

## تحديد عتبة التصنيف المثلى ديناميكياً في أنظمة الكشف المبكر عن الشذوذ القائمة على التعلم العميق

المهندس علي ياسين<sup>1</sup> - الدكتور كمال السلوم<sup>2</sup> - الدكتور وسيم رمضان<sup>3</sup>

### المخلص

لازال اكتشاف الشذوذ محط اهتمام الكثير من الباحثين لما له من دور في حل مشكلات العالم الحقيقية مثل الكشف عن الحالات غير الطبيعية ومنها عيوب الأنظمة الصناعية واكتشاف الاحتيال.. الخ.

يهدف البحث إلى بناء نظام عام للكشف المبكر عن حالات الشذوذ، يمكن استخدامه لتقليل من من الخسائر الكبيرة الناجمة عن وجود خلل في بيانات النظام. أهم المشاكل التي تواجه أنظمة كشف الشذوذ الحالية هي عدم قدرتها على (1) استخراج الميزات من بيانات متعددة عالية الأبعاد، (2) تحديد عتبة التصنيف الأمثل بشكل ديناميكي.

إعتمدَ البحث لحل المشاكل السابقة على شبكة الترميز الآلي ذات الذاكرة طويلة المدى (LSTM-Autoencoder) لاستخراج الميزات الأكثر أهمية، والتقاط التبعيات الزمنية لتسلسل الشذوذ ضمن سلاسل البيانات. بالإضافة لنظرية تشبيبيشيف (Chebyshev) لتحديد العتبة الأمثل ديناميكياً.

طُبِقَ النظام المُقترح على مشكلة كسر الورق في معامل صناعة الورق، حيث استطاع اكتشاف 53% من الحالات الشاذة قبل حدوثها بأربعة دقائق، ونسبة إندارات كاذبة لم تتجاوز 9%، والحد من تكاليف الخسارة بنسبة 54%. مما يؤكد قدرة النظام المُقترح على التكيف مع المسألة المطروحة في اكتشاف الشذوذ والتقليل من أثاره.

**الكلمات المفتاحية:** أنظمة كشف الشذوذ، استخراج الميزات، بيانات متعددة عالية الأبعاد، التبعيات الزمنية، عتبة التصنيف، شبكة الترميز الآلي ذات الذاكرة طويلة المدى، نظرية تشبيبيشيف.

<sup>1</sup> طالب دكتوراه - قسم هندسة البرمجيات - كلية الهندسة المعلوماتية - جامعة البعث - حمص - سوريا

<sup>2</sup> أستاذ دكتور - قسم هندسة البرمجيات - كلية الهندسة المعلوماتية - جامعة البعث - حمص - سوريا

<sup>3</sup> دكتور مشارك - قسم الاقتصاد الزراعي - كلية الزراعة - جامعة البعث - حمص - سوريا

# Selecting the Optimal Classification Threshold Dynamically in Early Anomaly Detection Systems Based on Deep Learning

Eng. Ali Yassin, Dr. Kamal Al-Salloum, Dr. Wassim Ramadan

## Abstract

Anomaly detection is still an important subject for many researchers. It has a crucial role in solving real-world problems such as detecting abnormal cases, including defects in industrial systems, and fraud detection...etc.

The research aims to build a general system for early anomaly detection, which can be used to reduce the large losses caused by abnormality in the system data.

The main problems facing current anomaly detection systems are their inability to 1) extract features from high-dimensional, multiple data, 2) selecting the optimal classification threshold dynamically.

In order to solve the problems above, the research employed an LSTM-Autoencoder network to extract the most important features and capture the temporal dependencies of the occurrence of anomalies within the data series. In addition to Chebyshev's theory for selecting the optimal threshold dynamically.

The proposed system was applied to the problem of paper breaks in paper factories, where it was able to detect 53% of the anomalies four minutes before its occurrence, with a rate of false alarms that did not exceed 9%, and reduce the costs of loss by 54%. This confirms the ability of the proposed system to adapt to the issue at hand in detecting anomalies and minimizing its effects

**Keywords:** Anomaly Detection Systems, Feature Extraction, Chebyshev Theorem, High-Dimensional Multiple Data, Time Dependencies, Threshold Classification, LSTM-Autoencoder.

---

## 1- مقدمة

يعدّ اتخاذ القرارات الصائبة عاملاً رئيسياً في نجاح أي عمل تجاري أو صناعي مهما يكن، لذلك تعدّ مشكلة دعم اتخاذ القرارات المناسبة في إدارة سلسلة التوريد (Supply Chain Management) مصدر قلق كبير لعدد من الدراسات الحديثة [1][2][3]. تتحكم إدارة سلسلة التوريد (SCM) بجميع العمليات التي تُحوّل المواد الخام إلى منتجات نهائية.

يوجد مجموعة من الإجراءات التي تُساعد في دعم اتخاذ القرارات ضمن سلاسل التوريد. يُعتبر كل من التنبؤ (Forecasting) واكتشاف الشذوذ (Anomaly Detection) مهمتين أساسيتين في إدارة سلاسل التوريد (SCM). حيث تُساعد عملية التنبؤ في إدارة المواعيد المحددة وتخطيط الإنتاج والتسعير. لكن تؤدي كميات البيانات الهائلة التي يتم إنشاؤها في كل مرحلة من مراحل سلسلة التوريد، إلى زيادة تحميل البيانات وصعوبة تمييز الإشارات المفيدة التي تُتيح اتخاذ قرارات ذات مغزى. في المقابل تسمّح أساليب اكتشاف الشذوذ بتحديد تلك الإشارات الشاذة الخارجة عن نطاق البيانات الطبيعية.

تهدف عملية الكشف عن الحالات الشاذة [4] إلى تحديد كل نقاط البيانات التي تسلك سلوكاً مختلفاً عن باقي نقاط المجموعة. يُمكن أن تنتج الحالات الشاذة عن خطأ في البيانات؛ ولكنها تدل أيضاً على عمليات أساسية جديدة لم تكن معروفة مسبقاً وغالباً ما تكون حرجة في مجموعة واسعة من التطبيقات. يوجد إسقاط كبير لمفهوم اكتشاف الشذوذ في تطبيقات العالم الحقيقي كتطبيقات اكتشاف الاحتيال المالي، الاختراقات الشبكية، الأمراض النادرة، اكتشاف العيوب الصناعية وهو ما يتم التركيز عليه في الدراسة الحالية. توجّهت العديد من شركات القطاع الصناعي مؤخراً لاستخدام أساليب كشف الشذوذ لتطوير خطوط الإنتاج. تعدّ معامل صناعة الورق من أكثر القطاعات توجّهاً نحو ذلك، حيث يُنتج العالم في الوقت الحالي نحو 400 مليون طن من الورق، لتصنيع الاحتياجات الأساسية من النقود، علب الورق المقوى، الإيصالات، ورق الزبدة، الخ. بينما يبلغ استهلاك الفرد في الجمهورية العربية السورية 12 كيلو غرام من الورق في العام الواحد. يتطلّب إنتاج اللب والورق عمليات معقدة ومتكاملة للغاية بوسائل كيميائية أو ميكانيكية

[[5]]. كما تَتَطَلَّبُ مراقبة الأنظمة باستمرار حتى يَتِمَّكَّنَ المشغلون من إدارة العمليات واكتشاف أي حوادث محتملة، قد تَتَسَبَّبُ في انقطاع إنتاجي مفاجئ. لِتَحَقِيقِ ذلك يَتِمُّ نشر مجموعة كبيرة من أجهزة الاستشعار في أجزاء مختلفة من معدات التصنيع، لقياس متغيرات العملية الهامة ومراقبة حالة النظام. تولد هذه المستشعرات كميات كبيرة من بيانات السلاسل الزمنية متعددة المتغيرات (Multivariate Time Series). تحتوي السلسلة على أكثر من متغير تابع (Dependent Variable). لا يَتَعَمَدُ كُلُّ متغير فقط على قيمه السابقة، بل يَتَعَمَدُ أيضاً على بعض المتغيرات الأخرى ضمن السلسلة[[6]]. تُعَبَّرُ السلاسل الناتجة عن قراءات المقاييس على طول خط الإنتاج.

يجب معالجة هذه القراءات الأولية وتحليلها بشكل مناسب، للحصول على معلومات مفيدة بشأن الحالة الطبيعية للنظام، حيث يَتِمُّ بناء أنظمة قادرة على تحليل وتفسير بيانات المراقبة، لاكتشاف أي فشل مُمَكِّن في المستقبل. تُدْعَى هذه الأنظمة أنظمة كشف الشذوذ.

يُمكن تعريف أنظمة كشف الشذوذ على أنها أجهزة أو برامج تقوم بمراقبة البيانات المتعلقة بمجال مُعين لتحديد الحالات الشاذة والمختلفة عن السلوك الطبيعي باستخدام طرائق معينة مثل التعلُّم العميق (Deep Learning).

وعلى النقيض من ذلك، تَتَمَثَّلُ أحد أكبر القضايا التي تواجه الأنظمة الحالية لكشف العيوب (الشذوذ) الصناعية؛ أنه يجب توقع فشل الآلة قبل حدوثه. حيث كانت تُركِّز الأنظمة التقليدية، على اكتشاف الأسباب المؤدية للحالات غير الطبيعية للنظام [[7]] في أقرب وقت ممكن بعد حدوثها. يُساعد الكشف المبكر ولو قبل بضع ثوانٍ على تجنب أي انقطاع مُحتمَل في سير عملية الإنتاج. فعلى سبيل المثال تُسَبَّبُ مشكلة كسر الورق (Paper Breaks) توقف الآلات لأكثر من ساعة عن كُلِّ عملية كسر، وتُقَدَّرُ تكلفة الخسارة الناتجة عن التوقف بحوالي 15 مليون دولار عن كُلِّ خط إنتاج[[8]].

