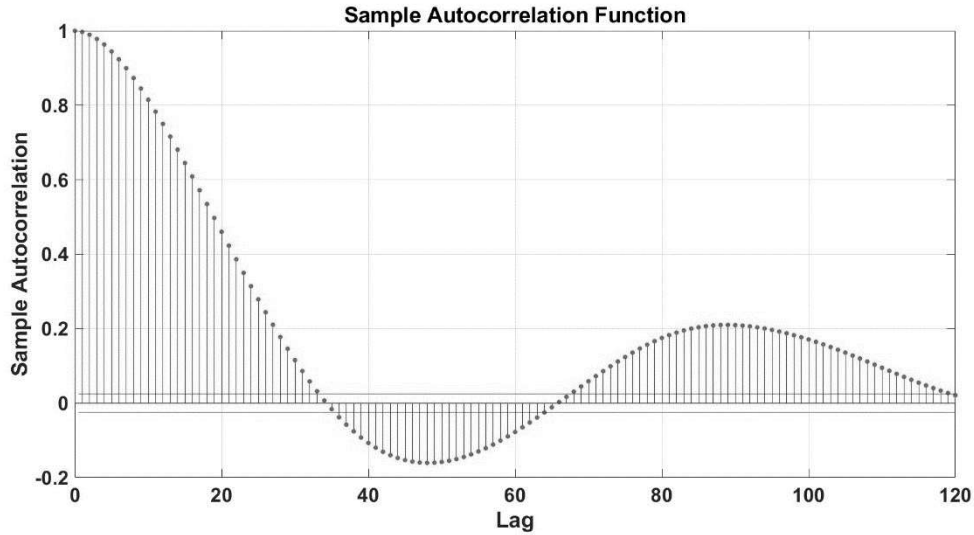
الجدول 3 . 1 : مثال لهيكلية نماذج الانحدار الذاتي

$x_t$	$x_{t-1}$	$x_{t-2}$	$x_{t-3}$	$x_{t-4}$	$x_{t-5}$	$x_{t-6}$
23	*	*	*	*	*	*
26	23	*	*	*	*	*
30	26	23	*	*	*	*
54	30	26	23	*	*	*
33	54	30	26	23	*	*
10	33	54	30	26	23	*
18	10	33	54	30	26	23
29	18	10	33	54	30	26
37	29	18	10	33	54	30
44	37	29	18	10	33	54
68	44	37	29	18	10	33
70	68	44	37	29	18	10
89	70	68	44	37	29	18
96	89	70	68	44	37	29
20	96	89	70	68	44	37

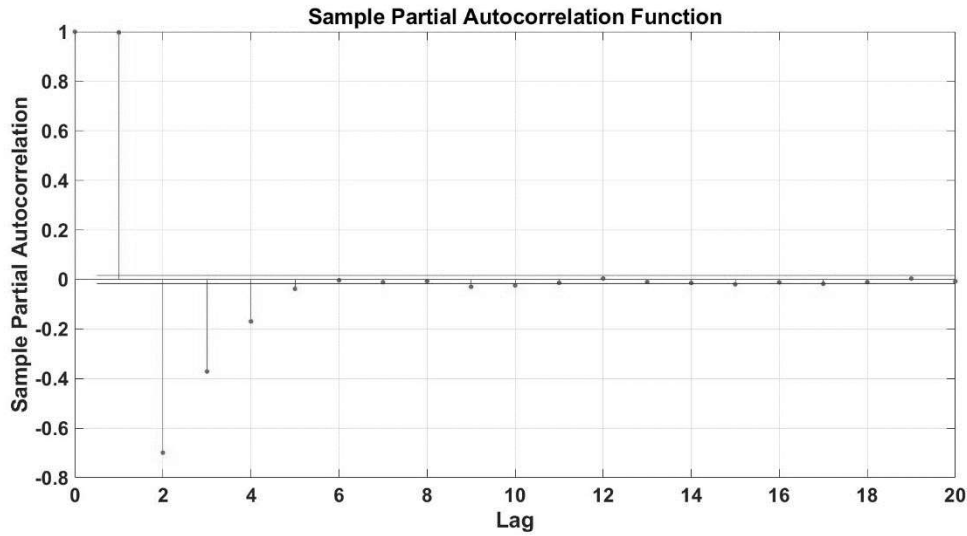
إذ أن  $x_t$  هي السلسلة الأصلية. وان  $x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-6}$  هي التخلفات الزمنية الستة الموجودة في نموذج (6) AR ، التي يعتمد على تأثيراتها متغير السلسلة الزمنية الأصل  $x_t$  إذ نلاحظ على سبيل المثال من خلال الجدول 3 . 1 في أعلاه أن المشاهدة التي قيمتها (44) في الصف الرابع من العمود الاول تتأثر بسابقتها الستة المحاطة بدوائر صغيرة. كما أنها تتأثر بالقيم أنفسها لسابقتها الموجودة في التخلفات الزمنية المقابلة لها. علما ان التخلف الزمني الأول مثلا يتخلف بازاحة زمنية واحدة عن متغير السلسلة الزمنية الأصل فيما تفقد المشاهدة الأولى، وكذلك يتم اقتصاص المشاهدة الأخيرة الزائدة عن عدد المشاهدات الأصلية وهكذا الأمر في باقي التخلفات الزمنية.

اعتمادا على مبدأ الانحدار الذاتي والارتباط الذاتي تم تحديد رتبة نموذج الانحدار الذاتي (AR) من خلال الدالتين (ACF) و (PACF)، بغض النظر عن استقرارية البيانات؛ لأنها تؤدي الى فقدان جزء كبير من

خصائص الانحدار الذاتي والارتباط الذاتي للسلسلة الزمنية الأصلية وبالتالي فقدان الخواص الأساسية التي تتميز بها السلسلة. ولعل من الممكن الاستدلال على أن أفضل نموذج للانحدار الذاتي (5) AR لعينة السلسلة الأولى كما في الشكل 3.3 في أدناه؛ وذلك لأن ACF يعطي نمطاً مضمحلاً تدريجياً مع بطء في الاضمحلال، مما يشير الى وجود عدم استقرارية. أما PACF فيعطي نمط الانقطاع الفجائي بعد (5) من التخلفات الزمنية كما في الشكل 3.4.



الشكل 3.3 : الارتباط الذاتي لعينة السلسلة الأولى (ACF)



الشكل 3.4 : الارتباط الذاتي الجزئي لعينة السلسلة الأولى (PACF)

بعد تحديد رتبة أنموذج الانحدار الذاتي لبيانات السلاسل الزمنية الخاصة بالتدريب قُدرت معاملات نماذج الانحدار الذاتي المعنوية لكل متغير سلسلة زمنية باستخدام برنامج MATLAB وان الجدول 2. 3 و الجدول 3. 3 في أدناه يبين معاملات الانحدار الذاتي المعنوية المقدره وقيم p-value لكل متغير سلسلة زمنية.

**الجدول 2. 3 : معاملات الانحدار الذاتي لسلاسل العينة الأولى**

السلسلة الأولى AR(5)		السلسلة الثانية AR(5)		السلسلة الثالثة AR(7)		السلسلة الرابعة AR(8)		السلسلة الخامسة AR(8)	
P-value	المعلمة	P-value	المعلمة	P-value	المعلمة	P-value	المعلمة	P-value	المعلمة
0	-1.384	0	-1.642	0	-0.777	0	-0.881	0	-1.046
0	0.097	0	0.542	0	-0.298	0	-0.33.	0	-0.089
0	0.142	0	0.123	0	-0.046	0	-0.009	0	0.006
0	0.127	0	0.004	0	-0.032	0	0.043	0	0.017
0	0.019	0	-0.019	0	0.029	0	0.043	0	0.011
0				0	0.079	0	0.070	0	0.023
0				0	0.047	0	0.042	0	0.030
0							0.021	0	0.047

**الجدول 3. 3 : معاملات الانحدار الذاتي لسلاسل العينة الثانية**

السلسلة الأولى AR(5)		السلسلة الثانية AR(5)		السلسلة الثالثة AR(6)		السلسلة الرابعة AR(7)		السلسلة الخامسة AR(2)	
P-value	المعلمة	P-value	المعلمة	P-value	المعلمة	P-value	المعلمة	P-value	المعلمة
0	-1.299	0	-1.384	0	-1.157	0	-1.128	0	-0.724
0	0.005	0	0.118	0	-0.113	0	-0.176	0	-0.275
0	0.108	0	0.194	0	0.144	0	0.110		
0	0.126	0	0.064	0	0.080	0	0.105		
0	0.060	0	0.011	0	0.075	0	0.048		
				0	-0.028	0	0.009		
							0.031		

من خلال الجدول 2. 3 و الجدول 3. 3 يتبين ان المعلمات في نماذج الانحدار الذاتي لمتغيرات السلاسل الزمنية المستخدمة معنوية بحسب قيم p-value المشار اليها في الجداول في أعلاه، وعلى ذلك فقد أُستُخدمت

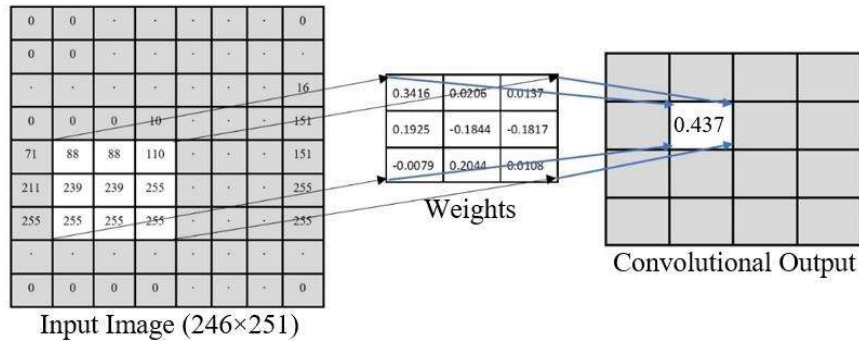
التخلفات الزمنية Lags المشار اليها في الجدول 3. 4 في أدناه لمتغير كل سلسلة زمنية لبيانات التدريب في تحديد هيكلية الأساليب التي أُستُخدمت في التصنيف.

الجدول 3. 4 : نماذج الارتباط الذاتي الانسب

السلسلة الأولى	السلسلة الثانية	السلسلة الثالثة	السلسلة الرابعة	السلسلة الخامسة
AR(5)	AR(5)	AR(7)	AR(8)	AR(8)
AR(5)	AR(5)	AR(6)	AR(7)	AR(2)

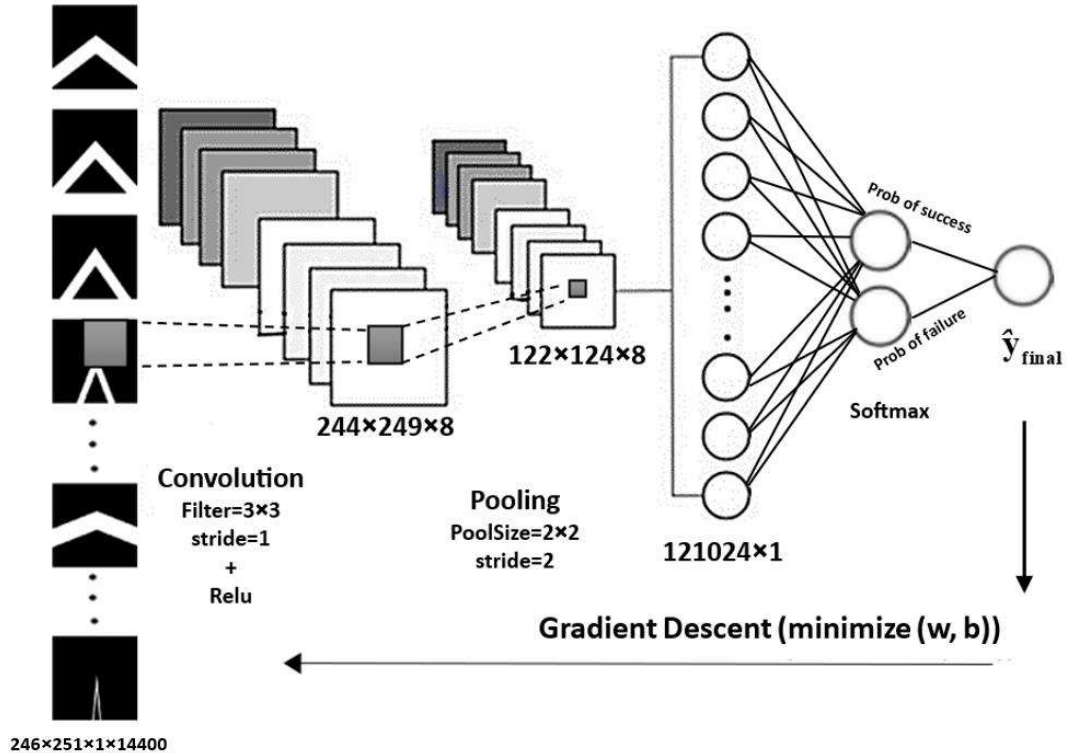
## 4.2 الشبكة العصبية الالتفافية (CNN) Convolutional Neural Network

- إن الاطار العام لطريقة CNN يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي.
1. تحويل المشاهدات من حالتها الرقمية الى مصفوفة واحدة رباعية الأبعاد تجمع الصور مع بعضها.
  2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.
  3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والإختبار.
  4. تحديد بنية CNN (هيكل الشبكة) بطبقة الادخال، والطبقة الخفية عدد (2)، وطبقة الإخراج. أي أن اعداد الطبقات على نحو عام هي (1-2-1).
  5. تحديد حجم المرشح (3×3) وعدد المرشحات بالاعتماد على الجدول 3. 4 وخطوة واحدة في الطبقة الأولى والشكل 3. 5 يوضح حجم الصورة والمرشح وعملية التفاف المرشح على جزء من الصورة.



الشكل 3. 5 : عملية التفاف المرشح على جزء من الصورة

6. جمع قيمة التحيز مع كل عنصر من عناصر المصفوفة (244×249) الناتج من الخطوة 5 والمعادلة (2.2) وتطبيق دالة ReLU عليه.
  7. طُبِّق متوسط التجميع Average Pooling بحجم (2×2) وخطوة (2) على ناتج دالة ReLU.
  8. جمع مخرجات متوسط التجميع في عمود واحد، عن طريق عملية التسوية، فتم الحصول على متجه يمثل الطبقة المتصلة بالكامل، ثم طُبِّقَت المعادلة (2.3) على مدخلات هذه الطبقة.
  9. تطبيق دالة (Softmax) على مخرجات الطبقة المتصلة بالكامل.
  10. طبقة الإخراج تمثل عملية تصنيف صور السلسلة الزمنية الى الفئة الحادة أو المنفرجة.
  11. تحديد خيارات التدريب، ومنها: عدد تكرار المحاولات (افتراضي)، ومعدل التعلم.
  12. انشاء الشبكة للتدريب.
  13. قياس دقة نموذج التصنيف بتطبيق المعادلة (2.40).
- والشكل 3. 6 في أدناه يوضح خوارزمية الشبكة العصبية الالتفافية التي طُبِّقَت على صور زوايا حركة الدودة.

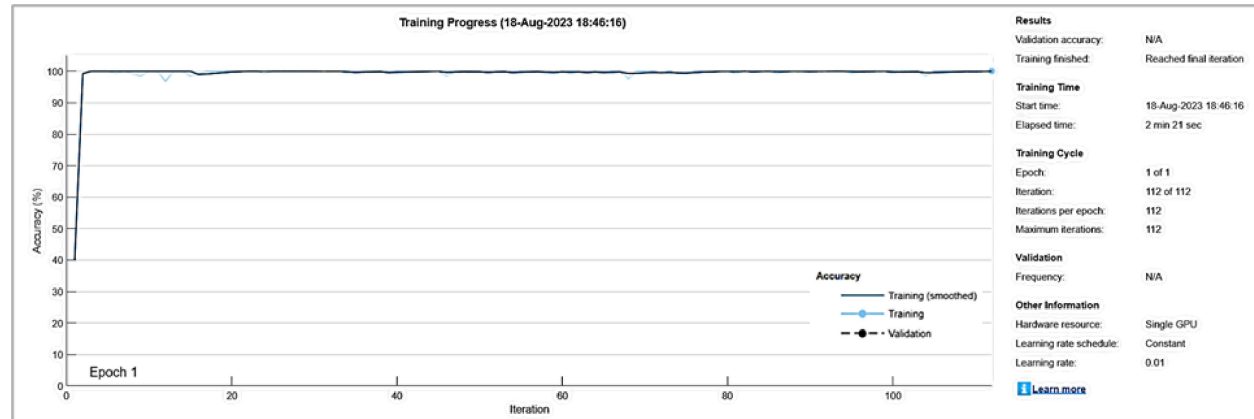


الشكل 3. 6 : عملية تدريب CNN لبيانات العينة الاولى من السلالة الرابعة

عند تقدم التدريب لأنموذج CNN يمكن مراقبة العملية من خلال الشكل الذي يظهر في أثناء تدريب الأنموذج، والذي سيتم ادراجه في أدناه تحت عنوان سلوك عملية التدريب وكذلك إدراج جدول يتضمن معلومات عن أداء الشبكة، كما سيُدْرَج جدول عملية التدريب لكل سلالة، وكل جدول منها يتضمن خمسة أعمدة: يمثل العمود الأول رقم التكرار. ويمثل العمود الثاني الوقت الإجمالي بالساعات والدقائق والثواني للوصول الى بداية التكرار اللاحق. ويمثل العمود الثالث دقة الأنموذج على بيانات التدريب، ويمثل العمود الرابع خطأ الأنموذج في بيانات التدريب. ويمثل العمود الخامس معدل التعلم.

**الجدول 3. 5 : معلومات عملية تدريب CNN لبيانات العينة الأولى من السلالة الأولى بمعدل تعلم 0.01**

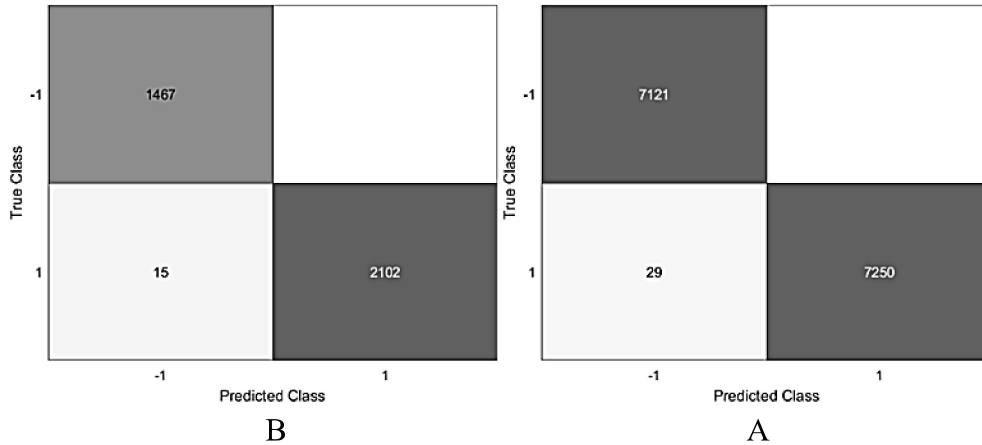
Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 00: 57	39.84%	0.76	0.01
50	00: 01: 30	100.00%	0	0.01
100	00: 02: 11	99.22%	0.08	0.01
112	00: 02: 21	100.00%	0	0.01



**الشكل 3. 7 : سلوك عملية تدريب CNN لبيانات العينة الأولى من السلالة الأولى بمعدل تعلم 0.01**

من خلال الجدول 3. 5 في أعلاه يتبين أن التدريب يبدأ بدقة 39.84% وخسارة 0.7684 عند انتهاء التكرار الأول. ويتبين من الشكل 3. 7 أن الدقة تتحسن مع تقدم التدريب لتصل الى أفضل نتيجة وهي 100% تقريبا. ان القيم الظاهرة في الجدول 3. 5 في أعلاه تمثل تكرارات معينة تم تحديدها من خلال البرنامج، وهي عبارة عن لمحة سريعة عن أداء جزء بسيط من تدريب CNN على بيانات التدريب. ولا تمثل هذه التكرارات افضل نتيجة من حيث (افضل دقة) و(اقل خسارة). ومن خلال جدول معلومات أداء الشبكة التي تمت الإشارة إليها

في أعلاه و الشكل 3. 7 الذي يمثل سلوك عملية التدريب، تبين زيادة دقة التصنيف تدريجياً، مثلما نلاحظ أن إنخفاض قيمة خطأ التصنيف يعني أن أداء CNN يتحسن في تصنيف البيانات مع التدريب. ولغرض تصور أداء CNN من خلال مقارنة نتائج التصنيف المتوقعة لصور زوايا حركة CE بالتصنيف الحقيقي، وذلك بتطبيق دالة مصفوفة الارتباك، تم تطبيق الدالة (confusionchart) على بيانات التدريب والاختبار لمتغيرات السلاسل الزمنية كما في الشكل 3. 8 في أدناه والذي يمثل نتائج تطبيق دالة مصفوفة الارتباك على بيانات التدريب A التي تتضمن أربعة مكونات حيث ان القيمة 7121 تمثل عدد المشاهدات التي صنف بشكل صحيح على انها إيجابية ويعبر عنها في المصفوفة (TP) والقيمة 7250 تمثل عدد المشاهدات التي صنف بشكل صحيح على انها سلبية وهي (TN) والقيمة 29 تمثل عدد المشاهدات التي صنف بشكل غير صحيح على انها سلبية وقد كانت في الواقع إيجابية وهي (FN) والمكون الذي لا يحتوي على قيمة فهو في الواقع صفر ويمثل عدد المشاهدات التي صنف بشكل غير صحيح على انها إيجابية وهي في الواقع سلبية وهي (FP) وبيانات الاختبار B والتي تتضمن أربعة مكونات ايضاً.



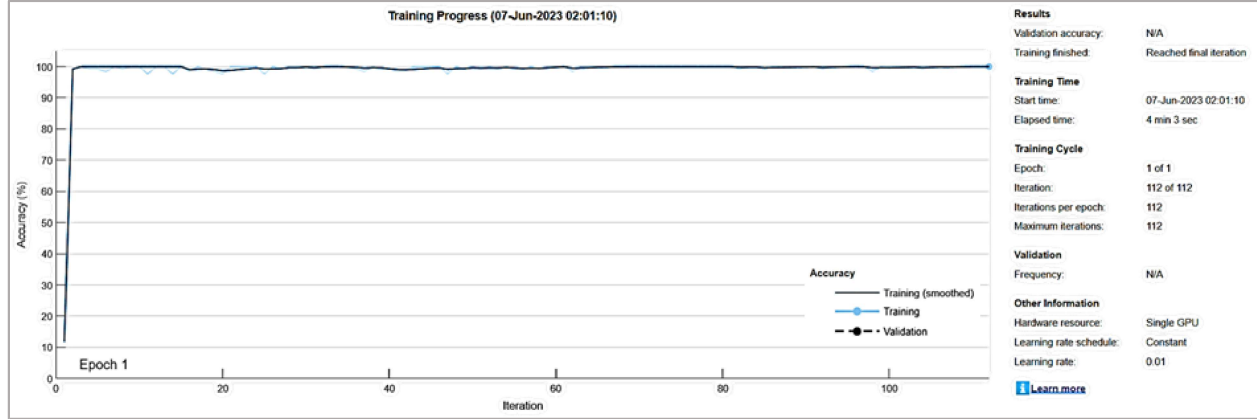
الشكل 3. 8: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب و B: الاختبار (العينة الاولى من

السلسلة الاولى) بمعدل تعلم 0.01

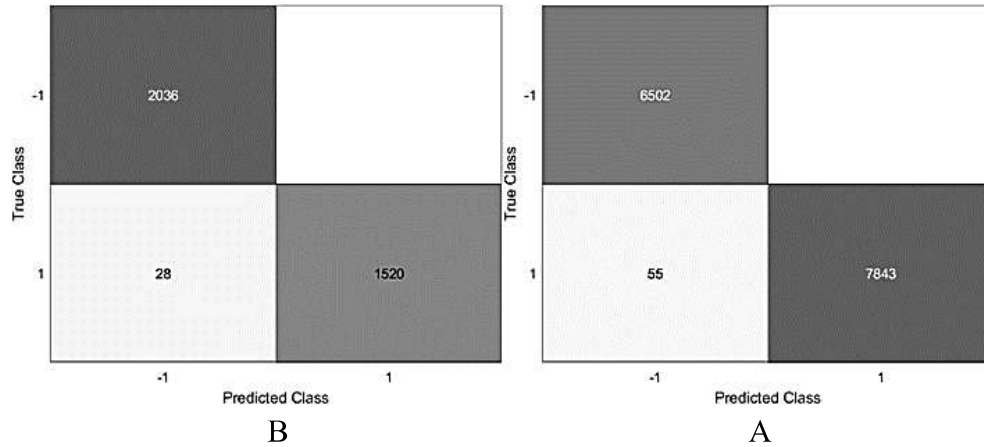
بالنسبة لبيانات العينة الثانية من السلسلة الأولى فان الجدول 3. 6 يمثل معلومات عملية تدريب CNN و الشكل 3. 9 يمثل سلوك عملية تدريب CNN و الشكل 3. 10 في ادناه يمثل مصفوفة الارتباك.

الجدول 3. 6 : معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 59	11.72%	1.32	0.01
50	00: 03: 07	100.00%	0	0.01
100	00: 03: 52	100.00%	0	0.01
112	00: 04: 03	100.00%	0	0.01



الشكل 3. 9 : سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.01 من خلال الجدول 3. 6. يتبين أن الأنموذج يبدأ بدقة 11.72% وخسارة 1.321 عند انتهاء التكرار الأول. ويتبين من الشكل 3. 9 أن الدقة تتحسن مع تقدم التدريب لتصل الى افضل نتيجة وهي 100% تقريباً.



الشكل 3. 10: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى) بمعدل تعلم 0.01

اما سلوك عملية التدريب لباقي السلالات وكذلك الجداول التي تتضمن معلومات عن أداء الشبكة. سيُدْرَج في الملاحق.

وطُبِّقَت معادلة (2.40) لقياس دقة نموذج التصنيف لبيانات التدريب والاختبار لمتغيرات السلاسل الزمنية وكانت النتائج كما في الجدول 7. 3 في أدناه.

### الجدول 7. 3: نتائج قياس دقة التصنيف بأسلوب CNN بمعدل تعلم 0.01.

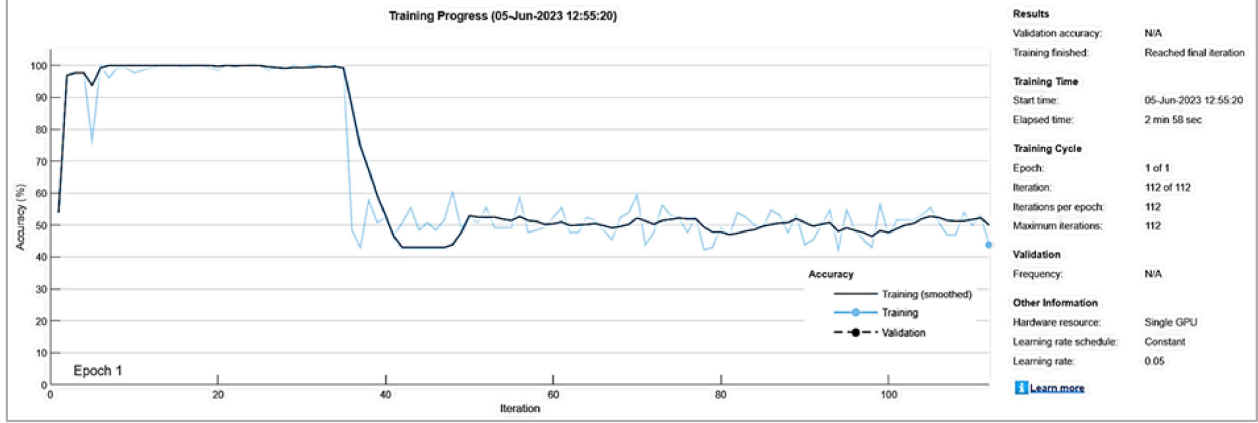
السلالة الأولى	السلالة الثانية	السلالة الثالثة	السلالة الرابعة	السلالة الخامسة		
99.80	99.56	99.74	99.6	99.78	بيانات التدريب	العينة الأولى
99.58	99.72	99.92	99.69	99.83	بيانات الاختبار	
99.62	99.67	99.74	99.53	99.55	بيانات التدريب	العينة الثانية
99.22	99.78	99.67	99.64	99.83	بيانات الاختبار	

ان نتائج قياس دقة نموذج CNN كما في الجدول 7. 3 تؤكد زيادة الدقة في التصنيف، وهي مطابقة تماما لما تم الحصول عليه عند تطبيق دالة مصفوفة الارتباك اذ أن النتائج للسلالات جميعاً كانت ممتازة في مرحلتي التدريب والاختبار.

ولمعرفة مدى تأثير معدل التعلم في أداء الأنموذج تم تدريب الشبكة العصبية الالتفافية بمعدلي تعلم: الأول كان بمقدار (0.01) الذي سبق أن عُرِضَت جداول وسلوك عملية التدريب وكذلك النتائج الخاصة به في أعلاه. وكان معدل التعلم الثاني بمقدار (0.05) الذي سيتم ادراج سلوك وجداول عملية التدريب والنتائج التي تم الحصول عليها باستخدام بيانات التدريب للسلالات الخمس عند استخدامه في ادناه.

### الجدول 8. 3: معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الأولى بمعدل تعلم 0.05.

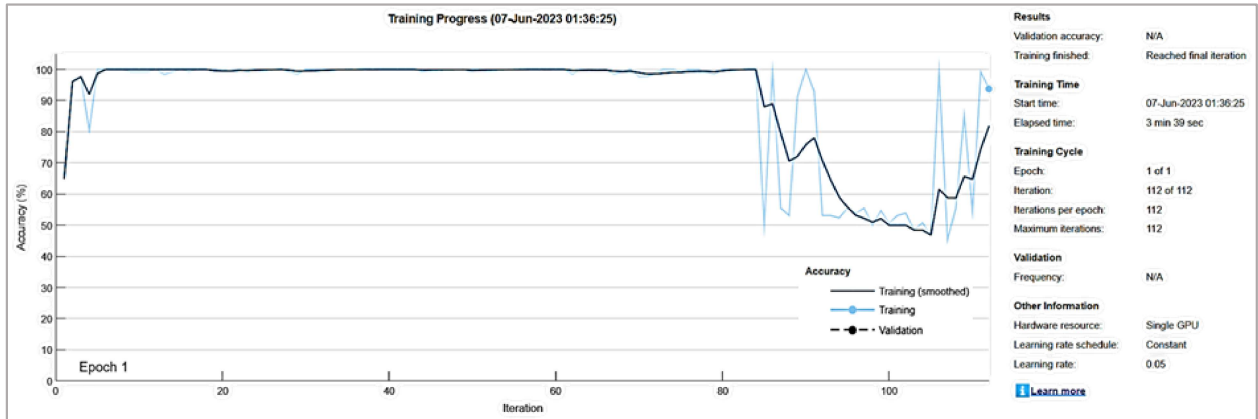
Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 02	53.91%	0.73	0.05
50	00: 01: 57	52.34%	0.69	0.05
100	00: 02: 47	46.88%	0.69	0.05
112	00: 02: 58	43.75%	0.70	0.05



الشكل 3. 11: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الأولى بمعدل تعلم 0.05. من خلال الجدول 3. 8 أعلاه فإن الأنموذج يبدأ بدقة 53.91% وخسارة 0.7396 عند انتهاء التكرار الأول. ويتبين من الشكل 3. 11 أن الدقة تتحسن مع تقدم التدريب لغاية التكرار 35 ثم تبدأ بالانخفاض لتصل الى معدل ثابت تقريبا ابتداءً من التكرار 40 حتى نهاية عملية التدريب.

الجدول 3. 9: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.05.

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 29	64.84%	0.70	0.05
50	00: 02: 12	99.22%	0.12	0.05
100	00: 02: 48	50.78%	0.70	0.05
112	00: 03: 39	93.75%	0.54	0.05



الشكل 3. 12: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الأولى بمعدل تعلم 0.05.

من خلال الجدول 3. 9 في أعلاه يتبين أن الأنموذج يبدأ بدقة %64.84 وخسارة 0.7087 عند انتهاء التكرار الأول. ويتبين من الشكل 3. 12 أن الدقة تتحسن مع تقدم التدريب لغاية التكرار 80 ثم حصل انخفاض وارتفاع في الدقة على نحو غير مستقر حتى نهاية عملية التدريب. وقد طُبِّقَت معادلة (2.40) لقياس دقة أنموذج التصنيف لبيانات التدريب والاختبار لمتغيرات السلاسل الزمنية وكانت النتائج كما في الجدول 3. 10 في أدناه.

