

# تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد

## ملخص البحث

مع ازدياد عدد مستخدمي شبكة الانترنت الناتج عن تطورها وتوسعها المتسارع، أدى ذلك إلى زيادة حركة المرور في هذه الشبكة، ولذلك، أصبحت مهمة تصنيف حركة مرور العناوين المنطقية IP traffic ضرورة أكثر من قبل، وقد انخفض استخدام طرق التصنيف التي تعتمد على فحص حمولة الحزمة packet payload والمنفذ port في معظم التطبيقات الحديثة، ولم تتطرق الأعمال البحثية لدراسة حركة المرور في شبكات الجيل الخامس، كما لم تتطرق إلى مشكلة الأبعاد الضخمة لمجموعات البيانات مما يجعل عملية التصنيف مكلفة. ولذلك، استخدمنا نموذجاً لتصنيف حركة المرور في شبكات الجيل الرابع وطورناه ليعمل في شبكات الجيل الخامس التي تعمل على حركة مرور أكثر كثافة، وذلك برفع دقة التصنيف بنسبة ٤% لتصبح 88.45% وتخفيض تكلفة التدريب والتنبؤ بنسبة تفوق ٩٨% بعد خفض الأبعاد باستخدام خوارزمية Fisher Score التي اختارت ٢٨ ميزة بدلاً من استخدام ٦٤ ميزة واستخدمنا نفس خوارزميات التصنيف في النموذج القديم المقترح وهي: Random Forest، K-nearest Neighbours، SVM، Decision Tree و Naive Bayes

**الكلمات المفتاحية:** العناوين المنطقية، حمولة الحزمة، خوارزمية Fisher Score، خوارزميات التصنيف .

# Improving The Performance of Classification Algorithms for Internet Protocol Traffic in 5G Networks Using Dimension Reduction Techniques

## Abstract:

With the increasing number of internet users due to its rapid development and expansion, there has been a surge in network traffic. Consequently, the classification of logical IP traffic addresses has become a pressing necessity. The utilization of classification methods relying on packet payload inspection and port examination has decreased in most modern applications. Research endeavors have not delved into studying traffic in 5G networks, nor have they addressed the challenge of massive dimensions in datasets, rendering the classification process costly. To address this, we employed a model for classifying traffic in 4G networks and enhanced it to operate efficiently in 5G networks, characterized by denser traffic. This was achieved by increasing the classification accuracy by 4% to reach 88.45% and reducing the training and prediction costs by an impressive 98% after dimension reduction using the Fisher Score algorithm, which selected 28 features instead of the original 64. The same classification algorithms were applied to the proposed old model, including Decision Tree, SVM, K-nearest Neighbours, Random Forest, and Naive Bayes.

Keywords: IP traffic, packet payload, Fisher Score algorithm, classification algorithms

## ١- مقدمة

أدى توسع سوق الهواتف المحمولة ورخص أسعارها وتطور شبكات الاتصالات إلى زيادة كبيرة جداً في عدد مستخدمي الانترنت حول العالم والذي أدى بدوره إلى زيادة هائلة في حركة المرور ضمن هذه الشبكات عالمياً، بالإضافة إلى انفجار صناعة التطبيقات التي اكتسحت جميع الهواتف المحمولة، كالبريد الإلكتروني، تصفح الويب، مكالمات الفيديو والصوت، وتطبيقات المحادثة وغيرها. وبالتالي، فقد أصبح تصنيف حركة مرور في شبكات الهاتف المحمول ضرورة ملحة لمزودات خدمة الانترنت وللحكومات، لأثره الملموس في تحسين نشاطات إدارة الشبكة كتحليل جودة الخدمة (QoS) لخدمة الانترنت، واكتشاف أي خطأ في الشبكة واكتشاف الهجمات الأمنية مبكراً وغيرها من النشاطات [1].

وقد وضع الباحثون طرقاً عديدة لتصنيف حركة المرور، واعتمدت هذه الطرق على تحديد رقم المنفذ في الحمولة، وقد تغيرت لاحقاً هذه الطرق المعتمدة على المنفذ بشكل جذري، وسبب ضعف هذه الطرق حالياً هي تغير رقم المنفذ بشكل ديناميكي في التطبيقات الحديثة فلم تعد التطبيقات تعمل باستخدام رقم منفذ محدد مسبقاً. أما الطرق التي تعتمد على فحص الحمولة *payload-based technique* فيتم فيها فحص محتوى الحمولة فلم تعد فعالة أيضاً بسبب استخدام تقنيات التشفير والتي تعيق تحليل الحمولة واكتشاف ما فيها من المعلومات [2].

حالياً، تستخدم تقنيات تعلم الآلة لتصنيف حركة المرور [3]، ومع ذلك، لم يتوجه الباحثون لتطبيق ذلك في شبكات الهاتف المحمول، ومع ذلك، فلم تتناول الدراسات الحديثة حقيقة تعقيد مجموعات البيانات في التدريب، الأمر الذي يؤثر على الأداء في مرحلتي التدريب

والتنبؤ. وهذا ما قدمناه في هذه الورقة البحثية وذلك باستخدام تقنيات تخفيض الأبعاد على مجموعة بيانات لحركة المرور في شبكات الهاتف المحمول بالجيل الرابع والخامس.

اعتمدنا على تطوير نموذج سابق لتصنيف حركة المرور في شبكة الجيل الرابع سنسميه في متن البحث بالنموذج القديم [4]، وتطويره لاستخدامه في شبكات الجيل الخامس باستخدام نفس الخوارزميات في النموذج والقديم وهي: SVM، Decision Tree، K-Nearest Neighbours، Random Forest و Naive Bayes. ركزت عملية التطوير على تخفيض أبعاد مجموعات من البيانات باستخدام خوارزمية Fisher score لتتخفف أبعاد مجموعة البيانات من ٦٤ إلى ٢٨ وتزداد دقة التصنيف لتصبح ٨٨,٤٥% في النموذج المقترح بدلاً من ٨٤,٩٩% في النموذج القديم وينخفض زمن التدريب لسته أضعاف وزمن التنبؤ بمقدار عشر أضعاف.