يُساعد الاعتماد على السلاسل الزمنية متعددة المتغيرات، في نمذجة الترابطات المثيرة للاهتمام والتنبؤ بحالات الشذوذ المستقبلية. سَاعَدَتِ التطورات السريعة في الآونة الأخيرة للشبكات العصبونية العميقة [[9]] في توفير أداة قوية للتعامل مع هذا النوع من

البيانات، كاخترالها في بُعد أقل. يمكن اعتبار شبكات الذاكرة طويلة المدى LSTMs (Long Short-Term Memory) الحل الأكثر فعالية للتسلسلات الزمنية، حيث تَمَيَّزَ عن الشبكات العصبونية التقليدية بامتلاكها ذاكرة تساعد على تذكر الأنماط لفترات طويلة من الزمن.

لكن على الجانب الأخر، تَتَطَلَّبُ الشبكات العصبونية العميقة قدراً كبيراً من البيانات الشاذة، وهذا غير ممكن في العديد من تطبيقات العالم الحقيقي، كما هو الحال بالنسبة لمشكلة البحث الحالية. حيث تشتمل مجموعة البيانات فقط على 124 نقطة زمنية لانهايار الآلات (كسر الورق) ضِمْنَ أكثر من 18000 نقطة زمنية، مما يجعل مجموعة البيانات غير متوازنة للغاية. يؤدي ذلك إلى صعوبة إنشاء نماذج ذات كفاءة عالية نظراً لعدم توفر بيانات شاذة كافية لعملية التدريب، وعلى الرغم من ذلك فإن التخفيض ولو بنسبة قليلة من أحداث كسر الورق سيعطي توفيراً كبيراً في تكلفة الإنتاج.

ومن جهة أخرى، تَتَطَلَّبُ أنظمة كشف الشذوذ القائمة على الشبكات العصبونية تحديد العتبة (Threshold) لفصل الحالات الشاذة والطبيعية ضِمْنَ البيانات، تَعَمَّدُ معظم الدراسات على تحديد قيمة العتبة تجريبياً وبشكل ثابت (Static)؛ ولكن من خلال الاعتماد على فرضية أن متجهات البيانات (Data Vectors) تتبَّع توزيع غاوس متعدد المتغيرات، فيمكن استخدام تابع احتمالي مثل Likelihood لتقدير قيمة العتبة [10]. بالمقابل يبدو في العديد من المواقف العملية أنه من الصعب الاحتفاظ بهذا الافتراض لأن توزيع البيانات (Data Distribution) غالباً ما يكون غير معروف. يمكن تَجَنُّبُ وضع أي افتراضات حول توزيع البيانات من خلال تطبيق طريقة غير معلمية (Nonparametric Method)، وبالتالي يُتيح لهذه الأنظمة تَعْيِيرُ قيمة العتبة بشكل ديناميكي (Dynamic) في حال حدوث اختلاف بحجم وتوزيع البيانات. تُعتبر نظرية تشيبشيف (Chebyshev's Theory) طريقة غير معلمية، فهي تحدد الحد الأدنى لنسبة الملاحظات (Observations) التي تقع ضِمْنَ عدد محدد من الانحرافات المعيارية عن المتوسط.

إِعْتَمَدَ البحث لحل المشاكل السابقة على استخدام تقنيات التعلُّم العميق وتحديداً شبكة الترميز الآلي ذات الذاكرة طويلة المدى (LSTM-Autoencoder) ونظرية

تشبيبيشيف؛ لبناء نظام ديناميكي للكشف عن الحالات الشاذة قبل حدوثها بوقت مناسب. حيث تم تجريبه على اكتشاف حالات فشل آلة الورق قبل حدوثها بأربعة دقائق. بالإضافة الى تحديد قيم البارامترات الفائقة (Hyperparameters) للنظام المقترح لتحقيق أعلى نسبة كشف لحالات كسر الورق.

## 2- أهمية وأهداف البحث

تساعد أنظمة كشف الشذوذ ضمن السلاسل الزمنية متعددة المتغيرات، على اكتشاف الحالات الشاذة قبل حدوثها. لكن بالمقابل يتطلب ذلك الاعتماد على تقنيات أكثر تعقيداً مثل بعض خوارزميات التعلم العميق؛ لقدرتها على استخراج الميزات الهامة من جهة، ومن جهة أخرى لقدرتها على تذكر الأنماط المعقدة والخفية لفترات طويلة.

يهدف البحث بشكل أساسي إلى بناء وتطوير نظام ديناميكي للكشف المبكر عن

الحالات الشاذة ضمن السلاسل الزمنية، وذلك من خلال تحقيق ما يلي:

- 1- توقع الحالات الشاذة ضمن مجموعة البيانات قبل حدوثها.
- 2- النقاط التبعيات الزمنية لتسلسل الشذوذ ضمن سلاسل البيانات.
- 3- إيجاد أفضل قيم البارامترات الفائقة للنظام المقترح.
- 4- إيجاد عتبة التصنيف الأمثل ديناميكياً، لتحقيق أفضل دقة كشف ممكنة.

## 3- مواد وطرائق البحث

### 3-1- البيانات البحثية

تمثل البيانات المستخدمة مجموعة بيانات حقيقية تتعلق بمشكلة كسر الورق. قُدمت مجموعة بيانات البحث ضمن مسابقة البيانات المقامة في معهد أنظمة المهندسين الصناعيين (Institute of Industrial and Systems Engineers) لعام 2019. تم جمع المشاهدات من أحد معامل صناعة اللب والورق [11] على مدار شهر واحد باستخدام مجموعة متنوعة من أجهزة الاستشعار. نُقيس هذه المستشعرات المواد الخام

(مثل كمية الألياف الليبية والمواد الكيميائية وغيرها)، ومتغيرات العملية (مثل نوع الشفرة وسرعة الدوران). تتضمّن مجموعة البيانات قراءات 61 مستشعر عند 18398 نقطة زمنية ضمّن فترات منتظمة (كل دقيقتين)، مع تحديد حالة النظام (طبيعي أو شاذ) عند كل نقطة. على الرغم من هذا العدد الكبير من القياسات، إلا أن حالات الفشل تحدث فقط عند 124 نقطة زمنية (0.67% من إجمالي الملاحظات) أثناء التشغيل مما يجعل من الصعب التنبؤ بالفشل قبل حدوثه؛ لكن أي تقليل لحالات الفشل بواسطة الاكتشاف المبكر، يوفر قدراً كبيراً من تكاليف الانتاج. يُلخص الجدول 1 مجموعة بيانات البحث.

الجدول 1 وصف مجموعة البيانات

Element	No.	Remark
Continuous ) المتغيرات المستمرة (variables)	59	X1 ~X27, X29 ~X60
(Categorical variables) المتغيرات الفئوية	2	X28 (8 categories), X61 (2 categories)
(Normal) الحالات الطبيعية	18274	2-minute intervals
(Abnormal) الحالات الشاذة	124	-

### 3-2- الحزم والمكتبات المستخدمة Used Packages and Libraries

تُمنّى بناء نظام الكشف عن الشذوذ في مجموعة البيانات السابقة باستخدام لغة بايثون Python وبالاعتماد على مجموعة من المكتبات البرمجية [12] وأهمها:

- **Keras**: توفر هذه الحزمة واجهات برمجية متناسقة وبسيطة من أجل التخاطب مع المستخدم النهائي وليس الآلة ، وتحتوي على مجموعة من النماذج (Models) مثل الشبكات العصبونية وأشجار القرار وتوابع التنشيط، كما تتميز بقابليتها للتوسع أي القدرة على إضافة نماذج جديدة. تكون المهمة الأساسية للمكتبة جعل التطبيق أكثر استجابة مع إمكانية إعطاء المستخدم المزيد من القدرة على التحكم به.

● **TensorFlow** : تُطبّق هذه الحزمة في العديد من المجالات كحساب المشتقات والمصفوفات الضخمة بالإضافة إلى استخدامها في توزيع العمليات الحاسوبية على وحدات المعالجة المركزية CPU، وكذلك على شبكة موزعة مكونة من مجموعة أجهزة بعيدة تتضمّن هذه المكتبة. يُستخدم TensorFlow بشكل أساسي في تعلم الآلة في الوقت الحالي.

### 3-3 - مقاييس الأداء Performance Metrics

يعتبر مقياس الدقة Accuracy المقياس الأكثر استخداماً والأكثر منطقية في تقييم أداء خوارزميات التصنيف وذلك عندما تكون مجموعات البيانات متوازنة. فهو يمثل النسبة بين عدد العينات المصنفة بشكل صحيح وعدد العينات الكلية. ولكن بالمقابل لا يمكن اعتباره مقياساً جيداً في البيانات غير المتوازنة (كما في حالة اكتشاف الشذوذ) لأنه يسبب ملائمة زائدة (Overfitting) لجهة صف الأغلبية.

يوجد مجموعة واسعة من المقاييس التي يمكن استخدامها لتقييم أداء أنظمة الكشف عن الشذوذ [13] ونذكر منها هنا فقط ما تم استخدامه لتقييم الأداء كونها الأكثر استخداماً في تقييم أنظمة اكتشاف الشذوذ:

#### 1. مصفوفة الارتباك Confusion Matrix

يُتنبأ النموذج بصف عينات البيانات وينسب إلى كل عينة تسميتها (Label) المتوقعة (إيجابية أو سلبية)، لتقع كل عينة في نهاية المطاف ضمن أحد الحالات التالية:

(1) الإيجابيات الحقيقية (True Positives (TP) تدل على العينات الإيجابية المتوقعة بشكل صحيح.

(2) السلبيات الحقيقية (True Negatives (TN) تدل على العينات السلبية المتوقعة بشكل صحيح.

(3) الإيجابيات الخاطئة (False Positives (FP) والسلبيات الخاطئة (False Negatives (FN) عندما يتعارض الصف الحقيقي مع الصف المتوقع.