### الجدول 3. 10: نتائج قياس دقة التصنيف بأسلوب CNN بمعدل تعلم 0.05.

السلاسل الأولى	السلاسل الثانية	السلاسل الثالثة	السلاسل الرابعة	السلاسل الخامسة		
50.55	52.12	59.36	48.92	53.12	بيانات التدريب	العينة الأولى
59.07	53.26	13.81	36.13	43.83	بيانات الاختبار	
54.85	53.34	49.14	51.56	59.61	بيانات التدريب	العينة الثانية
43.19	52.29	55.36	50.78	13.92	بيانات الاختبار	

ومن خلال سلوك التدريب، وجداول معلومات أداء الشبكة، وتقدم التدريب، فضلاً نتائج دقة التصنيف، تبين أن النتائج الخاصة بأداء الأنموذج في مرحلة التدريب كانت محصورة بين 48.92 و 59.61. ولكنها كانت في مرحلة الاختبار محصورة بين 13.81 و 59.07 أي ان هناك تبايناً كبيراً في أداء الأنموذج فيما يخص البيانات الجديدة، وهذا يدل على ان دقة الأنموذج تتأثر بتغير معدل التعلم؛ نستنتج بانه ومن خلال تقليل معدل التعلم يتم الحصول على نتائج جيدة في التصنيف بأسلوب CNN.

## 5.3 شبكة المدرك العصبية (PNN) Perceptron Neural Network

ان الاطار العام لطريقة PNN يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي:

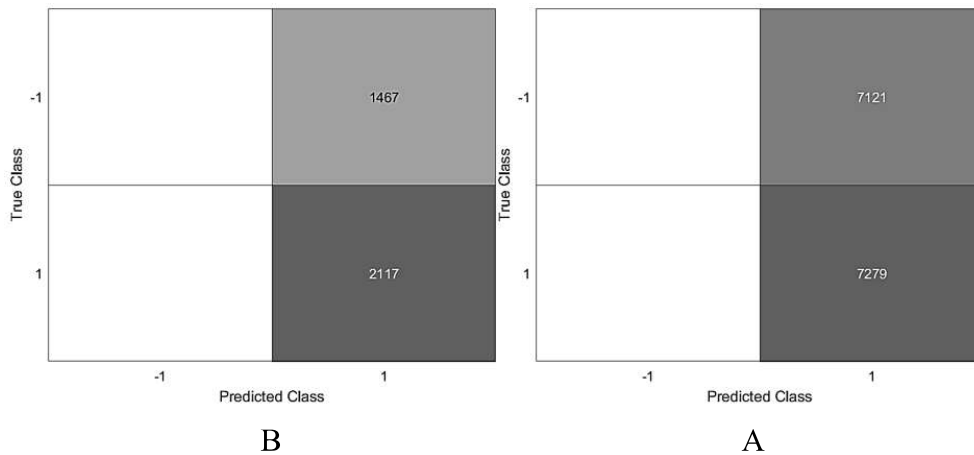
1. استخدام متغيرات الانحدار الذاتي الأمثل اعتمادا على الجدول 3. 4 لتحديد متغيرات الادخال لاسلوب PNN.
2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.
3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والاختبار.

4. اختيار إحدى دوال التحويل الخاصة بشبكة المدرك العصبية.

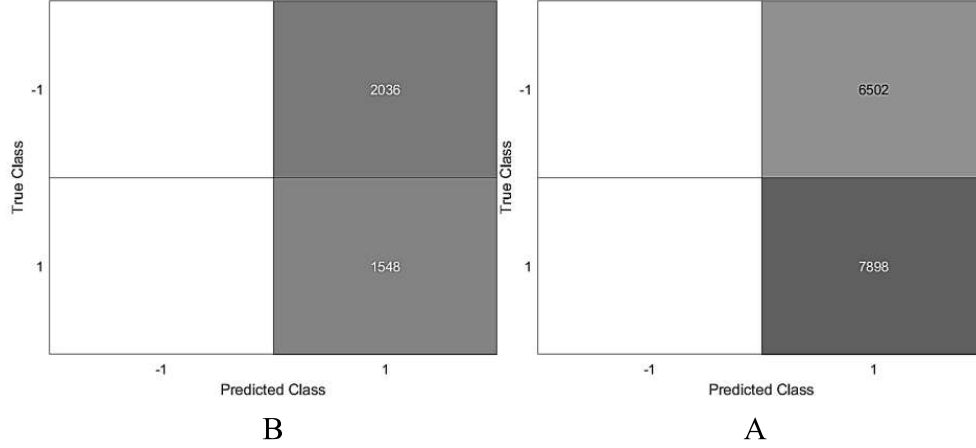
5. انشاء الشبكة للتدريب.

6. حساب دقة التصنيف الخاصة ببيانات الاختبار بتطبيق المعادلة (2.40).

تم تطبيق الدالة (confusionchart) على بيانات التدريب والاختبار لمتغيرات السلاسل الزمنية كما في الشكل 3 . 13 في أدناه والذي يمثل نتائج تطبيق دالة مصفوفة الارتباك على بيانات التدريب A التي تتضمن صفين وعمودين حيث ان المكون (أول صف وأول عمود) لا يحتوي على قيمة وهو في الواقع صفر يمثل عدد المشاهدات التي صنفت بشكل صحيح على انها إيجابية ويعبر عنها في المصفوفة (TP) والقيمة 7279 (ثاني صف وثاني عمود) تمثل عدد المشاهدات التي صنفت بشكل صحيح على انها سلبية وهي (TN) والمكون (ثاني صف وأول عمود) لا يحتوي على قيمة وهو صفر يمثل عدد المشاهدات التي صنفت بشكل غير صحيح على انها سلبية وقد كانت في الواقع إيجابية وهي (FN) والقيمة 7121 تمثل عدد المشاهدات التي صنفت بشكل غير صحيح على انها إيجابية وهي في الواقع سلبية وهي (FP) وبيانات الاختبار B والتي تتضمن أربعة مكونات أيضا.



الشكل 3 . 13 : مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)



الشكل 3 . 14 : مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)

اما بالنسبة لمصفوفة الارتباك لباقي السلالات ستدرج في الملاحق. وقد طُبِّقَت معادلة (2.40) لقياس دقة التصنيف لبيانات التدريب والاختبار لمتغيرات السلاسل الزمنية وكانت النتائج كما في الجدول 3 . 11 في أدناه.

الجدول 3 . 11: نتائج قياس دقة التصنيف لاسلوب PNN.

السلالة الاولى	السلالة الثانية	السلالة الثالثة	السلالة الرابعة	السلالة الخامسة		
50.5486	52.1181	59.3611	48.9167	46.8819	بيانات التدريب	العينة الأولى
59.0681	53.2645	13.8114	36.1328	56.1663	بيانات الاختبار	
54.8472	53.3403	41.1389	48.4375	40.3889	بيانات التدريب	العينة الثانية
43.1920	52.2879	55.3571	49.2188	86.0770	بيانات الاختبار	

من خلال نتائج قياس دقة التصنيف لشبكة المدرك العصبية لبيانات التدريب، يتضح انها تراوحت بين 40.3889 و 59.3611 للسلالات جميعاً. ومثلت العينة الأولى من السلالة الثالثة أعلى دقة تصنيف خاصة ببيانات التدريب. أما بيانات الاختبار فقد تراوحت دقة التصنيف بين 13.8114 و 86.0770 للسلالات جميعها. ومثلت العينة الثانية من السلالة الخامسة اعلى دقة في التصنيف. ويتضح من ذلك أن هناك تقلباً كبيراً وعدم استقرارية في أداء الأنموذج في مرحلة الاختبار.

## 6.3 الانحدار اللوجستي (LR) Logistic Regression

ان الاطار العام لطريقة LR يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي:

1. استخدام متغيرات الانحدار الذاتي الأمثل اعتمادا على الجدول 3. 4 لتحديد متغيرات الادخال لاسلوب .LR
2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.
3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والإختبار.
4. تدريب أنموذج الانحدار اللوجستي الثنائي على البيانات بواسطة الابعاز (fitglm) باستخدام المدخلات، وتتضمن متغيرات الادخال والإخراج ونوع التوزيع.
5. تقييم أداء الأنموذج على بيانات الاختبار بواسطة الابعاز (predict) باستخدام المدخلات، وتتضمن متغيرات الادخال فضلاً عن الأنموذج LR.
6. قياس دقة أنموذج التصنيف باستخدام المعادلة (2.40).

وقد طُبِقَ الانحدار المتدرج على متغيرات السلاسل الزمنية الخمسة للعينتين الأولى والثانية؛ للإبقاء على المعلمات المعنوية حسب. والهدف من استخدام الانحدار المتدرج هو أنه يوصلنا الى اقل عدد من المتغيرات المؤثرة على نحو معنوي في تصنيف المتغير المعتمد حيث يعمل على تحديد المتغيرات الأكثر تأثيرا في الانموذج والتي يمكن استخدامها للتنبؤ بالقيم في المستقبل. وكانت النتائج لأنموذج الانحدار اللوجستي كما هو من الجدول 3. 12 الى الجدول 3. 21 وتتضمن الجداول المذكورة أنفاً في العمود الأول منها التقديرات لمعامل الانحدار لمتغيرات الأنموذج لكل سلسلة زمنية، ويمثل العمود الثاني الانحراف المعياري أما العمود الثالث فيمثل قيمة t الجدولية والعمود الرابع يمثل قيمة p-value لكل متغير.

الجدول 3. 12: جدول تحليل الانحدار اللوجستي لبيانات العينة الأولى من السلسلة الأولى.

	Estimate	SE	t	pValue
Y	30.50482481	1.751975934	17.41166886	6.73E-68
x1	-0.468290908	0.027172671	-17.2338933	1.48E-66
x4	0.126163763	0.008660097	14.56840129	4.46E-48

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الأولى من السلسلة الأولى في المعادلة (3.1) في أدناه.

$$P(y = 1|x_1, x_4) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 30.504 - 0.4682x_1 + 0.1261x_4 \quad (3.1)$$

**الجدول 3. 13: جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الثانية.**

	Estimate	SE	t	pValue
Y	13.2790079	0.3814349	34.81330073	1.53E-265
x1	-0.274699263	0.0098547	-27.8749504	5.37E-171
x2	0.078792089	0.012321811	6.394521762	1.61E-10
x3	0.026316085	0.013422643	1.960574041	0.049928733
x4	0.0198389	0.007978319	2.486601408	0.012896981

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الأولى من السلالة الثانية في المعادلة (3.2) في أدناه.

$$P(y = 1|x_1, x_2, x_3, x_4) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 13.279 - 0.2746x_1 + 0.0788x_2 + 0.0263x_3 + 0.0198x_4 \quad (3.2)$$

**الجدول 3. 14: جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الثالثة.**

	Estimate	SE	t	pValue
Y	19.29895941	0.776763033	24.84536288	2.90E-136
x1	-0.179154004	0.011377711	-15.7460504	7.31E-56
x2	-0.069405926	0.010614564	-6.5387449	6.20E-11
x6	0.032263131	0.004904927	6.577698623	4.78E-11

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الأولى من السلالة الثالثة في المعادلة (3.3) في أدناه.

$$P(y = 1|x_1, x_2, x_6) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 19.298 - 0.1791x_1 - 0.0694x_2 + 0.0322x_6 \quad (3.3)$$

**الجدول 3. 15: جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الرابعة.**

	Estimate	SE	t	pValue
Y	28.99268466	1.475444673	19.65013342	5.77E-86
x1	-0.273184483	0.022115809	-12.3524526	4.72E-35
x2	-0.102509431	0.020345895	-5.03833484	4.70E-07
x8	0.048482238	0.005824379	8.324017754	8.50E-17

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الأولى من السلالة الرابعة في المعادلة (3.4) في أدناه.

$$P(y = 1|x_1, x_2, x_8) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 28.993 - 0.2731x_1 - 0.1025x_2 + 0.0485x_8 \quad (3.4)$$

**الجدول 3. 16: جدول تحليل الانحدار اللوجستي لبيانات العينة الاولى من السلالة الخامسة.**

	Estimate	SE	t	pValue
Y	43.33643015	2.764125088	15.67817257	2.13E-55
x1	-0.524526232	0.047571018	-11.0261722	2.86E-28
x2	-0.089022245	0.043958702	-2.0251336	0.042853647
x6	0.125664634	0.016399156	7.662872179	1.82E-14

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الأولى من السلالة الخامسة في المعادلة (3.5) في أدناه.

$$P(y = 1|x_1, x_2, x_6) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 43.336 - 0.5245x_1 - 0.089x_2 + 0.1257x_6 \quad (3.5)$$

**الجدول 3. 17: جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الأولى.**

	Estimate	SE	t	pValue
Y	29.50524697	1.709351031	17.26108121	9.24E-67
x1	-0.456578608	0.026505964	-17.2255045	1.71E-66
x4	0.123432924	0.008654179	14.26281068	3.73E-46

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الثانية من السلالة الأولى في المعادلة (3.6) في أدناه.

$$P(y = 1|x_1, x_4) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 29.505 - 0.4566x_1 + 0.1234x_4 \quad (3.6)$$

**الجدول 3. 18: جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثانية.**

	Estimate	SE	t	pValue
Y	15.181859	0.504490364	30.09345683	5.90E-199
x1	-0.239329025	0.008851398	-27.0385554	5.21E-161
x3	0.042612484	0.008735247	4.878222802	1.07E-06
x5	0.026757247	0.005382347	4.971297286	6.65E-07

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثانية في المعادلة (3.7) في أدناه.

$$P(y = 1|x_1, x_3, x_5) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 15.182 - 0.2393x_1 + 0.0426x_3 + 0.0268x_5 \quad (3.7)$$

**الجدول 3. 19: جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثالثة.**

	Estimate	SE	t	pValue
Y	18.78229869	0.956037267	19.64599011	6.26E-86
x1	-0.265751256	0.014365239	-18.4996055	2.08E-76
x4	0.056569818	0.006519431	8.677110135	4.06E-18

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الثانية من السلالة الثالثة في المعادلة (3.8) في أدناه.

$$P(y = 1|x_1, x_4) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 18.782 - 0.2658x_1 + 0.0566x_4 \quad (3.8)$$

**الجدول 3. 20: جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الرابعة.**

	Estimate	SE	t	pValue
Y	29.55179438	1.624184059	18.19485557	5.67E-74
x1	-0.386062007	0.021171576	-18.2349209	2.73E-74
x7	0.05499949	0.005115474	10.75159121	5.82E-27

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الثانية من السلالة الرابعة في المعادلة (3.9) في أدناه.

$$P(y = 1|x_1, x_7) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 29.552 - 0.386x_1 + 0.055x_7 \quad (3.9)$$

**الجدول 3. 21: جدول تحليل الانحدار اللوجستي لبيانات العينة الثانية من السلالة الخامسة.**

	Estimate	SE	t	pValue
Y	42.04329103	2.533803403	16.59295705	7.84E-62
x1	-0.370384237	0.04110276	-9.01117683	2.04E-19
x2	-0.106241121	0.035167173	-3.02103103	0.002519155

وتتمثل معادلة الانحدار اللوجستي لبيانات العينة الثانية من السلالة الخامسة في المعادلة (3.10) في أدناه.

$$P(y = 1|x_1, x_2) = \frac{e^{Y'}}{1 + e^{Y'}}; Y' = 42.043 - 0.3703x_1 + 0.1062x_2 \quad (3.10)$$

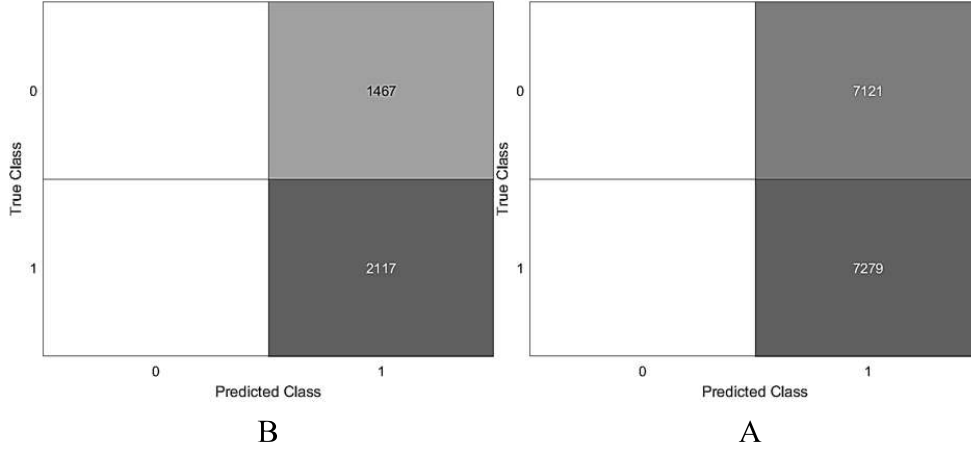
ان متغيرات الانحدار اللوجستي لكل سلسلة زمنية والموجودة من الجدول 3. 12 الى الجدول 3. 21 في أعلاه تمثل المتغيرات المعنوية حسب، أما المتغيرات المهملة التي لم تظهر في الجداول المذكورة أنفا فتعد متغيرات غير معنوية، أي انها لا تؤثر منطقياً في النتائج. وكانت نتائج قياس دقة نموذج التصنيف لبيانات التدريب والاختبار بأسلوب الانحدار اللوجستي كما في الجدول 3. 22 في أدناه.

**الجدول 3. 22: دقة التصنيف للسلالات الخمس لبيانات التدريب والاختبار باستخدام نموذج LR.**

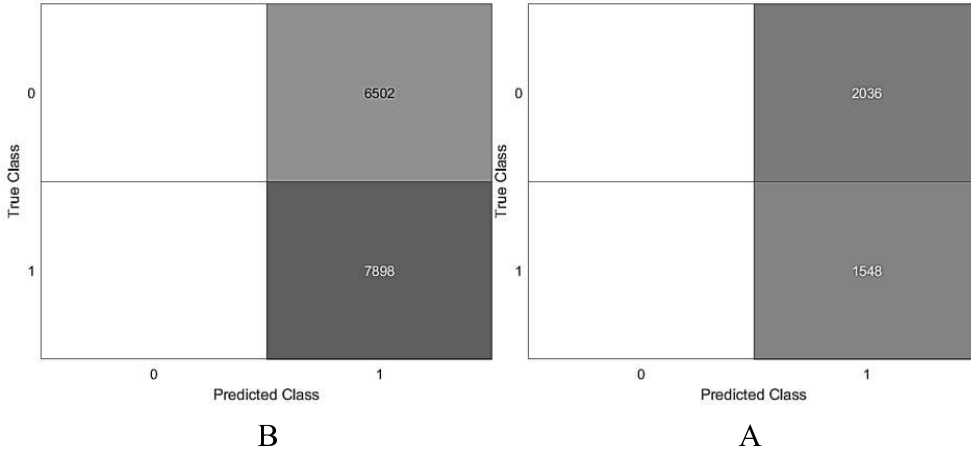
السلالة الأولى	السلالة الثانية	السلالة الثالثة	السلالة الرابعة	السلالة الخامسة		
50.5486	52.1181	59.3611	48.9167	46.8819	بيانات التدريب	العينة الأولى
59.0681	53.2645	13.8114	36.1328	56.1663		
54.8472	53.3403	49.1389	48.4375	40.3889	بيانات التدريب	العينة الثانية
43.1920	52.2879	55.3571	49.2188	86.0770		

تم تطبيق الدالة (confusionchart) على بيانات التدريب والاختبار لمتغيرات السلاسل الزمنية كما في الشكل 3. 15 في أدناه والذي يمثل نتائج تطبيق دالة مصفوفة الارتباك على بيانات التدريب A التي تتضمن صفين وعمودين حيث ان المكون (أول صف وأول عمود) لا يحتوي على قيمة وهو في الواقع صفر يمثل عدد المشاهدات التي صنفت بشكل صحيح على انها إيجابية ويعبر عنها في المصفوفة (TP) والقيمة 7279 (ثاني

صف وثاني عمود) تمثل عدد المشاهدات التي صنفتم بشكل صحيح على أنها سلبية وهي (TN) والمكون (ثاني صف وأول عمود) لا يحتوي على قيمة وهو صفر يمثل عدد المشاهدات التي صنفتم بشكل غير صحيح على أنها سلبية وقد كانت في الواقع إيجابية وهي (FN) والقيمة 7121 تمثل عدد المشاهدات التي صنفتم بشكل غير صحيح على أنها إيجابية وهي في الواقع سلبية وهي (FP) وبيانات الاختبار B والتي تتضمن أربعة مكونات أيضا.



الشكل 3 . 15 : مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)



الشكل 3 . 16 : مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)

أما بالنسبة لمصفوفة الارتباك لباقي السلالات فانها ستدرج في الملاحق. ومن خلال نتائج قياس دقة التصنيف لأنموذج الانحدار اللوجستي في الجدول 3 . 22 في أعلاه لبيانات التدريب يتضح انها تتراوح بين 40.3889

و 59.3611 للسلاطات جميعاً، وقد مثلت العينة الأولى من السلاطة الثالثة أعلى دقة تصنيف خاصة ببيانات التدريب. أما بيانات الاختبار فقد تراوحت دقة التصنيف بين 13.8114 و 86.0770 للسلاطات جميعها، وقد مثلت العينة الثانية من السلاطة الخامسة أعلى دقة في التصنيف وهذا يدل على عدم استقرارية أداء أنموذج التصنيف وتبين من جدول دقة التصنيف ان نتائج الدقة بأسلوب LR كانت مطابقة للنتائج التي تم الحصول عليها باستخدام أسلوب PNN.

### 7.3 آلة المتجه الداعم (SVM) Support Vector Machine

إن الاطار العام لطريقة SVM يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي:

1. استخدام متغيرات الانحدار الذاتي الأمثل اعتمادا على الجدول 3. 4 لتحديد متغيرات الادخال لاسلوب SVM.
2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.
3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والاختبار.
4. بناء افضل أنموذج SVM باستخدام بيانات التدريب لغرض التصنيف الثنائي، واختيار إحدى دوال النواة (RBF, Linear, Polynomial) بوساطة الايعاز (fitcsvm)، باستخدام المدخلات وتتضمن متغيرات الادخال والإخراج.
5. استخدام الأنموذج في الخطوة السابقة لتصنيف البيانات في فترة الاختبار بوساطة الايعاز (predict) باستخدام المدخلات، وتتضمن متغيرات الادخال فضلاً عن الأنموذج SVM.
6. قياس دقة أنموذج التصنيف بتطبيق المعادلة (2.40).

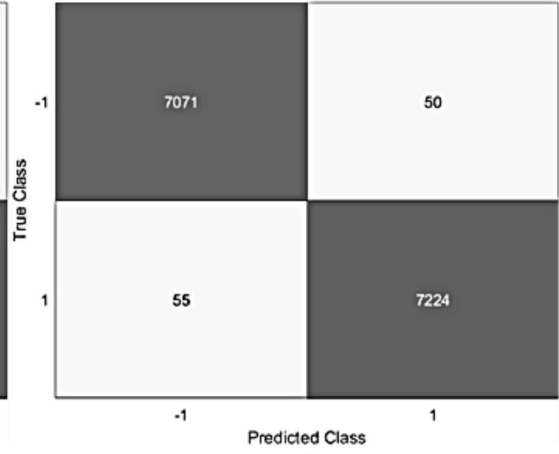
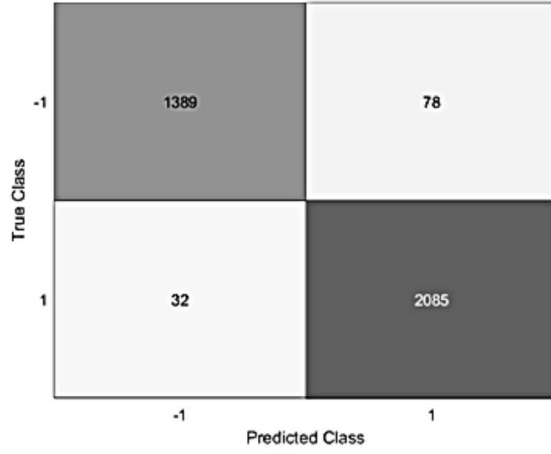
بعد قياس دقة تصنيف آلة المتجه الداعم باستخدام عدة أنواع من دوال النواة تم الحصول على نتائج دقة التصنيف للسلاطات الخمسة من العينة الأولى كما في الجدول 3. 23 في أدناه اذ أنه يعرض نوع دالة النواة المستخدمة (RBF, Linear, Polynomial) في تطبيق طريقة آلة المتجه الداعم، وتمثل القيم الموجودة نسبة دقة التصنيف الثنائي عند تطبيق الأنموذج على بيانات التدريب والاختبار للسلاطات الخمس.

الجدول 3. 23: نتائج أداء أنموذج SVM لسلاسل العينة الأولى

Polynomial	Linear	RBF	العينة الأولى	
99.188	99.222	99.271	بيانات التدريب	السلسلة الأولى
97.182	97.461	96.931	بيانات الاختبار	
97.16	97.132	97.132	بيانات التدريب	السلسلة الثانية
97.349	97.349	97.294	بيانات الاختبار	
98.458	98.375	98.604	بيانات التدريب	السلسلة الثالثة
99.637	99.637	99.609	بيانات الاختبار	
99.007	98.819	99.021	بيانات التدريب	السلسلة الرابعة
98.298	98.382	98.382	بيانات الاختبار	
99.431	99.424	99.444	بيانات التدريب	السلسلة الخامسة
99.554	99.609	99.554	بيانات الاختبار	

ولغرض مطابقة نتائج التصنيف التي تم الحصول عليها في الجدول 3. 23 طُبِّقَت دالة مصفوفة الارتباك، وقد طُبِّقَت الدالة (confusionchart) على بيانات التدريب والاختبار للعينة الأولى من متغيرات السلاسل الزمنية كما في الشكل 3. 17 في أدناه، وقد تضمَّن كل شكل من هذه الاشكال نتائج تطبيق دالة مصفوفة الارتباك على بيانات التدريب A لدوال النواة التي تم استخدامها وكل دالة تتضمن صفين وعمودين حيث ان المكون (أول صف وأول عمود) يمثل عدد المشاهدات التي صنفت بشكل صحيح على انها إيجابية ويعبر عنها في المصفوفة (TP) والمكون (ثاني صف وثاني عمود) يمثل عدد المشاهدات التي صنفت بشكل صحيح على انها سلبية وهي (TN) والمكون (ثاني صف وأول عمود) يمثل عدد المشاهدات التي صنفت بشكل غير صحيح على انها سلبية وقد كانت في الواقع إيجابية وهي (FN) والمكون (أول صف وثاني عمود) يمثل عدد المشاهدات التي صنفت بشكل غير صحيح على انها إيجابية وهي في الواقع سلبية وهي (FP) وبيانات الاختبار B التي تتضمن أربعة مكونات ايضا.

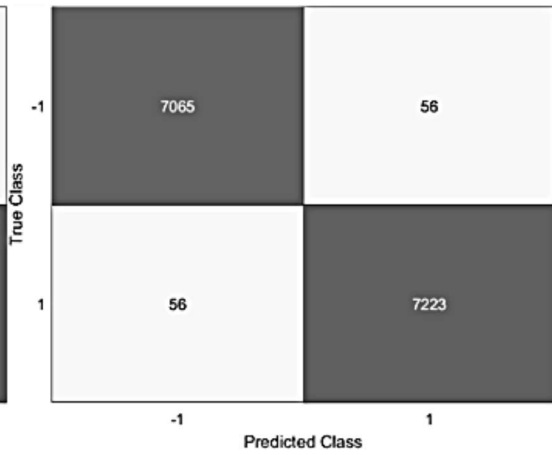
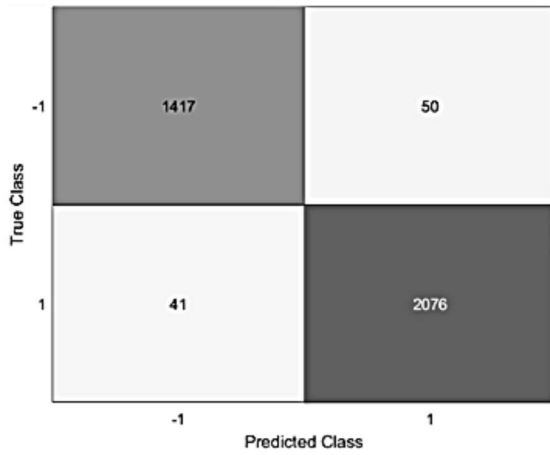
RBF



B

A

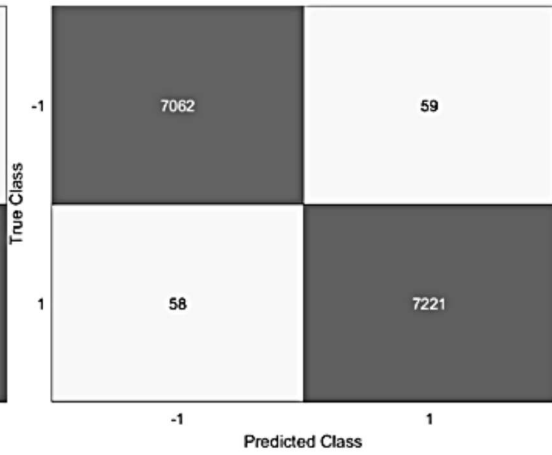
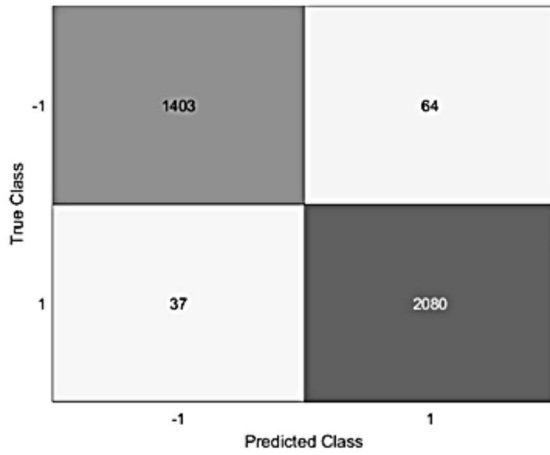
Linear



B

A

Polynomial



B

A

الشكل 3. 17: مصفوفة الارتباك لانموذج SVM لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الاولى)

ومن خلال الجدول 3. 23 في أعلاه وكذلك مصفوفة الارتباك في الشكل 3. 17 في أعلاه يمكن استنتاج ان الأنموذج يعمل بصورة جيدة على دوال النواة المستخدمة جميعاً فقد كانت نسبة الدقة متقاربة جدا بين الدوال المستخدمة جميعاً فكانت دقة التصنيف باستخدام الدوال كلها لبيانات التدريب محصورة بين 97% الى 99% تقريبا إلا نتائج الدقة لبيانات الاختبار باستخدام دالة RBF فقد كانت محصورة بين 96% و 99% أما نتائج الدقة لبيانات الاختبار باستخدام دالتي Linear و Polynomial فقد كانت أدق بدرجة واحدة حسب في العينة الأولى من السلالة الأولى فكانت محصورة بين 97% و 99% وهو فرق بسيط جدا لا يكاد يُذكر. أما نتائج دقة التصنيف للسلالات الخمسة من العينة الثانية فقد كانت كما في الجدول 3. 24 في أدناه، الذي يعرض نوع دالة النواة المستخدمة في تطبيق أسلوب آلة المتجه الداعم، وتمثل القيم الموجودة نسبة دقة التصنيف الثنائي عند تطبيق الأنموذج على بيانات التدريب والاختبار للسلالات الخمس.

**الجدول 3. 24: نتائج أداء أنموذج SVM لسلالات العينة الثانية**

Polynomial	Linear	RBF	العينة الثانية	
99.222	99.208	99.208	بيانات التدريب	السلالة الأولى
98.298	98.326	98.242	بيانات الاختبار	
98.042	97.938	97.951	بيانات التدريب	السلالة الثانية
98.103	98.186	98.159	بيانات الاختبار	
99.208	99.083	99.188	بيانات التدريب	السلالة الثالثة
98.884	98.689	98.912	بيانات الاختبار	
99.09	99.014	99.076	بيانات التدريب	السلالة الرابعة
99.219	99.191	99.135	بيانات الاختبار	
99.292	99.285	99.319	بيانات التدريب	السلالة الخامسة
99.582	99.554	99.637	بيانات الاختبار	

ومن خلال الجدول 3. 24 في أعلاه يمكن استنتاج ان الأنموذج يعمل بصورة جيدة على دوال النواة المستخدمة جميعاً، فقد كانت نسبة الدقة متقاربة جدا بين الدوال المستخدمة جميعاً فكانت دقة التصنيف باستخدام الدوال كلها لبيانات التدريب محصورة بين 97% الى 99% تقريبا وكانت بيانات الاختبار محصورة بين 98% و 99% للدوال المستخدمة كلها. وفي ادناه الجدول 3. 25 الذي يتضمن الوقت الإجمالي لمراحل عملية التدريب كلها ولكل نواة مستخدمة في أنموذج SVM.