## ٢- الهدف من البحث

يهدف هذا البحث إلى تحسين أداء خوارزميات التصنيف لحركة المرور في شبكات الجيل الخامس وذلك بتخفيض أبعاد مجموعات البيانات المستخدمة لتخفيض زمن التدريب وزمن التنفيذ ورفع دقة التصنيف.

## ٣- الدراسة النظرية

### ٣-١ تعلم الآلة

تعتبر تعلم الآلة (Machine Learning) أحد أهم فروع الذكاء الاصطناعي الذي شهد تطوراً هائلاً في العقود الأخيرة. يعني تعلم الآلة أن الأنظمة والبرامج يمكنها التعلم والتحسين من خلال التفاعل مع البيانات والخبرات السابقة بدلاً من الاعتماد على البرمجة التقليدية. يتيح تعلم الآلة للأنظمة الاصطناعية فهم الأنماط والتقارير المتكررة في البيانات، ومن ثم اتخاذ قرارات وتنفيذ المهام بكفاءة أعلى ودقة أفضل [5]. ومن بين الخوارزميات الشائعة

في تعلم الآلة: خوارزميات التصنيف [6]، خوارزميات التجميع [7]، وخوارزميات التعلم [8].

### ٢-٣ خوارزميات تخفيض الأبعاد

تخفيض الأبعاد (Dimensionality Reduction) هو عملية تقليل عدد السمات أو المتغيرات في مجموعة البيانات، مع الحفاظ على المعلومات الهامة والبارزة. تلعب خوارزميات تخفيض الأبعاد دورًا هامًا في تعلم الآلة، حيث تساهم في تحسين أداء النماذج وتقليل التعقيد الحسابي وتحسين فهم البيانات. واحدة من الخوارزميات الشهيرة في تخفيض الأبعاد هي Fisher Score، والتي تستخدم لاختيار السمات الهامة والمميزة في مجموعة البيانات. تهدف خوارزمية Fisher Score إلى تصنيف السمات بناءً على فاعليتها في تمييز الفئات المختلفة [9].

### ١-٢-٣ خوارزمية Fisher Score

تم تطوير هذه الطريقة باسم Ronald A. Fisher، وهي تستند إلى مفهوم "تحليل التباين" (Analysis of Variance) وهو مفهوم إحصائي [10]. تقوم خوارزمية Fisher Score بحساب درجة الفصل (Separation) بين متوسط القيم في كل فئة وتباين البيانات داخل الفئات [x1].

يتم حساب Fisher Score لكل سمة باستخدام الصيغة التالية (المعادلة ١):

$$Fisher\ Score = \frac{(Mean\ of\ Class\ 1 - Mean\ of\ Class\ 2)^2}{Variance\ of\ Class\ 1 + Variance\ of\ Class\ 2} \quad (1)$$

حيث تكون هناك فئتين، على سبيل المثال، فئة إيجابية وفئة سلبية. بعد حساب Fisher Score لجميع السمات في البيانات، يمكن ترتيب السمات بتنازلي حسب قيم Fisher Score. السمات ذات قيم عالية لـ Fisher Score تكون أكثر أهمية في التمييز بين الفئات.

يمكن استخدام القائمة المرتبة للاختيار من بين السمات الأكثر أهمية لتشكيل مجموعة جديدة من البيانات، وهذا يُسمى تخفيض الأبعاد. يمكن أيضاً استخدام *Fisher Score* لتحديد مدى أثر السمات المختارة على النموذج بعد تدريبه [11].

### ٣-٣ تصنيف حركة المرور للإنترنت

تم إجراء العديد من الأعمال البحثية لتصنيف حركة المرور، مع مراعاة أنواع مختلفة من تطبيقات الإنترنت. اقترح العديد من الباحثين تقنيات تصنيف مختلفة في هذا المجال وتنقسم ضمن ما يلي:

#### ٣-٣-١ التصنيف على أساس رقم المنفذ

في هذه الطريقة يتم تسجيل منافذ تطبيقات الإنترنت في هيئة الأرقام المخصصة للإنترنت أو (IANA) Internet Assigned Number Authority. بعد ذلك، يتم تصنيف حركة مرور IP باستخدام قائمة IANA لأرقام المنافذ المسجلة. على سبيل المثال، ترد أرقام المنافذ لبعض تطبيقات الإنترنت المسجلة في IANA في الجدول ١، وكما ذكرنا سابقاً فهي طريقة غير فعالة نظراً لطبيعة التطبيقات الجديدة التي تستخدم أرقام منافذ متغيرة [12].

الجدول 1: أرقام المنافذ المحددة من IANA لبعض تطبيقات الإنترنت

رقم المنفذ	التطبيق
21	FTP
23	Telnet
25	SMTP
53	DNS
80	HTTP
194	IRC

#### ٣-٣-٢ التصنيف على أساس الحمولة

يُعرف هذا النهج أيضاً باسم تقنية الفحص العميق للحزم Deep Packet Inspection (DPI). في هذا النهج، يتم تحليل محتويات الحمولة، ويتم البحث عن التوقيع signature الدقيق للتطبيقات المعروفة. كان هذا هو البديل الأول للنهج الذي يعتمد على رقم المنفذ.

تم تطوير هذه التقنية خصيصاً لتطبيقات P2P. لكن تقنية التصنيف هذه لها عيوب كثيرة، مما يجعلها غير مقبولة على نطاق واسع. أولاً، هذا النهج غير قادر على تصنيف حركة الإنترنت التي لا تتوفر لها التوقعات. وبالتالي، تتضمن هذه الطريقة التحديث المستمر لنمط التوقع للتطبيقات الجديدة. وتتطلب أجهزة مكلفة للغاية للبحث عن أنماط في حمولة الحزمة. نظراً لأن حمولة الحزمة بأكملها تحتاج إلى تحليل، وبالتالي فإن سعة التخزين وقوة الحوسبة عالية جداً مطلوبة. علاوة على ذلك، يتم تشفير حمولة الحزمة في الوقت الحاضر باستخدام تقنيات تشفير مختلفة، مما يؤدي إلى منع فحص حمولة الحزمة وبالتالي تصبح هذه التقنية غير فعالة [13].