يُمكن تلخيص الحالات السابقة ضمن مصفوفة  $2 \times 2$  تسمى مصفوفة الارتباك

$$.M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}$$

## 2. الدقة Precision والاستذكار Recall

يُمكن تمثيل العينات الإيجابية والسلبية في مصفوفة الارتباك على إنها عينات ذات صلة أو غير ذات صلة بالبيانات الشاذة على التوالي. بالتالي يُعبر مقياس الاستذكار (Recall) -مقياس الدقة في صف العينات الإيجابية، معدل الإيجابيات الحقيقية (TPR)- عن نسبة العينات ذات الصلة التي تم استرجاعها بشكل صحيح، بينما يُعرف مقياس الدقة (Precision) على أنه نسبة العينات المسترجعة ذات الصلة. يوفر كل من زوج دقة واستذكار رؤية مفيدة في سلوك المصنف، ويؤدي الجمع بينهما إلى إنشاء مقاييس تقييم ذات فعالية وعلى وجه التحديد مقياس F1 والذي يمثل المتوسط التوافقي بين كل من الدقة والاستذكار.

## 3. مقياس F1 Score

يُعدّ F1 Score في التحليل الإحصائي مقياساً لدقة الاختبار، فهو المتوسط التوافقي للدقة (Precision) والاستذكار (Recall). تحسب قيمة F1 بالاعتماد على مصفوفة الارتباك بالشكل التالي:

$$F1\ score = \frac{2 * TP}{2 * TP + FP + FN} = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$

تمثل القيمة 1 أفضل نتيجة، بينما 0 أسوأها.

## 4. معدل الإيجابيات الخاطئة False Positive Rate

يُقَيِّس معدل الإيجابيات الخاطئة (FPR) النسبة بين عدد الحالات السلبية التي صنفت على أنها إيجابية (الإيجابيات الخاطئة FP) وإجمالي عدد الحالات السلبية بغض النظر عن تصنيفها.

$$FPR = \frac{FP}{FP + FN} \quad (2)$$

### 3-4 - نظرية تشيبشيف (Chebyshev's Theory)

إعتمدَ البحث لتحديد قيمة عتبة التصنيف ديناميكياً في نظام كشف الشذوذ المُقترح على اختبارات غير معلميه، تقادياً لوضع أي فرضية حول توزيع البيانات. ركزَ البحث الحالي على استخدام نظرية تشيبشيف كطريقة غير معلميه.

تنص النظرية [14] على أنه يوجد على الأقل نسبة  $1 - \frac{1}{k^2}$  من البيانات تقع ضمن  $k$  انحراف معياري عن المتوسط  $\mu \pm k\sigma$ ، حيث  $k > 1$ . وذلك من أجل مجموعة واسعة من التوزيعات الاحتمالية.

تطبق النظرية على متجهات الخطأ  $(e_t)$  للبيانات الطبيعية، لتحديد المجال الذي تقع ضمنه معظم هذه المتجهات. تمثل متجهة الخطأ الفرق بين السلسلة الزمنية المتوقعة والحقيقة في الزمن  $t$ .

$$e_t = ||\hat{x}_t - x_t|| \quad (3)$$

يختار النظام قيمة عتبة التصنيف ضمن المجال المُحدّد، حيث تم ضبط القيمة الافتراضية للعتبة على متوسط المجال، بينما القيمة العليا على الحد الأعلى للمجال.

### 4 - الدراسة المرجعية

قدّمت أدبيات الدراسات السابقة العديد من الطرائق والخوارزميات لمعالجة مسائل التصنيف ضمن السلاسل الزمنية أحادية المتغير (Univariate Time Series). حيث اعتمدت منهجية هجينة لتحسين دقة مصنّفات السلاسل الزمنية باستخدام نماذج ماركوف المخفية (Hidden Markov Models) [15]. كما تم مقارنة ثلاث منهجيات من خوارزميات أشجار القرار (Decision Tree) الشائعة، وهي الغابات العشوائية (Random Forest)، AdaBoost، Multi Boost [16] في اكتشاف حالات عدم انتظام ضربات القلب ضمن السلاسل البيولوجية الزمنية (Biomedical Time-Series)، وأشارت النتائج إلى تفوق AdaBoost على المصنّفات الأخرى. أُستخدِمت

مؤخراً الشبكات العصبونية الالتفافية (Convolutional Neural Networks) في مسائل تصنيف السلاسل الزمنية أحادية المتغير [17]، بهدف دمج كل من مهام التصنيف واستخراج الميزات (Feature Extraction) في إطار عمل واحد.

لُقَّت السلاسل الزمنية متعددة المتغيرات في الآونة الأخيرة اهتماماً كبيراً لدى العديد من الباحثين. حيث تمَّ معالجة التحديات التي تواجهها [18] ومن أهمها: (1) تدفقات البيانات عالية الأبعاد، (2) الحالات الشاذة الطارئة على عمل النظام، وذلك من خلال بناء نظام مراقبة تدفق بيانات حقيقية. إستطاع النظام المُقترح استخراج الميزات من خلال دراسة الارتباطات بين عناقد (Cluster) البيانات، واكتشاف الحالات الشاذة باستخدام تحليل المكونات الرئيسية (Principal Component Analysis)، لكن بالمقابل لم يتمَّ تحديد عتبة التصنيف بشكل ديناميكي، فضلاً عن عدم استخدام خوارزميات التعلُّم العميق.

حقَّقت خوارزميات التعلُّم العميق نتائج بارزة في مجال تصنيف السلاسل الزمنية متعددة المتغيرات. تَفَوَّقت الشبكات العصبونية العميقة على الأساليب القائمة على الميزات في تعلُّم ميزات السلاسل الزمنية [19]، تمَّ تطوير شبكة عصبونية التفافية متعددة القنوات (Multi-Channels Convolution Networks) لتصنيف السلاسل الزمنية ضمن مجموعتين من بيانات العالم الحقيقي، حيث تتعلَّم الشبكة الميزات من سلاسل زمنية أحادية المتغير في كل قناة، ومن ثم يتمَّ جمع المعلومات من جميع القنوات لتمثيل الميزات المهمة في الطبقة النهائية. أظهرت النتائج التجريبية قدرة التعلُّم العميق على تعلُّم الميزات الأكثر أهمية، مما يُساعد على تحسين أداء التصنيف. بالمقابل لم تستخدم الدراسة تقنيات ضبط البارامترات الفائقة، بالإضافة لتحديد عتبة التصنيف بشكل يدوي.

تمَّ مقارنة أربع نماذج مختلفة للتعلُّم العميق [20] وهي: الشبكات العصبونية الاصطناعية (Artificial Neural Networks)، والشبكات العصبونية التكرارية

(Recurrent Neural Networks)، والوحدة المتكررة ذات البوابات ( Gated Recurrent Units)، والذاكرة طويلة المدى (LSTM) في مسائل كشف الاحتيال المالي القائم على التسلسل. أظهرت النتائج تفوق LSTM على النماذج الأخرى، وتوصلت الدراسة إلى تحديد الطوبولوجيا المناسبة لكل نموذج باستخدام ضبط البارامترات الفائقة. بالمقابل لم تتطرق الدراسة إلى موضوع تحديد عتبة التصنيف بشكل ديناميكي.

أستُخدمت الذاكرة طويلة المدى (LSTM) لبناء نموذج كشف الشذوذ ضمن طلبات المستهلك في إداة سلسلة التوريد [21]]. حَقَّقَت الشبكة تفوقاً على خوارزمية آلة شعاع الدعم (Support Vector Machine) في دقة الكشف، كما اغتمدت الدراسة على طريقة توقع الكميات (Quantiles Estimation) لتحديد قيمة العتبة. على الرغم من تفوق النموذج المقترح في كشف الحالات الشاذة، إلا إنه من المُمكِن تطويره من خلال التحكم بتدفق البيانات واختزالها قدر الإمكان، من خلال بناء نموذج هجين بين LSTM وأحد الشبكات العصبونية العميقة مثل Autoencoder و CNN.

تمَّ بناء عدة نماذج هجينة باستخدام شبكات عصبونية عميقة لتصنيف السلاسل الزمنية متعددة المتغيرات، ومن أهمها شبكة الذاكرة طويلة المدى (LSTM) والشبكة الالتفافية (CNN) وشبكة الترميز الآلي (Autoencoder). أستخدمت الشبكة (CNN-LSTM) لبناء نموذج اكتشاف الأنماط غير الطبيعية [22] في حركة بيانات السلاسل الزمنية عبر الويب. حيث تُستخدم CNN لتقليل أبعاد الميزات المكانية ( Spatial Features)، بينما يكون الهدف من LSTM نمذجة معلومات الوقت. حَقَّقَ النموذج المقترح أداء غير مسبوق للكشف عن الحالات الشاذة حتى مع البيانات المتشابهة جداً. بالرغم من تفوق هذه الطريقة، إلا أن دقتها تميل للانخفاض عند اكتشاف القيم الشاذة التي يتم إنشاؤها عند حدود نوافذ البيانات، كما هناك تأخير زمني في كشف الحالات الشاذة.

إستطاع النموذج (LSTM-CNN) التفوق على جميع النماذج الأخرى [23] في التعامل مع السلاسل الزمنية. حيث لا يحتاج للكثير من عمليات المعالجة المسبقة للبيانات، وذو كفاءة عالية في وقت الاختبار. تمّ اختبار النموذج على أكثر من 30 مجموعة بيانات متنوعة ومن ضمنها بيانات صناعية. لكن على الجانب الأخر، لم يتمّ ضبط البارامترات الفائقة، بالإضافة لعدم تحديد العتبة للنموذج بشكل ديناميكي.

تمّ الاعتماد على خصائص كل من شبكتي Autoencoder و LSTM لبناء نموذج هجين لكشف الاحتيال القائم على التسلسل ضمن مجموعة بيانات حقيقية وصناعية [24]. أظهرت النتائج أن لدمج الشبكتين السابقتين تأثير كبير في زيادة دقة الكشف مقارنة مع استخدام كل شبكة على حدة. حيث يتميز النموذج (LSTM-Autoencoder) بقدرته على التعامل مع البيانات عالية الأبعاد من جهة، ومن جهة تذكر الأنماط المعقدة لفترات أطول. بالمقابل لم تعتمد الدراسة على أي منهجية لتحديد العتبة بشكل ديناميكي.