**الجدول 3. 25: الوقت المستغرق للخطوات والمراحل جميعاً لعملية تدريب أنموذج SVM.**

الوقت الاجمالي بالثواني			السلالة	العينة
Polynomial	Linear	RBF		
7236.6713	3894.1786	289.1281	الاولى	الاولى
8678.0949	3693.7996	800.9617	الثانية	
6784.0285	8700.3261	617.998	الثالثة	
4874.0224	11478.0307	497.3169	الرابعة	
5618.2476	7023.8522	457.5835	الخامسة	
5360.6439	5982.6252	264.9977	الاولى	الثانية
10994.6678	5671.9683	617.8609	الثانية	
5265.9835	8378.881	246.141	الثالثة	
4750.1375	7618.1478	233.3545	الرابعة	
4770.6209	2469.67	409.7457	الخامسة	

ومن خلال الجدول 3. 25 يتبين أن هناك فرقاً في الوقت المستغرق في تدريب الأنموذج بين دوال النواة المستخدمة وقد كانت دالة نواة RBF هي الأفضل؛ نظراً لأن الوقت المستغرق للخطوات والمراحل جميعاً لعملية تدريب الأنموذج هو الأقل في السلالات كلها.

## 8.3 K الجار الأقرب (K-Nearest Neighbor)

إن الاطار العام لطريقة KNN يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي:

1. استخدام متغيرات الانحدار الذاتي الأمثل اعتمادا على الجدول 3. 4 لتحديد متغيرات الادخال لاسلوب .KNN
2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.
3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والاختبار.
4. تحديد مقياس المسافة وعدد الجيران الاقرب.
5. تدريب أنموذج KNN بواسطة اليعاز (fitcknn) باستخدام المدخلات، وتتضمن متغيرات الادخال والإخراج وعدد الجيران.
6. استخدام الأنموذج الذي تم تدريبه لتصنيف بيانات الاختبار باستخدام الدالة (predict) في المدخلات، وتتضمن متغيرات الادخال فضلاً عن الأنموذج KNN.
7. قياس دقة أنموذج التصنيف بتطبيق المعادلة (2.40).

بعد قياس دقة أنموذج KNN باستخدام عدة أنواع من مقاييس المسافة تم الحصول على نتائج دقة التصنيف للسلاطات الخمس كما في الجدول 3. 26 و الجدول 3. 27 في أدناه، التي تتضمن نتائج أداء أنموذج KNN بحسب مقياس المسافة المستخدم في تطبيق أسلوب الجار الأقرب. وتمثل القيم الموجودة نسبة دقة التصنيف الثنائي عند تطبيق الأنموذج على بيانات التدريب والاختبار للسلاطات الخمس وللعينتين الأولى والثانية.

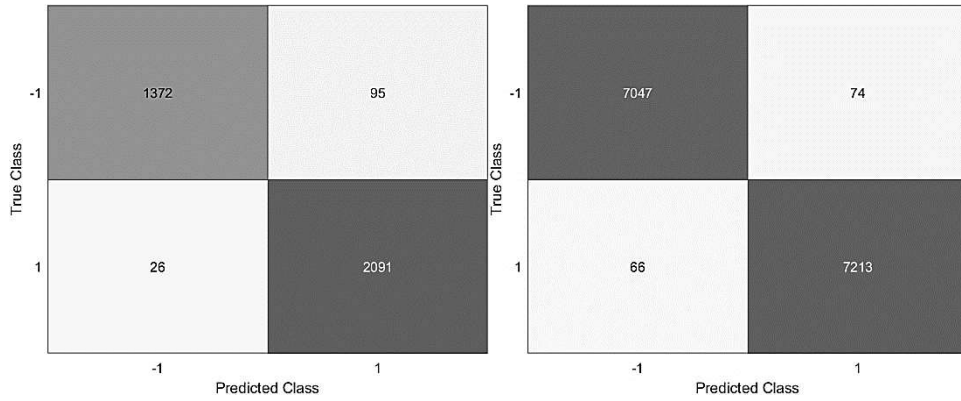
الجدول 3. 26: نتائج أداء أنموذج KNN لسلاطات العينة الأولى

Correlation	Cosine	Jaccard	Hamming	CityBlock	Euclidean	Chebychev	العينة الأولى	
65.931	84.104	96.035	96.035	98.944	99.014	99.028	بيانات التدريب	السلاطة الأولى
63.309	71.345	96.68	96.68	96.791	96.68	96.624	بيانات الاختبار	
73.188	82.035	87.451	87.451	96.5	96.583	96.688	بيانات التدريب	السلاطة الثانية
63.616	77.372	91.63	91.63	96.931	97.042	96.959	بيانات الاختبار	
62.424	76.153	97.167	97.167	98.236	98.257	98.326	بيانات التدريب	السلاطة الثالثة
26.284	70.201	91.323	91.323	99.526	99.526	99.47	بيانات الاختبار	
61.66	76.66	97.382	97.382	98.66	98.729	98.708	بيانات التدريب	السلاطة الرابعة
58.789	81.334	97.907	97.907	98.242	98.27	98.214	بيانات الاختبار	
58.181	76.271	97.826	97.826	98.861	98.882	98.847	بيانات التدريب	السلاطة الخامسة
51.535	74.554	98.438	98.438	99.135	99.219	99.051	بيانات الاختبار	

الجدول 3. 27: نتائج أداء أنموذج KNN لسلاسل العينة الثانية

Correlation	Cosine	Jaccard	Hamming	CityBlock	Euclidean	Chebychev	العينة الثانية	
65.076	83.722	96.222	96.222	99	99.056	99.104	بيانات التدريب	السلاطة الأولى
56.473	78.599	95.871	95.871	98.438	98.549	98.465	بيانات الاختبار	
68.347	81.264	92.674	92.674	97.569	97.694	97.806	بيانات التدريب	السلاطة الثانية
71.903	84.18	91.239	91.239	97.517	97.684	97.824	بيانات الاختبار	
58.868	82.813	97.674	97.674	98.764	98.792	98.785	بيانات التدريب	السلاطة الثالثة
55.134	82.952	96.987	96.987	98.521	98.577	98.661	بيانات الاختبار	
61.972	81.222	97.174	97.174	98.701	98.778	98.813	بيانات التدريب	السلاطة الرابعة
59.542	79.381	97.852	97.852	99.051	99.079	99.051	بيانات الاختبار	
59.611	68.556	97.493	97.493	99.292	99.285	99.285	بيانات التدريب	السلاطة الخامسة
13.923	40.151	68.666	68.666	99.582	99.498	99.526	بيانات الاختبار	

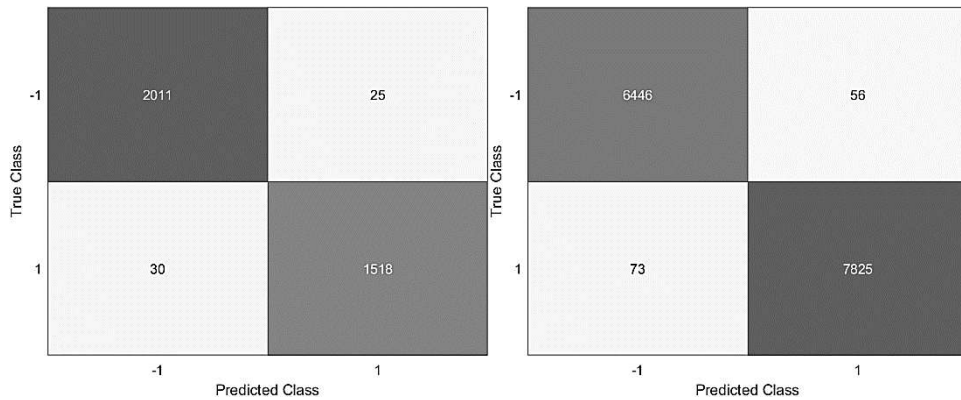
طبقت دالة مصفوفة الارتباك (confusionchart) على بيانات التدريب والاختبار للعيينة الأولى والثانية من متغيرات السلاسل الزمنية ولجميع السلاسل كما في الشكل 18. 3 و الشكل 19. 3 في ادناه.



B

A

الشكل 18. 3: مصفوفة الارتباك لانموذج KNN لبيانات A: التدريب وB: الاختبار (العيينة الاولى من السلسلة الاولى)



B

A

الشكل 19. 3: مصفوفة الارتباك لانموذج KNN لبيانات A: التدريب وB: الاختبار (العيينة الثانية من السلسلة الاولى)

ومن خلال الجدول 26. 3 و الجدول 27. 3 في أعلاه يتضح ان هناك اختلافاً بين نتائج التصنيف في طريقة KNN بحسب مقياس المسافة المستخدم وكذلك البيانات المستخدمة، وعليه يجب استخدام المقياس المناسب الذي يتلاءم مع البيانات المستخدمة لتحقيق افضل النتائج. وتمكن ملاحظة ان مقياس المسافة (Chebychev) هو الذي يحقق افضل أداء بالمقارنة بالمقاييس الأخرى. وتمكن كذلك ملاحظة ان المقياس (Euclidean) له نتائج مشابهة لنتائج المقياس (Chebychev) في عدة سلاسل.

### 9.3 طريقة النواة الغاوسية (GK) Gaussian Kernel Method

ان الاطار العام لطريقة GK يتضمن تنفيذ عدة خطوات متسلسلة وكما يأتي:

1. استخدام متغيرات الانحدار الذاتي الأمثل اعتمادا على الجدول 3. 4 لتحديد متغيرات الادخال لاسلوب GK.

2. تحديد الفئتين الإيجابية والسلبية لمتغير الهدف بصفتين للزوايا الحادة والمنفرجة.

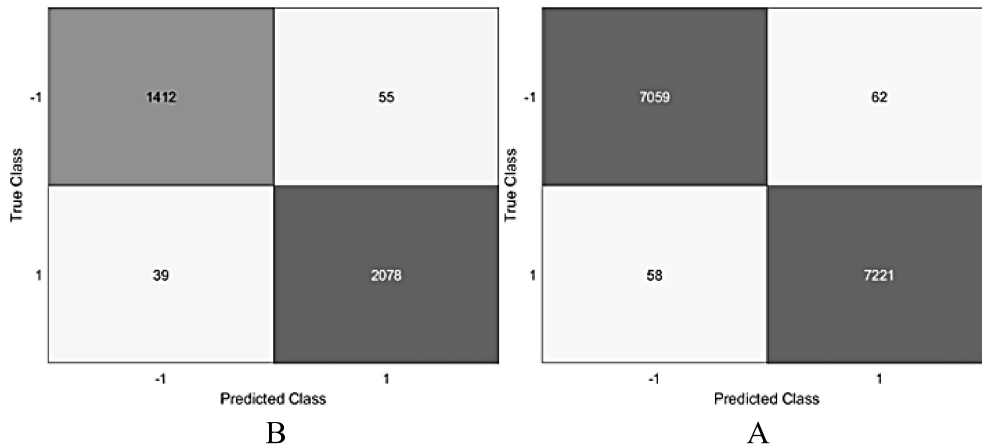
3. تقسيم مشاهدات السلسلة الزمنية على مجموعتين للتدريب والاختبار.

4. تدريب أنموذج GK بواسطة الابعاز (fitkernel) باستخدام المدخلات، وتتضمن متغيرات الادخال والإخراج.

5. استخدام الأنموذج الذي تم تدريبه لتصنيف بيانات الاختبار بواسطة الابعاز (predict) باستخدام المدخلات، وتتضمن متغيرات الادخال فضلاً عن الأنموذج GK.

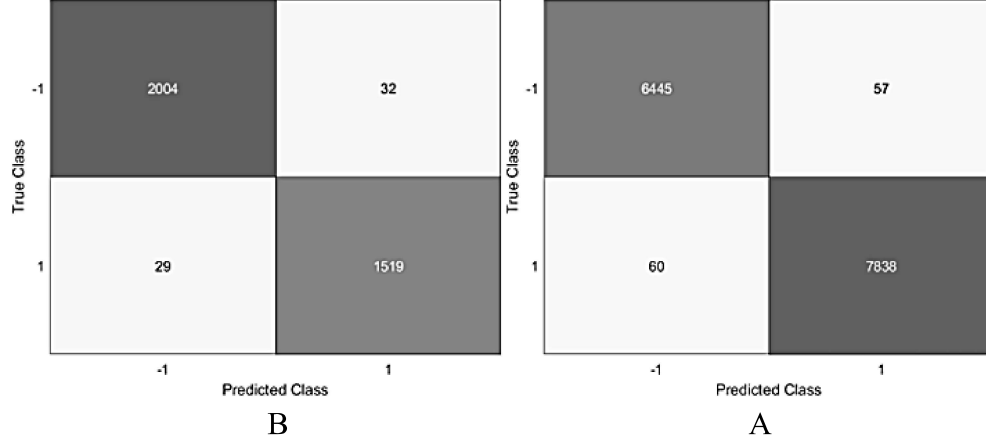
6. قياس دقة أنموذج التصنيف بتطبيق المعادلة (2.40).

ولغرض تصور أداء الأنموذج من خلال المقارنة بين نتائج التصنيف المتوقعة لصور زوايا حركة دودة الربداء الرشيقة والتصنيف الحقيقي، وذلك بتطبيق دالة مصفوفة الارتباك، فقد طُبِّقَت الدالة (confusionchart) على بيانات التدريب والاختبار لمتغيرات السلاسل الزمنية كما في الشكل 3. 20 و الشكل 3. 21 في أدناه وتضمّن كل شكل من هذه الاشكال نتائج تطبيق دالة مصفوفة الارتباك على بيانات التدريب A وبيانات الاختبار B.



الشكل 3. 20: مصفوفة الارتباك لانموذج GK لبيانات A: التدريب وB: الاختبار (العينة الاولى من

السلاسل الاولى)



الشكل 3. 21: مصفوفة الارتباك لانموذج GK لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الاولى)

وكانت نتائج قياس دقة أنموذج التصنيف لبيانات التدريب والاختبار كما في الجدول 3. 28 في أدناه.

الجدول 3. 28: دقة التصنيف للسلالات الخمس لبيانات التدريب والاختبار باستخدام اسلوب GK.

السلالة الخامسة	السلالة الرابعة	السلالة الثالثة	السلالة الثانية	السلالة الأولى		
99.4236	98.9306	98.5833	97.1806	99.1667	بيانات التدريب	العينة الأولى
99.5815	98.4096	99.6373	97.2377	97.3772	بيانات الاختبار	
68.5208	99.1319	99.1250	97.9792	99.1875	بيانات التدريب	العينة الثانية
35.9933	99.2467	98.8281	98.1585	98.2980	بيانات الاختبار	

من خلال الجدول 3. 28 في أعلاه يتبين أن أداء الأنموذج في مرحلة التدريب قد حقق نتائج جيدة في السلالات جميعاً بنسبة تتراوح بين ما يقرب من 97% و 99%، ولكنها كانت في العينة الثانية للسلالة الخامسة اقل دقة بنسبة 68% تقريبا. أما في مرحلة الاختبار فقد كان أداء الأنموذج جيدا في السلالات جميعاً ما عدا العينة الثانية من السلالة الخامسة فقد كانت نسبة الدقة ضعيفة.

استخدمت في هذه الرسالة ستة طرائق تصنيف احداها تختص بتصنيف الصور وهي طريقة CNN وقد استخدمت بيانات CE عن طريق تحويل مشاهداتها الى صور والتي تمثل متغير الادخال، وبالنسبة لطرائق

التصنيف (PNN, LR, SVM, KNN, GK) تم استخدامها في حالة البيانات الرقمية عند استخدام AR أساسا لتحديد عدد المتغيرات الداخلة اليها، و الجدول 3. 29 في ادناه يقدم نتائج أداء طرائق التصنيف التي تم تطبيقها على متغيرات السلاسل الزمنية للعيينة الأولى والعيينة الثانية وللسلاسل الخمس لدودة الريداء الرشيقة حيث ان النتائج مقسمة الى تدريب واختبار لكل طريقة والأرقام المستخدمة هي نسبة دقة التصنيف في كل مرحلة (تدريب واختبار)، ويتبين من خلال هذه الأرقام ان التصنيف الصوري يحقق نتائج دقة ممتازة للعينتين الأولى والثانية في مرحلتي التدريب والاختبار وهو الأفضل من بين جميع الطرائق، وكذلك طريقتي SVM و KNN أيضا حققتا نتائج ممتازة في حالة البيانات الرقمية، وبعد ذلك دقة التصنيف لطريقة GK، وأخيرا كانت نتائج دقة التصنيف هي الأقل بالنسبة لطريقتي PNN و LR حيث حققتا نفس النتائج.

الجدول 3. 29: نتائج الدقة لجميع طرائق التصنيف على بيانات التدريب والاختبار ولجميع السلالات.

العينة الثانية						العينة الأولى						
PNN	LR	GK	KNN	SVM	CNN	PNN	LR	GK	KNN	SVM	CNN	السلالة الأولى
54.84	54.84	99.18	99.10	99.20	99.62	50.54	50.54	99.16	99.02	99.27	99.80	التدريب
43.19	43.19	98.29	98.46	98.24	99.22	59.06	59.06	97.37	96.62	96.93	99.58	الاختبار
PNN	LR	GK	KNN	SVM	CNN	PNN	LR	GK	KNN	SVM	CNN	السلالة الثانية
53.34	53.34	97.97	97.80	97.95	99.67	52.11	52.11	97.18	96.68	97.13	99.56	التدريب
52.28	52.28	98.15	97.82	98.15	99.78	53.26	53.26	97.23	96.95	97.29	99.72	الاختبار
PNN	LR	GK	KNN	SVM	CNN	PNN	LR	GK	KNN	SVM	CNN	السلالة الثالثة
41.13	41.13	99.12	98.78	99.18	99.74	59.36	59.36	98.58	98.32	98.60	99.74	التدريب
55.35	55.35	98.82	98.66	98.91	99.67	13.81	13.81	99.63	99.47	99.60	99.92	الاختبار
PNN	LR	GK	KNN	SVM	CNN	PNN	LR	GK	KNN	SVM	CNN	السلالة الرابعة
48.43	48.43	99.13	98.81	99.07	99.53	48.91	48.91	98.93	98.70	99.02	99.6	التدريب
49.21	49.21	99.24	99.05	99.13	99.64	36.13	36.13	98.40	98.21	98.38	99.69	الاختبار
PNN	LR	GK	KNN	SVM	CNN	PNN	LR	GK	KNN	SVM	CNN	السلالة الخامسة
40.38	40.38	68.52	99.28	99.31	99.55	46.88	46.88	99.42	98.84	99.44	99.78	التدريب
86.07	86.07	35.99	99.52	99.63	99.83	56.16	56.16	99.58	99.05	99.55	99.83	الاختبار

## الفصل الرابع

### الاستنتاجات

#### 1.4 الاستنتاجات Conclusions

أُستخدِمَ في هذه الدراسة الأسلوب CNN الذي يختص دون غيره بتصنيف الصور بوصفه أسلوباً مقترحاً لتحسين نتائج دقة التصنيف لبيانات السلسلة الزمنية لدودة اليربداء الرشيقية CE، وذلك عن طريق تحويل مشاهدات السلسلة الزمنية الى صور. وقد أُستخدِمَت عينتان من البيانات تحتوي كل منهما على خمس سلاطات، كما أُستخدِمَ معيار دقة التصنيف لبيان جودة التصنيف. وقد أوضحت النتائج تفوق CNN على الطرائق المستخدمة جميعاً، التي تم تدريبه عليها، وبدا أن أدائه كان ثابتاً ومتجانساً على بيانات الاختبار. وتبين كذلك تفوق اسلوبي SVM و KNN بوصف كل منهما اسلوباً تصنيفياً في حالة البيانات الرقمية عند استخدام AR أساساً لتحديد عدد المتغيرات الداخلة الى SVM و KNN ثم طريقة النواة GK وبعده طريقتا LR و PNN وقد أمكننا استنتاج ما يأتي:

1. يمكن التخلص من عدم التجانس في بيانات السلاسل الزمنية التي تحتوي على عدد مشاهدات كبير جداً، من خلال تشكيل هذه البيانات وتحويلها الى صور؛ ليسهل التعامل معها لأغراض التصنيف أو التنبؤ، باستخدام الطرائق التي تتعامل مع الصور، ومنها الشبكة العصبية الالتغافية؛ وذلك للحصول على دقة تصنيف مثلى، وكذلك إمكانية استخدام SVM و KNN مع هذا النوع من بيانات السلاسل الزمنية في حالتها الرقمية قبل التحويل الى صور والحصول على دقة تصنيف عالية.
2. يجب اختيار معدل التعلم المناسب في أسلوب CNN لان له تأثيراً كبيراً في تقليل الخسارة وزيادة دقة التصنيف، وذلك من خلال تقليل معدل التعلم.
3. يجب مراعاة الوقت عند اختيار دالة النواة في SVM اذ ان نتائج دقة التصنيف كانت متقاربة لدوال النواة التي تم استخدامها، إلا ان الوقت المستغرق لدالة نواة RBF كان الأقل خلال فترة التدريب.

## 2.4 التوصيات Recommendation:

1. يوصي الباحث بدراسة السلاسل الزمنية متعددة المتغيرات ذات المشاهدات العالية الأبعاد، استكمالاً لما درسناه، وذلك من خلال السلاسل الزمنية أحادية المتغير ذات الأبعاد العالية.
2. يوصي الباحث بدراسة السلاسل الزمنية عالية الأبعاد، القابلة للتحويل الى صور ملونة. أي عندما يكون للبعد الثالث من متغيرات الإدخال أهمية تتحقق من خلال العمق اللوني للصور، واستكمالاً لهذه الدراسة التي عنيت بأن تكون الصور بتدرج رمادي حسب.

## References المصادر

1. Alkema, M. J., Hunter-Ensor, M., Ringstad, N., & Horvitz, H. R. (2005). Tyramine functions independently of octopamine in the *Caenorhabditis elegans* nervous system. *Neuron*, 46(2), 247-260 .
2. Bilbao, A., Patel, A. K., Rahman, M., Vanapalli, S. A & ,Blawdziewicz, J. (2018). Roll maneuvers are essential for active reorientation of *Caenorhabditis elegans* in 3D media. *Proceedings of the National Academy of Sciences*, 115(16), E3616-E3625 .
3. Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods*: Springer science & business media.
4. Chen, W.-W., Tang, W., Hamerton, E. K., Kuo, P. X., Lemieux, G. A., Ashrafi, K., & Cicerone, M. T. (2023). Identifying lipid particle sub-types in live *Caenorhabditis elegans* with two-photon fluorescence lifetime imaging. *Frontiers in Chemistry*, 11 .
5. Demuth, H., Beale, M., & Hagan, M. (1992). Neural network toolbox. *For Use with MATLAB. The MathWorks Inc, 2000* .
6. Du, K.-L., & Swamy, M. N. (2013). *Neural networks and statistical learning*: Springer Science & Business Media.
7. Fawaz, H. I. (2020). Deep learning for time series classification. *arXiv preprint arXiv:2010.00567* .
8. Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). *Deep neural network ensembles for time series classification*. Paper presented at the 2019 International Joint Conference on Neural Networks (IJCNN).
9. Fejfar, J., Šťastný, J., & Cepl, M. (2012). Time series classification using k-Nearest neighbours, Multilayer Perceptron and Learning Vector Quantization algorithms. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 2, 69-72 .
10. Gabrié, M., Ganguli, S., Lucibello, C., & Zecchina, R. (2023). Neural networks: from the perceptron to deep nets. *arXiv preprint arXiv:2304.06636* .
11. Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887* .
12. García Garvía, A., Puchalt, J. C., Layana Castro, P. E., Navarro Moya, F., & Sánchez-Salmerón, A.-J. (2021). Towards lifespan automation for *Caenorhabditis elegans* based on deep learning: analysing convolutional and recurrent neural networks for dead or live classification. *Sensors*, 21(14), 4943 .
13. Glorot, X., & Bengio, Y. (2010). *Understanding the difficulty of training deep feedforward neural networks*. Paper presented at the Proceedings of the thirteenth international conference on artificial intelligence and statistics.
14. Gould, W., Pitblado, J., & Sribney, W. (2006). *Maximum likelihood estimation with Stata*: Stata press.
15. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.
16. Ikirigo, S. (2022). *Classification of caenorhabditis elegans genotypes using locomotory behavioural patterns*. University of Southampton .
17. Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International journal of engineering research and applications*, 3(5), 605-610 .

18. Jade, A. M., Division, C. E. a. P. D., Laboratory, N. C., & India. (2006). Applications of support vector machines to process engineering systems .
19. Janghorbani, A., Arasteh, A., & Moradi, M. H. (2011). *Prediction of acute hypotension episodes using logistic regression model and support vector machine: A comparative study*. Paper presented at the 2011 19th Iranian Conference on Electrical Engineering.
20. Javer, A., Brown, A. E., Kokkinos, I., & Rittscher, J. (2018). *Identification of C. elegans strains using a fully convolutional neural network on behavioural dynamics*. Paper presented at the Proceedings of the European Conference on Computer Vision (ECCV) Workshops.
21. Jiang, W. (2020). Time series classification: Nearest neighbor versus deep learning models . *SN Applied Sciences*, 2(4), 721 .
22. Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 173, 24-49 .
23. Labbé, N., Lee, S.-H., Cho, H.-W., Jeong, M. K., & André, N. (2008). Enhanced discrimination and calibration of biomass NIR spectral data using non-linear kernel methods. *Bioresource technology*, 99(17), 8445-8452 .
24. Le, Q., Sarlós, T., & Smola, A. (2013). *Fastfood-approximating kernel expansions in loglinear time*. Paper presented at the Proceedings of the international conference on machine learning.
25. Liu, H., Guo, M., Xue, T., Guan, J., Luo, L., & Zhuang, Z. (2016). Screening lifespan-extending drugs in *Caenorhabditis elegans* via label propagation on drug-protein networks. *BMC systems biology*, 10(4), 509-519 .
26. Liu, L.-M. (2006). *Time Series Analysis and Forecasting* (2nd ed.). Illinois, USA: Scientific Computing Associates Corp.
27. Luque, A., Carrasco, A., Martín, A., & de Las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216-231 .
28. Maertens, T., Schöll, E., Ruiz, J., & Hövel, P. (2021). Multilayer network analysis of *C. elegans*: Looking into the locomotory circuitry. *Neurocomputing*, 427, 238-261 .
29. Neapolitan, R. E., & Jiang, X. (2018). *Artificial intelligence: With an introduction to machine learning*: CRC Press.
30. Newell Stamper, B. L., Cypser, J. R., Kechris, K., Kitzenberg, D. A., Tedesco, P. M., & Johnson, T. E. (2018). Movement decline across lifespan of *Caenorhabditis elegans* mutants in the insulin/insulin-like signaling pathway. *Aging cell*, 17(1), e12704 .
31. Ng, A. (2000). CS 229Lecture notes. *CS229 Lecture notes*, 1(1), 1-3 .
32. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458* .
33. Prasatha, V., Alfeilate, H. A. A., Hassanate, A., Lasassmehe, O., Tarawnehf, A. S., Alhasanatg, M. B., & Salmane, H. S. E. (2017). Effects of distance measure choice on knn classifier performance-a review. *arXiv preprint arXiv:1708.04321*, 56 .
34. Rithesh, R. (2017). SVM-KNN: a novel approach to classification based on SVM and KNN . *International Research Journal of Computer Science*, 4 .(8)
35. Sagheer, A., & Zidan, M. (2013). Autonomous quantum perceptron neural network. *arXiv preprint arXiv:1312.4149* .

36. Schölkopf, B., Guyon, I., & Weston, J. (2003). *Statistical learning and kernel methods in bioinformatics*. Paper presented at the NATO Advanced Study Institute on Artificial Intelligence and Heuristic Methods in Bioinformatics 2001.
37. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*: Cambridge university press.
38. Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *Towards Data Sci*, 6(12), 310-316 .
39. Smola, A. J., Vishwanathan, S., & Hofmann, T. (2005). *Kernel methods for missing variables*. Paper presented at the International Workshop on Artificial Intelligence and Statistics.
40. Steinbach, M., & Tan, P.-N. (2009). kNN: k-nearest neighbors. *The top ten algorithms in data mining*, 151-162 .
41. Stephens, G. J., Johnson-Kerner, B., Bialek, W., & Ryu, W. S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS computational biology*, 4(4), e1000028 .
42. Stephens, G. J., Johnson-Kerner, B., Bialek, W., & Ryu, W. S. (2010). From modes to movement in the behavior of *Caenorhabditis elegans*. *PloS one*, 5(11), e13914 .
43. Theobald, O. (2017). *Machine learning for absolute beginners: a plain English introduction* (Vol. 157): Scatterplot press.
44. Uhlemann, J., Cawley, O., & Kakouli-Duarte, T. (2020). *Nematode Identification using Artificial Neural Networks*. Paper presented at the DeLTA.
45. Véstias, M. P. (2021). Convolutional neural network *Encyclopedia of Information Science and Technology, Fifth Edition* (pp. 12-26): IGI Global.
46. Wang, J., ZHAO, P., HOI, S. C., Zhuang, J., & Liu, Z.-y. (2013). Large scale online kernel classification .
47. Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E., & Schafer, W. R. (2014). A database of *c. elegans* behavioral phenotypes. *Nature Methods*, 10(9), 877–879 .
48. Zhan, M., Crane, M. M., Entchev, E. V., Caballero, A., Fernandes de Abreu, D. A., Ch'ng, Q., & Lu, H. (2015). Automated processing of imaging data through multi-tiered classification of biological structures illustrated using *Caenorhabditis elegans*. *PLoS computational biology*, 11(4), e1004194 .
49. Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1), 162-169 .