### ٣-٣-٣ التصنيف باستخدام تعلم الآلة

تعتمد هذه التقنية على تدريب نموذج التعلم الآلي باستخدام ميزات إحصائية متنوعة مستقلة عن حمولة الحزمة ثم استخدام هذا النموذج المدرب لتصنيف حركة المرور. من المزايا الرئيسية التي يوفرها هذا النهج أن فحص رقم منفذ الحزمة أو حمولة الحزمة غير مطلوب. في الوقت الحاضر، تُستخدم العديد من تقنيات التعلم الآلي عموماً لتحقيق مهمة تصنيف حركة المرور [14].

## ٤- الأعمال السابقة

### ٤-١ التصنيف على أساس رقم المنفذ

قام الباحثون [15] بتقييم نهج التعريف القائم على المنفذ لتصنيف حركة مرور الشبكة في بروتوكولات المراسلة الخاصة بهم، مثل HTTP و FTP. قدم تقييمهم التجريبي نتائج بدقة لا تزيد عن ٧٠٪، كما قام باحثون آخرون [16] بتقييم التصنيف القائم على المنفذ لحركة مرور الانترنت الخاصة بهم في مختبرات جامعتهم. وأكدوا أن الحل المطبق لم يتمكن من تحديد ٣٠٪ - ٧٠٪ من البيانات التي تم جمعها.

#### ٤-٢ التصنيف على أساس الحمولة

اقترح باحثون آخرون [17] إطارًا يسمى Lightweight DPI (LW-DPI) لتصنيف حركة مرور الشبكة عن طريق فحص محتوى عدد محدود من الحزم أو جزء صغير من حمولة حزمة معينة. اعتمد التقييم على حركة مرور الشبكة التي تم جمعها من مزودي خدمة الإنترنت التجاريين ومختبرات الجامعة المحلية، وحقت تجاربهم دقة تصنيف تصل إلى ٩٩٪.

كما اقترحت مجموعة من الباحثين [18] طريقة BitProb والتي تستخدم مفهوم توقيعات البت الاحتمالية وهي عبارة عن طول سلسلة ثنائي n-bit مستخرج من تدفقات الشبكة لتصنيف حركة مرور الشبكة. يتم إنشاء توقيع n-bit لكل تدفق من خلال مراقبة الحزم القليلة الأولى. بعد ذلك، يتم تغذية n-bit إلى آلة انتقال الحالة المسماة Probabilistic Counting Deterministic Automata (PCDA) لحساب احتمالية توقيع البتات وتصنيفها إلى بروتوكولات مؤقتة. تم استخدام ثلاث مجموعات بيانات تغطي ٢٠ بروتوكولاً واستطاعوا تخفيض نتائج التصنيف الخاطئ.

من جانب آخر، توجه بعض الباحثين إلى تطوير خوارزميات تعتمد على السياق والتي حققت أداءً يضاهي أداء نماذج الذكاء الاصطناعي، فقد قام الباحثون [19] بتصميم مصنف MediaSense وهي خوارزمية قادرة على تصنيف حركة المرور المشفرة الخاصة بالملتميديا في الزمن الحقيقي بدقة تضاهي أساليب التعلم العميق.

#### ٤-٣ التصنيف باستخدام تعلم الآلة

لم تتطرق الأعمال التي استخدمت تقنيات تعلم الآلة إلى تخفيض الأبعاد إذ ركزت على رفع دقة التصنيف دون الأخذ بعين الاعتبار سرعة التدريب وسرعة اتخاذ القرار عند تشغيل النموذج في البيئة الحقيقية.

اقترحت إحدى الدراسات [20] تصنيف حركة المرور في شبكات الجيل الخامس باستخدام التعلم العميق لتخصيص الموارد وذلك بفحص حركة الشبكة لتوقع الطلب المستقبلي والذي

استطاع تحقيق دقة تصنيف وصلت إلى ٩٧,٣٩٧٥%. كما استخدم باحثون آخرون [21] تقطيع الشبكة وخوارزميات التعلم العميق لتوفير القدرة لمشغلي الشبكات على إنشاء شبكات أو شرائح افتراضية متعددة على نفس البنية التحتية المادية. تنتم هذه الشرائح بأنها مستقلة ومخصصة مما يسمح بإدارتها ديناميكياً وفقاً للمتطلبات المحددة بين مشغلي الشبكات والمستخدمين. يصنف النموذج حركة المرور بالتنبؤ بنوع الشريحة الدقيق لكل مستخدم.

لقد استخدم الباحثون [22] ثلاث أنواع من خوارزميات التعلم العميق وهي Deep Belief Network و Convolution Neural Network و Multi-layer Perceptron لتصنيف حركة مرور الشبكة. ودمجوا نتائج جميع المصنفات الثلاثة باستخدام تقنية Bayesian decision fusion. أظهرت النتائج التجريبية على مجموعة بيانات التي أنشئوها أن الطريقة المقترحة تعمل على تحسين دقة التصنيف وتعمل بشكل جيد على أنواع حركة مرور الشبكة المختلفة بمتوسط دقة تصنيف وصل إلى ٩٧%. كما استخدم باحثون آخرون [23] تقنية جديدة في التعليم شبه الخاضع للإشراف وهي federated semi-supervised learning وذلك لتصنيف حركة مرور الشبكة وحقق النموذج دقة تصنيف تصل إلى ٩٧,٨١%.