تمّ ملاحظة التفوق الكبير من خلال الدراسات السابقة للشبكات العصبونية العميقة في مجال اكتشاف الشذوذ ضمن السلاسل الزمنية متعددة المتغيرات على باقي الخوارزميات، كما أكدت الدراسات على ضرورة التوجه لبناء نماذج هجينة بين هذه الشبكات لتحقيق أعلى دقة كشف. لكن وبنفس الوقت تمّ ملاحظة القصر الحاصل فيما يتعلق ببناء هذه النماذج ضمن اتجاهين هُما: (1) ضبط البارامترات الفائقة، (2) تحديد عتبة التصنيف بشكل ديناميكي. استوجب كل ذلك المزيد من البحث لبناء نظام ديناميكي للكشف عن الحالات الشاذة قبل حدوثها بوقت مناسب. يكون قادر على اختزال الميزات عالية الأبعاد، والتقاط التبعيات الزمنية لتسلسل الشذوذ ضمن سلاسل البيانات. بالإضافة لتحديد عتبة التصنيف الأمثل بشكل ديناميكي. لتحقيق ذلك كان لابد من الاعتماد على النماذج الهجينة بين شبكات التعلم العميق، وإيجاد أفضل قيم البارامترات الفائقة لها.

## 5- الخوارزميات المقترحة للدراسة

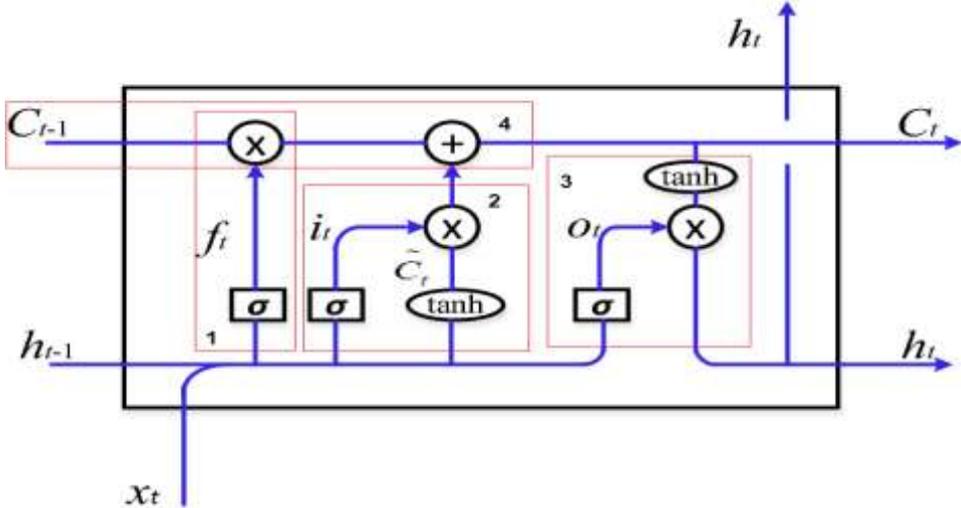
بناء على ما تمّ ملاحظته من الدراسات المرجعية السابقة، قد تمّ التركيز في هذا البحث على دراسة بعض خوارزميات التعلّم العميق التي أعطت أفضل أداء في اكتشاف الشذوذ.

### 5-1- شبكة الذاكرة طويلة المدى Long Short-Term Memory

تعدّ شبكة LSTM أحد أنواع الشبكات العصبونية المتكررة، والتي تسمح بالاحتفاظ بالتبعية طويلة المدى بين البيانات عند نقطة زمنية معينة، انطلاقاً من عدة خطوات زمنية سابقة. تحتوي LSTM على ما يسمى حالة الخلية (Cell State) وهي تمثل ذاكرة الشبكة، ويتمّ في كلّ خطوة الاحتفاظ بالمعلومات الضرورية والتخلص من المعلومات غير ذات الصلة [25] عن طريق بوابة النسيان (Forget Gate)، وبوابة الإدخال (Input Gate)، وبوابة الإخراج (Output Gate).

#### 5-1-1- بنية وآلية عمل شبكة الذاكرة طويلة المدى

يوضح الشكل 1 بنية شبكة LSTM [26] والتي تتكون من ثلاث بوابات.



الشكل 1 بنية شبكة LSTM

#### 1. بوابة النسيان Forget Gate

يُمثل المربع الأول (1) بوابة النسيان: يَنَمُّ في هذه البوابة الإجابة عن السؤال التالي: ماهي حالة الخلية السابقة التي يجب الاحتفاظ بها أو نسيانها؟

$$f_t = \sigma(w_f [h_{t-1}, x_t]) + b_f \quad (4)$$

حيث،  $f_t$ : مُؤَيَّر (Tensor) يستخدم كمرشح (Filter) للمعلومات الموجودة في حالة الخلية،  $\sigma$ : تابع Sigmoid،  $x_t$ : حالة الدخل الجديد في اللحظة  $t$ ،  $h_{t-1}$ : الخرج السابق لحالة الخلية في الزمن  $t - 1$ .

يكون خرج البوابة إما 0 أو 1، حيث تُقابل القيمة 0 نسيان الحالة الداخلية (Internal State) بشكل كامل، بينما تُقابل القيمة 1 تمرير الحالة بدون تغيير.

### 2. بوابة التحديث (الادخال) Update Gate

يُمثل المربع الثاني (2) بوابة الادخال: يَنَمُّ في هذه البوابة تحديد القيم المرشحة (Candidate) للدخل الحالي للاحتفاظ بها في حالة الخلية الجديدة.

$$c_t = \tanh(w_c [h_{t-1}, x_t]) + b_c \quad (5)$$

$$i_t = \sigma(w_i [h_{t-1}, x_t]) + b_i \quad (6)$$

يُستخدم التابع Tanh من أجل عملية تنظيم (Regulates) النموذج. حيث يُنتج عن التابع متجهة تحتوي على كل القيم (الحالات) التي يمكن إضافتها إلى حالة الخلية الجديدة، بينما يكون الهدف من استخدام تابع Sigmoid تحديد ماهي المعلومات الحالية التي يتم الاحتفاظ بها أو نسيانها.

### 3. بوابة الخرج Output Gate

يَنَمُّ في المربع الرابع (4) تحديث حالة الخلية السابقة  $c_{t-1}$  إلى حالة الخلية الجديدة  $c_t$ ، من خلال ضرب الخلية السابقة بخرج بوابة النسيان، ومن ثم إضافة القيم المرشحة الجديد الناتجة عن بوابة الإدخال.

$$c_t = f_t * c_{t-1} + i_t * c_t \quad (7)$$

حيث: تمثل  $c_t * i_t$  قيم مرشحة (خرج بوابة الإدخال)،  $f_t$  خرج بوابة النسيان

يُمثل المربع الثالث (3) بوابة الإخراج: يَتَمُّ في هذه البوابة الإجابة عن السؤال

التالي: ماذا سوف يكون خرج حالة الخلية الجديدة  $h_t$  ؟

$$O_t = \sigma(w_o [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = O_t * \tanh(c_t) \quad (9)$$

## 5-2-2 - شبكة الترميز الآلي Autoencoder Network

شبكة عصبونية متعدّدة الطبقات، غير خاضعة للإشراف (Unsupervised)، ذات تغذية متقدمة (Feed-Forward). تتكون من طبقة إدخال، وطبقة إخراج، ومُرَمِّز (Encoder)، ومفكك ترميز (Decoder)، وفضاء كامن (Latent Space). يَتَمُّ تغذية دخل الشبكة بشعاع مكوّن من ميزات (Features) مجموعة بيانات الدخل، ومن ثمّ تَعْمَلُ الشبكة على إعادة إنتاج هذا الشعاع في طبقة الخرج. يَتَمُّ تدريب الشبكة بالاعتماد خوارزميات الانتشار الخلفي (Backpropagation)، وتابع خسارة (Loss Function) لحساب الخسارة الناتجة عن إعادة بناء شعاع الدخل في طبقة الخرج **Error!** [27]. [Reference source not found].

## 5-2-1 - بنية وآلية عمل شبكة الترميز الآلي

نُفرض بدايةً عند تصميم بنية شبكة Autoencoder وجود مشكلة عنق الزجاجة (Bottleneck) في المنتصف مما يفرض تمثيل مضغوط للمدخلات الأصلية؛ لذلك تُعتبر الشبكة مناسبة جداً في مجموعات البيانات ذات الأبعاد الكبيرة حيث تَعْمَلُ على اختزال البيانات في بُعد أصغر مع الاحتفاظ بالمعلومات الرئيسية لهيكلية البيانات.

كما ذكرنا سابقاً أن الشبكة تتكون من ثلاث مكونات رئيسية **Error!** **Reference source not found.** هي: المُرَمِّز (Encoder)، الترميز (Code)، مفكك الترميز (Decoder). كما تحتاج إلى ثلاث أشياء: طريقة الترميز، طريقة فك الترميز، وتابع الخسارة.

يُعتبر المُرَمِّز ومفكك الترميز شبكة عصبونية ذات تغذية متقدمة، ومتصلة بالكامل (Fully-Connected) فيما بينها. يقوم المُرَمِّز بضغط البيانات الأصلية إلى بُعد أصغر، بينما يَعْمَلُ مفكك الترميز بفك ترميز البيانات المضغوطة وإعادة بناءها إلى بعدها الأصلي. يُمثل الترميز طبقة واحدة في الشبكة بحجم اختياري يَتَمُّ تحديده قبل عملية

التدريب (يسمى أيضاً عنق الزجاجاة). حيث يتَّحَكَم حجم هذه الطبقة ببعْد المدخلات المضغوطة التي يَنُم تغذيتها إلى وحدة فك الترميز. تُمرر بيانات الإدخال عبر المُرمِّز لضغط البيانات كمرحلة أولى، ومن ثمَّ إلى الترميز لضغطها إلى بُعد أصغر، بينما يقوم مفكك الترميز الذي له بنية مماثلة للمُرمِّز بإعادة إنتاج الدخل في طبقة الخرج بالاعتماد على البيانات المضغوطة ضِمَّنْ طبقة الترميز. يجب التنويه أن يكون لطبقتي الدخل والخرج نفس عدد العصبونات، بينما يمكن تعديل أحجام الطبقات في الوسط كما نريد.