الملاحق

Appendices

## الملحق A

### تحويل درجات زوايا CEM الى أشكال رسومية

```
n=length(x);
k=1;
for k=1:n
    i=x(k);
    str = ['The Corner is ',num2str(i), '°'];
    labels={' ',' '};
    explode=[1 1];
    figure(k), pie([(i/360) (1-
(i/360))],explode,labels),colormap([0 0 0; 0 0 0]),
title(str), print(figure(k),string(k),'-dpng')
    close
    H(k,:)=180+(i*0.5)-i;
    k=k+1;
end
k=1;
for k =1:n
    C(:,:, :,k)=imread(string(k),'png');
end
delete *.png
for k =1:n
    B1=C(:,:, :,k);
    B2= imrotate(B1,H(k,:), 'bilinear', 'crop');
    imshow(B2)
    print(string(k),'-dpng')
    close
end
k=1;
for k =1:n
    D(:,:, :,k)=imread(string(k),'png');
end
delete *.png
for k =1:n
    B3=D(:,:, :,k);
    B4= imcrop(B3,[510 330 250 245]);
    B5 = rgb2gray(B4);
    E(:,:, :,k)=B5;
    imshow(B5)
    print(string(k),'-dpng')
    close
end
```

## B الملحق

### CNN تطبيق

```
numClasses = 2;
layers = [imageInputLayer([246 251 1]);
         convolution2dLayer(3,5) %5 filters of size 3x3
         batchNormalizationLayer
         reluLayer();
         averagePooling2dLayer(2, 'Stride', 2)
         fullyConnectedLayer(numClasses);
         softmaxLayer();
         classificationLayer()];
options = trainingOptions('sgdm', 'MaxEpochs',
1, 'InitialLearnRate', 0.01, 'Plots', 'training-
progress', 'Verbose', true);
[cnn, info] = trainNetwork(ETrain, categorical(RTrain),
layers, options);
predictedLabels = classify(cnn, ETrain);
accuracy = sum(predictedLabels == categorical(RTrain)) /
numel(RTrain);
predictedLabels2 = classify(cnn, ETest);
accuracy2 = sum(predictedLabels2 == categorical(RTest)) /
numel(RTest);
Confusion_Tett_maxCold =
confusionchart(categorical(RTrain), predictedLabels);
Confusion_Tett_maxCold2 =
confusionchart(categorical(RTest), predictedLabels2);
```

## C الملحق

### PNN تطبيق

```
pnn = perceptron;
net = train(pnn, XTrain', YTrain');
YPred_tr = net(XTrain');
YPred_ts = net(XTest');
n_tr=length(YPred_tr);
for i=1:n_tr
if YPred_tr(i)<0
YPred_tr_Bin(i)=-1;
else
YPred_tr_Bin(i)=1;
end
end
n_tr=length(YPred_tr);
for i=1:n_tr
if YTrain(i)==YPred_tr_Bin(i)
Fit_tr(i)=1;
else
Fit_tr(i)=0;
end
end
SM_tr=sum(Fit_tr);
Accr_tr=SM_tr/n_tr*100;
n_ts=length(YPred_ts);
for i=1:n_ts
if YPred_ts(i)<0
YPred_ts_Bin(i)=-1;
else
YPred_ts_Bin(i)=1;
end
end
n_ts=length(YPred_ts);
for i=1:n_ts
if YTest(i)==YPred_ts_Bin(i)
Fit_ts(i)=1;
else
Fit_ts(i)=0;
end
end
SM_ts=sum(Fit_ts);
Accr_ts=SM_ts/n_ts*100;
```

## D الملحق

### LR تطبيق

```
LR=fitglm(XTrain,YTrain,'Distribution','binomial');
LR =
stepwiseglm(XTrain,YTrain,'constant','upper','linear','Distribution','binomial');
OutTrain1 = predict(LR,XTrain);
OutTest1 = predict(LR,XTest);
disp(LR)
n_tr=length(OutTrain1);
for i=1:n_tr
if OutTrain1(i)<0
YPred_tr_Bin(i)=0;
else
YPred_tr_Bin(i)=1;
end
end
n_tr=length(OutTrain1);
for i=1:n_tr
if YTrain(i)==YPred_tr_Bin(i)
Fit_tr(i)=1;
else
Fit_tr(i)=0;
end
end
SM_tr=sum(Fit_tr);
Accr_tr=SM_tr/n_tr*100;
n_ts=length(OutTest1);
for i=1:n_ts
if OutTest1(i)<0
YPred_ts_Bin(i)=0;
else
YPred_ts_Bin(i)=1;
end
end
n_ts=length(OutTest1);
for i=1:n_ts
if YTest(i)==YPred_ts_Bin(i)
Fit_ts(i)=1;
else
Fit_ts(i)=0;
end
end
```

```
end  
SM_ts=sum(Fit_ts);  
Accr_ts=SM_ts/n_ts*100;
```

## E الملحق

### SVM تطبيق

```
svm = fitcsvm(XTrain,YTrain,'KernelFunction','rbf',...
'OptimizeHyperparameters','auto','HyperparameterOptimizatio
nOptions',opts) %linear or polynomial
OutTrain2 = predict(svm,XTrain);
OutTest2 = predict(svm,XTest);
n_tr=length(YTrain);
for i=1:n_tr
if YTrain(i)==OutTrain2(i)
Fit_tr(i)=1;
else
Fit_tr(i)=0;
end
end
SM_tr=sum(Fit_tr);
Accr_tr=SM_tr/n_tr*100
n_ts=length(YTest);
for i=1:n_ts
if YTest(i)==OutTest2(i)
Fit_ts(i)=1;
else
Fit_ts(i)=0;
end
end
SM_ts=sum(Fit_ts);
Accr_ts=SM_ts/n_ts*100
```

## F الملحق

### KNN تطبيق

```
knn =  
fitcknn(XTrain,YTrain,'Distance','cityblock','NumNeighbors'  
,100);  
predtr = predict(knn, XTrain);  
predts = predict(knn, XTest);  
  
n_tr=length(YTrain);  
for i=1:n_tr  
if YTrain(i)==predtr(i)  
Fit_tr(i)=1;  
else  
Fit_tr(i)=0;  
end  
end  
SM_tr=sum(Fit_tr);  
Accr_tr=SM_tr/n_tr*100  
  
n_ts=length(YTest);  
for i=1:n_ts  
if YTest(i)==predts(i)  
Fit_ts(i)=1;  
else  
Fit_ts(i)=0;  
end  
end  
SM_ts=sum(Fit_ts);  
Accr_ts=SM_ts/n_ts*100  
  
Confusion_Tett_maxCold = confusionchart(YTrain,predtr);
```

## G الملحق

### GK تطبيق

```
GK
=fitkernel(XTrain,YTrain,'OptimizeHyperparameters','auto',
...
'HyperparameterOptimizationOptions',struct('AcquisitionFunc
tionName','expected-improvement-plus'))
predtr = predict(GK, XTrain);
predts = predict(GK, XTest);

n_tr=length(YTrain);
for i=1:n_tr
if YTrain(i)==predtr(i)
Fit_tr(i)=1;
else
Fit_tr(i)=0;
end
end
SM_tr=sum(Fit_tr);
Accr_tr=SM_tr/n_tr*100

n_ts=length(YTest);
for i=1:n_ts
if YTest(i)==predts(i)
Fit_ts(i)=1;
else
Fit_ts(i)=0;
end
end
SM_ts=sum(Fit_ts);
Accr_ts=SM_ts/n_ts*100
```

المحقق H

البيانات

1.H. جزء من بيانات العينة الأولى (السلالة الأولى)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	16.61	29	8.31	57	15.39	85	16.75	113	10.72
2	16.90	30	8.32	58	15.62	86	16.39	114	10.57
3	17.38	31	8.44	59	15.73	87	16.05	115	10.49
4	17.35	32	8.48	60	17.08	88	15.73	116	10.59
5	17.68	33	8.53	61	18.83	89	15.41	117	10.61
6	18.07	34	8.67	62	19.82	90	15.11	118	10.80
7	18.44	35	9.06	63	21.09	91	14.82	119	11.26
8	18.84	36	9.15	64	22.65	92	14.54	120	11.37
9	19.34	37	9.29	65	23.85	93	14.27	121	11.57
10	18.97	38	9.43	66	24.04	94	14.01	122	11.86
11	18.25	39	9.48	67	24.08	95	13.76	123	11.96
12	17.34	40	9.71	68	25.20	96	13.52	124	12.31
13	16.47	41	10.32	69	24.69	97	12.77	125	12.69
14	15.60	42	10.53	70	23.50	98	12.23	126	13.07
15	15.03	43	11.02	71	23.27	99	11.54	127	13.39
16	14.17	44	10.99	72	22.94	100	11.41	128	13.76
17	13.09	45	11.32	73	23.54	101	11.53	129	13.91
18	12.20	46	11.56	74	22.83	102	11.71	130	13.95
19	11.25	47	12.20	75	22.86	103	11.79	131	13.72
20	10.66	48	12.71	76	22.31	104	11.82	132	13.49
21	10.06	49	13.02	77	21.29	105	11.85	133	13.26
22	9.58	50	13.27	78	20.64	106	11.85	134	13.05
23	9.02	51	13.73	79	20.32	107	11.82	135	12.84
24	8.63	52	14.99	80	19.47	108	11.70	136	12.76
25	8.54	53	14.61	81	19.00	109	11.48	137	12.37
26	8.31	54	15.13	82	18.66	110	11.44	138	12.05
27	8.34	55	15.28	83	17.96	111	10.92	139	11.89
28	8.32	56	15.50	84	17.91	112	10.78	140	11.91

2.H. جزء من بيانات العينة الأولى (السلالة الثانية)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	147.43	29	109.09	57	63.68	85	38.76	113	49.87
2	146.81	30	105.14	58	62.84	86	40.84	114	51.09
3	146.18	31	105.05	59	59.66	87	36.60	115	55.51
4	145.52	32	104.91	60	59.64	88	30.19	116	56.36
5	144.85	33	105.91	61	57.87	89	38.00	117	57.05
6	144.14	34	102.80	62	54.27	90	39.89	118	58.22
7	143.42	35	100.95	63	53.28	91	40.77	119	58.84
8	142.85	36	99.78	64	50.55	92	46.92	120	51.02
9	141.66	37	98.02	65	45.41	93	47.40	121	47.35
10	140.71	38	97.11	66	37.81	94	47.80	122	45.20
11	138.04	39	95.21	67	34.83	95	41.65	123	45.30
12	137.31	40	94.50	68	32.24	96	43.57	124	39.74
13	135.27	41	93.15	69	29.97	97	47.33	125	36.31
14	133.93	42	90.11	70	27.97	98	42.47	126	38.72
15	132.15	43	89.35	71	26.21	99	38.63	127	37.64
16	132.90	44	87.63	72	24.64	100	35.84	128	35.21
17	133.21	45	86.42	73	23.24	101	35.83	129	32.57
18	131.36	46	85.44	74	21.99	102	33.53	130	31.21
19	130.82	47	80.08	75	20.85	103	30.92	131	39.36
20	129.55	48	77.51	76	25.22	104	33.78	132	41.63
21	126.99	49	75.03	77	29.21	105	36.82	133	43.69
22	124.67	50	72.78	78	34.53	106	40.09	134	45.40
23	123.87	51	70.22	79	31.81	107	43.87	135	47.72
24	120.09	52	68.17	80	31.94	108	49.93	136	47.15
25	118.54	53	67.55	81	33.88	109	47.30	137	37.79
26	117.56	54	66.77	82	35.57	110	40.06	138	32.11
27	115.88	55	65.10	83	37.54	111	43.61	139	30.17
28	113.57	56	64.12	84	39.14	112	42.50	140	29.44

3.H. جزء من بيانات العينة الأولى (السلالة الثالثة)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	60.70	29	47.72	57	90.22	85	131.74	113	142.58
2	64.38	30	51.43	58	95.28	86	128.46	114	143.32
3	60.97	31	50.75	59	98.96	87	125.06	115	139.08
4	60.37	32	59.39	60	103.15	88	125.89	116	137.40
5	68.00	33	64.11	61	106.20	89	124.20	117	136.60
6	65.51	34	72.65	62	107.27	90	123.16	118	135.92
7	62.80	35	78.76	63	108.17	91	123.69	119	135.63
8	68.10	36	69.96	64	111.92	92	120.03	120	130.34
9	72.35	37	68.91	65	113.28	93	120.56	121	133.92
10	69.04	38	71.36	66	114.48	94	120.81	122	132.91
11	67.37	39	67.44	67	118.59	95	120.93	123	131.11
12	70.12	40	67.74	68	123.41	96	121.23	124	128.61
13	66.58	41	66.73	69	125.72	97	124.73	125	131.31
14	66.94	42	70.04	70	128.81	98	126.96	126	133.59
15	65.98	43	72.35	71	128.47	99	126.38	127	138.78
16	64.46	44	70.58	72	130.01	100	127.37	128	137.26
17	57.54	45	69.96	73	131.75	101	130.26	129	136.61
18	51.45	46	71.05	74	132.85	102	131.27	130	141.11
19	47.86	47	68.53	75	134.52	103	133.12	131	139.55
20	42.82	48	64.98	76	132.57	104	134.09	132	140.19
21	38.32	49	62.88	77	142.22	105	135.20	133	138.61
22	40.94	50	65.02	78	143.35	106	135.85	134	141.37
23	43.42	51	60.50	79	140.90	107	137.54	135	138.40
24	40.76	52	72.40	80	141.16	108	139.56	136	132.77
25	41.60	53	77.92	81	138.22	109	140.21	137	133.39
26	45.95	54	80.42	82	135.05	110	141.80	138	138.54
27	42.59	55	83.91	83	135.38	111	140.79	139	138.52
28	48.03	56	88.38	84	131.60	112	142.62	140	129.89

4.H. جزء من بيانات العينة الأولى (السلالة الرابعة)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	169.13	29	171.61	57	171.46	85	170.22	113	167.27
2	169.33	30	171.68	58	171.49	86	170.24	114	167.12
3	169.37	31	171.74	59	171.41	87	170.24	115	166.85
4	169.68	32	171.68	60	171.43	88	170.16	116	166.57
5	169.62	33	171.74	61	171.45	89	169.87	117	166.31
6	169.72	34	171.80	62	171.36	90	169.71	118	166.16
7	169.89	35	171.76	63	171.27	91	169.63	119	165.76
8	170.09	36	171.68	64	171.12	92	169.40	120	165.19
9	170.50	37	171.87	65	171.02	93	169.23	121	165.37
10	170.43	38	171.82	66	170.78	94	169.07	122	164.87
11	170.73	39	171.79	67	170.71	95	168.95	123	164.67
12	170.62	40	171.85	68	170.54	96	168.49	124	164.49
13	170.91	41	171.79	69	170.65	97	168.34	125	164.51
14	171.21	42	171.71	70	170.60	98	168.19	126	164.21
15	171.16	43	171.75	71	170.53	99	168.26	127	164.08
16	171.37	44	171.76	72	170.42	100	168.14	128	164.05
17	171.38	45	171.73	73	170.33	101	168.15	129	164.17
18	171.66	46	171.74	74	170.42	102	168.21	130	164.37
19	171.43	47	171.74	75	170.25	103	168.03	131	164.17
20	171.68	48	171.74	76	170.32	104	168.18	132	164.28
21	171.71	49	171.54	77	170.20	105	168.09	133	164.37
22	171.67	50	171.59	78	170.20	106	168.20	134	164.47
23	171.77	51	171.60	79	170.10	107	168.13	135	164.16
24	171.67	52	171.61	80	170.05	108	168.04	136	164.26
25	171.50	53	171.66	81	170.32	109	167.96	137	164.42
26	171.62	54	171.61	82	170.31	110	167.63	138	164.13
27	171.69	55	171.50	83	170.17	111	167.43	139	164.72
28	171.75	56	171.48	84	170.31	112	167.51	140	164.78

5.H. جزء من بيانات العينة الأولى (السلالة الخامسة)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	56.22	29	45.62	57	65.66	85	108.91	113	115.97
2	57.60	30	45.76	58	65.97	86	109.69	114	117.91
3	55.75	31	46.36	59	65.49	87	110.45	115	119.98
4	53.99	32	47.08	60	66.40	88	113.86	116	120.78
5	52.29	33	47.11	61	67.15	89	112.35	117	121.63
6	49.00	34	48.17	62	68.40	90	111.82	118	122.44
7	48.22	35	49.26	63	69.68	91	113.96	119	123.24
8	49.33	36	50.39	64	70.92	92	116.21	120	123.87
9	50.48	37	51.56	65	72.17	93	118.37	121	125.22
10	48.81	38	52.77	66	73.44	94	120.45	122	125.05
11	46.57	39	54.02	67	72.97	95	118.89	123	124.88
12	49.07	40	55.30	68	72.70	96	118.77	124	124.70
13	45.05	41	56.63	69	76.56	97	118.74	125	127.29
14	44.87	42	54.41	70	76.11	98	118.72	126	125.22
15	44.69	43	55.08	71	75.65	99	118.69	127	123.05
16	43.56	44	55.76	72	75.20	100	120.39	128	122.73
17	45.73	45	56.46	73	79.31	101	119.17	129	117.35
18	44.98	46	56.90	74	81.96	102	117.91	130	118.05
19	44.25	47	56.61	75	84.64	103	114.56	131	118.74
20	43.54	48	63.60	76	87.36	104	115.47	132	119.42
21	43.54	49	62.27	77	92.80	105	116.37	133	114.61
22	43.07	50	64.64	78	96.14	106	115.73	134	114.35
23	42.60	51	66.04	79	102.37	107	117.15	135	114.08
24	42.80	52	65.75	80	102.80	108	114.97	136	113.81
25	42.79	53	65.27	81	102.99	109	112.67	137	113.93
26	42.03	54	64.79	82	104.97	110	114.34	138	114.76
27	45.34	55	64.31	83	106.92	111	114.89	139	114.53
28	45.48	56	65.36	84	108.13	112	115.43	140	114.31

6.H. جزء من بيانات العينة الثانية (السلالة الأولى)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	24.02	29	88.69	57	102.62	85	31.02	113	74.04
2	24.08	30	90.00	58	100.39	86	31.17	114	75.08
3	23.37	31	92.92	59	105.33	87	31.14	115	71.55
4	22.96	32	98.31	60	101.82	88	32.22	116	80.40
5	23.32	33	103.19	61	109.84	89	32.40	117	85.31
6	24.42	34	104.57	62	108.10	90	32.51	118	90.37
7	25.92	35	107.88	63	103.06	91	32.30	119	95.63
8	26.75	36	101.82	64	100.55	92	32.60	120	90.40
9	28.70	37	101.19	65	98.50	93	31.87	121	97.71
10	28.01	38	101.89	66	88.76	94	30.94	122	95.57
11	29.05	39	103.34	67	78.68	95	30.11	123	98.66
12	29.84	40	104.35	68	71.86	96	30.09	124	105.34
13	28.95	41	104.02	69	67.24	97	32.63	125	110.00
14	28.71	42	104.26	70	56.45	98	33.40	126	116.16
15	30.02	43	104.50	71	50.35	99	36.44	127	121.50
16	30.60	44	105.32	72	48.20	100	39.87	128	123.55
17	31.80	45	105.26	73	43.52	101	42.32	129	126.06
18	36.15	46	105.31	74	40.45	102	45.00	130	123.72
19	39.66	47	105.08	75	35.49	103	49.44	131	128.06
20	43.41	48	103.90	76	34.51	104	52.57	132	129.30
21	49.08	49	103.74	77	32.57	105	56.45	133	131.25
22	50.72	50	102.90	78	32.00	106	59.49	134	131.31
23	62.02	51	102.73	79	30.83	107	62.48	135	133.52
24	64.48	52	101.75	80	29.66	108	64.43	136	133.55
25	69.01	53	109.87	81	29.72	109	72.79	137	126.83
26	76.98	54	108.25	82	30.71	110	75.37	138	122.11
27	77.49	55	105.54	83	30.36	111	79.70	139	117.84
28	80.60	56	105.04	84	31.15	112	75.34	140	114.80

7.H. جزء من بيانات العينة الثانية (السلالة الثانية)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	76.67	29	39.48	57	83.11	85	63.86	113	73.23
2	75.99	30	35.46	58	84.29	86	60.56	114	74.16
3	74.76	31	34.54	59	86.25	87	57.47	115	78.59
4	74.85	32	33.73	60	87.17	88	54.57	116	84.59
5	72.07	33	33.05	61	89.08	89	51.87	117	85.41
6	71.31	34	30.88	62	89.99	90	49.35	118	88.03
7	69.60	35	30.56	63	92.90	91	47.01	119	92.83
8	67.19	36	29.60	64	93.81	92	45.87	120	93.82
9	66.98	37	34.45	65	94.72	93	46.69	121	95.86
10	66.00	38	34.60	66	95.63	94	42.50	122	99.83
11	65.23	39	36.09	67	94.55	95	45.36	123	103.34
12	64.83	40	37.69	68	94.46	96	47.07	124	107.66
13	62.76	41	39.42	69	94.37	97	47.55	125	108.07
14	61.83	42	42.47	70	94.28	98	49.29	126	114.24
15	59.29	43	46.11	71	95.19	99	50.16	127	115.45
16	58.56	44	50.26	72	95.10	100	51.12	128	117.84
17	58.42	45	54.97	73	93.01	101	53.15	129	120.09
18	57.47	46	56.21	74	91.82	102	54.04	130	123.01
19	56.68	47	55.67	75	89.12	103	56.57	131	124.98
20	56.01	48	56.27	76	87.84	104	58.89	132	130.29
21	54.39	49	59.92	77	86.03	105	60.65	133	134.86
22	53.81	50	64.42	78	84.94	106	61.05	134	138.60
23	53.70	51	69.00	79	80.63	107	63.03	135	141.49
24	50.13	52	73.65	80	77.79	108	64.43	136	145.62
25	48.82	53	78.54	81	73.08	109	66.09	137	144.20
26	45.22	54	80.59	82	70.51	110	70.80	138	145.13
27	43.64	55	81.47	83	69.13	111	71.55	139	147.16
28	40.59	56	82.74	84	66.18	112	72.36	140	146.98

8.H. جزء من بيانات العينة الثانية (السلالة الثالثة)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	37.82	29	56.18	57	48.68	85	38.46	113	132.41
2	42.26	30	48.43	58	45.46	86	45.57	114	128.64
3	48.62	31	43.11	59	39.10	87	47.47	115	133.38
4	56.90	32	34.37	60	39.06	88	52.44	116	132.70
5	60.41	33	33.13	61	30.92	89	49.60	117	138.82
6	66.88	34	35.54	62	28.72	90	48.48	118	134.85
7	75.03	35	37.13	63	27.56	91	44.71	119	131.03
8	77.53	36	34.97	64	27.29	92	44.18	120	133.89
9	78.46	37	33.72	65	29.96	93	43.68	121	141.01
10	84.21	38	34.14	66	26.82	94	43.19	122	139.12
11	81.82	39	35.14	67	29.15	95	42.59	123	140.67
12	87.66	40	37.34	68	31.11	96	42.46	124	140.12
13	86.37	41	40.37	69	31.44	97	40.84	125	141.55
14	88.10	42	45.64	70	32.79	98	44.30	126	142.33
15	88.47	43	58.83	71	33.96	99	43.47	127	149.53
16	94.81	44	68.15	72	32.99	100	42.75	128	155.84
17	111.06	45	79.06	73	33.25	101	42.65	129	159.35
18	112.15	46	85.82	74	31.73	102	45.09	130	162.72
19	124.34	47	89.81	75	30.66	103	49.54	131	163.04
20	120.53	48	98.30	76	29.98	104	47.48	132	164.02
21	119.63	49	103.56	77	29.33	105	66.63	133	163.77
22	116.16	50	105.29	78	27.87	106	80.26	134	162.26
23	115.81	51	100.66	79	29.27	107	87.32	135	162.53
24	108.84	52	100.31	80	29.38	108	103.92	136	161.83
25	92.96	53	90.76	81	28.52	109	113.61	137	161.63
26	88.87	54	84.39	82	33.09	110	123.40	138	161.11
27	76.69	55	74.90	83	34.04	111	127.44	139	162.53
28	65.62	56	63.18	84	36.59	112	127.53	140	161.88

9.H. جزء من بيانات العينة الثانية (السلالة الرابعة)

Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	159.16	29	166.08	57	162.69	85	156.09	113	168.30
2	158.71	30	166.14	58	161.90	86	156.30	114	168.51
3	158.91	31	166.36	59	160.77	87	157.55	115	168.71
4	158.73	32	166.75	60	159.74	88	158.13	116	168.90
5	158.50	33	167.09	61	158.08	89	157.62	117	169.09
6	159.10	34	167.13	62	156.56	90	158.81	118	168.90
7	158.90	35	167.07	63	154.65	91	158.98	119	169.06
8	159.35	36	167.22	64	153.29	92	159.86	120	169.21
9	159.85	37	167.24	65	152.41	93	159.56	121	169.42
10	160.19	38	167.23	66	153.54	94	159.64	122	169.31
11	159.81	39	167.02	67	153.24	95	159.99	123	169.43
12	159.61	40	166.91	68	153.71	96	160.65	124	169.36
13	160.29	41	166.92	69	152.30	97	160.26	125	169.51
14	160.60	42	166.95	70	152.66	98	161.12	126	169.30
15	161.00	43	167.23	71	152.14	99	162.00	127	169.29
16	160.42	44	166.88	72	152.14	100	162.53	128	169.60
17	160.80	45	166.30	73	151.69	101	162.95	129	169.87
18	160.83	46	166.22	74	150.81	102	163.46	130	169.66
19	160.85	47	166.22	75	150.34	103	163.83	131	169.60
20	161.22	48	166.13	76	150.72	104	164.57	132	168.68
21	161.93	49	165.74	77	151.21	105	165.01	133	168.84
22	162.43	50	165.69	78	151.64	106	165.38	134	169.71
23	162.49	51	164.70	79	151.09	107	165.74	135	168.59
24	163.32	52	165.41	80	151.71	108	166.57	136	169.58
25	164.06	53	164.84	81	152.57	109	167.07	137	169.63
26	164.58	54	164.35	82	153.82	110	167.45	138	169.17
27	165.49	55	164.11	83	153.45	111	167.86	139	169.30
28	165.49	56	163.53	84	155.96	112	168.08	140	168.33

10.H. جزء من بيانات العينة الثانية (السلالة الخامسة)

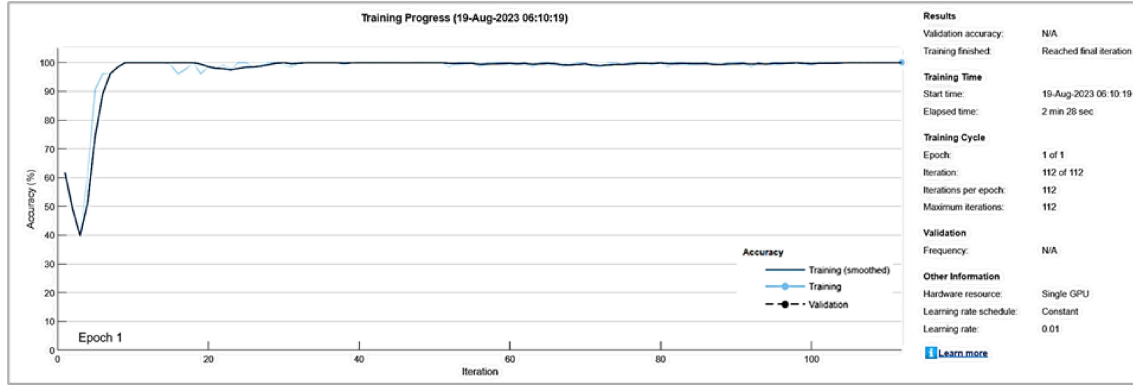
Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)	Frame (time)	CEM (angle)
1	158.74	29	157.69	57	157.69	85	157.84	113	160.20
2	159.12	30	157.80	58	157.96	86	157.35	114	159.43
3	159.48	31	157.91	59	157.60	87	157.96	115	159.68
4	159.84	32	158.02	60	157.90	88	157.81	116	160.01
5	160.18	33	157.37	61	158.20	89	157.67	117	159.43
6	160.38	34	157.38	62	158.49	90	157.53	118	159.39
7	160.62	35	158.03	63	158.08	91	158.08	119	160.05
8	159.60	36	157.53	64	158.05	92	157.42	120	159.39
9	160.07	37	157.82	65	158.43	93	157.82	121	159.90
10	160.20	38	157.15	66	158.85	94	158.03	122	159.71
11	160.65	39	157.10	67	158.51	95	157.57	123	159.73
12	160.68	40	157.68	68	158.41	96	158.96	124	159.75
13	161.01	41	157.38	69	158.75	97	157.98	125	159.76
14	160.11	42	157.28	70	158.73	98	157.99	126	159.27
15	159.82	43	157.56	71	158.44	99	158.32	127	159.62
16	159.72	44	157.83	72	158.14	100	158.60	128	159.53
17	159.61	45	158.09	73	157.83	101	158.25	129	159.45
18	159.51	46	157.40	74	157.76	102	158.84	130	159.36
19	158.99	47	157.19	75	158.18	103	159.65	131	159.55
20	159.00	48	157.19	76	158.24	104	159.36	132	159.03
21	158.48	49	156.75	77	158.35	105	158.76	133	158.75
22	158.39	50	157.69	78	158.18	106	159.21	134	158.64
23	159.49	51	157.38	79	158.11	107	159.64	135	158.55
24	159.16	52	157.69	80	158.11	108	160.05	136	158.83
25	158.82	53	157.99	81	158.10	109	159.16	137	158.70
26	158.47	54	158.28	82	158.09	110	159.15	138	159.34
27	158.77	55	157.57	83	158.57	111	158.77	139	159.20
28	158.24	56	158.17	84	158.19	112	159.75	140	159.16

# المحقق I

مصفوفات الارتباك وسلوك ومعلومات عملية تدريب CNN بمعدل

تعلم 0.01

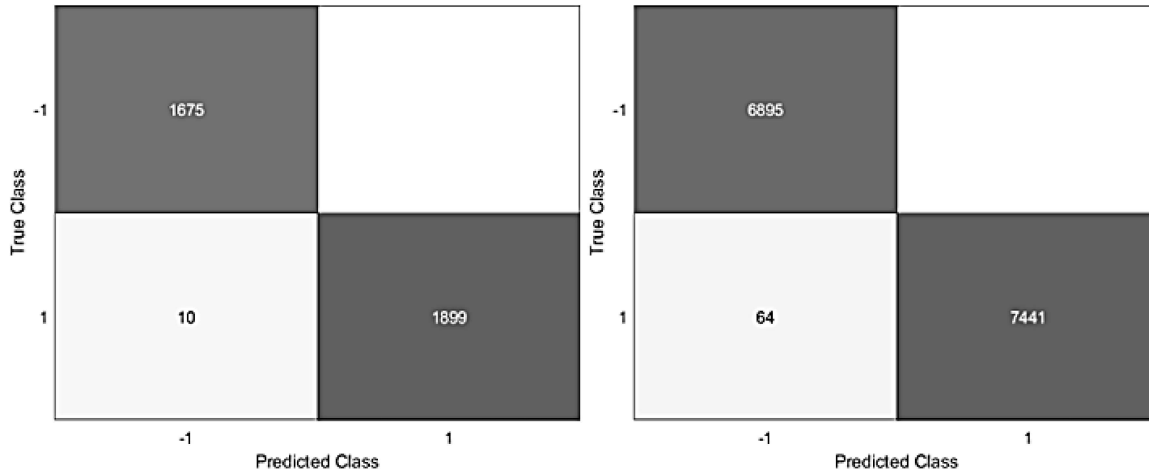
للسلات الثانية والثالثة والرابعة والخامسة



الشكل I . 1: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الثانية بمعدل تعلم 0.01.

الجدول I . 1: معلومات عملية تدريب CNN لبيانات العينة الأولى من السلالة الثانية بمعدل تعلم 0.01

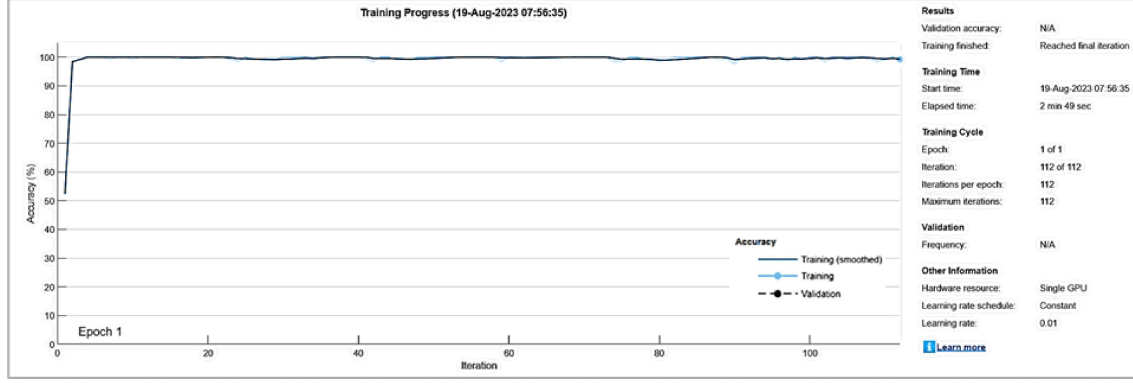
Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 00: 51	61.72%	0.70	0.01
50	00: 01: 37	100.00%	0	0.01
100	00: 02: 17	99.22%	0.01	0.01
112	00: 02: 28	100.00%	0	0.01



B

A

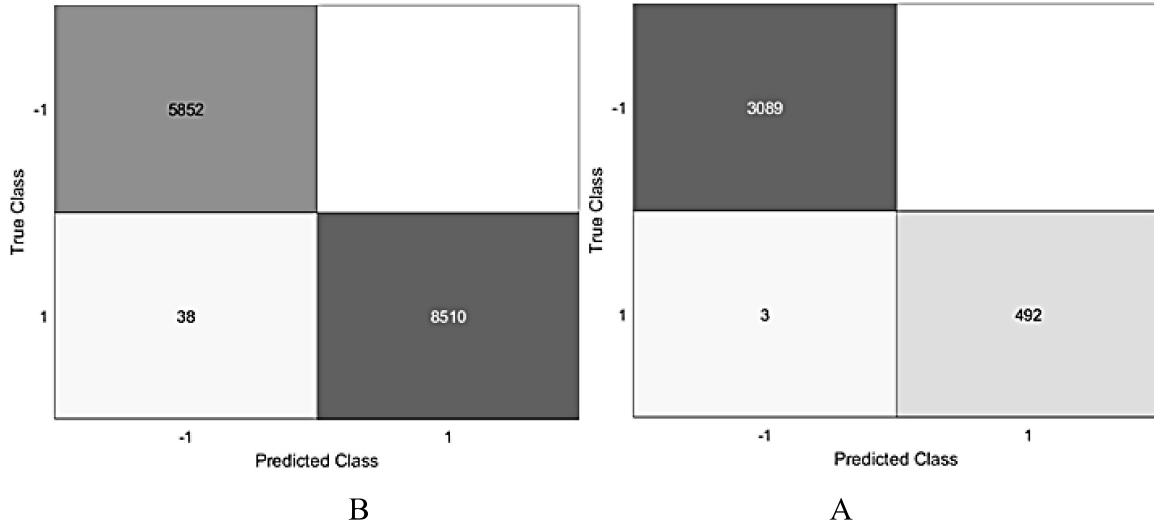
الشكل I . 2: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثانية) بمعدل تعلم 0.01



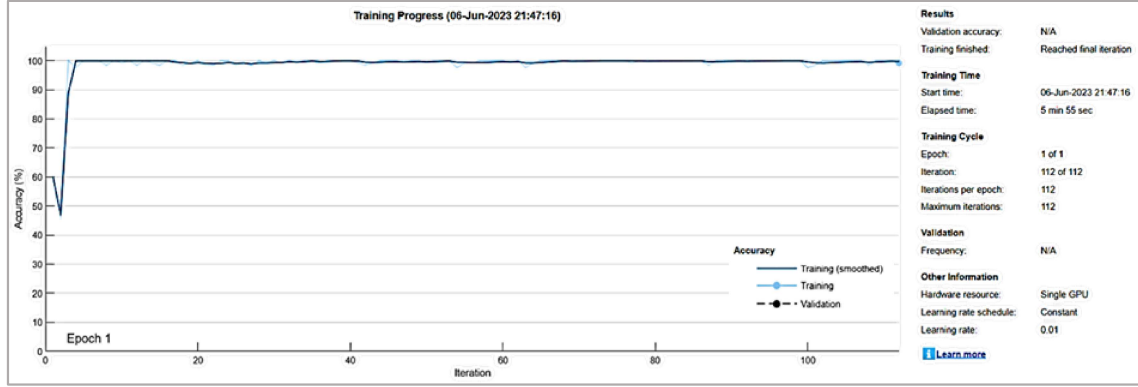
الشكل I 3: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الثالثة بمعدل تعلم 0.01.