أنشأ الباحثون [24] مجموعة بيانات حية لحركة الإنترنت تألفت من ٢٣ ميزة وتتألف من حركات المرور لتطبيقات DNS و WWW و P2P و FTP و Telnet. كما قاموا بتطبيق مجموعة من خوارزميات تعليم الآلة Naive Bayes و Bayes Net و decision tree و SVM لتصنيف حركة مرور الشبكة. حققت شجرة القرار أعلى دقة تصنيف بلغت ٧٨,٩١%. كما صنف الباحثون [25] حركة المرور باستخدام تعلم الآلة وذلك باستخدام نوعين مختلفين من مجموعات البيانات لحركة مرور تتضمن مجموعة من التطبيقات ك تطبيقات البريد الإلكتروني و WWW و CHAT و FTP. استخدموا مجموعة من خوارزميات تعلم الآلة وهي: K-Nearest Neighbor و Naive Bayes و RBF و Bayes Net. وحققت Bayes Net أعلى دقة تصنيف. بالإضافة إلى عمل مشابه

للعمل السابق، طور الباحثون [26] مجموعة بيانات لحركة الإنترنت لمدة ثانيتين ودقيقتين على التوالي لالتقاط الحزم لثمانية تطبيقات من تطبيقات الإنترنت. واستخدموا خمس خوارزميات تعلم آلة لتصنيف حركة المرور وهي: C4.5 و RBF و MLP و Naïve و Bayes Net و Bayes Net. وحقت Bayes Net أعلى دقة تصنيف أيضاً.

#### ٤-٤؛ ملخص الدراسات المرجعية

قدمت الأعمال الأولى طرقاً لتصنيف إدارة حركة المرور باستخدام خوارزميات تصنيف حسب المنفذ أو الحمولة، ثم ظهرت العديد من الأعمال التي استخدمت تقنيات متنوعة للذكاء الاصطناعي والتي ركزت في معظمها على تحسين دقة التصنيف وتوفير تكلفة التصنيف عن طريق تحسين دقة التصنيف باختيار التقنيات الأنسب، ولم تتطرق هذه الأعمال إلى تخفيض تكلفة هذه العملية بمعالجة مجموعات البيانات الضخمة وتخفيض أبعادها وهو ما عملنا عليه في هذا البحث واستطعنا تخفيض تكلفة التدريب والتنبؤ مع رفع دقة التصنيف كما سيظهر في قسم النتائج.

#### ٥- النموذج المقترح

اعتمد النموذج المقترح الذي يعمل على تصنيف حركة المرور في شبكات الجيل الرابع على مجموعة بيانات تم إنشاؤها ومعالجتها لتدريب أشهر خوارزميات التصنيف عليها واختيار الخوارزمية الأنسب وفق النتائج.

تحتوي مجموعة البيانات على مجموعة متنوعة من الحقول التي توفر رؤية شاملة حول حركة المرور في الشبكة. من بين هذه الحقول، يمكن التمييز عن العديد من المعلومات المهمة. يشمل ذلك حجم البيانات المرسل والمستقبل، وزمن التأخير بين الحزم، ومعدلات البيانات الكلية للتدفق، بالإضافة إلى معلومات حول المصدر والوجهة والمنفذ المستخدمة. يعكس هذا التنوع في الحقول قدرة المجموعة على تقديم رؤية شاملة حول سلوك حركة المرور، مما يمكن الخوارزميات من استخدام هذه المعلومات لتحسين عمليات التصنيف

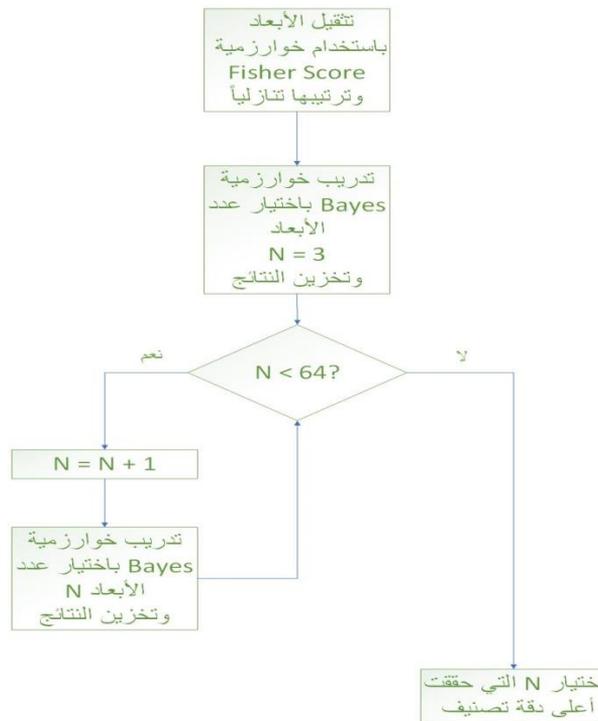
وتحليل الأنماط في الشبكة بشكل فعال، ولكن ليست جميع الحقول على نفس الدرجة من الأهمية في اتخاذ القرار عند التصنيف، ومع ذلك، يعمل النموذج القديم على استخدامها جميعاً في عملية التصنيف وهو ما يعني ارتفاع تكلفة التصنيف لمعالجتها، وانخفاض دقة التدريب نظراً لكثرة الأبعاد المستخدمة.

ولذلك، عملنا على تطوير النموذج ليعمل في شبكات الجيل الخامس وذلك بتخفيض تكلفة التصنيف نظراً لكثافة حركة المرور في هذا النوع من الشبكات مقارنة بالجيل الرابع، وذلك بتخفيض الأبعاد من ٦٤ إلى ٢٨ وهو ما أدى إلى انخفاض زمني التدريب والتنبؤ بنسبة تفوق ٩٨% وتحسن دقة التصنيف بنسبة ٤% كما سيظهر في قسم النتائج. يوضح الشكل ١ مراحل بناء النموذج القديم والتعديلات التي قمنا بها - المرحلة الرابعة- لتقديم النموذج المقترح، ويوضح الشكل ٢ المخطط التفصيلي لعملية تخفيض الأبعاد واختيار عدد الأبعاد الأفضل.