### 5-2-2- اكتشاف الشذوذ باستخدام شبكة الترميز الآلي

يَنُم الكشف عن الشذوذ باستخدام شبكة الترميز الآلي، من خلال تَعَلُّم الأنماط والعلاقات المعقدة ضِمَّنْ البيانات الطبيعية. تُصَبَّح الشبكة بعد عملية التدريب قادرة على إعادة بناء البيانات الطبيعية بشكل جيد للغاية، بينما تقشل في القيام بذلك مع البيانات الشاذة. حيث تُعتبر نقاط البيانات التي لها خطأ إعادة بناء (Reconstruction Error) أعلى من حدِّ معين نقاط شاذة.

### 5-3- شبكة الترميز الآلي ذات الذاكرة طويلة المدى LSTM Autoencoder

يُمكن تنظيم شبكة LSTM Autoencoder بالاعتماد على بنية مشابهة لشبكة Autoencoder، تُسمى LSTM Encoder-Decoder [28] والتي يكون فيها كل من المُرمِّز ومفكك الترميز هو شبكة LSTM. يكون الهدف من استخدام LSTM ضِمَّنْ المُرمِّز ومفكك الترميز هو التقاط التبعية الزمنية (Temporal Dependencies) ضِمَّنْ سلاسل البيانات.

تُساعد شبكة LSTM Autoencoder على تقليل أبعاد البيانات واختزالها في بُعد أصغر مع الاحتفاظ بالمعلومات الرئيسية لهيكلية البيانات باستخدام خصائص شبكة الترميز الآلي، وتَعَلُّم الأنماط المعقدة داخل الترتيب الزمني لتسلسل الإدخال باستخدام شبكة الذاكرة طويلة المدى. كما تساعد في المحافظة على التبعية بين مكونات متجهة الخرج المتوقع.

### 5-3-1- اكتشاف الشذوذ باستخدام شبكة LSTM AUTOENCODER

تُشاهد الشبكة التسلسلات الطبيعية فقط أثناء عملية التدريب وتتعلم إعادة بنائها، مما يجعلها قادرة على اكتشاف الشذوذ في السلاسل الزمنية متعددة المتغيرات. فعندما يتم تغذيتها بتسلسل غير طبيعي (شاذ) لا تكون قادرة على إعادة بنائها بشكل جيد. أن تدريب الشبكة بهذه الطريقة له معنى عملي لأن البيانات الشاذة ليست متاحة دائماً، كما أنه من المستحيل تغطية جميع انماط هذه البيانات.

### 6- التجارب والنتائج Experiments and Results

يُستعرض في هذا القسم بدايةً وصفاً تجريبياً لتصميم التجارب التي أجراها البحث لبناء أنظمة الكشف المبكر عن الشذوذ (اكتشاف كسر الورق) باستخدام خوارزميات التعلم العميق المقترحة، وبالاعتماد على أساليب ضبط البارامترات الفائقة لتحديد أفضل قيم البارامترات، ونظرية تشبيشيف لتحديد قيمة عتبة التصنيف بشكل ديناميكي. يتبع التوصيف استعراض النتائج وتقييم الأداء.

### 6-1- بناء النموذج Model Creation

تتضمن مراحل بناء النموذج مجموعة من الإجراءات وهي كما يلي:

#### 1. تحضير البيانات Data Preprocessing

يتم معالجة وتحضير البيانات قبل عملية تطبيق النموذج المقترح. يجب بدايةً تقسيم البيانات بأكملها إلى مجموعات تدريب واختبار (Training and Test)، ونظراً لأن الشبكات المقترحة ضمن البحث، تحتاج للبيانات الطبيعية فقط أثناء عملية التدريب، لذا فأولاً يتم وضع جميع العينات الشاذة ضمن بيانات الاختبار، وبعد ذلك تقسم البيانات الطبيعية إلى 80% للتدريب و20% للاختبار (مجموعة الاختبار: 20% من البيانات طبيعية + كل البيانات الشاذة).

وبعد عملية التَّقْسِيم، يُنمَّ تَقْيِيس (Standardization) بيانات التدريب، ثم تَقْيِيس بيانات الاختبار باستخدام المتوسط والانحراف المعياري لبيانات التدريب بعد تَقْيِيسها، مما يُبْقِي بيانات الاختبار غير معروفة أثناء النمذجة.

يُهدَف البحث إلى اكتشاف الشذوذ قبل حدوثه، تَنَمُّثُ أحد الأساليب لتحقيق ذلك بإزاحة تصنيف النقطة الزمنية المقابلة لحالة الشذوذ إلى للأعلى بمقدار  $s$ ، بحيث تكون المصنفات قادرة على التنبؤ بحالة النظام الحالية ( $c_t$ ) قبل  $s$  وحدة زمنية.

$$c_t \leftarrow c_{t+s}, \text{ where } s = 1, 2, \dots \quad (10)$$

تُمنَّ تَعْيِين  $s = 2$  ضِمَّنَ البحث الحالي، مما يعني أن النموذج قادر على توقع الشذوذ قبل حدوثه بمقدار 4 دقائق (يفصل بين كل قراءة دقيقتين). تُطبَّق عملية الإزاحة على كُلِّ من بيانات التدريب والاختبار.

## 2. النافذة الزمنية Time Window

يحتاج النموذج حتى يَتَمَكَّنَ من النقاط التبعيات الزمنية خلال فترة معينة، أن يَعْتَمِدَ على جميع النقاط الزمنية ضِمَّنَ تلك الفترة. لذا يَتَمَّ الاحتفاظ بمجموعة من النقاط الزمنية المتتالية باستخدام مفهوم النوافذ الزمنية (Time Windows). تَضِمَّنَ النافذة بحجم  $m$ ، النقاط الزمنية المقابلة لأخر  $m$  قراءة وصولاً للوقت الحالي  $t$ ، يَتَمَّ تصنيف مثل النافذة (Window Instance)، بالاعتماد على تصنيف آخر نقطة زمنية في النافذة ( $c_t$ ).

يُسْتَخْدَمُ مثل النافذة في تدريب الشبكات العصبونية، بالإضافة لمعالجة مشكلة البيانات غير المتوازنة (Imbalanced Data)، من خلال إنشاء مثل النافذة عند جميع النقاط الزمنية المقابلة لحالة الشذوذ ( $c_t = 1$ )، وبعض النقاط الزمنية المقابلة للحالة الطبيعية ( $c_t = 0$ )، مما يُقلِّصُ الفرق بين صفوف البيانات.

تُمنَّ تَطْبِيقِ حجوم مختلفة للنافذة على بيانات التدريب والاختبار ضِمَّنَ البحث وهي  $m = 3, 5, 7, 9$ ، وبالتالي فإن الشبكة تكون قادرة على التقاط التبعيات الزمنية ضِمَّنَ 6، 10، 14، 18 دقيقة على الترتيب.

### 3. ضبط البارامترات الفائقة Tuning Hyperparameters

تَعْنِي عملية ضبط (أو تَحْسِين) البارامترات الفائقة إيجاد مجموعة قِيم البارامترات التي تُحَقِّق أفضل أداء للنموذج. رَكَزَ هذا البحث على استخدام طريقة البحث العشوائي لضبط البارامترات الفائقة، حيث أن البحث العشوائي (Random Search) [29] Error! Reference source not found. أكثر كفاءة في تَحْسِين البارامترات من البحث الشبكي (Grid Search) من الناحية التجريبية والنظرية.

يَعْمَل البحث العشوائي بشكل أفضل عندما يكون فضاء البحث عالي الأبعاد، ويحتوي على عدد كبير من التركيبات المختلفة للبارامترات الفائقة. حيث تَبَحَّث تقنية البحث العشوائي ضِمْنَ مجموعة عشوائية من هذه التركيبات لاختيار أفضل القيم للبارامترات الفائقة الخاصة بالنموذج المُقْتَرَح، وبالتالي فإن الوقت المُستَعْرَق للعثور على المجموعة الصحيحة يكون أقل مع عدد أقل من التكرارات. فيما يلي وصفاً مختصراً لجميع البارامترات الفائقة التي أَعْتَمَدَهَا البحث من أجل تطوير عمل النموذج المُقْتَرَح.

#### • تابع التنشيط (Activation Function)

يُحَدَد آية تحويل الأوزان الموزونة (Weights Weighted) للمدخلات إلى قيمة محددة تمثل خرج العُقد. يوجد مجموعة واسعة من توابع التنشيط أهمها:  
(1) Logistic (Sigmoid): يحدد قِيمَةَ الخرج ضمن المجال  $[0, 1]$ ، والصيغة الرياضية للتابع هي بالشكل التالي:

$$1.0/(1.0 + e^{-x}) \quad (11)$$

(2) Hyperbolic Tangent (Tanh): يحدد قِيمَةَ الخرج ضمن المجال  $[-1, 1]$ ، والصيغة الرياضية للتابع هي بالشكل التالي:

$$(e^x - e^{-x})/(e^x + e^{-x}) \quad (12)$$

(3) Rectified Linear Activation (ReLU): يحافظ على قِيمَ الإدخال الموجبة، بينما يكون الخرج مساوياً للصفر من أجل قِيمَ الإدخال السالبة. الصيغة الرياضية للتابع هي بالشكل التالي:

$$\max(0.0, x) \quad (13)$$

• **معدل التعلّم (Learning Rate)**

يُشير إلى مقدار خطوة الانتشار الخلفي عند تحديث الأوزان. يتم ضبطه بقيمة صغيرة جداً.

• **حجم الدفعة (Batch Size)**

يمثل عدد العينات التي يتم تغذية الشبكة بها في كل مرة. أن الحجم النموذجي هو 32، ويكون الحجم من مضاعفات العدد 2.

• **معدل الحذف (Dropout Rate)**

يُفيد في إزالة بعض العقد ضمن الطبقة والتي تكون زائدة وغير مجدية.