الجدول I 2: معلومات عملية تدريب CNN لبيانات العينة الأولى من السلالة الثالثة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 07	52.34%	0.75	0.01
50	00: 01: 47	100.00%	0	0.01
100	00: 02: 35	100.00%	0	0.01
112	00: 02: 49	99.22%	0.12	0.01



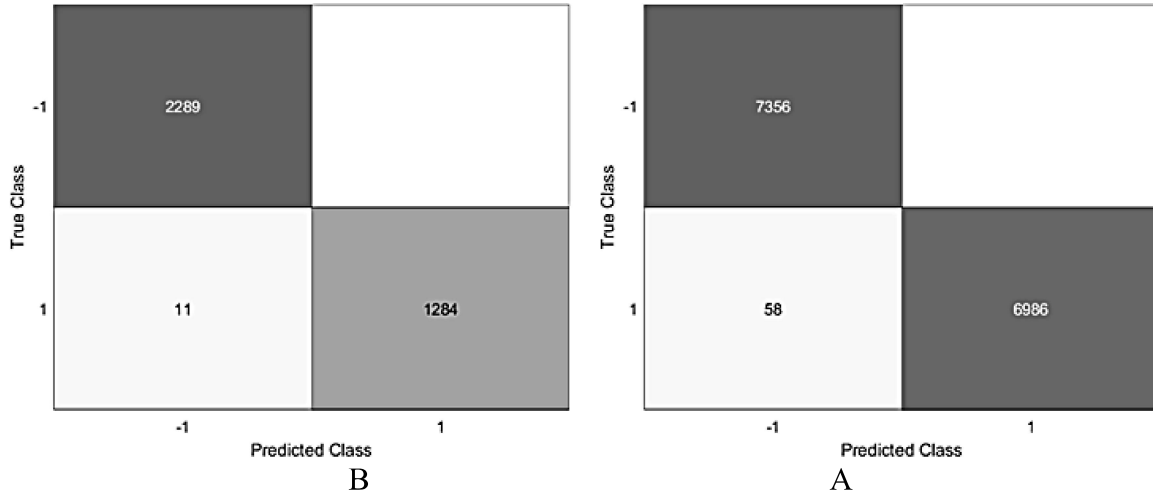
الشكل I 4: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثالثة) بمعدل تعلم 0.01



الشكل I 5: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الرابعة بمعدل تعلم 0.01.

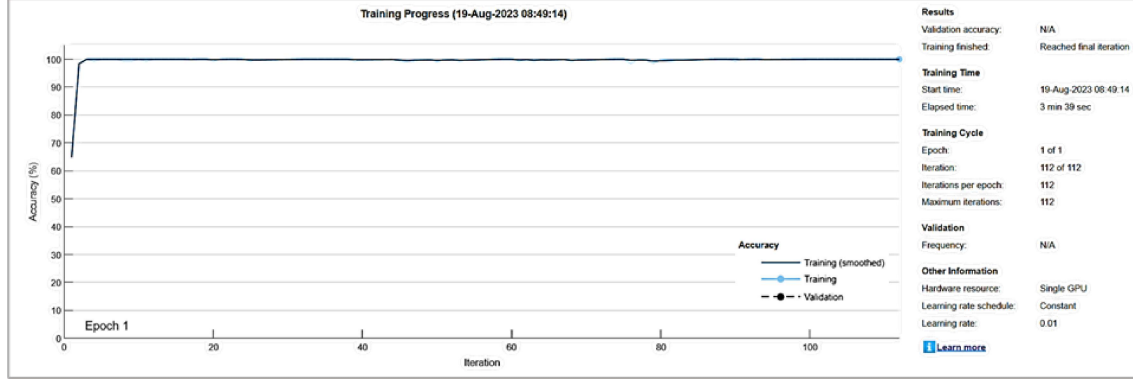
الجدول I 3: معلومات عملية تدريب CNN لبيانات العينة الأولى من السلالة الرابعة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 02: 21	60.16%	0.64	0.01
50	00: 03: 22	99.22%	0.12	0.01
100	00: 05: 27	97.66%	0.37	0.01
112	00: 05: 55	99.22%	0.12	0.01



الشكل I 6: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من

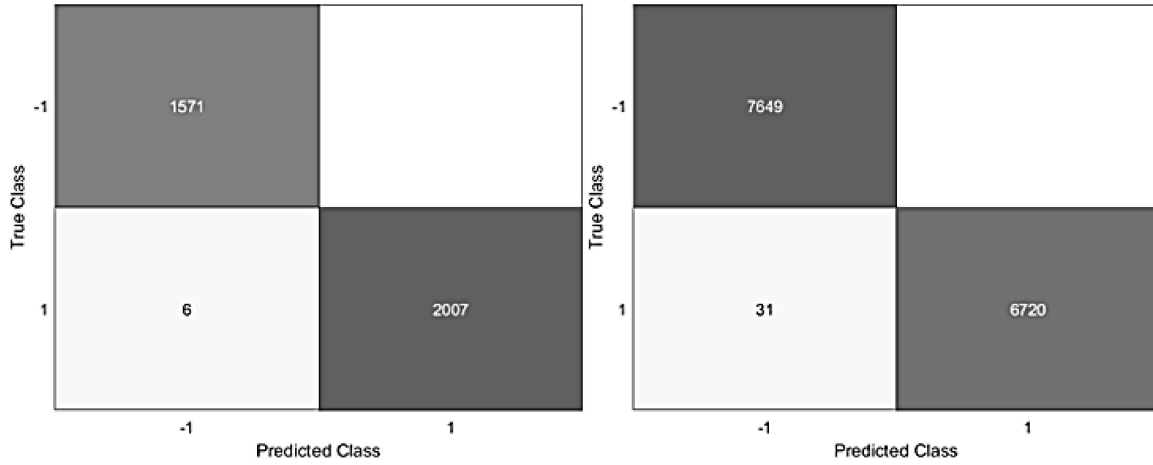
السلالة الرابعة) بمعدل تعلم 0.01



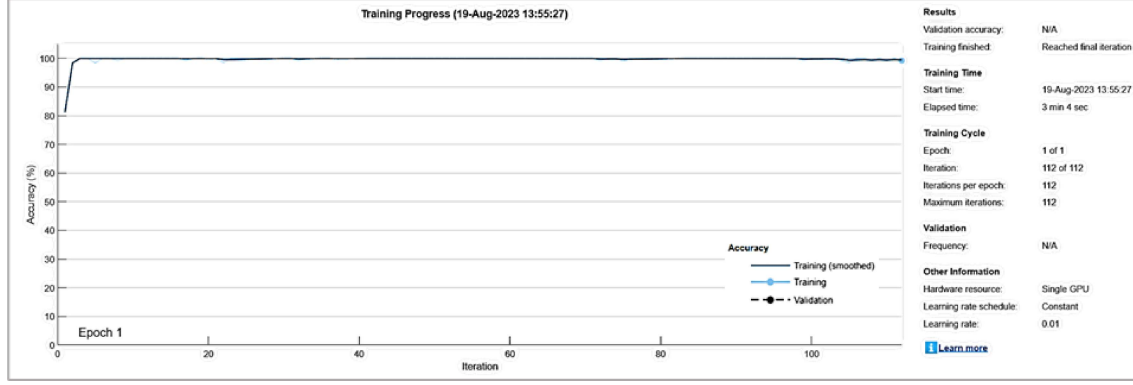
الشكل I 7: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الخامسة بمعدل تعلم 0.01.

الجدول I 4: معلومات عملية تدريب CNN لبيانات العينة الأولى من السلالة الخامسة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 46	64.84%	0.56	0.01
50	00: 02: 27	99.22%	0.10	0.01
100	00: 03: 27	100.00%	0	0.01
112	00: 03: 39	100.00%	0	0.01



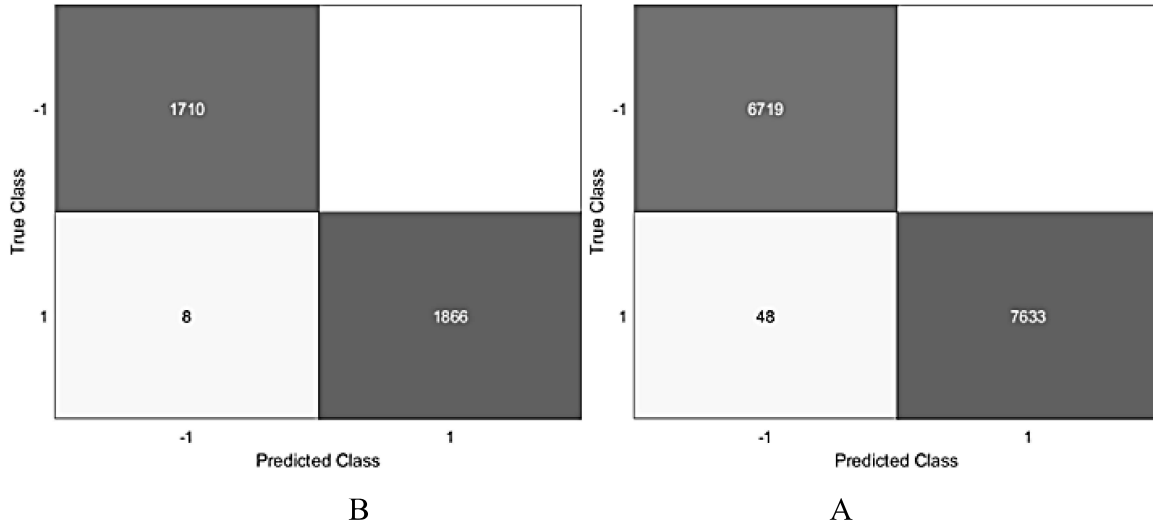
الشكل I 8: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الخامسة) بمعدل تعلم 0.01



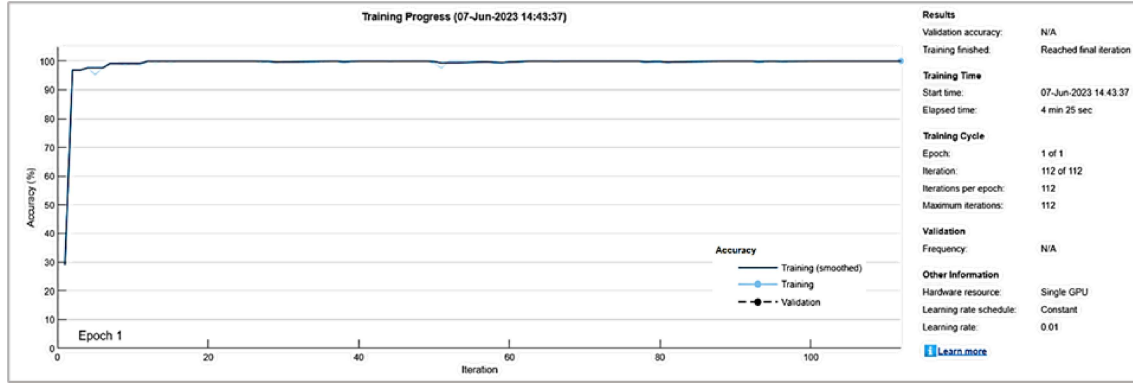
الشكل I 9: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الثانية بمعدل تعلم 0.01.

الجدول I 5: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الثانية بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 00: 57	81.25%	0.46	0.01
50	00: 01: 51	100.00%	0	0.01
100	00: 02: 51	100.00%	0	0.01
112	00: 03: 04	99.22%	0.12	0.01



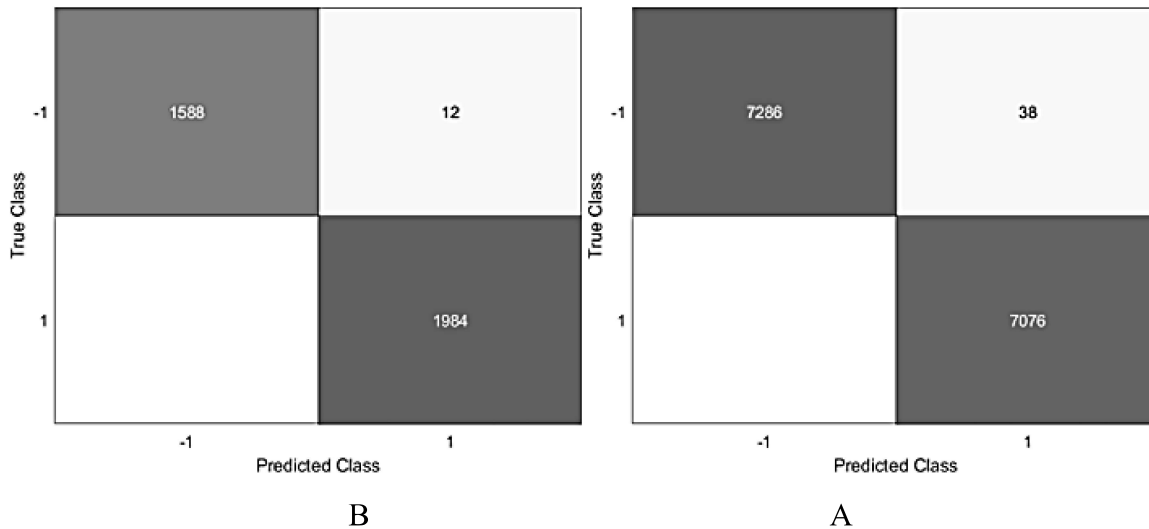
الشكل I 10: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثانية) بمعدل تعلم 0.01



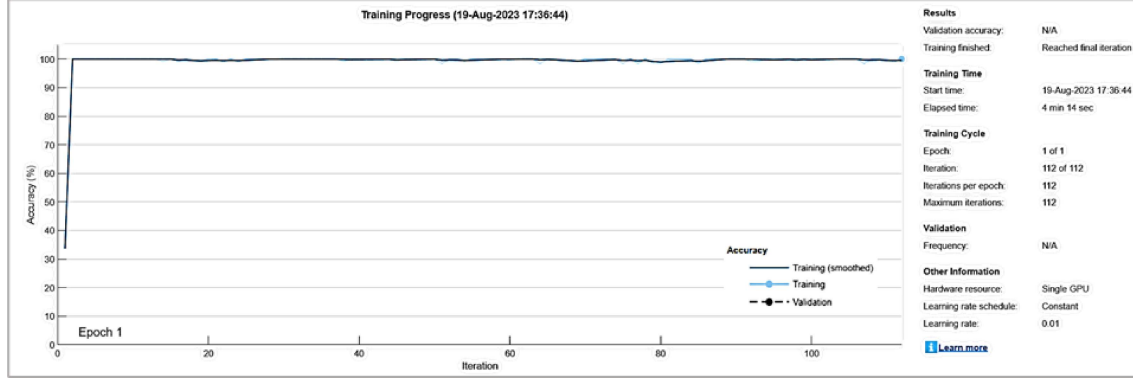
الشكل I 11: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الثالثة بمعدل تعلم 0.01.

الجدول I 6: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الثالثة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 02: 11	52.34%	0.71	0.01
50	00: 03: 01	100.00%	0	0.01
100	00: 04: 11	99.22%	0.12	0.01
112	00: 04: 33	100.00%	0	0.01



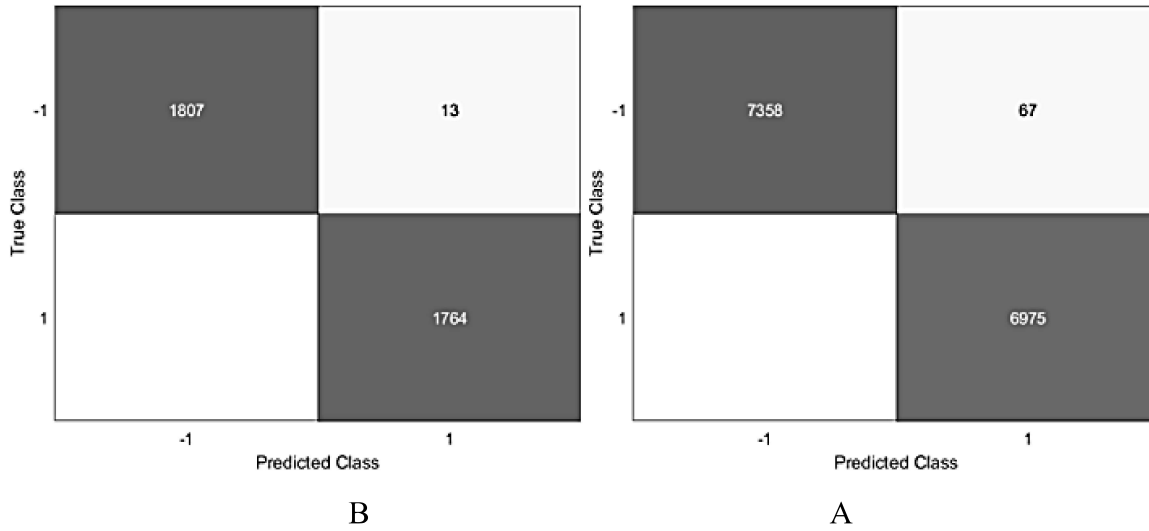
الشكل I 12: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثالثة) بمعدل تعلم 0.01



الشكل I 13: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الرابعة بمعدل تعلم 0.01.

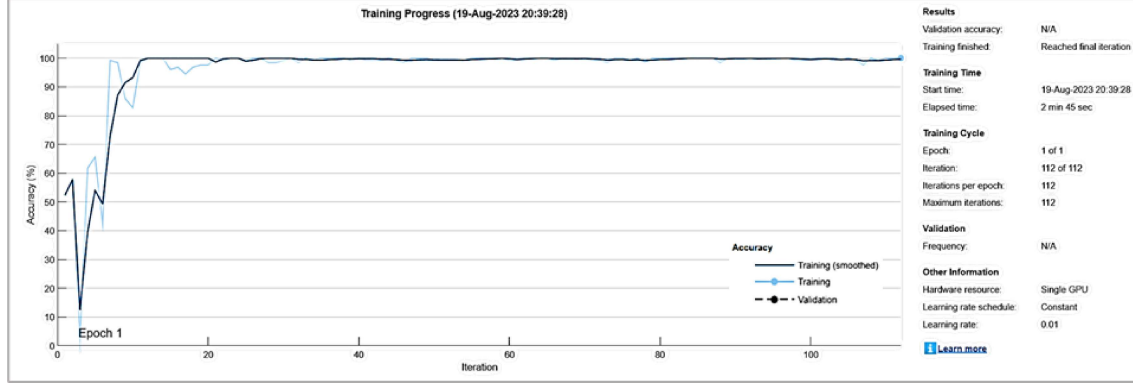
الجدول I 7: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الرابعة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 47	33.59%	1.23	0.01
50	00: 02: 43	100.00%	0	0.01
100	00: 04: 00	100.00%	0	0.01
112	00: 04: 14	100.00%	0	0.01



الشكل I 14: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من

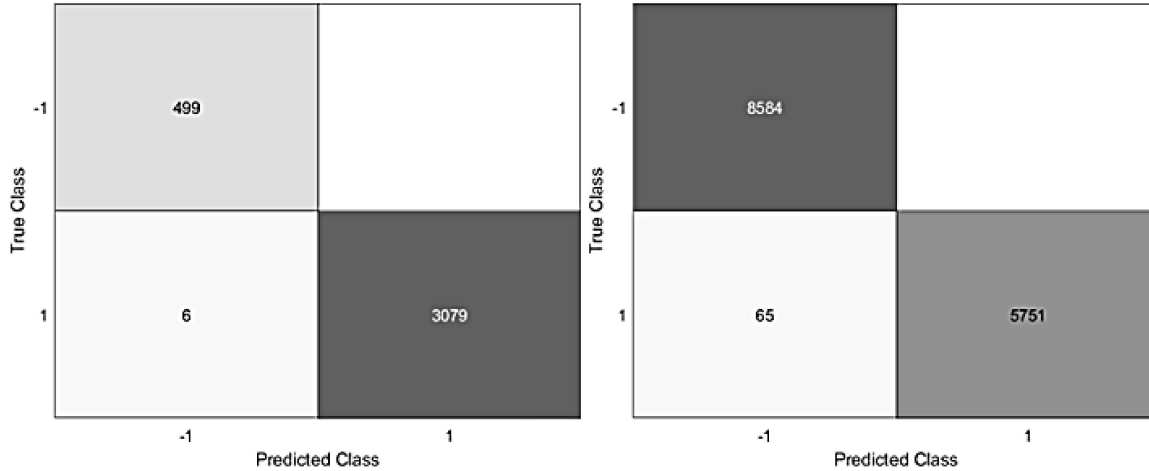
السلالة الرابعة) بمعدل تعلم 0.01



الشكل I. 15: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الخامسة بمعدل تعلم 0.01.

الجدول I. 8: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الخامسة بمعدل تعلم 0.01

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 24	52.34%	0.70	0.01
50	00: 01: 54	99.22%	0.03	0.01
100	00: 02: 36	99.22%	0.05	0.01
112	00: 02: 45	100.00%	0	0.01



B

A

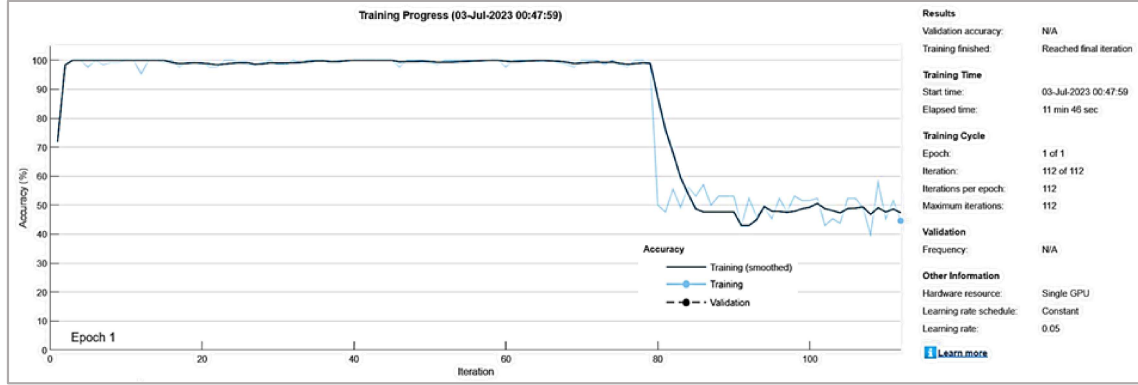
الشكل I. 16: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الخامسة) بمعدل تعلم 0.01

# المحقق J

مصفوفات الارتباك وسلوك ومعلومات عملية تدريب CNN بمعدل

تعلم 0.05

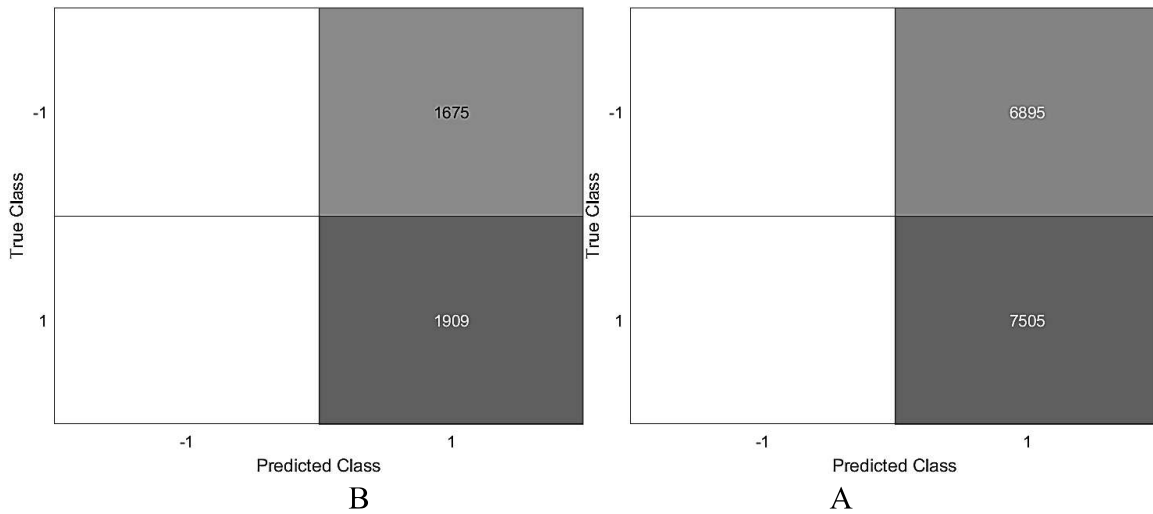
للسلالات الثانية والثالثة والرابعة والخامسة



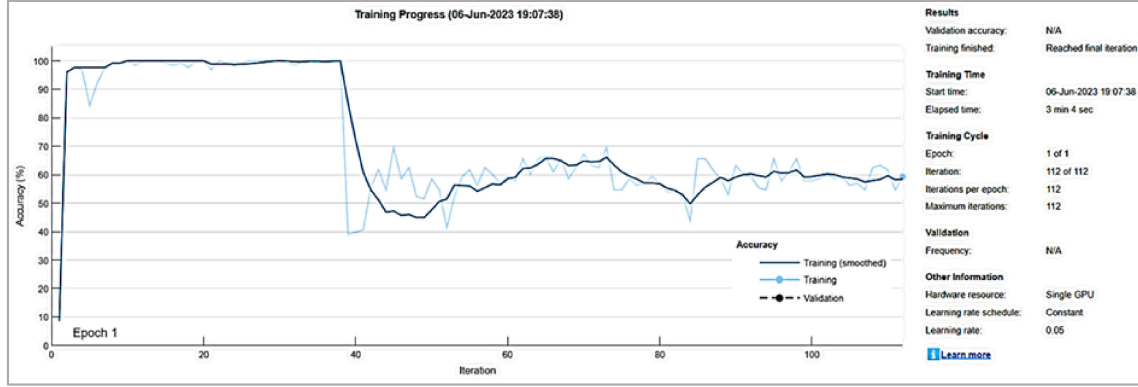
الشكل J . 1: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الثانية بمعدل تعلم 0.05.

الجدول J . 1: معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الثانية بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 04: 36	71.88%	0.48	0.05
50	00: 07: 31	99.22%	0.12	0.05
100	00: 10: 22	51.56%	0.69	0.05
112	00: 11: 46	44.53%	0.69	0.05



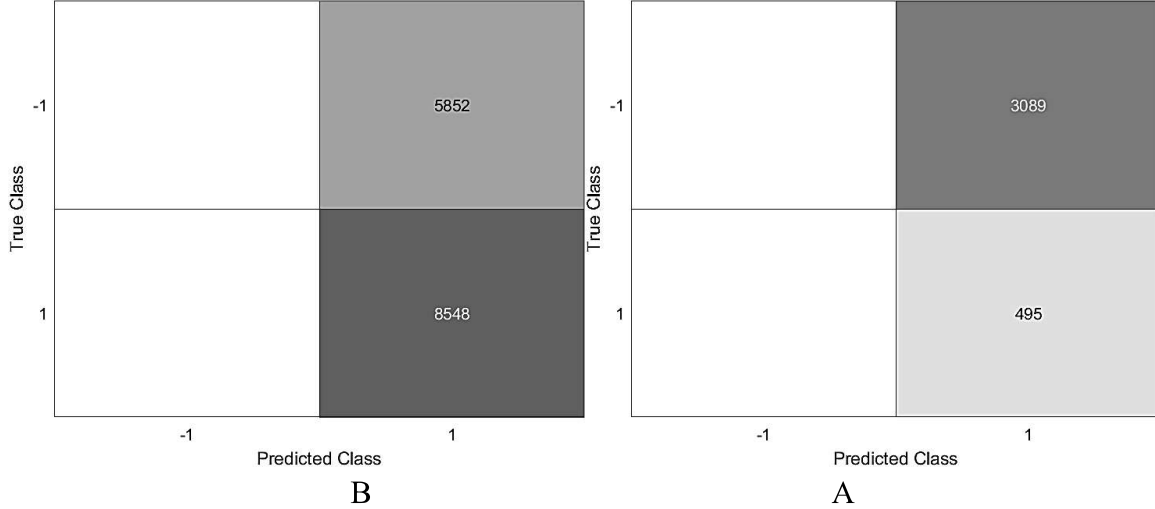
الشكل J . 2: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثانية) بمعدل تعلم 0.05



الشكل J . 3: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الثالثة بمعدل تعلم 0.05.

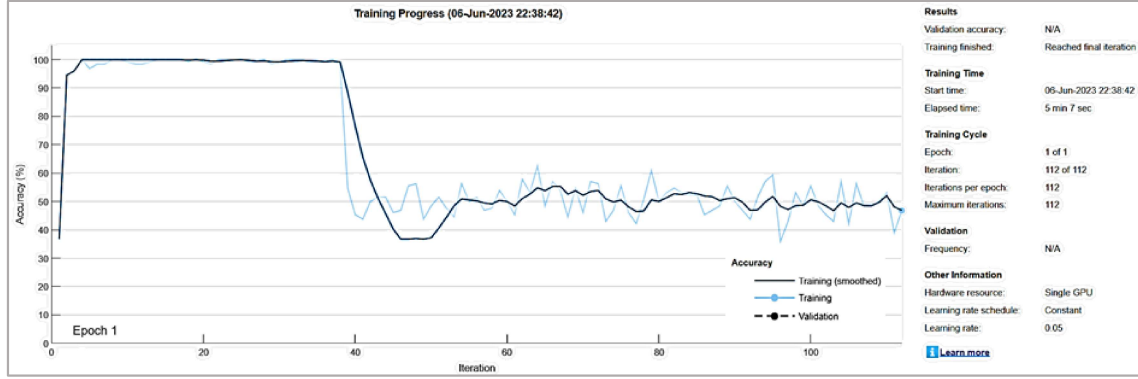
الجدول J . 2: معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الثالثة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 04	8.59%	1.39	0.05
50	00: 01: 50	58.59%	0.67	0.05
100	00: 02: 37	57.81%	0.68	0.05
112	00: 03: 04	59.38%	0.67	0.05



الشكل J . 4: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من

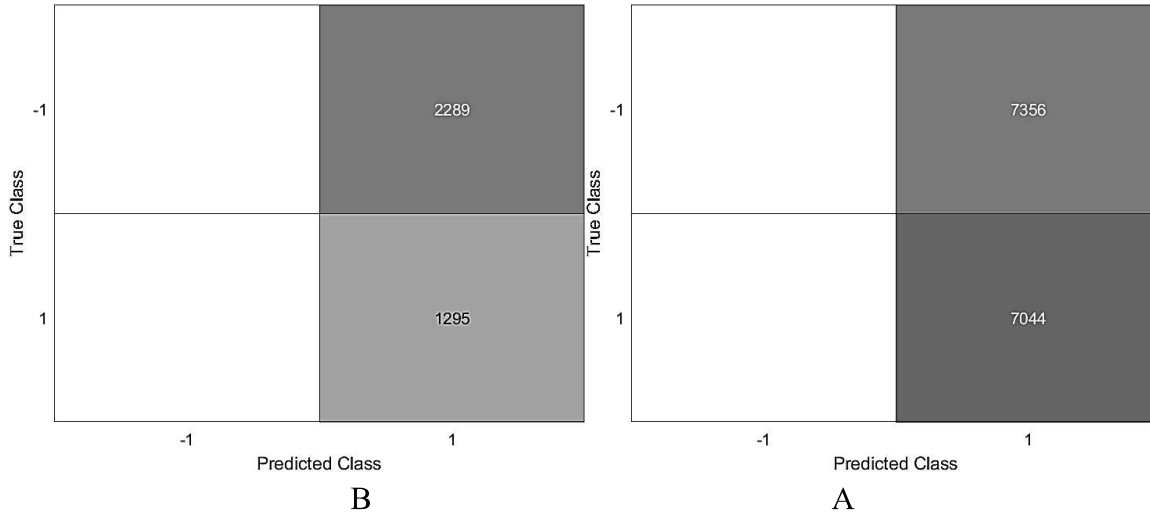
السلالة الثالثة) بمعدل تعلم 0.05



الشكل J . 5: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الرابعة بمعدل تعلم 0.05.