الشكل ١: مقارنة بين مراحل بناء النموذج القديم والنموذج المقترح الذي تشكل بعد تخفيض الأبعاد

## تحسين أداء خوارزميات تصنيف حركة مرور عناوين الإنترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد



الشكل 2: خوارزمية اختيار عدد الأبعاد الأمثل

### ٥-١ جمع بيانات حركة المرور

استخدمنا البيانات التي جمعها الباحثون في النموذج القديم [4] وتم تجميعها بالتقاط بيانات حركة مرور الإنترنت في الوقت الفعلي لشبكة الجيل الخامس باستخدام أداة التقاط حزم شائعة تسمى Wireshark. وتتضمن حركة المرور على الإنترنت لمدة ٣٠ ثانية لستة أنواع من التطبيقات: البريد الإلكتروني والمراسلة الفورية و P2P و VOIP ووسائط الويب وتطبيقات WWW من خلال الاتصال بشبكة الجيل الخامس باستخدام نقطة اتصال من بيانات الجوال.، وهذه المجموعة محفوظة بتنسيق .pcap.

### ٥-٢ استخلاص الميزات

قام الباحثون في النموذج القديم بتحويل ملف pcap الذي يتضمن حركة المرور بصيغتها الأساسية واستخلاص المعلومات المهمة منها لتصبح بشكل جدول ليكون قابلاً لتمثيله على

خوارزميات التصنيف كما في الشكل ٣، تتألف مجموعة البيانات بعد عملية استخلاص الميزات من ٦٤ ميزة، وهو عدد كبير ويجب تخفيضه للحصول على أداء أسرع ودقة تصنيف أفضل من التي قدمها الباحثون الذين أنشأوا مجموعة البيانات وهذا ما استطعنا تحقيقه في هذا العمل.

```
d4 c3 b2 a1 02 00 04 00
00 00 00 00 00 00 00 00
00 00 04 00 01 00 00 00
00 45 d4 5e 18 8e 0c 00
42 00 00 00 42 00 00 00
00 1e ec 26 d2 ac 26 02
06 49 6b 31 08 00 45 02
00 34 30 8c 40 00 72 06
81 7f 2e 69 63 a3 c0 a8
04 02 cf 3a 00 50 8d a5
ee 7b 00 00 00 00 80 c2
20 00 ac 29 00 00 02 04
05 78 01 03 03 08 01 01
04 02 00 45 d4 5e 2c 77
0d 00 36 00 00 00 36 00
00 00 00 1e ec 26 d2 ac
```

عملية استخلاص الميزات

#	A	B	C	D	E	F
1	flow	src	src_post	dst	dst_post	feudationI
2	1157<->17292.168.0.1	58167	2.217.166.	443	6.824893	
3	2956<->17292.168.0.1	52296	2.217.167.	443	9.236139	
4	40724<->2092.168.0.1	60724	32.88.149.	53	0.012017	
5	1725<->17292.168.0.1	60725	2.217.161.	443	0.399743	
6	1197<->21692.168.0.1	49197	6.58.200.1	443	6.239398	
7	344<->74.92.168.0.1	64344	125.200.1	443	0.105584	
8	3347<->2092.168.0.1	53347	32.88.149.	53	0.013218	
9	348<->17292.168.0.1	53348	2.217.166.	443	0.483113	
10	5353<->22492.168.0.2	5353	224.0.251	5353	1	
11	3186<->74.92.168.0.1	53186	4.125.24.16	443	15.08241	

جدول للمعلومات المستخلصة  
من حركة المرور

حركة المرور بالترميز الست عشري  
في ملف حركة المرور الملتقط

الشكل 3: عملية استخلاص الميزات لتحويل ملف pcap إلى جدول يمكن استخدامه في خوارزميات الذكاء الاصطناعي

### ٣-٥ معالجة البيانات

قام الباحثون الذين وضعوا النموذج القديم بتنظيف البيانات من الأسطر الناقصة والتخلص من الحقول غير الضرورية.

### ٤-٥ تخفيض الأبعاد

وهي المرحلة التي قمنا بها لتحسين نموذجهم، لقد قمنا بتخفيض الأبعاد التي تبلغ ٦٤ وذلك لتسريع عملية التصنيف وتحسين دقتها. لقد استخدمنا خوارزمية fisher score وهي من أفضل الخوارزميات الممتازة لتخفيض الأبعاد وذلك لقدرتها على تمييز المتغيرات باستخدام التباين الداخلي والخارجي في الإحصاء فيسمح بتحديد المتغيرات ذات الأهمية العالية في عملية التصنيف، ولتقليل الأبعاد بشكل فعال والحفاظ على المعلومات الأكثر أهمية دون التأثير على جودة عملية التصنيف. وكذلك لتحسين أداء النماذج، فعندما يتم تركيز النماذج

تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد

على المتغيرات الرئيسية، سيصبح النموذج أكثر قدرة على اكتشاف العلاقات والأنماط. وبالتالي، تتحسن دقة وفعالية عملية التصنيف. بالإضافة إلى سهولة التنفيذ والتطبيق والوصول إلى نتائج التخفيض منطقية قابلة للفهم.

استطعنا تخفيض الأبعاد باستخدام خوارزمية fisher score (الأشكال ٤-٥ والجدول ٢) لتصبح ٢٨ بدلاً من ٦٤ وتحسنت دقة التصنيف في خوارزميات التصنيف المختارة من ١% إلى ٤% واستطعنا تخفيض وقت التدريب والتنبؤ كما سيتوضح ذلك في القسم الخاص بالنتائج.