• **مرحلة (Epoch)**

يمثل مقدار الوقت الذي تحتاجه الشبكة من أجل أن تتدرب على مجموعة البيانات بأكملها.

• **العقد (Nodes)**

تمثل حجم العصبونات في طبقات الشبكة. يتم ضبط الحجم ضمن النموذج المُقترح إلى النصف في كل طبقة تالية في المُرّمز، ومن ثمّ مضاعفته رجوعاً في مفكك الترميز. فعلى سبيل المثال، من أجل حجم 32 وشبكة مكونة من ثلاث طبقات ضمن المُرّمز ومفكك الترميز، تكون العقد بالشكل التالي [32,16,8,8,16,32]

4. **بناء النموذج**

كما تمّ ذكره يتم بناء نموذج الكشف المبكر عن الحالات الشاذة، باستخدام شبكة الترميز الآلي ذات الذاكرة طويلة المدى (LSTM–Autoencoder). من ثمّ مقارنته مع شبكة الترميز الآلي التقليدية (Autoencoder).

تمّت تسمية النماذج المُقترحة للكشف الديناميكي المبكر عن الشذوذ، كما هو موضح في الجدول التالي.

الجدول 2 أسماء النماذج المُقترحة ضمن البحث

Name	Description
LSTM–AEDT	LSTM–Autoencoder with a Dynamic Threshold

### 5. تحديد العتبة Threshold Selection

تُستخدم عتبة التصنيف ( $T$ ) للفصل بين الحالات الشاذة والطبيعية. إن اختيار قيمة عتبة مناسبة دوراً مهماً في عمَل أنظمة الكشف، يتم تحديد العتبة ديناميكياً ضمن النظام المقترح. تُصنف حالة النظام الحالية  $c_t$  ذو متجهة الخطأ  $e_t$  بالاعتماد على قيمة العتبة المحددة بالشكل التالي:

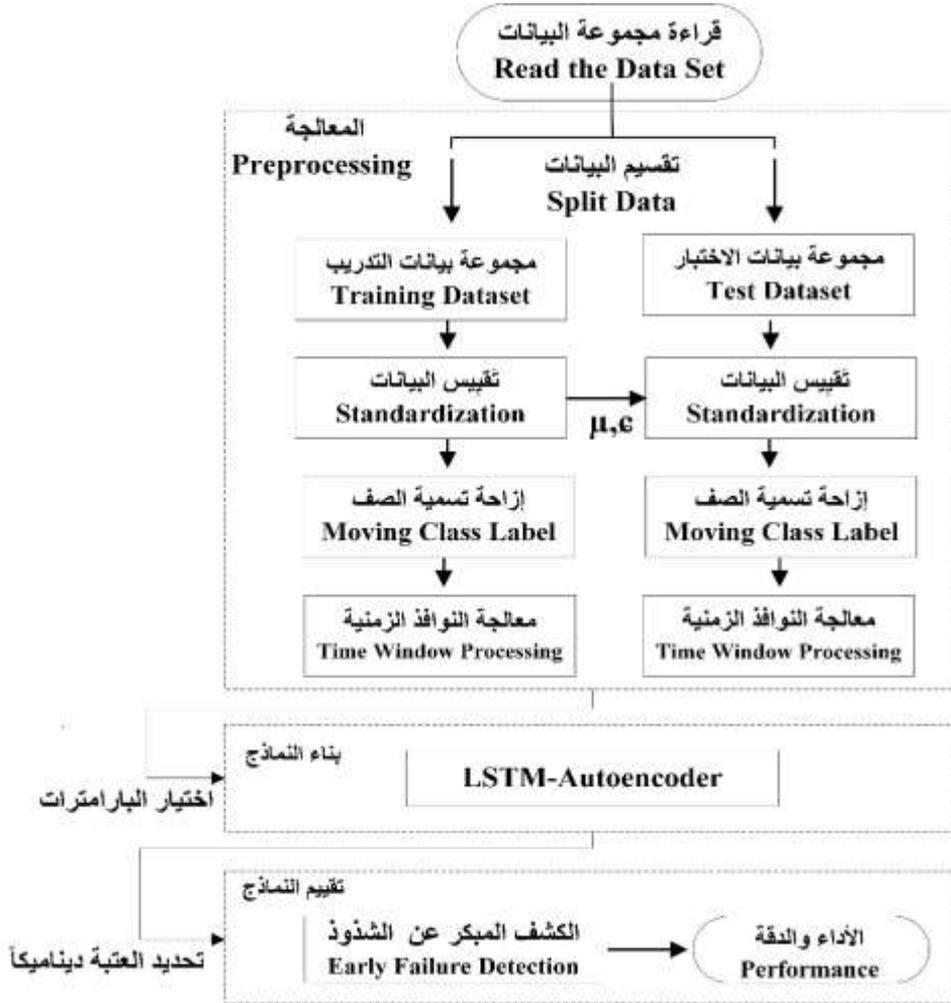
$$c_t \rightarrow \text{normal if } e_t \leq T$$

$$c_t \rightarrow \text{anomaly if } e_t > T$$

### 6. اختبار النموذج

المرحلة الأخيرة وسيتم مناقشتها في القسم التالي.

يُلخّص المخطط التالي مراحل العمل لبناء نظام كشف شذوذ مبكر (كشف فواصل الورق كمثال تطبيقي على النموذج).



الشكل 2 النظام المقترح لاكتشاف الشذوذ

## 2-6-2- تقييم الأداء Performance Evaluation

### 1-2-6- نموذج LSTM-AEDT

رَكَزَ البحث على دراسة أداء النموذج المقترح باستخدام حجوم مختلفة للنافذة الزمنية، وذلك لمعرفة مدى تأثيرها على كُلاً من نسبة الكشف وزمن التنفيذ. حيث يتم بدايةً إيجاد قيم البارامترات الفائقة اللازمة لبناء النموذج، ومن ثم مناقشة النتائج وتفاضل العتبة بالاعتماد على المقاييس المقترحة.

## 1 ضبط البارامترات الفائقة

تُظهر النتائج المُبيّنة في الجدول 3 أفضل تكوينات البارامترات الفائقة لنموذج LSTM-AEDT، باستخدام ضبط البارامترات الفائقة، ومن أجل حجوم النافذة الزمنية المختلفة.

الجدول 3 البارامترات الفائقة لنموذج LSTM-AEDT

Hyperparameters Configurations						Window Size
Neurons	#Epochs	#Batch	Learn Rate	Dropout	Activation	
32	200	256	0.001	0.0	Tanh	3
32	50	128	0.01	0.0	Relu	5
32	50	32	0.01	0.0	Tanh	7
32	50	32	0.001	0.2	Tanh	9

## 2 تفاضل العتبة (Threshold Tradeoff)

يتم التفاضل بين قيم العتبة حسب طبيعة المسألة، فإذا كان الهدف من نظام الكشف إسترجاع أكبر قدر ممكن من الحالات الشاذة، دون التركيز على دقة التصنيف، يتم اختيار العتبة التي تُحقق أفضل قيمة لمقياس Recall. أما عندما يكون دقة التصنيف عاملاً مهماً، فإنه يتم اختيار العتبة التي تُحقق أفضل قيمة لمقياس Precision. يمكن جمع الحالتين السابقتين (إسترجاع أكبر قدر من الحالات الشاذة، وأفضل دقة تصنيف) من خلال اختيار العتبة التي تُحقق أفضل قيمة لمقياس F1Score.

بالرُجوع إلى نظام كشف كسر الورق المُقترح ضمن البحث، يكون الهدف غالباً هو إسترجاع أكبر قدر ممكن من حالات الكسر، تقادياً لتوقف الآلات لساعات طويلة، حيث يبقى زمن فحص الحالة والتأكد من صحتها، أقل تكلفة مادية وزمنية من حدوث الكسر، لكن بالمقابل يجب ألا تتجاوز نسبة الإيجابيات الخاطئة (FPR) حدّ معين موصى به.

• من أجل حجم نافذة  $m = 3$

تُظهر النتائج المُبيّنة في الجدول 4 أداء نموذج LSTM-AEDT في اكتشاف الشذوذ (كسر الورق). نلاحظ من خلال الجدول أن النموذج يقوم بتحديد عتبة التصنيف بشكل ديناميكي ضمنَ المجال  $[0.2, 1.7]$ ، حيث أن القيمة الافتراضية للعتبة و1.7 هي القيمة العظمى.

الجدول 4 نتائج نموذج LSTM-AEDT من أجل  $m = 3$

Threshold	F1Score	Recall	Precision	FPR	Run Time
Default:0.2	0.54	<b>0.59</b>	0.54	0.74	2.5 min
0.5	<b>0.56</b>	0.54	0.58	0.27	
0.8	0.53	0.52	0.56	0.18	
1.1	0.54	0.52	<b>0.60</b>	0.13	
1.4	0.54	0.52	0.59	0.11	
Max:1.7	0.53	0.51	<b>0.60</b>	<b>0.11</b>	

تصل أفضل قيمة لمقياس Recall إلى 59% عند العتبة  $(T = 0.2)$ ، وقيمة Precision إلى 60% عند العتبة  $(T = 1.1, 1.7)$ ، أما بالنسبة لمقياس F1Score فإن أفضل قيمة تصل إلى 56% عند العتبة  $(T = 0.5)$ ، بينما كانت أقل نسبة للإيجابيات الخاطئة FPR هي 11% عند العتبة  $(T = 1.7)$ .

استطاع نموذج LSTM-AEDT المُقترح من أجل حجم نافذة زمنية  $m = 3$  كشف 51% من حالات الكسر قبل 4 دقائق من حدوثها ( $\text{Recall} = 0.51$ ) وبنسبة إنذارات خاطئة 11% ( $\text{FPR} = 0.11$ ) عند العتبة  $(T = 1.7)$ ، علماً أن زمن التنفيذ 2.5 دقيقة.

• من أجل حجم نافذة  $m = 5$

تُظهِر النتائج المُبيّنة في الجدول 5 أداء نموذج LSTM-AEDT من أجل حجم نافذة  $m = 5$ . حيث نلاحظ أن النموذج يقوم بتحديد عتبة التصنيف بشكل ديناميكي ضمن المجال  $[0.2, 1.4]$ ، علماً أن القيمة الافتراضية للعتبة و1.4 هي القيمة العظمى.