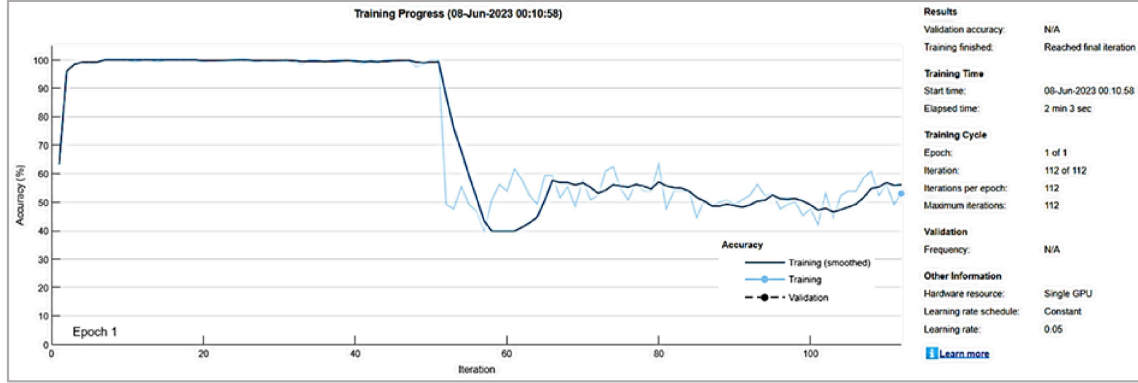
الجدول J . 3: معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الرابعة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 15	36.72%	0.91	0.05
50	00: 03: 18	48.44%	0.69	0.05
100	00: 04: 34	55.47%	0.69	0.05
112	00: 05: 07	46.88%	0.69	0.05



الشكل J . 6: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من

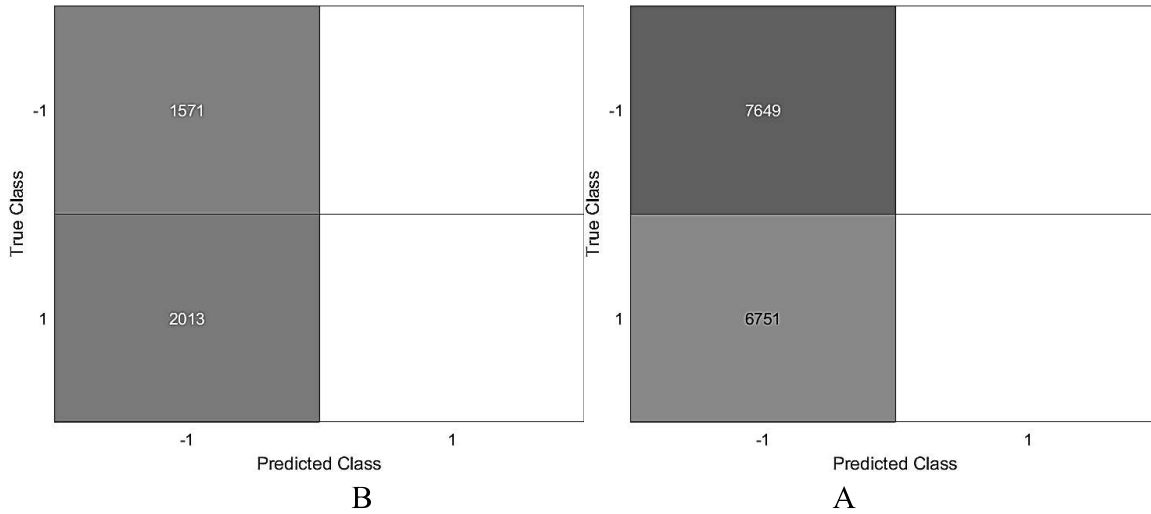
السلالة الرابعة) بمعدل تعلم 0.05



الشكل J . 7: سلوك عملية تدريب CNN لبيانات العينة الاولى من السلالة الخامسة بمعدل تعلم 0.05.

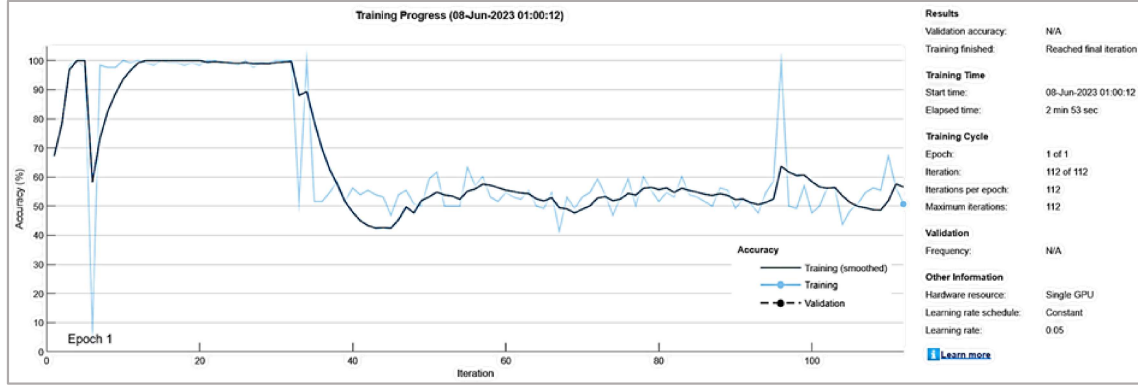
الجدول J . 4: معلومات عملية تدريب CNN لبيانات العينة الاولى من السلالة الخامسة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 00: 54	63.28%	0.60	0.05
50	00: 01: 22	100.00%	0	0.05
100	00: 01: 55	47.66%	0.69	0.05
112	00: 02: 03	53.12%	0.69	0.05



الشكل J . 8: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من

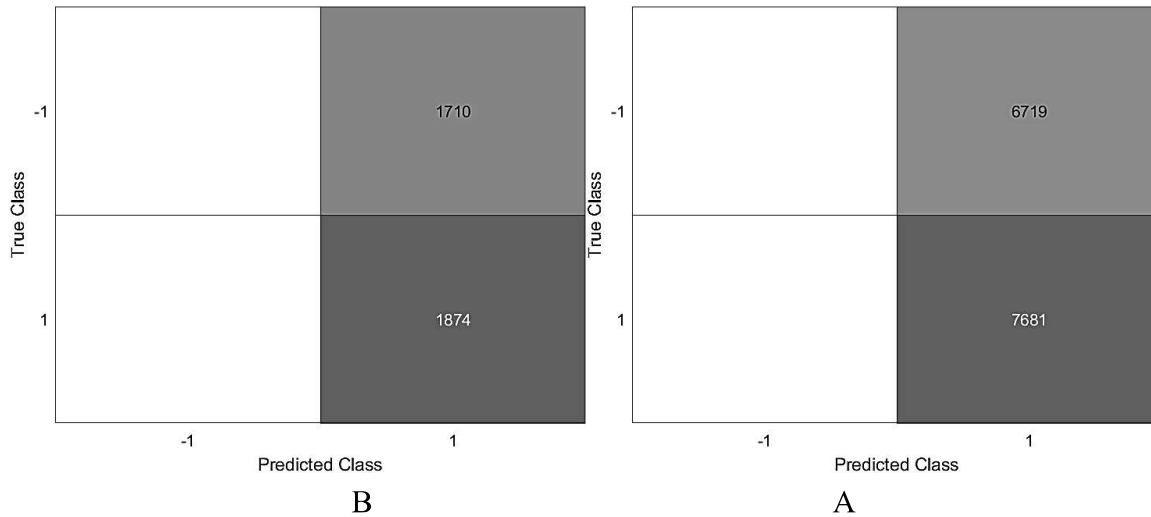
السلالة الخامسة) بمعدل تعلم 0.05



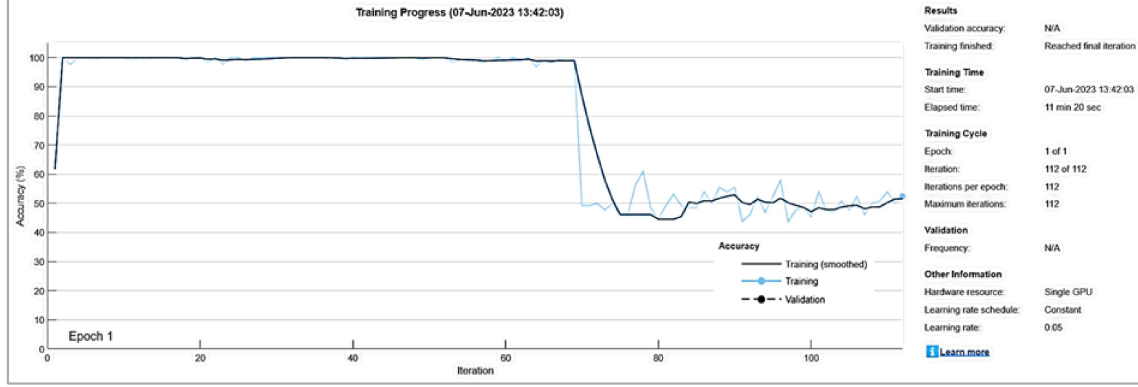
الشكل J . 9: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الثانية بمعدل تعلم 0.05.

الجدول J . 5: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الثانية بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 32	67.19%	0.68	0.05
50	00: 02: 05	59.38%	0.68	0.05
100	00: 02: 43	47.66%	0.69	0.05
112	00: 02: 53	50.78%	0.69	0.05



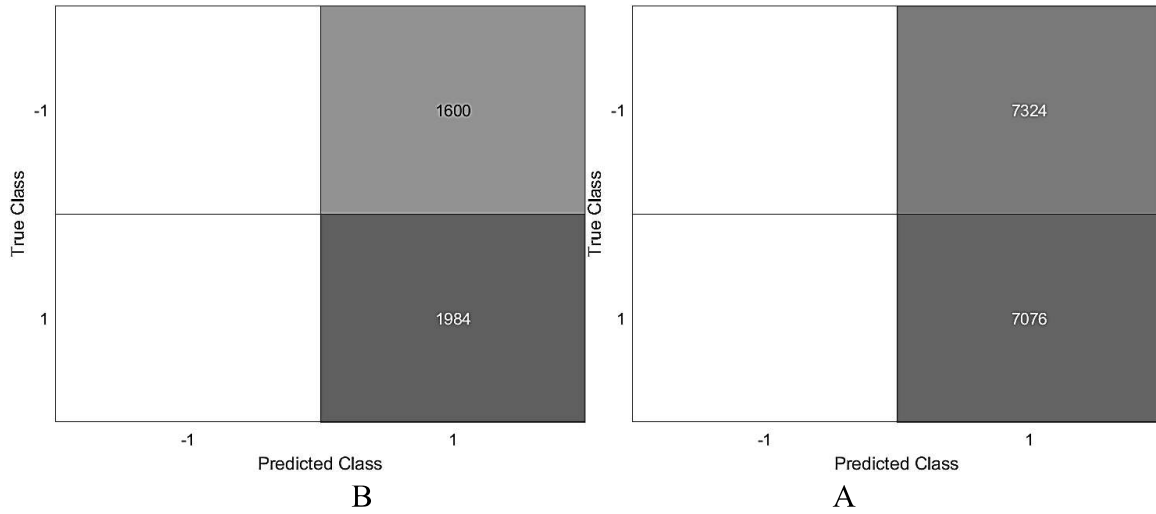
الشكل J . 10: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثانية) بمعدل تعلم 0.05



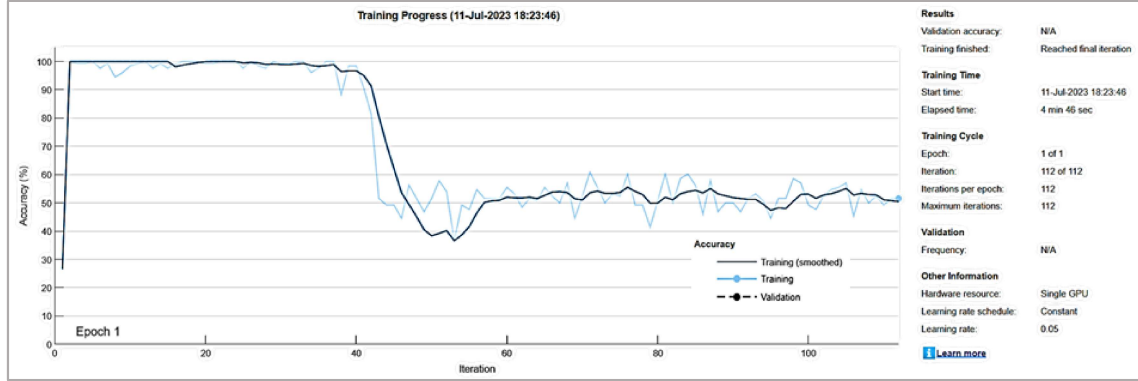
الشكل J . 11: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الثالثة بمعدل تعلم 0.05.

الجدول J . 6: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الثالثة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 03: 53	61.72%	0.56	0.05
50	00: 06: 59	100.00%	0	0.05
100	00: 10: 31	45.31%	0.69	0.05
112	00: 11: 20	52.34%	0.69	0.05



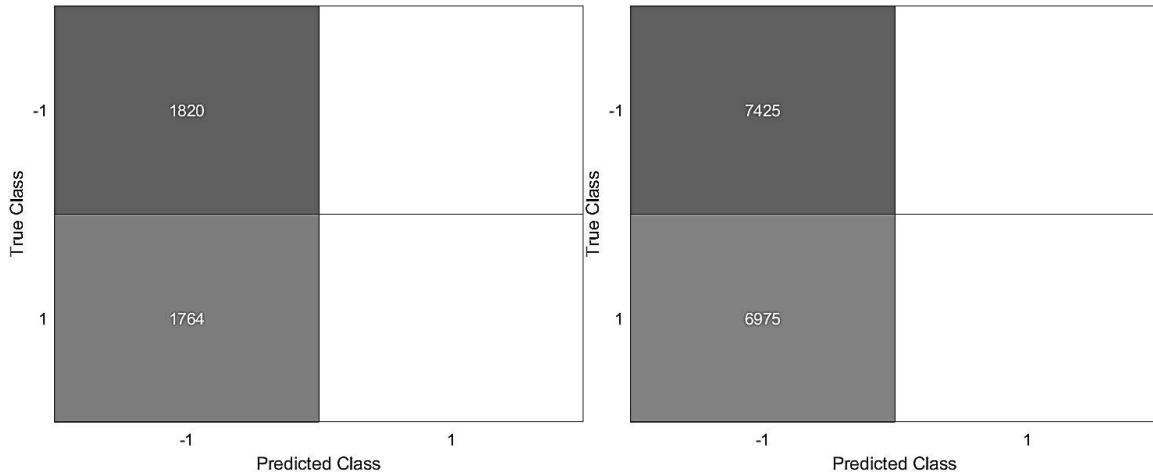
الشكل J . 12: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثالثة) بمعدل تعلم 0.05



الشكل J . 13: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الرابعة بمعدل تعلم 0.05.

الجدول J . 7: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الرابعة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 00: 49	26.56%	0.90	0.05
50	00: 01:47	51.56%	0.69	0.05
100	00: 04: 14	49.22%	0.69	0.05
112	00: 04: 46	51.56%	0.69	0.05

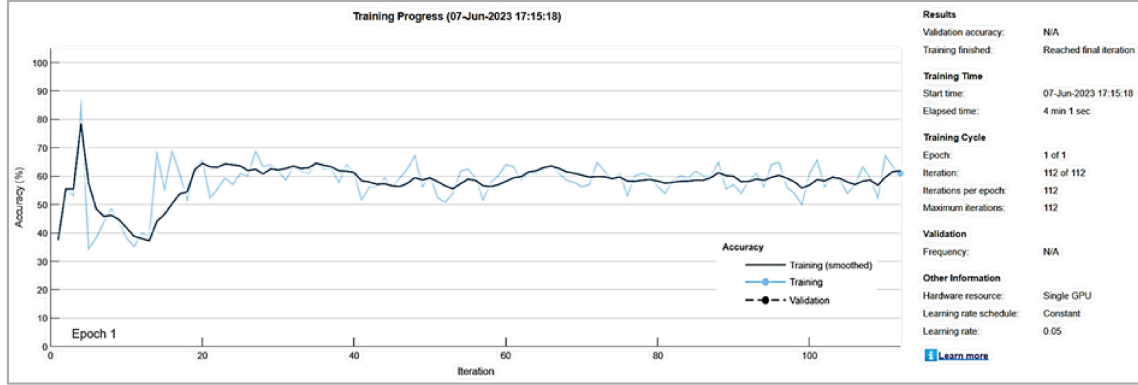


B

A

الشكل J . 14: مصفوفة الارتباك لـ نموذج CNN لبيانات A: التدريب و B: الاختبار (العينة الثانية من

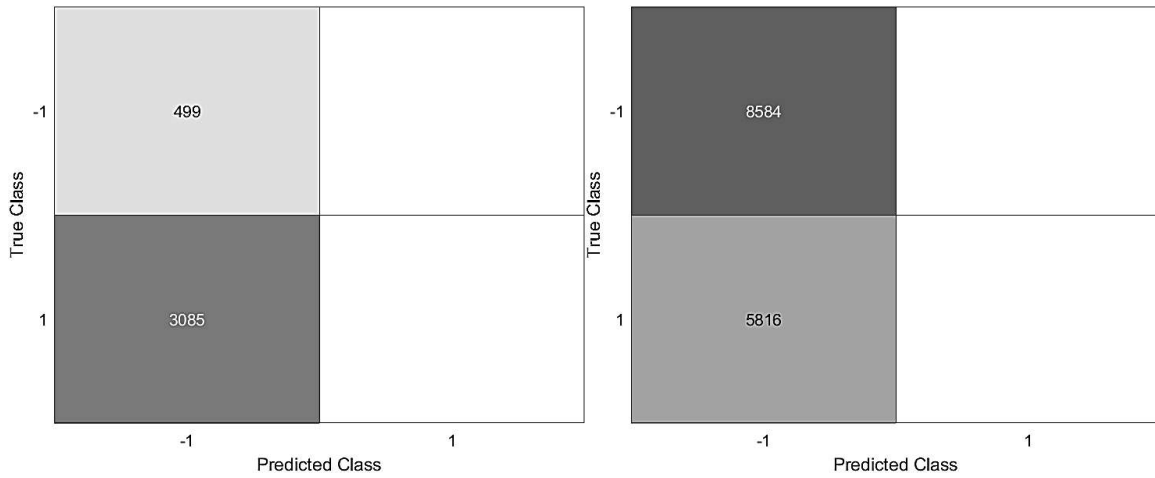
السلالة الرابعة) بمعدل تعلم 0.05



الشكل J . 15: سلوك عملية تدريب CNN لبيانات العينة الثانية من السلالة الخامسة بمعدل تعلم 0.05.

الجدول J . 8: معلومات عملية تدريب CNN لبيانات العينة الثانية من السلالة الخامسة بمعدل تعلم 0.05

Iteration	(hh: mm: ss)	Accuracy	Loss	L-Rate
1	00: 01: 52	37.50%	1.02	0.05
50	00: 02: 45	60.16%	0.67	0.05
100	00: 03: 48	60.94%	0.67	0.05
112	00: 04: 01	60.94%	0.67	0.05



B

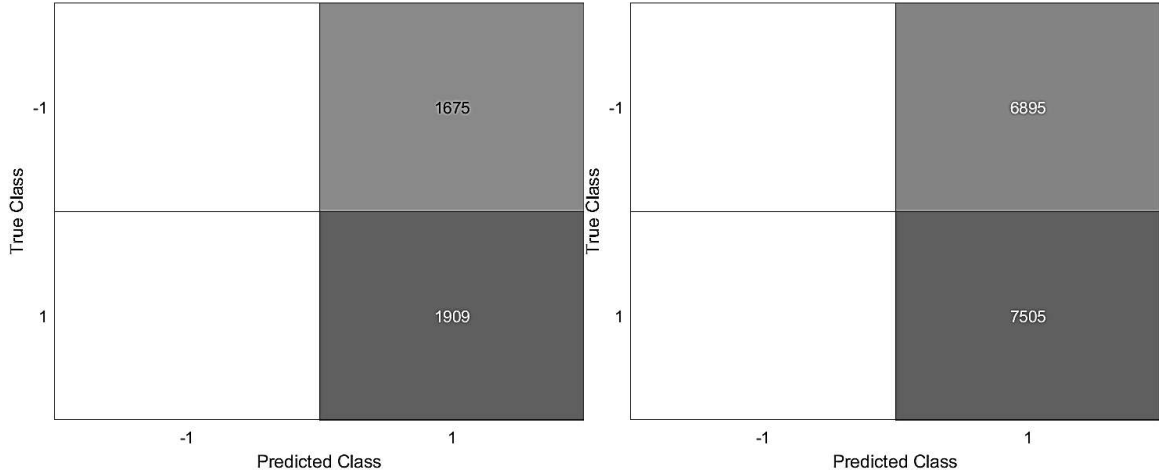
A

الشكل J . 16: مصفوفة الارتباك لانموذج CNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الخامسة) بمعدل تعلم 0.05

# K الملقق

مصفوفات الارتباك لامتودج PNN

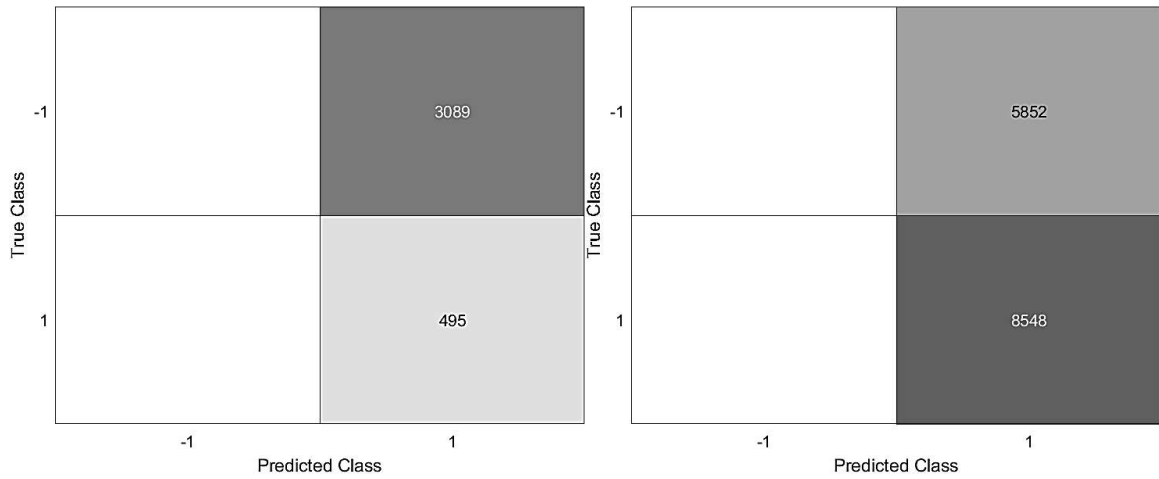
للسلالات الثانية والثالثة والرابعة والخامسة



B

A

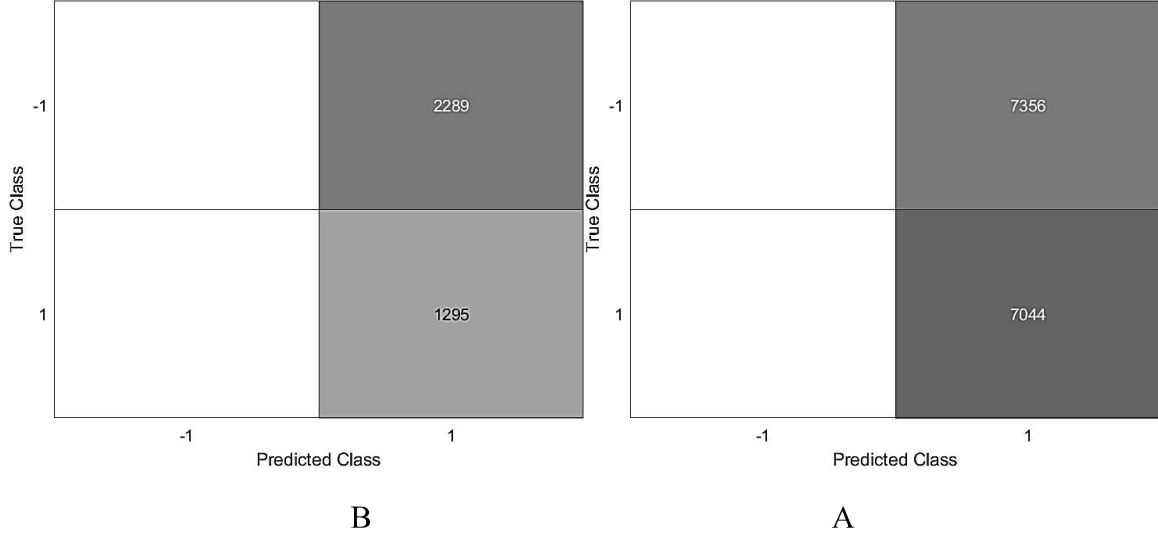
الشكل K . 1: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثانية)



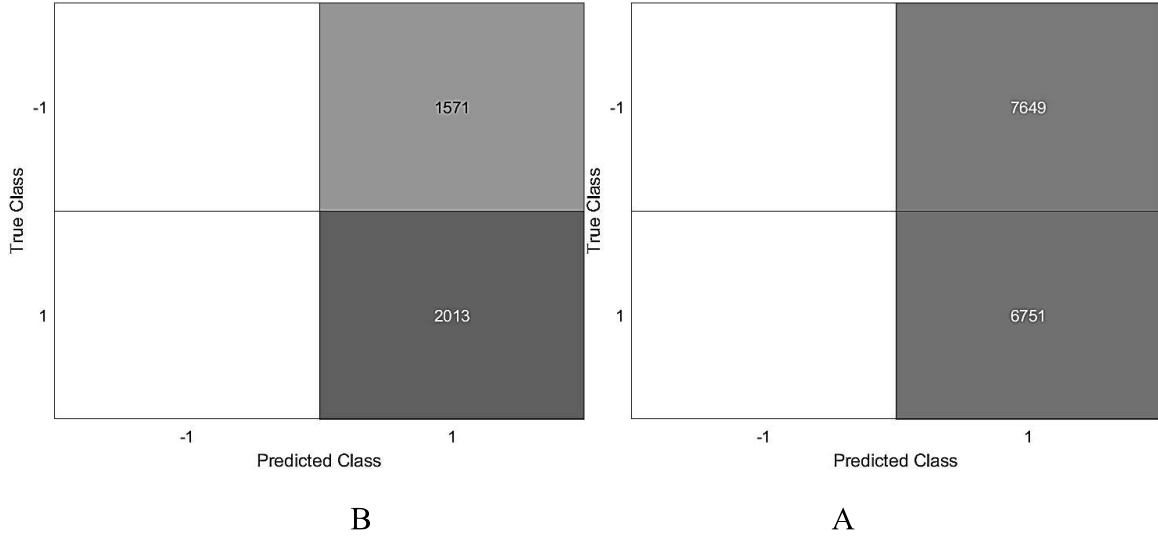
B

A

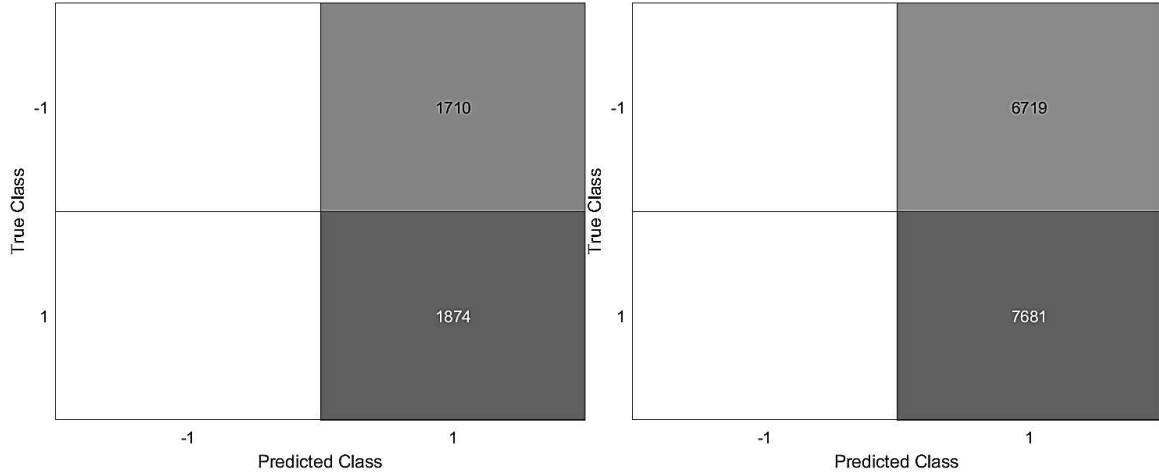
الشكل K . 2: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثالثة)



الشكل K . 3: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الرابعة)



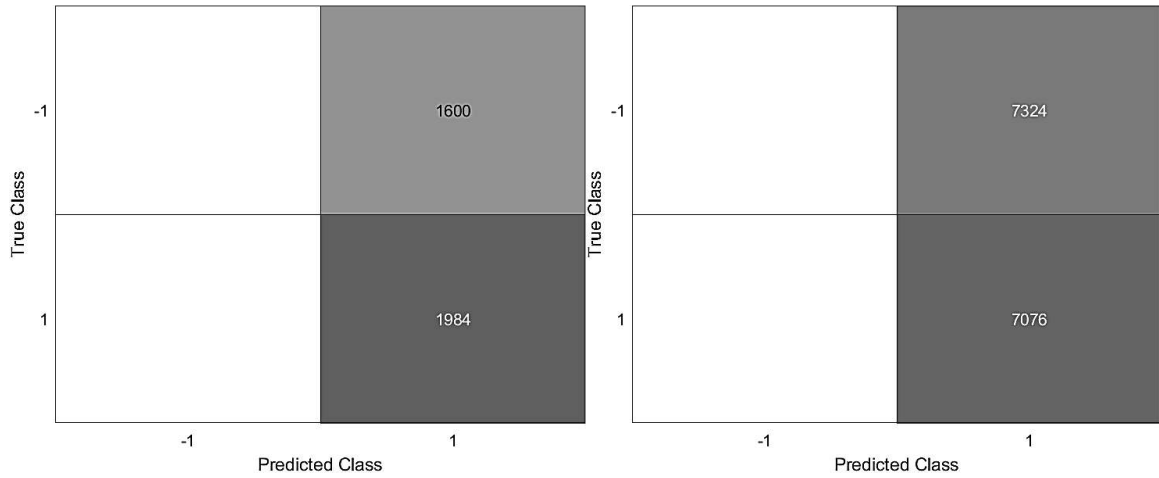
الشكل K . 4: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الخامسة)



B

A

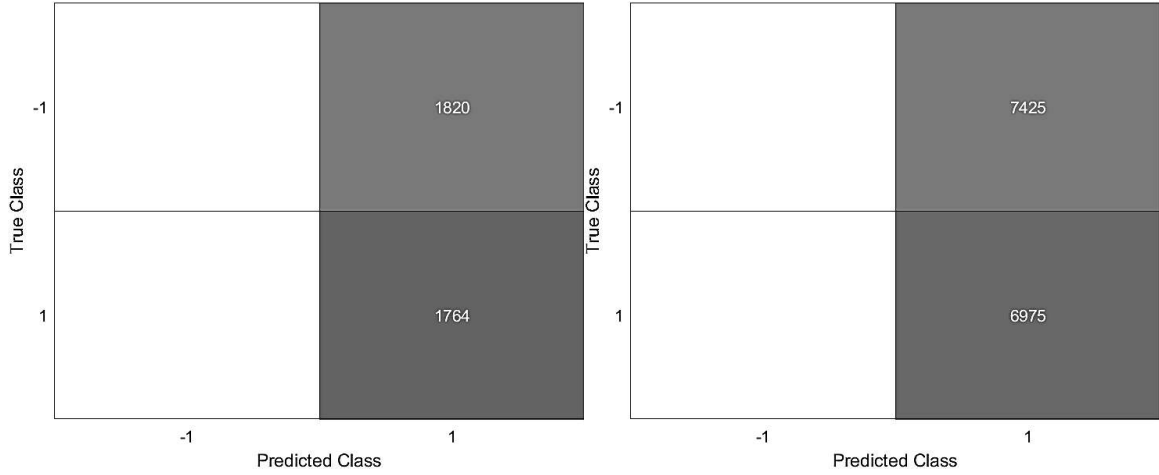
الشكل K . 5: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثانية)