الجدول 2 : ميزات مجموعة البيانات، الميزات المظللة بالأخضر هي التي تم اختيارها بعد التصنيف والمميزات المظللة بالأحمر هي التي تم استبعادها

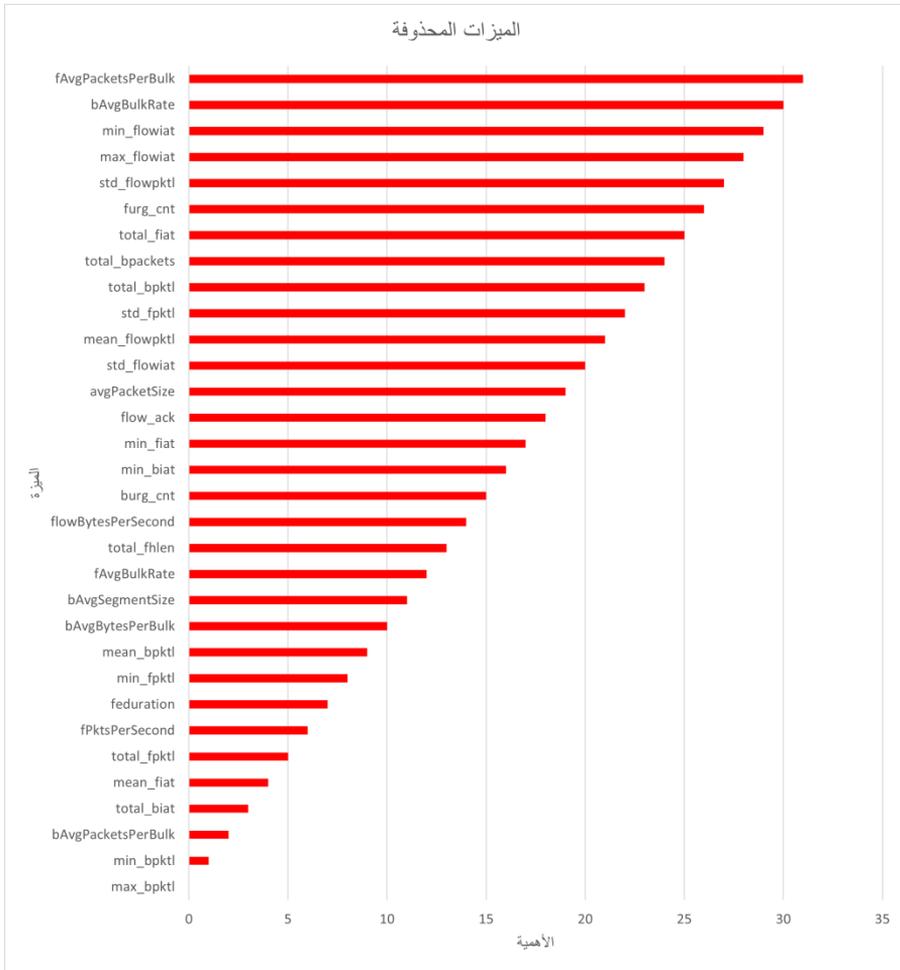
flow	total_biat	bAvgSegmentSize	avgPacketSize
src	mean_fiat	fAvgBulkRate	std_flowiat
src_port	total_fpktl	total_fhlen	mean_flowpktl
dst	fPktsPerSecond	flowBytesPerSecond	std_fpktl
dst_port	federation	burg_cnt	total_bpktl
max_bpktl	min_fpktl	min_biat	total_bpackets
min_bpktl	mean_bpktl	min_fiat	total_fiat
bAvgPacketsPerBulk	bAvgBytesPerBulk	flow_ack	furg_cnt
std_flowpktl	min_flowpktl	std_biat	max_fpktl
max_flowiat	flow_cwr	flow_psh	mean_fpktl
min_flowiat	downUpRatio	mean_flowiat	bpsh_cnt
bAvgBulkRate	flow_rst	fpsh_cnt	flow_fin
fAvgPacketsPerBulk	flow_ece	std_bpktl	flowPktsPerSecond
fAvgBytesPerBulk	mean_biat	max_biat	bPktsPerSecond

fAvgSegmentSize	std_fiat	total_fpackets	total_bhlen
max_flowpktl	flow_urg	max_fiat	flow_syn



الشكل 4: قيمة عامل الأهمية للميزات المختارة وتتراوح قيمها بين ٣٢ و ٥٩

## تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد



الشكل 5: قيمة عامل الأهمية للميزات التي تم استبعادها وتتراوح قيمها بين 0 و 32

### ٥-٥ تدريب النموذج

استخدمنا خوارزميات التصنيف التي استخدمها الباحثون في النموذج القديم وهي تتوافق مع طبيعة البحث، فتصنيف حركة المرور يتطلب خوارزميات تصنيف فعالة في الزمن الحقيقي، وهي أربع خوارزميات: Naive Bayes و Random Forest و Decision tree و SVM.

## ٥-٥ مقاييس الأداء

وتتضمن مقياس عملية تخفيض الأبعاد و ٥ مقاييس عملية التصنيف.

### المحصلة Score

وهي رقم موجب يبدأ من صفر لتحديد أهمية الميزة عند تخفيض الأبعاد، وهو نتيجة التحليل الإحصائي التي تقوم به خوارزمية تخفيض الأبعاد Fisher score. تشير القيم المتدنية إلى الميزات التي يجب استبعادها والقيم المرتفعة إلى القيم المؤثرة في عملية التصنيف [27]. واستناداً لنتائج عملية التدريب يتم تحديد القيمة المعيارية للمحصلة والتي يتم بناء عليها اختيار أو استبعاد الميزات.

### الدقة Accuracy

الدقة هي مقياس التقييم الأكثر استخداماً في مسائل التصنيف. حيث تقيس نسبة نقاط البيانات المصنفة بشكل صحيح من إجمالي عدد نقاط البيانات. يتم حسابها على النحو التالي:

$$\text{Accuracy} = (\text{TP} + \text{TN} + \text{FP} + \text{FN}) / (\text{TP} + \text{TN})$$

حيث TP هو عدد الإيجابيات الحقيقية، TN هو عدد السلبيات الحقيقية، FP هو عدد الإيجابيات الخاطئة، و FN هو عدد السلبيات الخاطئة [28].