الجدول 5 نتائج نموذج LSTM-AEDT من أجل  $m = 5$

Threshold	F1Score	Recall	Precision	FPR	Run Time
Default:0.2	0.53	<b>0.61</b>	0.54	0.84	
0.5	<b>0.56</b>	0.55	0.57	0.41	
0.8	<b>0.56</b>	0.53	0.58	0.22	3.2 min
1.1	0.55	0.52	<b>0.60</b>	0.15	
Max:1.4	0.54	0.52	<b>0.60</b>	<b>0.10</b>	

تصل أفضل قيمة لمقياس Recall إلى 61% عند العتبة  $(T = 0.2)$ ، وقيمة Precision إلى 60% عند العتبة  $(T = 1.1, 1.4)$ ، أما بالنسبة لمقياس F1Score فإن أفضل قيمة تصل إلى 56% عند العتبة  $(T = 0.5, 0.8)$ ، بينما كانت أقل نسبة للإيجابيات الخاطئة FPR هي 10% عند العتبة  $(T = 1.4)$ .

حقّق نموذج LSTM-AEDT المُقترح من أجل حجم نافذة زمنية  $m = 5$  كشف 52% من حالات الكسر قبل 4 دقائق من حدوثها  $(\text{Recall} = 0.52)$  وبنسبة إنذارات خاطئة 10%  $(\text{FPR} = 0.1)$  عند العتبة  $(T = 1.4)$ ، بالمقابل ارتفع زمن تنفيذ النموذج إلى 3.2 دقيقة.

• من أجل حجم نافذة  $m = 7$

تُظهِر النتائج المُبيّنة في الجدول 6 أداء نموذج LSTM-AEDT من أجل حجم نافذة  $m = 7$ . حيث نلاحظ أن النموذج يقوم بتحديد عتبة التصنيف بشكل ديناميكي ضمن المجال  $[0.2, 1.4]$ ، علماً أن القيمة الافتراضية للعتبة و1.4 هي القيمة العظمى.

الجدول 6 نتائج نموذج LSTM-AEDT من أجل  $m = 7$

Threshold	F1Score	Recall	Precision	FPR	Run Time
Default:0.2	0.54	<b>0.62</b>	0.54	0.82	
0.5	0.55	0.54	0.55	0.41	
0.8	0.55	0.53	0.57	0.28	5.3 min
1.1	<b>0.56</b>	0.53	<b>0.59</b>	0.21	
Max:1.4	0.54	0.53	0.58	<b>0.09</b>	

تصل أفضل قيمة لمقياس Recall إلى 62% عند العتبة  $(T = 0.2)$ ، وقيمة Precision إلى 59% عند العتبة  $(T = 1.1)$ ، أما بالنسبة لمقياس F1Score فإن أفضل قيمة تصل إلى 56% عند العتبة  $(T = 1.1)$ ، بينما كانت أقل نسبة للإيجابيات الخاطئة FPR هي 9% عند العتبة  $(T = 1.4)$ .

حقّق نموذج LSTM-AEDT من أجل حجم نافذة زمنية  $m = 7$ ، كشف 53% من حالات الكسر قبل 4 دقائق من حدوثها ( $\text{Recall} = 0.53$ ) وبنسبة إنذارات خاطئة 9% ( $\text{FPR} = 0.09$ ) عند العتبة  $(T = 1.4)$ ، بينما ارتفع زمن تنفيذ النموذج إلى 5.3 دقيقة.

• من أجل حجم نافذة  $m = 9$

تُظهر النتائج المُبيّنة في الجدول 7 أداء نموذج LSTM-AEDT من أجل حجم نافذة  $m = 9$ . حيث نلاحظ أن النموذج يقوم بتحديد عتبة التصنيف بشكل ديناميكي ضمن المجال  $[0.2, 2.0]$ ، علماً أن القيمة الافتراضية للعتبة و 2.0 هي القيمة العظمى.

الجدول 7 نتائج نموذج LSTM-AEDT من أجل  $m = 9$

Threshold	F1Score	Recall	Precision	FPR	Run Time
Default:0.2	0.53	<b>0.6</b>	0.54	0.81	
0.5	<b>0.57</b>	0.55	0.57	0.4	
0.8	0.56	0.53	0.59	0.24	
1.1	<b>0.57</b>	0.53	0.6	0.19	<b>9.7 min</b>
1.4	0.56	0.53	<b>0.61</b>	0.16	
1.7	0.55	0.52	0.6	0.13	
Max:2.0	0.52	0.50	0.57	<b>0.12</b>	

تصل أفضل قيمة لمقياس Recall إلى 60% عند العتبة  $(T = 0.2)$ ، وقيمة Precision إلى 61% عند العتبة  $(T = 1.4)$ ، أما بالنسبة لمقياس F1Score فإن أفضل قيمة تصل إلى 57% عند العتبة  $(T = 0.5, 1.1)$ ، بينما كانت أقل نسبة للإيجابيات الخاطئة FPR هي 12% عند العتبة  $(T = 2.0)$ .

حقّق نموذج LSTM-AEDT من أجل حجم نافذة زمنية  $m = 9$  كشف 50% من حالات الكسر قبل 4 دقائق من حدوثها ( $\text{Recall} = 0.5$ ) وبنسبة إنذارات خاطئة 12% ( $\text{FPR} = 0.12$ ) عند العتبة  $(T = 2.0)$ ، بينما ارتفع زمن تنفيذ النموذج إلى 9.7 دقيقة.

أكدت التجارب أن لحجم النافذة دوراً مهماً في تحسين أداء النموذج، لكن من جهة أخرى، يجب المحافظة على حدّ معين لحجم النافذة، حيث أن النموذج يسلك سلوكاً

عكسياً بعد حجم مُعيّن. حَقَّقَ LSTM-AEDT أفضل أداء له من أجل حجم نافذة  $m = 7$ ، من حيث نسبة الكشف (Recall =53%)، والإنذارات الخاطئة (FPR =9%). بينما سلك النموذج بعد حجم نافذة  $m = 7$  سلوكاً معاكساً، فمن أجل حجم نافذة  $m = 9$ ، إنخَفَصَت نسبة الكشف إلى 50%، وازْتَمَعَت الإنذارات الخاطئة إلى 12%.

### (3) تحليل التكاليف والفوائد (Cost Benefit Analysis)

تمَّ تَحْلِيل مساهمة النموذج المُقْتَرَح في الحدّ من الخسارة المالية ضِمَّن الصناعة (مشكلة كسر الورق). حيث تمَّ إحتِسَاب الربح على أساس الاستنكار، والخسارة الناتجة على أساس الإيجابيات الخاطئة، علماً أنه قُدرت تكلفة كُلّ عملية كسر بحوالي 10,000 دولار أمريكي، بينما تكلفة كل إنذار خاطئ بحوالي 100 دولار أمريكي فقط [11]. يُبيِّن الجدول 8 مقدار الخسارة الذي يوفره النموذج المُقْتَرَح من أجل نافذة  $m = 7$  في كل عام.

الجدول 8 تحليل التكلفة والفائدة لنموذج LSTM-AEDT

Item	Gain (by TPR)	Loss (by FPR)	Remark
تكلفة	10,000\$	100\$	-
عدد الحالات	124x12month =1488	(2 minx30x24 hr.x365 d)- 1488=524112	1 year
معدل الحالات	Recall=53%	FPR=9%	Test Result
تكلفة/ سنة	7,886,400\$	-47,170\$	
الإجمالي	7,839,230\$		

نُلاحظ أن نموذج LSTM-AEDT المُقترح إستطاع مع عدد ليس بالكبير من عمليات الكشف المبكر، أن يوفر قدراً كبيراً من تكاليف الخسارة وصلت لحوالي 8 مليون دولار سنوياً عن كل خط إنتاج، وبمعدل 54% من الخسارة الإجمالية.

### 6-2-2- نموذج AEDT

يهدف البحث من خلال استخدام شبكة الترميز الآلي التقليدية في بناء نموذج الكشف، إلى تقديم مقارنة بين الأنظمة القائمة على المنطق الزمني مثل شبكة الترميز الآلي ذات الذاكرة طويلة المدى، والأنظمة التي تعتمد على دراسة النقطة الزمنية الحالية فقط مثل شبكة الترميز الآلي.

#### 1) ضبط البارامترات الفائقة

تُظهر النتائج المُبيّنة في الجدول 9 أفضل تكوينات البارامترات الفائقة لنموذج AEDT، باستخدام ضبط البارامترات الفائقة.

#### الجدول 9 البارامترات الفائقة لنموذج AEDT

Hyperparameters Configurations						
Neurons	#Epochs	#Batch	Learn Rate	Dropout	Hidden Activation	Out Layer Activation
32	200	64	0.001	0.4	Tanh	Linear

#### 2) تفاضل العتبة (Threshold Tradeoff)

تُظهر النتائج المُبيّنة في الجدول 10 أداء نموذج AEDT في اكتشاف الشذوذ لمشكلة كسر الورق. نلاحظ من خلال الجدول أن النموذج يقوم بتحديد عتبة التصنيف بشكل ديناميكي ضمن المجال [0.2,1.4]، حيث أن 0.2 هي القيمة الافتراضية و1.4 هي القيمة العظمى.