B

A

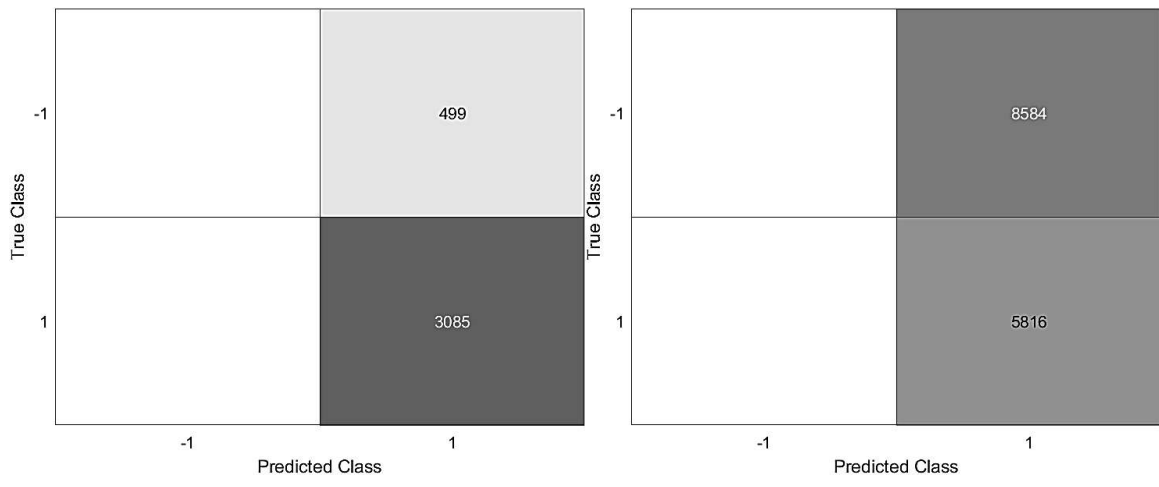
الشكل K . 6: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثالثة)



B

A

الشكل K . 7: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الرابعة)



B

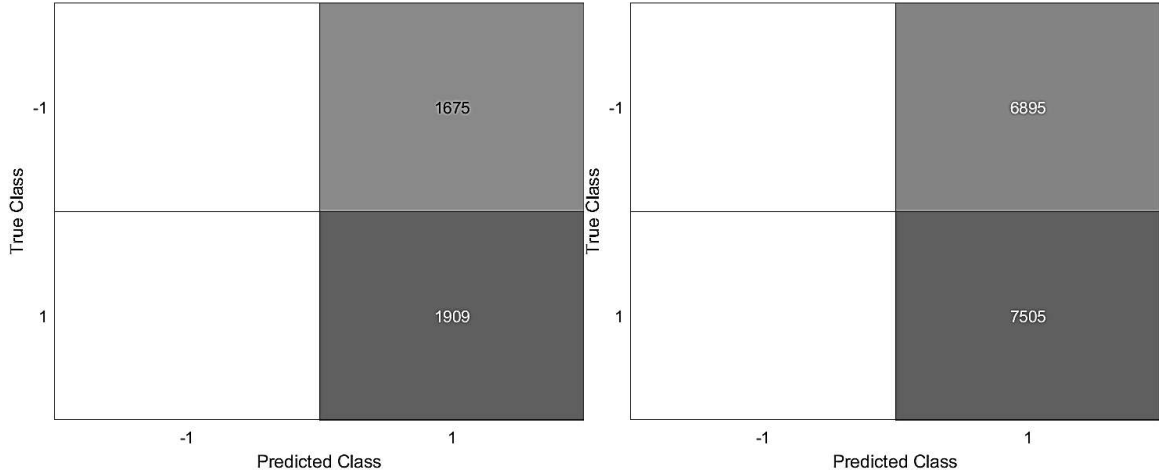
A

الشكل K . 8: مصفوفة الارتباك لانموذج PNN لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الخامسة)

# المحقق L

مصفوفات الارتباك لانموذج LR

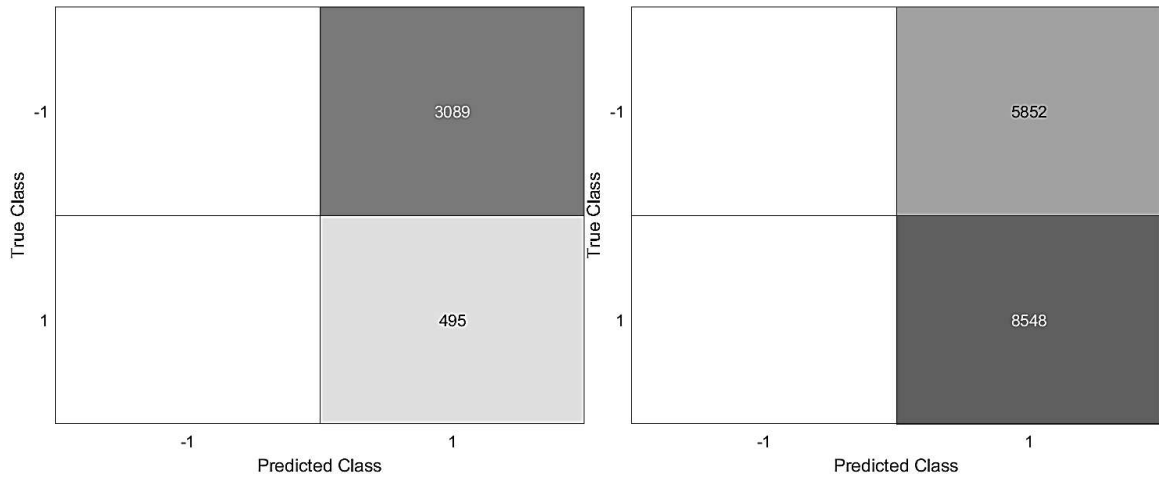
للسلاات الثانية والثالثة والرابعة والخامسة



B

A

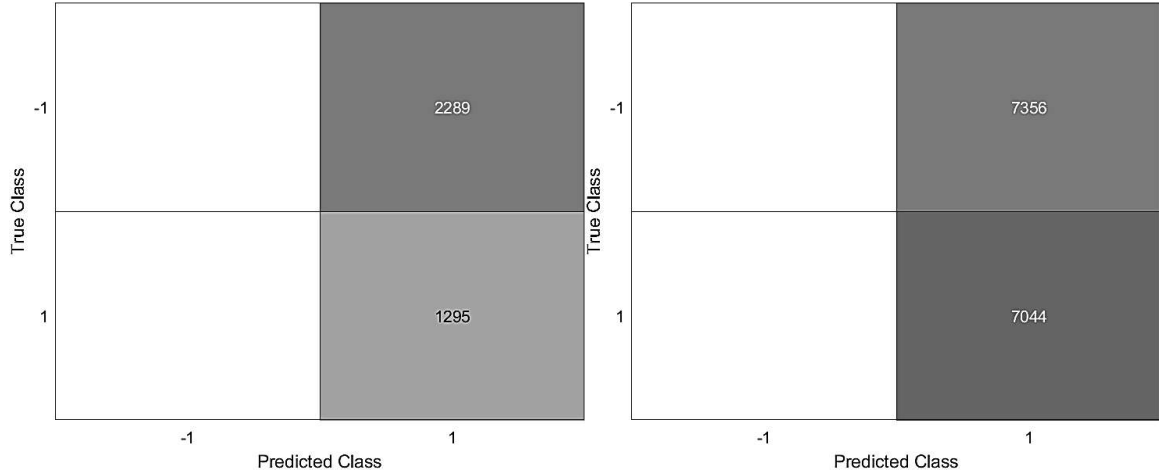
الشكل L . 1: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثانية)



B

A

الشكل L . 2: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثالثة)



B

A

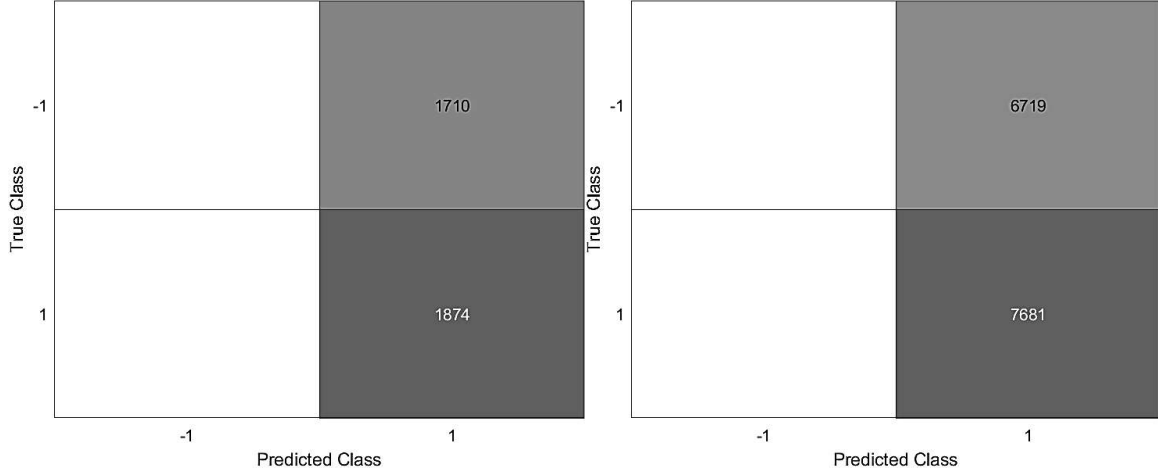
الشكل L . 3: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الرابعة)



B

A

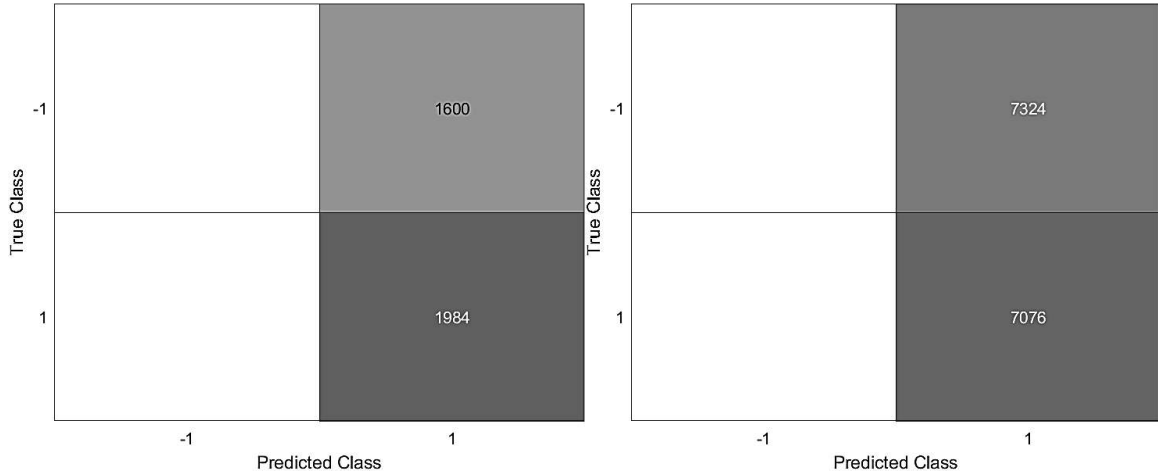
الشكل L . 4: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الخامسة)



B

A

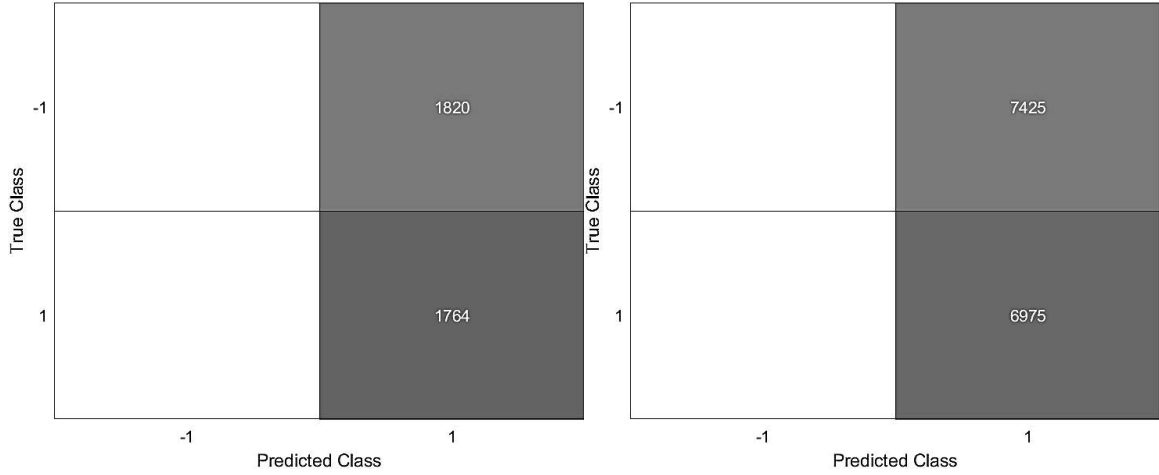
الشكل L . 5: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثانية)



B

A

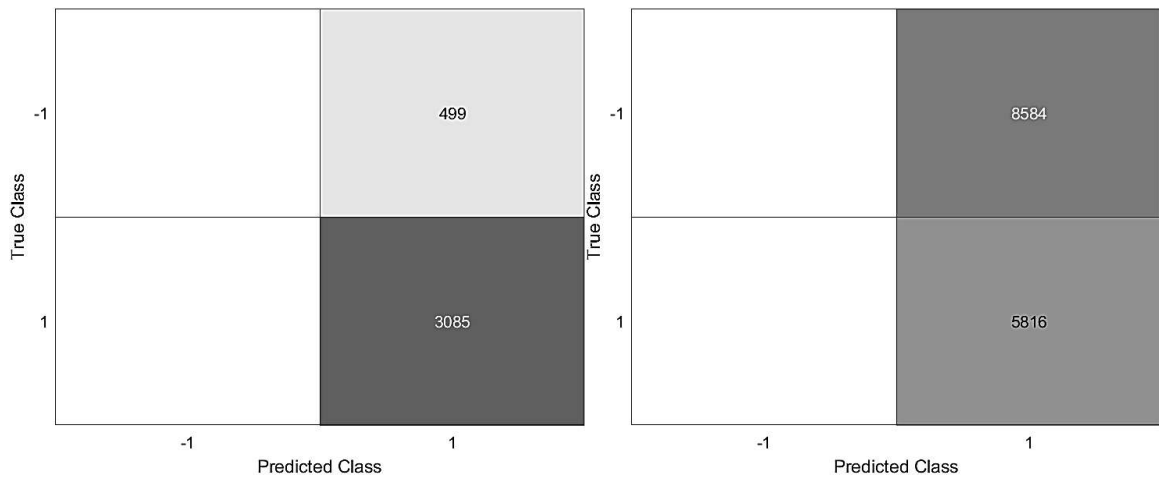
الشكل L . 6: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثالثة)



B

A

الشكل L . 7: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الرابعة)



B

A

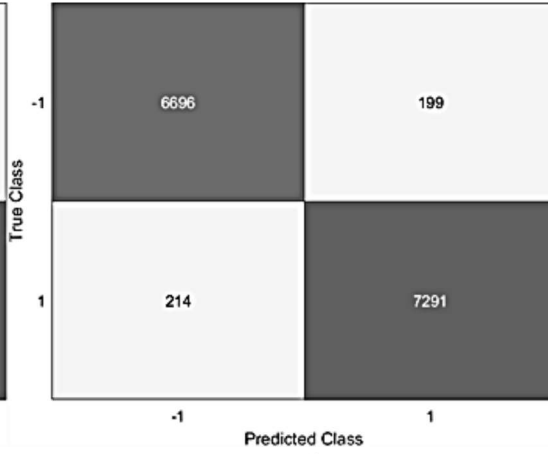
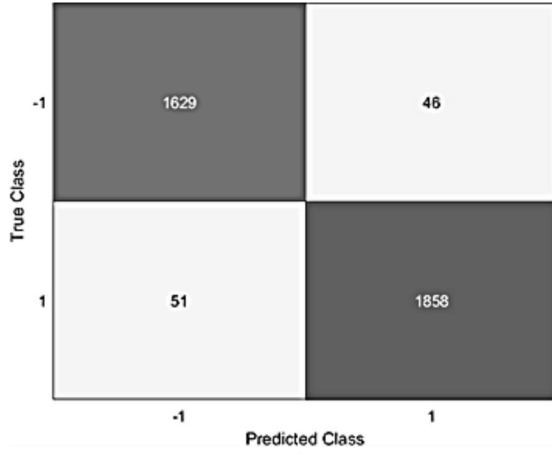
الشكل L . 8: مصفوفة الارتباك لانموذج LR لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الخامسة)

# M المطبق

مصفوفات الارتباك لانموذج SVM

للسلاسل الثانية والثالثة والرابعة والخامسة

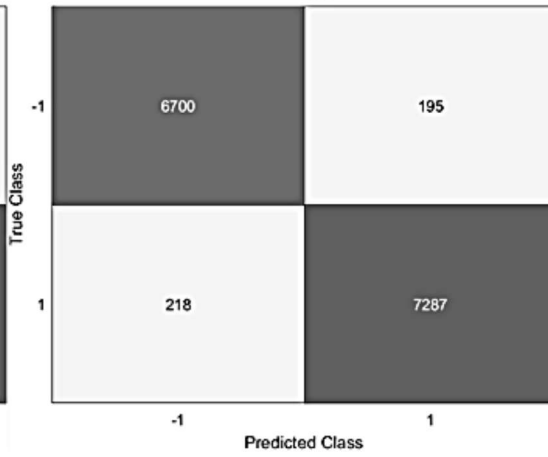
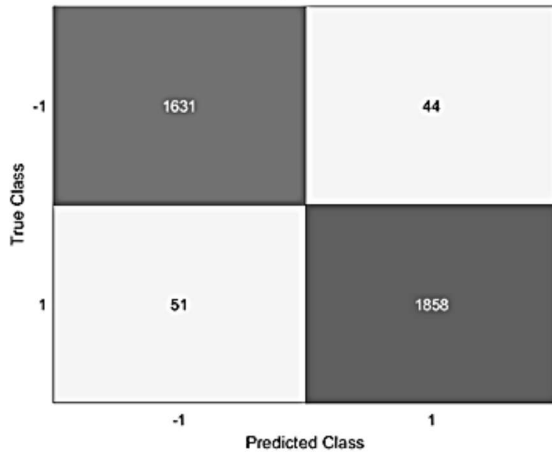
RBF



B

A

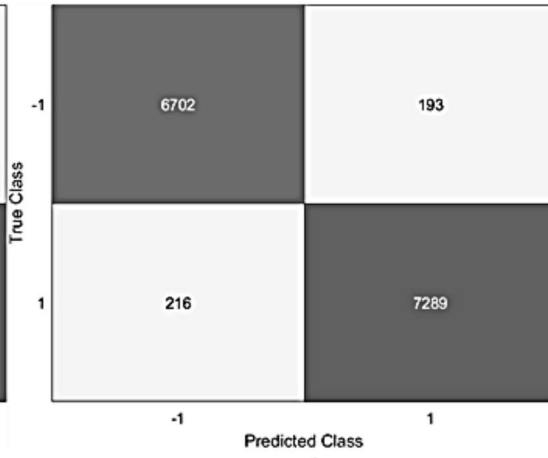
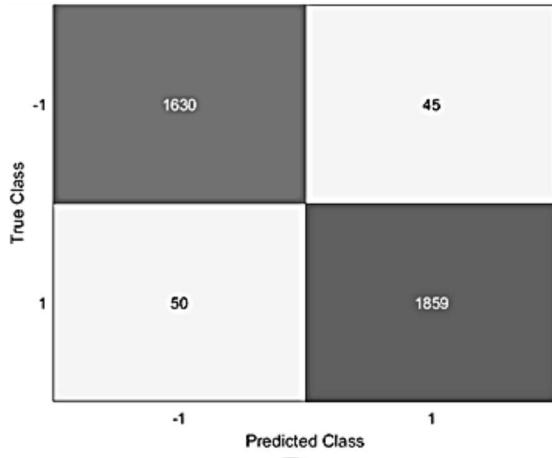
Linear



B

A

Polynomial

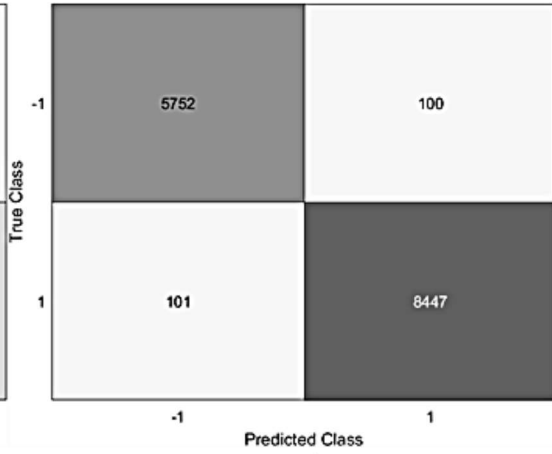
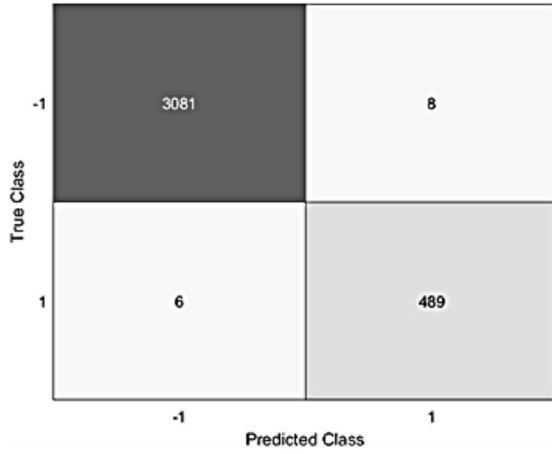


B

A

الشكل M . 1: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثانية)

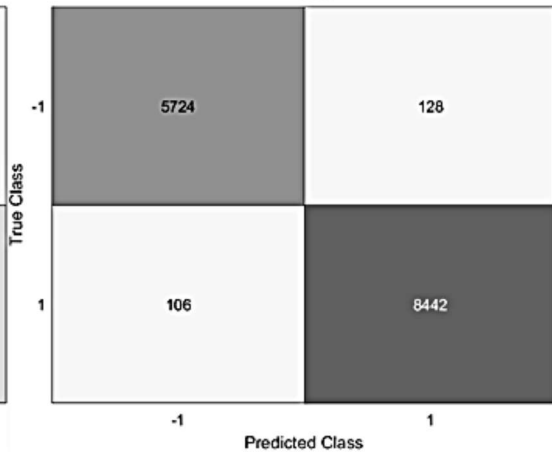
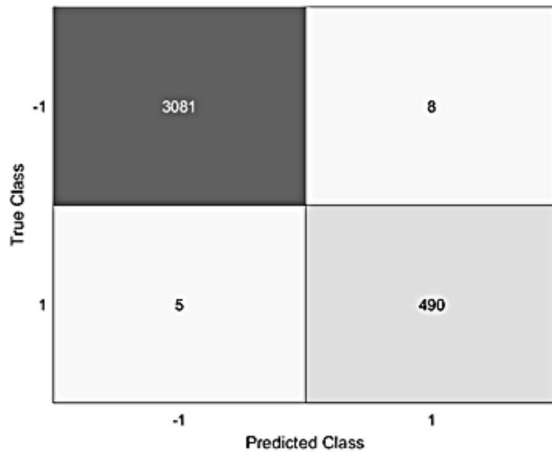
RBF



B

A

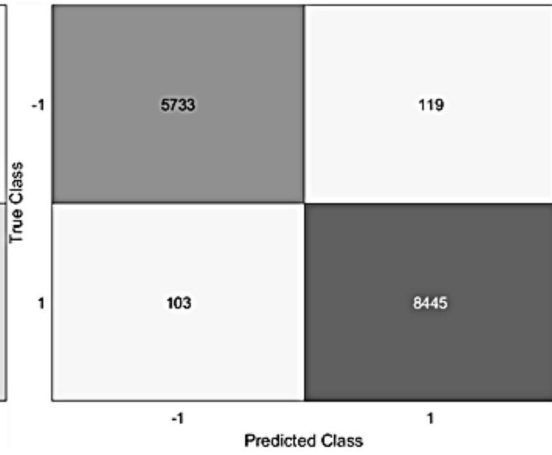
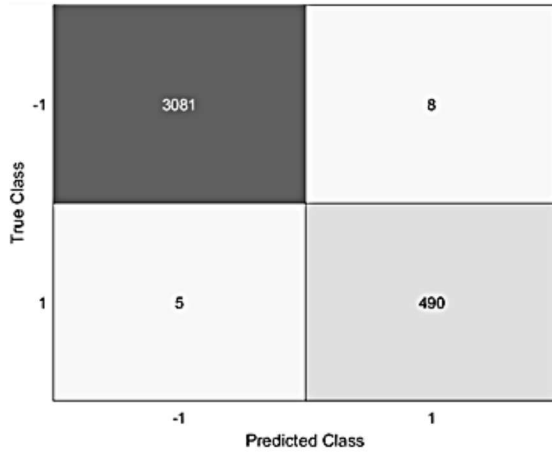
Linear



B

A

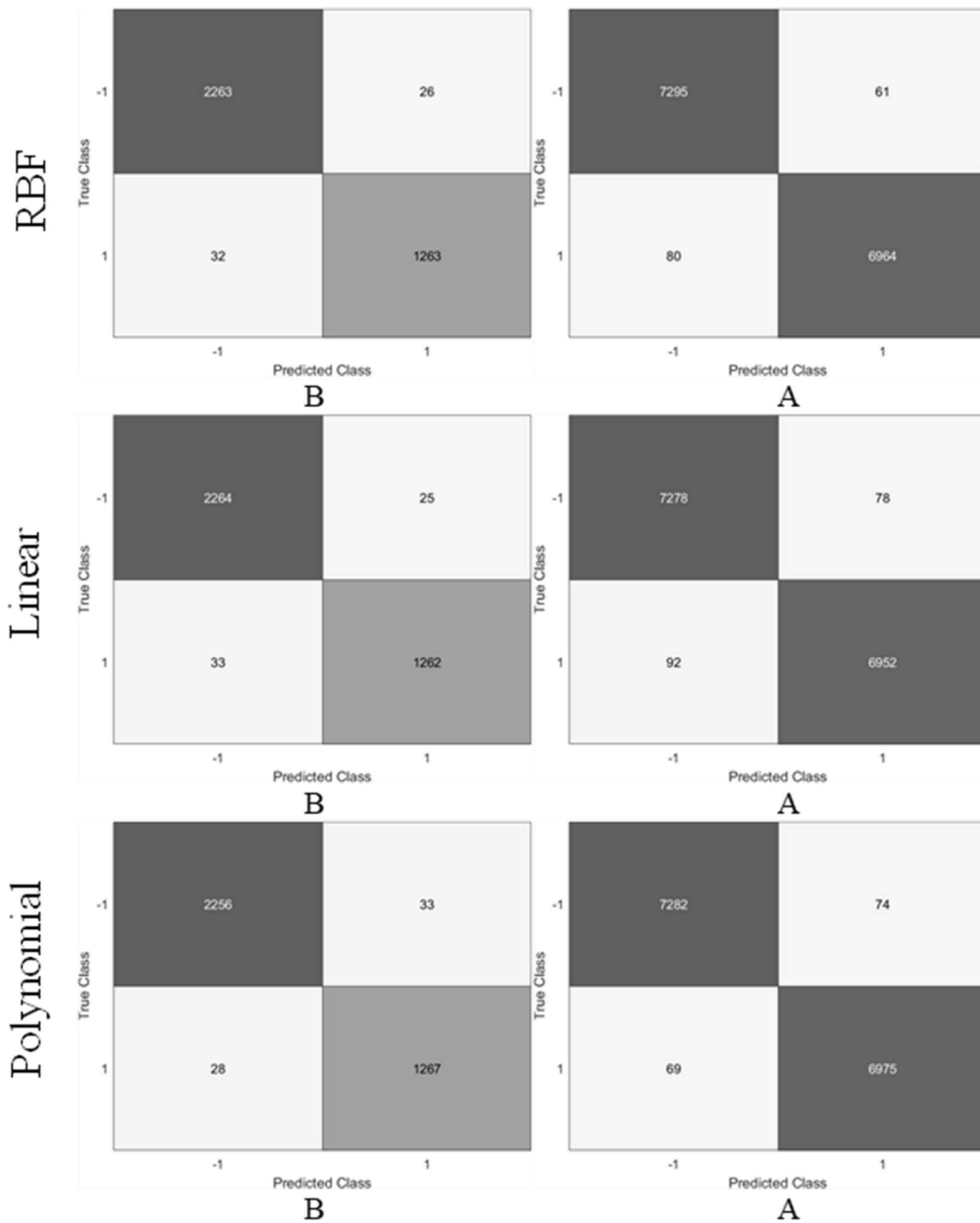
Polynomial



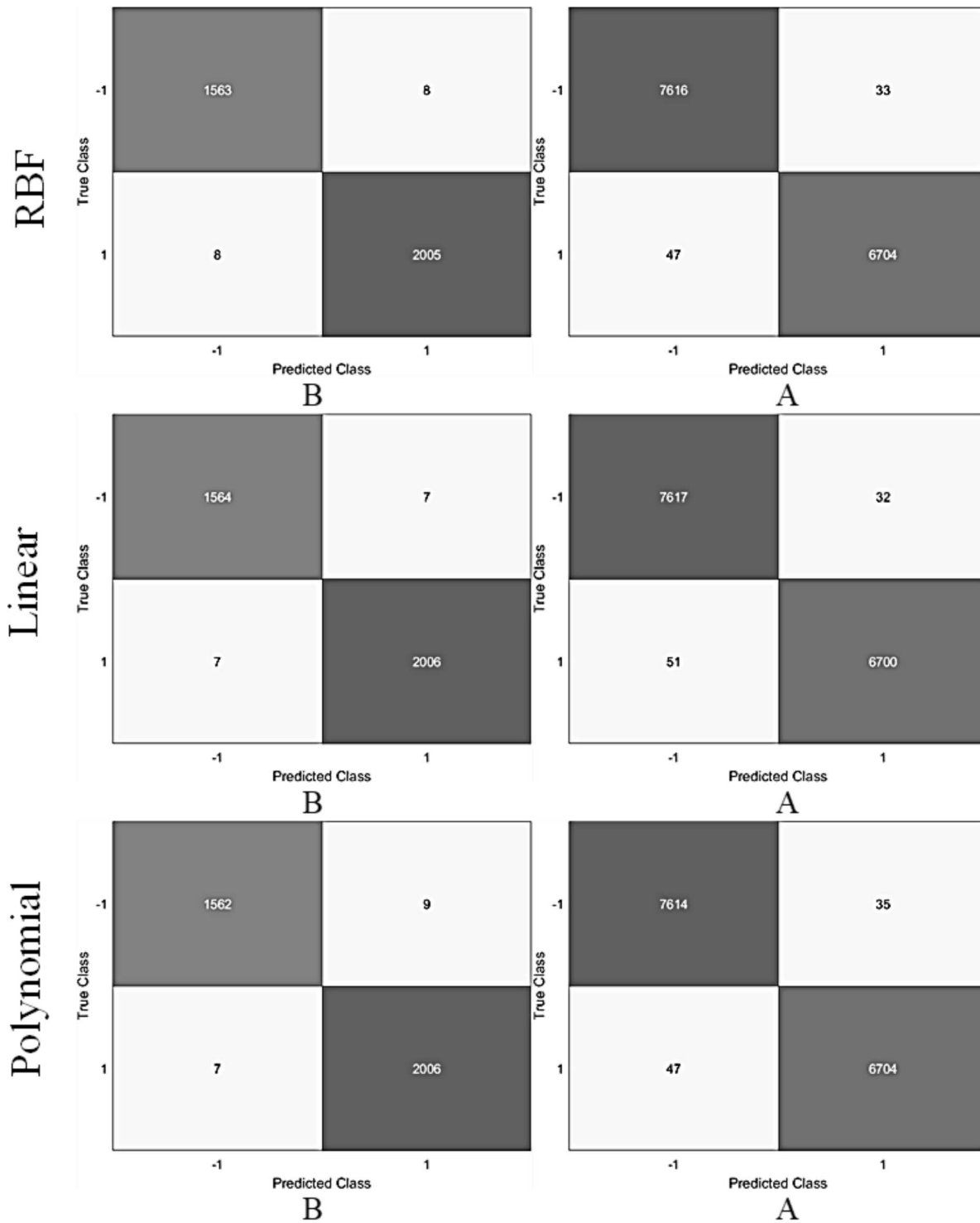
B

A

الشكل M . 2: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الثالثة)