### الإنضباط Precision

الإنضباط يعرف بأنه نسبة التنبؤات الإيجابية الحقيقية من جميع التوقعات الإيجابية. حيث يقيس قدرة النموذج على التنبؤ بالعينات الإيجابية بشكل صحيح. يتم حسابه على النحو التالي [28]:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

### الاستدعاء Recall

الاستدعاء Recall، المعروف أيضاً باسم الحساسية، هو نسبة التنبؤات الإيجابية الصحيحة من جميع العينات الإيجابية الفعلية. كذلك يقيس قدرة النموذج على تحديد العينات الإيجابية بشكل صحيح. يتم حسابه على النحو التالي [28]:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

### درجة F1 (F1-Score)

مقياس درجة F1 هي المتوسط التوافقي للدقة و الاستدعاء. كذلك، تعتبر درجة F1 مقياس شائع الاستخدام عندما يكون كل من الدقة والاستدعاء مهمين. علاوة على ذلك يتم حساب درجة F1 على النحو التالي [28]:

$$\text{F-score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

### مصفوفة الإرباك Confusion Matrix

مصفوفة الإرباك هي جدول يلخص أداء نموذج التصنيف. على سبيل المثال، تعرض مصفوفة الإرباك عدد الإيجابيات الصحيحة والسلبيات الصحيحة و الإيجابيات الخاطئة و السلبيات الخاطئة. من ناحية أخرى تعد هذه المصفوفة مفيدة لتصور أداء النموذج وتحديد مجالات التحسين [29].

	Positive	Negative	
الصفة المتوقعة Predicted Label	True Positive (TP_	False Positive (FP)	Positive

False Negative (FN)	True Negative (TN)
------------------------	-----------------------

Negative

الصفة الحقيقية True Label

## ٦-٥ التنفيذ العملي للنموذج

٦-٥-١ تصدير ملفات حركة المرور:

وهي المرحلة الأولى قبل عملية التدريب، والتي تتطلب تصدير ملفات حركة المرور إلى صيغة CSV ليتمكن استخدامها مع مكتبات بايثون المتخصصة بالكفاءة الاصطناعي، وذلك باتباع الخطوات التالية:

١. استعراض ملف pcap في Wireshark

٢. اختيار "CSV" File > Export Packet Dissections > as

٣. اختيار الموقع والاسم للملف الناتج CSV وحفظ الملف

٦-٥-٢ استيراد المكتبات:

وفي هذه المرحلة يتم استيراد مكتبات مهمة مثل pandas للتعامل مع البيانات، numpy للعمليات الرياضية، ومكتبات أخرى للتصوير البياني وتنفيذ نماذج التعلم الآلي.

٦-٥-٣ تحميل البيانات والتحضير:

وهي بداية عملية التدريب، وفيها يتم تنفيذ ما يلي:

- قراءة ملف CSV الذي يحتوي على بيانات حركة الشبكة باستخدام pandas
- معالجة البيانات وحذف الأسطر غير المكتملة وحذف بعض الأعمدة غير الضرورية.

٦-٥-٤ تثقيف المدخلات وتحديد أهميتها:

تطبيق خوارزمية Fisher Score على مجموعة البيانات لتحديد تثقيف كل مدخل من المدخلات في مجموعة البيانات.

٦-٥-٥ تحديد عدد الأبعاد الأمثل بعد التخفيض:

وفي هذه المرحلة يتم تحديد عدد الأبعاد الأمثل وفق ما يلي (راجع الشكل ٢):

- ترتيب المدخلات وفق تثقيلها وتنفيذ عملية التدريب (الخطوة رقم ٦) لخوارزمية واحدة فقط وهي Naïve Bayes باستخدام أثقل ثلاثة مدخلات ثم أثقل أربعة مدخلات وهكذا حتى ٦٤ مدخل، والغرض من هذه العملية التكرارية تحدد عدد الأبعاد الذي يحقق دقة تصنيف مساوية أو تفوق دقة التصنيف في النموذج القديم.
- مقارنة النتائج التي تظهر في كل تكرار مع دقة التصنيف للنموذج القديم، وتظهر النتائج أن عدد الأبعاد ٢٧ حقق أفضل نتيجة.
- بعد تحديد العدد الأمثل، يتم استخدامه في تدريب الخوارزميات الأربعة.

٦-٥-٦ التدريب والتقييم:

وهي عملية التدريب التقليدية لخوارزميات التصنيف الأربعة (Naive Bayes، Random Forest، Decision Tree، SVM) وتبدأ بتغذية الخوارزميات بالمدخلات التي تم اختيارها من المرحلة السابقة وإظهار النتائج ثم حساب زمني التدريب والتنبؤ ودقة التصنيف لكل خوارزمية.

## ٦- النتائج والمناقشة

### ٦-١ نتائج تدريب النموذج قبل تخفيض الأبعاد

استخدم الباحثون في النموذج القديم خوارزميات التصنيف الأربعة: Naive Bayes، Random Forest، Decision tree، و SVM على مجموعة البيانات الأصلية والتي تتألف من ٦٤ ميزة وقد حققت خوارزمية Random Forest أفضل دقة تصنيف وصلت إلى ٨٧%. وقمنا بإعادة عمليات التدريب التي أجراها الباحثون على نفس مجموعة البيانات الموجودة على مستودع الكود المصدري الخاص بالورقة البحثية [30] وقد حققت خوارزمية

Random Forest أعلى دقة تصنيف ولكنها وصلت إلى ٨٤,٩٩% ولم تصل إلى ٨٧%  
(الجدول ٣-٤).