### الجدول 10 نتائج نموذج AEDT

Threshold	F1Score	Recall	Precision	FPR	Run Time
Default:0.2	0.49	<b>0.56</b>	0.50	0.84	
0.5	<b>0.52</b>	0.51	0.53	0.42	
0.8	<b>0.52</b>	0.49	0.55	0.25	1.3 min
1.1	0.51	0.49	0.56	0.15	
Max:1.4	0.51	0.49	<b>0.57</b>	<b>0.14</b>	

تُصل أفضل قيمة لقياس Recall إلى 56% عند العتبة ( $T = 0.2$ )، وقيمة Precision إلى 57% عند العتبة ( $T = 1.4$ )، أما بالنسبة لقياس F1Score فإن أفضل قيمة تصل إلى 52% عند العتبة ( $T = 0.5, 0.8$ )، بينما كانت أقل نسبة للإيجابيات الخاطئة FPR هي 14% عند العتبة ( $T = 1.4$ ).  
 استطاع نموذج AEDT كشف 49% من حالات الكسر قبل 4 دقائق من حدوثها (Recall = 0.49) بنسبة إنذارات خاطئة 14% (FPR = 0.14) عند العتبة ( $T = 1.4$ )، بينما وصل زمن تنفيذ النموذج إلى 1.3 دقيقة.

### (3) تحليل التكاليف والفوائد (Cost Benefit Analysis)

يُبين الجدول 11 مقدار الخسارة الذي يوفره النموذج AEDT في كل عام.

#### الجدول 11 تحليل التكلفة والفائدة لنموذج AEDT

Item	Gain (by TPR)	Loss (by FPR)	Remark
تكلفة	10,000\$	100\$	-
عدد الحالات	124x12month =1488	(2 minx30x24 hr.x365 d)- 1488=524112	1 year
معدل الحالات	Recall=49%	FPR=14%	Test Result

تكلفة/ سنة	7,291,200\$	-73,375\$
الإجمالي	7,217,825\$	

إستطاع النموذج، أن يوفر قدراً جيداً من تكاليف الخسارة وصلت لحوالي 7 مليون دولار سنوياً عن كل خط إنتاج، وبمعدل 46% من الخسارة الإجمالية. أكّدت نتائج التجارب العملية لهذا البحث، قدرة النظام المُقترح على اكتشاف الحالات الشاذة قبل حدوثها بوقت مناسب، والتقاط التبعيات الزمنية لتسلسل الشذوذ ضمن سلاسل البيانات، وتحديد عتبة التصنيف ديناميكياً باستخدام أساليب غير معلميه. إستطاع نموذج LSTM-AEDT عند تجريبه على اكتشاف حالات كسر الورق، كشف 53% من الحالات قبل حدوثها بأربعة دقائق، وبنسبة إنذارات خاطئة لم تتجاوز 9%، كما حدّ من تكاليف الخسارة بنسبة وصلت إلى 54%، ومتقوفاً بذلك على نموذج AEDT الذي إستطاع كشف 49% من الحالات الشاذة، والحدّ من تكاليف الخسارة بنسبة 46%.

## 7- الخاتمة والاستنتاجات والتوصيات

ركّز البحث على معالجة مشكلة هامة تواجه أنظمة كشف الشذوذ القائمة على التعلّم العميق، وهي تحديّد عتبة التصنيف ديناميكياً. إستطاع البحث أن يتوصل إلى بناء نظام ديناميكي للكشف عن الحالات الشاذة قبل حدوثها بوقت مناسب. تمّت تسمية هذا النظام بـ: LSTM-AEDT، ويتمتع النظام بقدرته على دراسة الارتباطات بين البيانات عبر الزمن، واستخراج الميزات الأكثر أهمية (اختزل أبعاد البيانات)، بالإضافة إلى تحديّد عتبة التصنيف ديناميكياً. اعتمدَ نظام LSTM-AEDT المُقترح على شبكة الترميز الآلي ذات الذاكرة طويلة المدى، لالتقاط التبعيات الزمنية لتسلسل الشذوذ ضمن سلاسل البيانات باستخدام مفهوم النوافذ الزمنية. بالإضافة لنظرية تشبيثيف من أجل تحديّد المجال الذي يقع ضمنه معظم متجهات الخطأ للبيانات الطبيعية. بالتالي يستطيع النظام اختيار العتبة

الأفضل ضمنَ المجال المُحدّد ديناميكياً، دون وضع فرضيات حول توزيع البيانات التي يقوم بمعالجتها. كما إعتَمَدَ على تقنيات ضبط البارامترات الفائقة لتحقيق أفضل أداء له. تَوَصَّلَت نتائج البحث إلى أن أفضل أداء لنموذج LSTM-AEDT يكون عند اختيار حجم نافذة  $m = 7$ ، وضبط البارامترات الفائقة بالشكل التالي: عدد العصبونات: 32، عدد الدورات: 50، حجم الدفعة: 32، معدل التعلم: 0.01، معدل الحذف: 0.0، تابع التنشيط: Tanh، وعدد طبقات مخفية: 6 طبقات، حيث بَلَغَت نسبة الكشف 53% من الحالات قبل حدوثها، ونسبة إنذارات خاطئة لم تتجاوز 9% عند العتبة ( $T = 1.4$ ). كما استَطَاعَ الحد من الخسارة الإجمالية لِكُلِّ خط إنتاج، بحوالي 8 مليون دولار سنوياً وبمعدل 54%.

كما تَقَوَّقَ نموذج LSTM-AEDT على نموذج AEDT، الذي حَقَّقَ أفضل أداء له بنسبة كشف 49%، ونسبة إنذارات الخاطئة 14% عند العتبة ( $T = 1.4$ )، بينما بَلَغَت نسبة الحد من تكاليف الخسار 46%.

تُؤَكِّد نتائج البحث على أهمية استخدام خوارزميات التعلُّم العميق القائم على الزمن عند التعامل مع بيانات السلاسل الزمنية، حيث حَقَّقَت أداء أفضل من الخوارزميات التي تَعْتَمِدُ فقط على النقطة الزمنية الحالية. كما تُؤَكِّد على ضرورة اتباع أساليب غير معلمية لإيجاد قيم عتبة التصنيف ديناميكياً. وبناء على النتائج التي تم التوصل إليها يمكن أن نقدم بعض المُقْتَرَحَات:

- محاولة التوسع في الدراسة لتشمل مجالات أخرى في كشف الشذوذ.
- استخدام شبكات عصبونية عميقة أخرى بهدف المقارنة.

## المراجع

- [1] Dolgui, A., Ivanov, D., Potryasaev, S., Sokolov, B., Ivanova, M., & Werner, F. (2020). Blockchain-oriented dynamic modelling of smart contract design and execution in the supply chain. *International Journal of Production Research*, 58(7), 2184-2199.
- [2] Ivanov, D., & Dolgui, A. (2020). A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0. *Production Planning & Control*, 1-14.
- [3] What is Supply Chain Management? [Online] Available: <https://www.oracle.com> [Accessed 29 Dec 2021]
- [4] Zenati, H., Romain, M., Foo, C. S., Lecouat, B., & Chandrasekhar, V. (2018, November). Adversarially learned anomaly detection. In 2018 IEEE International conference on data mining (ICDM) (pp. 727-736).
- [5] Bajpai, P. (2015). Basic overview of pulp and paper manufacturing process. In *Green chemistry and sustainability in pulp and paper industry* (pp. 11-39). Springer, Cham.
- [6] Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *International Journal of Information Management*, 57, 102282
- [7] Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.
- [8] Pathak, P., & Sharma, C. (2021). Processes and problems of pulp and paper industry: an overview. *Physical Sciences Reviews*.
- [9] Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 116, 237-245.

- [10] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. arXiv preprint arXiv:1607.00148.
- [11] Ranjan, C., Reddy, M., Mustonen, M., Paynabar, K., & Pourak, K. (2018). Dataset: rare event classification in multivariate time series. arXiv preprint arXiv:1809.10717.
- [12] Python 3.10.0 documentation [Online] Available: <https://docs.python.org/3/> [Accessed 1 May. 2021]
- [13] Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1), 1-17.
- [14] Alsmeyer, G. "Chebyshev's Inequality." (2011): 239-240.
- [15] Esmael, B., Arnaout, A., Fruhwirth, R. K., & Thonhauser, G. (2012, December). Improving time series classification using Hidden Markov Models. In 2012 12th International Conference on Hybrid Intelligent Systems (HIS) (pp. 502-507). IEEE.
- [16] Jović, A., Brkić, K., & Bogunović, N. (2012, August). Decision tree ensembles in biomedical time-series classification. In Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium (pp. 408-417). Springer, Berlin, Heidelberg.
- [17] Cui, Z., Chen, W., & Chen, Y. (2016). Multi-scale convolutional neural networks for time series classification. arXiv preprint arXiv:1603.06995.
- [18] Zhang, C., Yan, H., Lee, S., & Shi, J. (2018). Multiple profiles sensor-based monitoring and anomaly detection. *Journal of Quality Technology*, 50(4), 344-362.
- [19] Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014, June). Time series classification using multi-channels deep convolutional neural networks. In International conference on web-age information management (pp. 298-310). Springer, Cham.
- [20] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018, April). Deep learning detecting fraud in credit card transactions. In 2018 Systems and Information Engineering Design Symposium (SIEDS) (pp. 129-134). IEEE.
- [21] Tran, K. P., Du Nguyen, H., & Thomassey, S. (2019). Anomaly detection using long short term memory networks and its applications in supply chain management. *IFAC-PapersOnLine*, 52(13), 2408-2412.
- [22] Kim, T. Y., & Cho, S. B. (2018). Web traffic anomaly detection using C-LSTM neural networks. *Expert Systems with Applications*, 106, 66-76.
- [23] Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2019). Multivariate LSTM-FCNs for time series classification. *Neural Networks*, 116, 237-245.
- [24] Nordling, C. (2020). Anomaly Detection in Credit Card Transactions using Autoencoders.
- [25] Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep Learning Methods for Credit Card Fraud Detection. arXiv preprint arXiv:2012.03754.

- [26] Marie-Sainte, S. L., Alamir, M. B., Alsaleh, D., Albakri, G., & Zouhair, J. (2020, July). Enhancing Credit Card Fraud Detection Using Deep Neural Network. In Science and Information Conference (pp. 301-313). Springer, Cham
- [27] Yu, S. and Principe, J.C., 2018. Understanding autoencoders with information theoretic concepts. arXiv preprint arXiv:1804.00057.
- [28] Ghrib, Z., Jaziri, R., & Romdhane, R. (2020, July). Hybrid approach for anomaly detection in time series data. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.
- [29] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of machine learning research, 13(2).