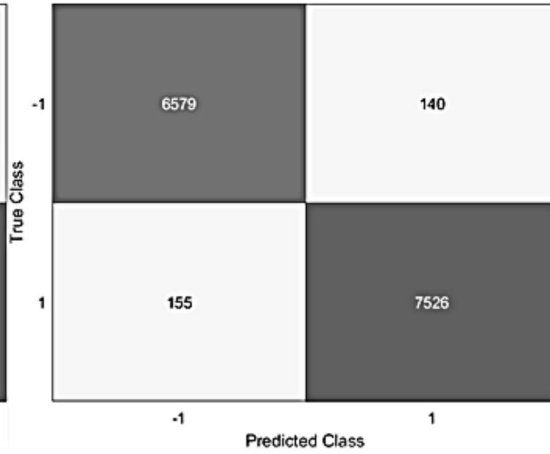
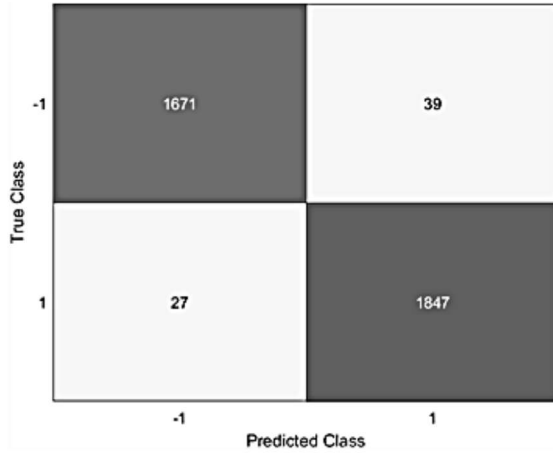


الشكل M . 3: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الرابعة)



الشكل M . 4: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الاولى من السلالة الخامسة)

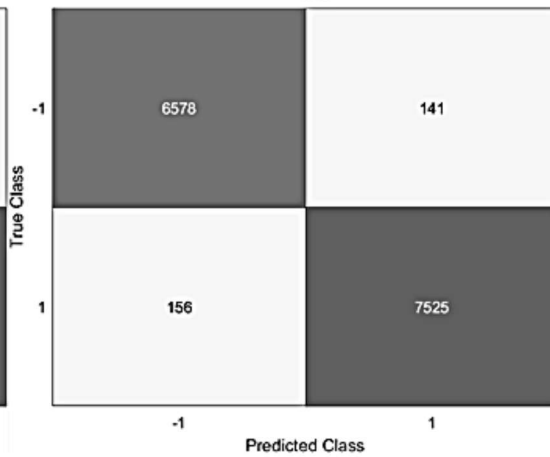
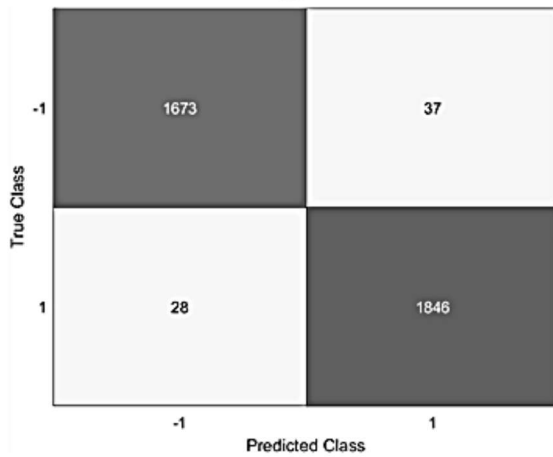
RBF



B

A

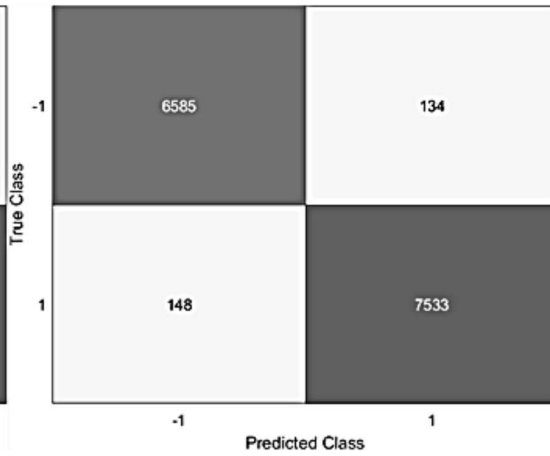
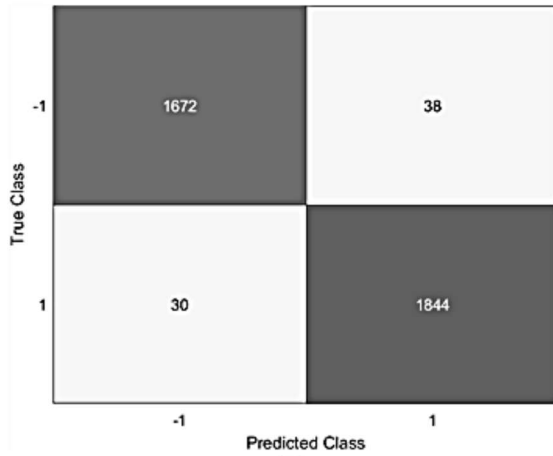
Linear



B

A

Polynomial

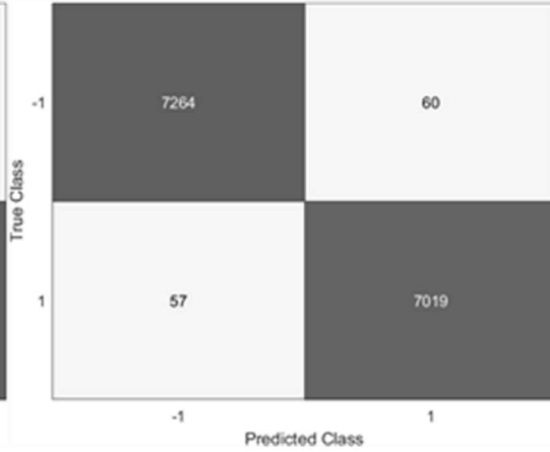
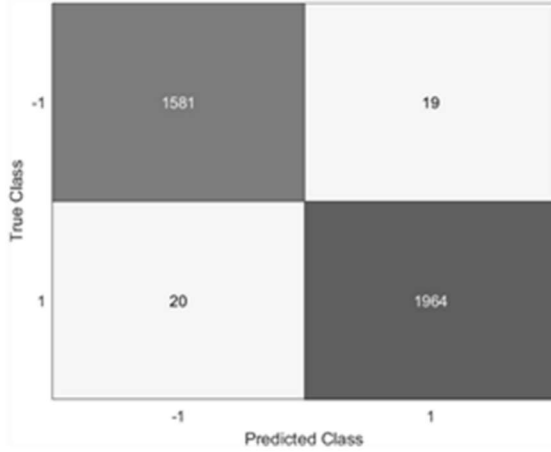


B

A

الشكل M . 5: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثانية)

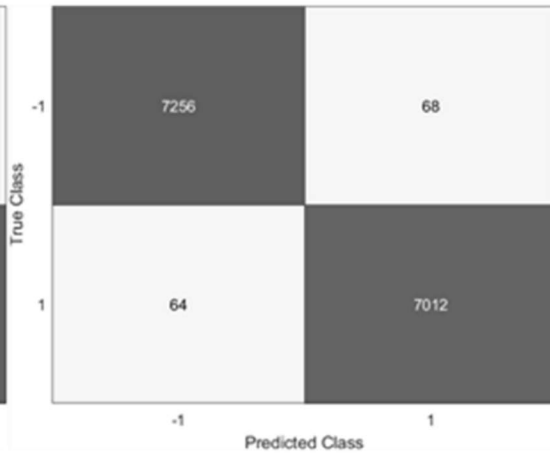
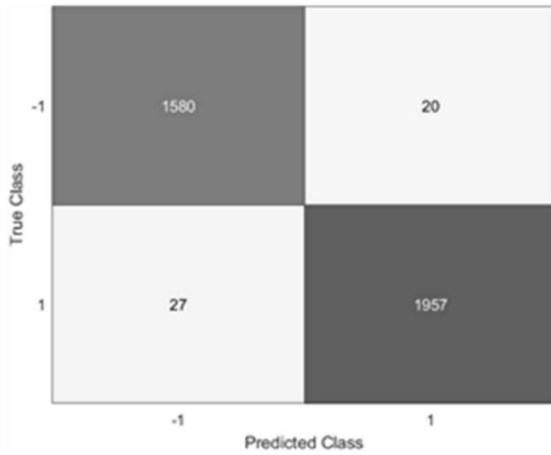
RBF



B

A

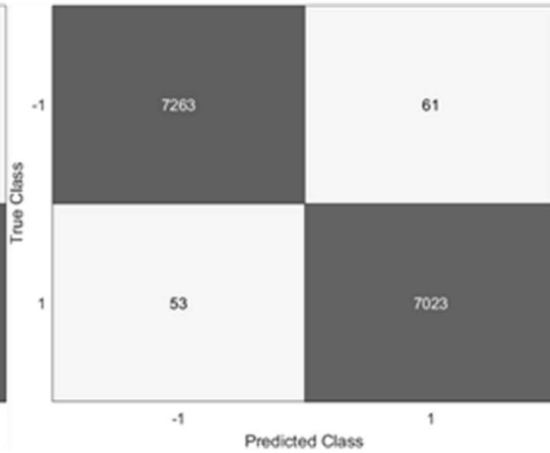
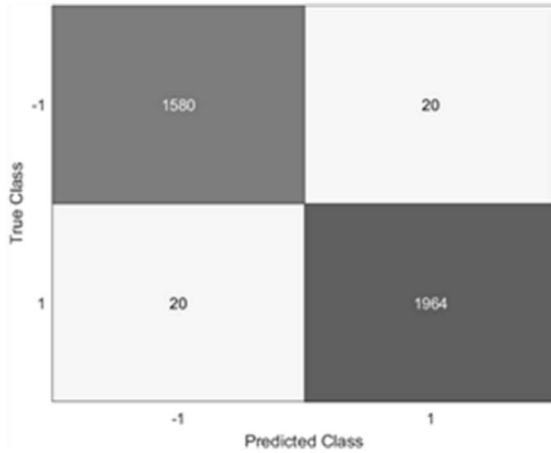
Linear



B

A

Polynomial

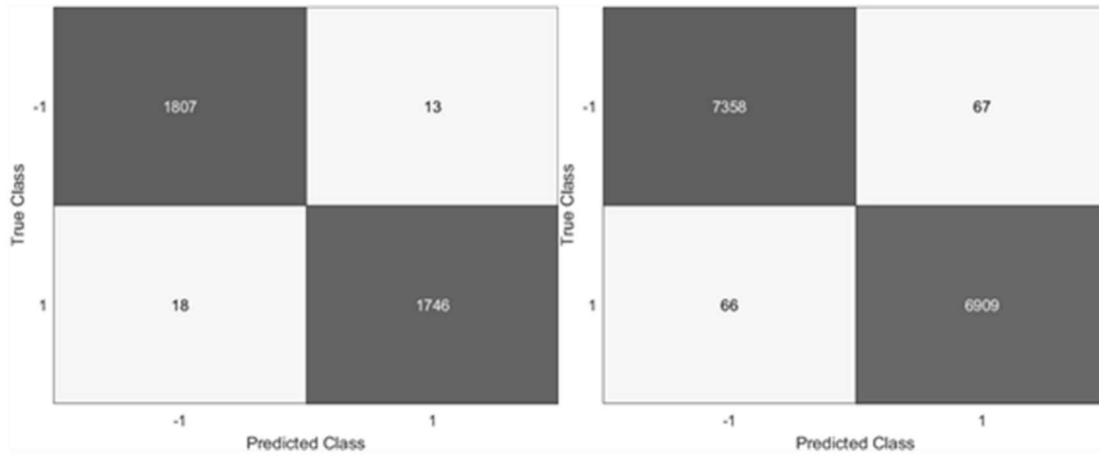


B

A

الشكل M . 6: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الثالثة)

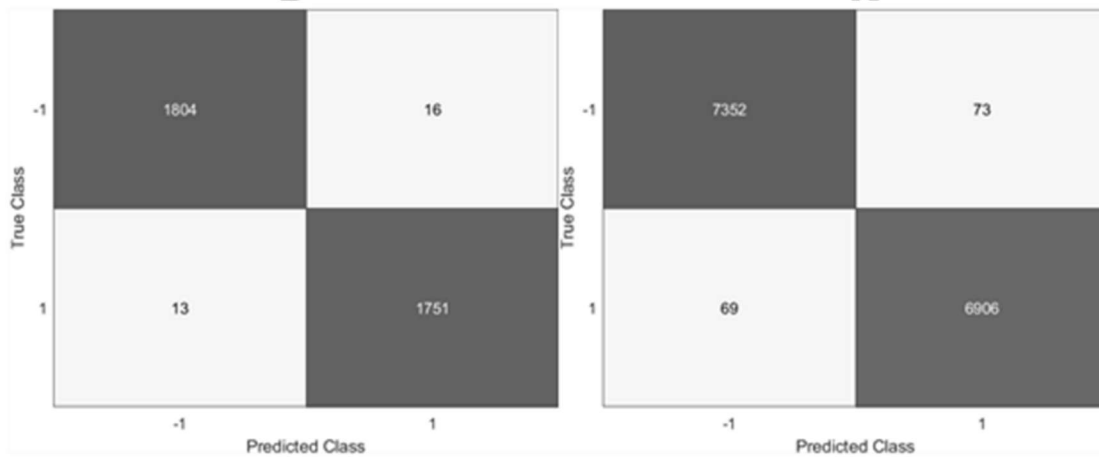
RBF



B

A

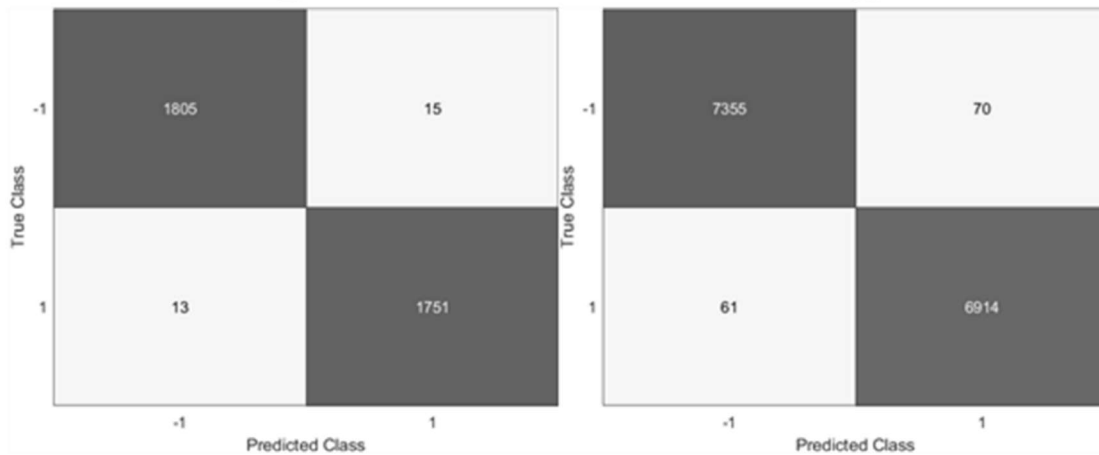
Linear



B

A

Polynomial

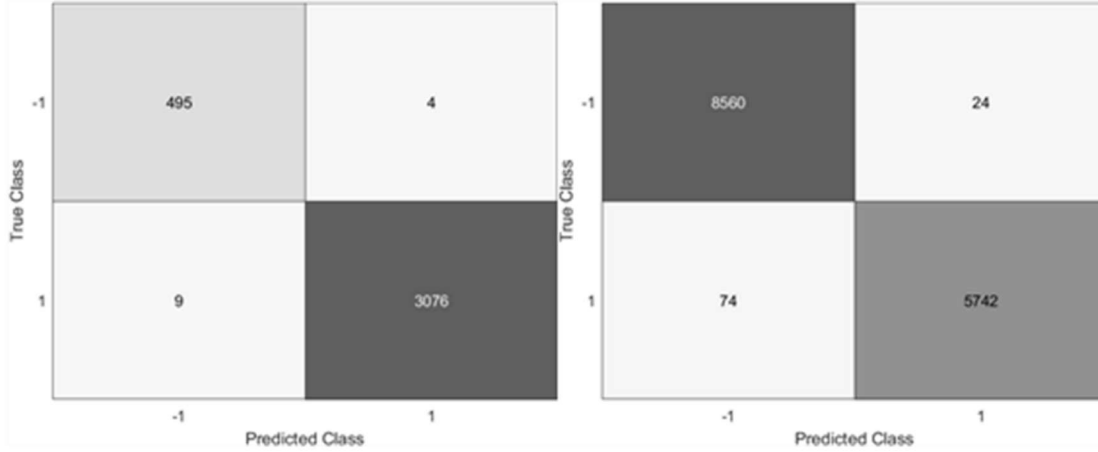


B

A

الشكل M . 7: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الرابعة)

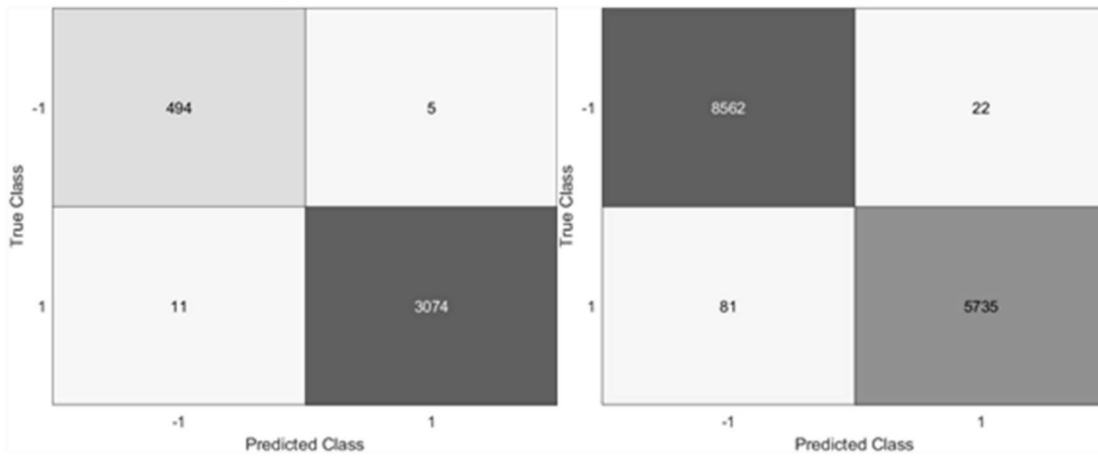
RBF



B

A

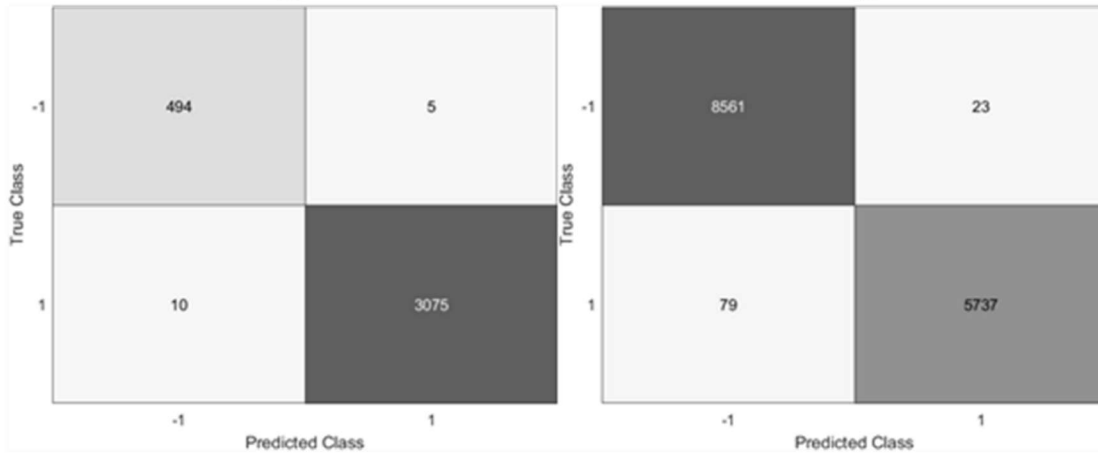
Linear



B

A

Polynomial



B

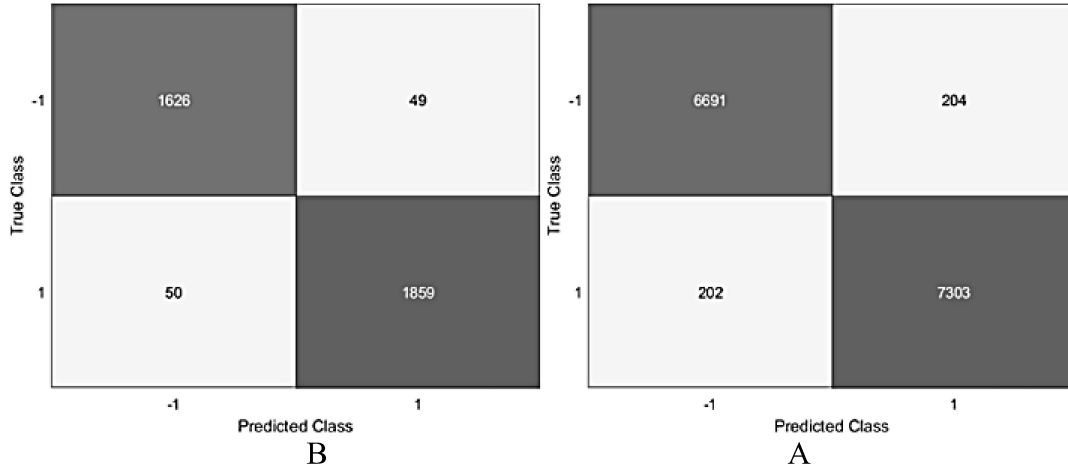
A

الشكل M . 8: مصفوفة الارتباك لبيانات A: التدريب وB: الاختبار (العينة الثانية من السلالة الخامسة)

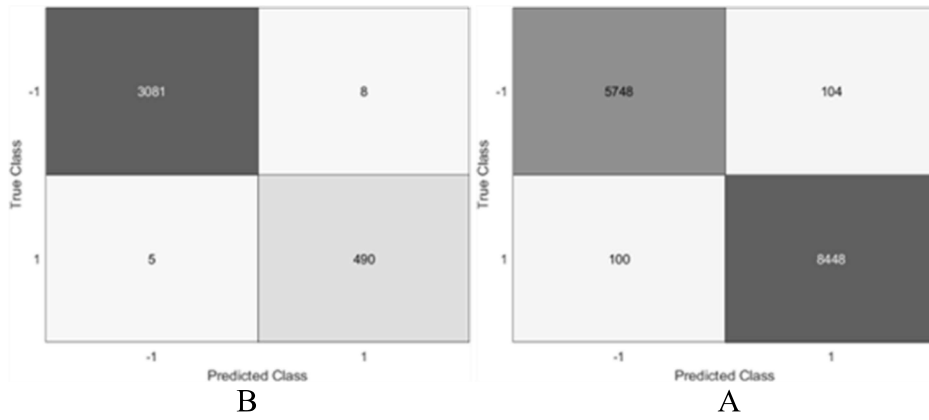
# N المطبق

مصفوفات الارتباك لانموذج GK

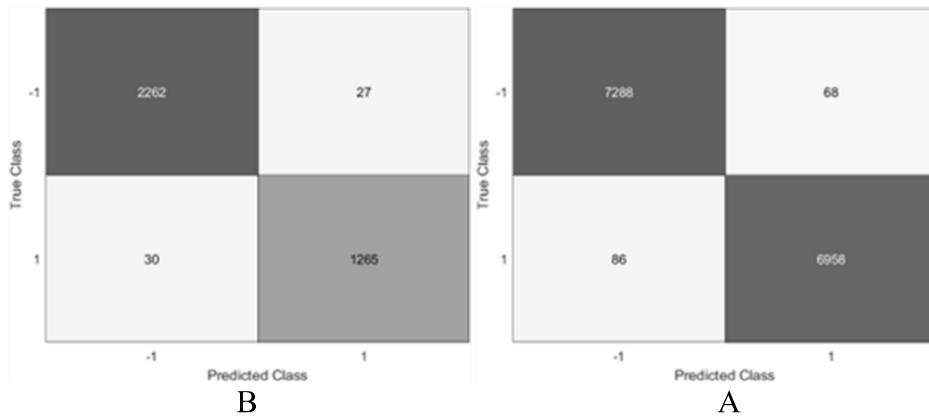
للسلات الثانية والثالثة والرابعة والخامسة



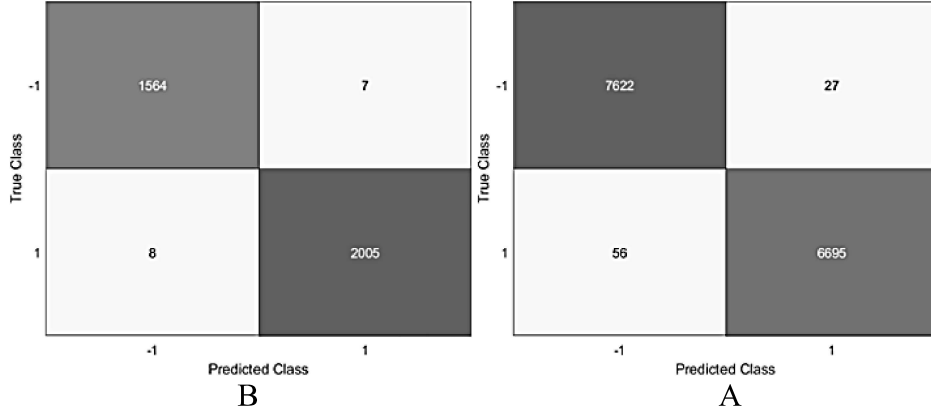
الشكل N . 1: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الاولى من السلالة الثانية)



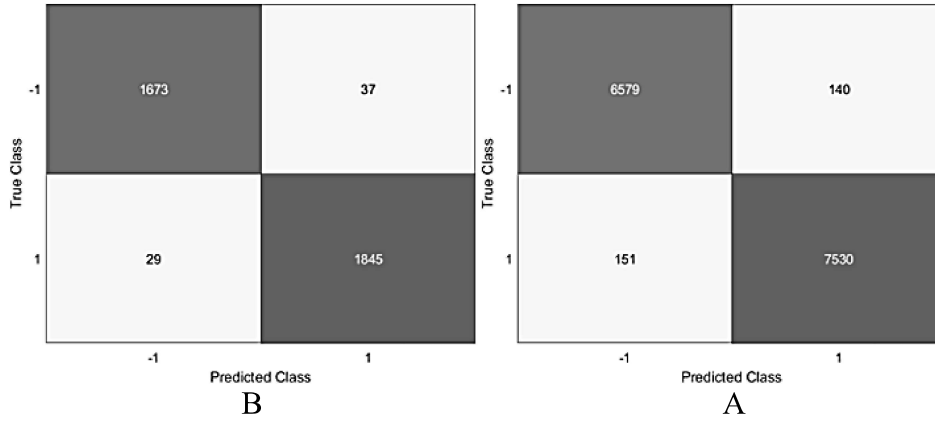
الشكل N . 2: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الاولى من السلالة الثالثة)



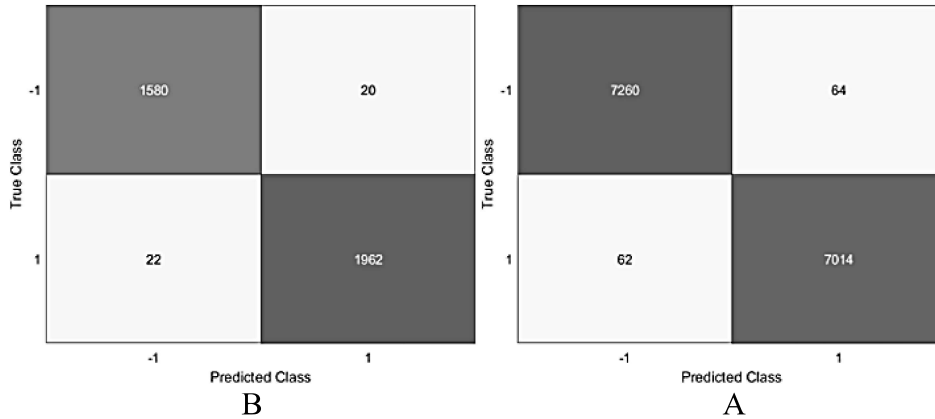
الشكل N . 3: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الاولى من السلالة الرابعة)



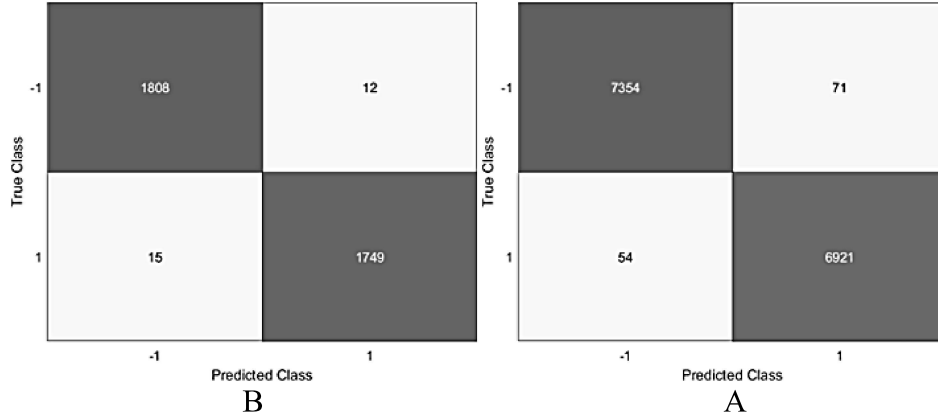
الشكل N . 4: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الاولى من السلالة الخامسة)



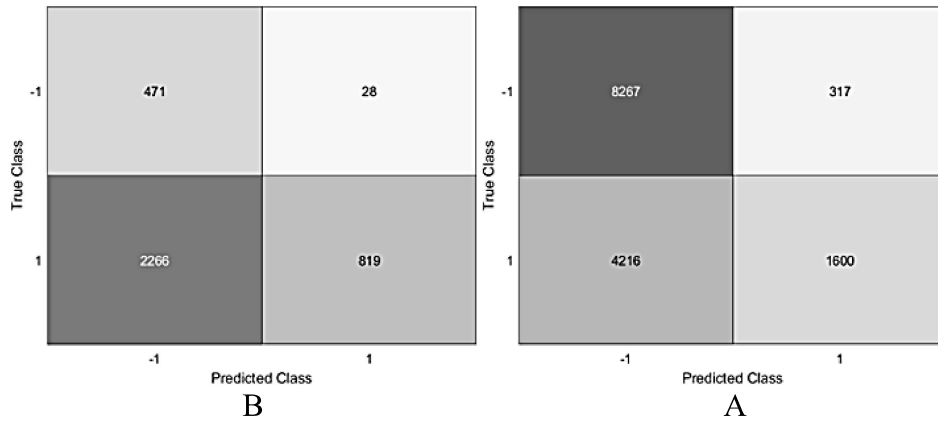
الشكل N . 5: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الثانية من السلالة الثانية)



الشكل N . 6: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الثانية من السلالة الثالثة)



الشكل N . 7: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الثانية من السلالة الرابعة)



الشكل N . 8: مصفوفة الارتباك لبيانات التدريب والاختبار (العينة الثانية من السلالة الخامسة)

## **Abstract**

In this thesis, time series were classified using Logistic Regression (LR), Perceptron Neural Network (PNN), Support Vector Machine (SVM), K-nearest neighbor (KNN), and Gaussian Kernel Method (GK). Autoregressive (AR) variables were utilized through the rank of the AR (p) model in determining the input variables for the methods used. For the purpose of comparison and increasing accuracy in classification, the Convolutional Neural Network (CNN) was used as one of the special deep learning techniques in image classification, and for the purpose of creating the input variable for the CNN method, the time series data used were converted from their digital form and formed two-dimensional images. This study used time series data for the worm *Caenorhabditis elegans* (CE) with its wave movement angles, which will represent the case of the study. The data used in this thesis is from long time series with a very large number of observations, in addition to the multiplicity of characteristics and characteristics that the time series experiences, which may be a reason for the heterogeneity of the data, which may lead to classifications that are not accurate. Using the CNN method will lead to accurate results; This is because it is suitable for dealing with study data to solve heterogeneity problems through the visual representation of digital data. By comparing the results between the methods used, it was found that the CNN method outperforms the rest of the other methods used. Therefore, it is possible to conclude that using the CNN method that relies on binary image classification leads to accurate classification results compared to other methods that rely on binary digital classification.

**Ministry of Higher Education and  
Scientific Research  
University of Mosul  
College of Computer Science and  
Mathematics  
Department of Statistics and Informatics**



# **Time Series Classification based on Statistical Methods and Deep Neural Network**

**A Thesis Submitted to the Council of the College of  
Computer Science and Mathematics  
University of Mosul  
as a Partial Fulfillment of Requirements  
for the Degree of Master of Science  
in  
Statistics**

**By  
Omar Akram Mohammed Saeed**

**Supervised by  
Assist. Prof. Dr. Osamah Basheer Hannon**

---

**2024 A.D.**

**1445 A.H.**