الجدول 3: نتائج تدريب النموذج على مجموعة البيانات الأصلية قبل تخفيض الأبعاد باستخدام خوارزمية Random Forest

Correctly Classified Instances	1614	84.99%
Incorrectly Classified Instances	285	١٥,٠٠%
Mean absolute error	0.12	
Root mean squared error	0.21	
Total Number of Instances	1899	
Training Time	1.38368 s	
Prediction Time	0.10853 s	
TP Rate	85%	
FP Rate	13.6%	
Precision	84.9%	
Recall	85%	
F-Measure	83.7%	

الجدول 4: مصفوفة الارتباك للنموذج المدرب على مجموعة البيانات الأصلية قبل تخفيض الأبعاد باستخدام خوارزمية Random Forest

a	b	c	d	e	f	<-- classified as
١٥٦	٣	٠	٣	١٦	٤٥	a = EMAIL
٣	٢٩	٠	٦	١٨	٣٦	b = INSTANT MESSAGING
٠	٠	١١٦	٠	٠	٢	c = P2P

تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام  
تقنيات تخفيض الأبعاد

٤	٣	٠	١٠٠	٩	٦٨	d = VOIP
٧	٣	٠	٥	١٦٧	٤٨	e = WEB MEDIA
١	٢	٠	٣	٠	١٠٤٦	f = WWW

٦-٢ نتائج تدريب النموذج بعد تخفيض الأبعاد

قمنا بترتيب الميزات وفقاً لأهميتها في عملية التصنيف كما وضحنا في قسم "معالجة البيانات" وقد حصلت الميزات على محصلة تراوحت من صفر إلى ٥٩. أجرينا مجموعة من عمليات التدريب بدءاً من ثلاثة ميزات وتم زيادة الميزات تدريجياً حتى لاحظنا أن أول ٢٦ ميزة قد حققت نفس النتائج التي حققها النموذج قبل التصنيف (الجدول ٥-٦). تابعنا زيادة عدد الميزات المدخلة إلى النموذج واستطعنا تحقيق نتائج أفضل باستخدام أول ٢٨ ميزة، وبناءً على عمليات التدريب، اخترنا القيمة المعيارية للمحصلة ٣٢، فكل ميزة محصلتها أكبر من أو تساوي ٣٢ تم اختيارها، وكل ميزة دون ٣٢ تم استبعادها.

الجدول 5: نتائج تدريب النموذج على مجموعة البيانات الأصلية بعد تخفيض الأبعاد باستخدام خوارزمية *Random Forest*

Correctly Classified Instances	١٦٧٩	88.٤٥%
Incorrectly Classified Instances	220	١١,٥٥%
Mean absolute error	0.06	
Root mean squared error	0.17	
Total Number of Instances	1899	
Training Time	0.02049 s	
Prediction Time	0.00075 s	
TP Rate	88.4%	
FP Rate	6.6%	
Precision	88.1%	
Recall	88.4%	

F-Measure	88.1%
-----------	-------

الجدول 6: مصفوفة الإرتباك للنموذج المدرب على مجموعة البيانات الأصلية بعد تخفيض الأبعاد باستخدام خوارزمية Random Forest

a	b	c	d	e	f	←-- classified as
١٧٢	٢	٣	٤	٩	٣٣	a = EMAIL
٥	٥٦	٠	١٠	٧	١٤	b = INSTANT MESSAGING
١	٢	١١٠	٢	٢	١	c = P2P
٤	٥	٣	١٣٢	١٥	٢٥	d = VOIP
٨	٧	٢	٦	١٩١	١٦	e = WEB MEDIA
١٨	٧	٢	٣	٤	١٠١٨	f = WWW

### ٦-٣ المقارنة

لقد حققت مجموعة البيانات الجديدة على نفس الخوارزميات الأربعة المستخدمة مسبقاً نتائج أفضل فرفعت دقة التصنيف بنسبة ٤% بعد تخفيض الأبعاد بنسبة ٥٦,٢٥% وخفضت وقت التدريب والتنبؤ بنسبة تفوق ٩٨%. ونظراً لاعتماد خوارزمية Random Forest في كلا العمليين كأفضل خوارزمية، سنقارن النموذجين قبل التخفيض وبعده مع هذه الخوارزمية (الجدول ٧).

الجدول 7: مقارنة النموذج المدرب بخوارزمية Random Forest قبل تخفيض الأبعاد وبعده تخفيض الأبعاد في مجموعة البيانات

النموذج بعد التخفيض	النموذج قبل التخفيض	معيار المقارنة
88.45%	84.99%	دقة التصنيف
١١,٥٥%	١٥,٠٠%	أخطاء التصنيف
0.06	0.12	Mean absolute error
0.17	0.21	Root mean squared error
0.02049 s	1.38368 s	زمن التدريب

تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد

0.00075 s	0.10853 s	زمن التنبؤ
88.4%	85%	معدل الإيجابيات الصحيحة TP Rate
6.6%	13.6%	معدل الإيجابيات الكاذبة FP Rate
88.1%	84.9%	الإنضباط Precision
88.4%	85%	الاستدعاء Recall
88.1%	83.7%	درجة (F1-Score) F1

### ٧- التطلعات المستقبلية

نقترح العمل على إحدى المحاور التالية في الأعمال المستقبلية:

- ١- استخدام مجموعات بيانات حديثة في أعمال التطوير المستقبلية.
- ٢- تطبيق مجموعة من تقنيات تخفيض الأبعاد ومقارنة أدائها لاختيار التقنية الأفضل.
- ٣- استخدام تقنيات التعلم العميق بعد تخفيض الأبعاد.
- ٤- دراسة تأثير عملية التخفيض الأبعاد على أمان البيانات.
- ٥- دراسة تأثير عملية تخفيض الأبعاد على استهلاك الطاقة.

## 8- الخلاصة

استعرضنا في هذه الورقة البحثية مشكلة تصنيف حركة مرور الانترنت في شبكات الجيل الخامس، وتلخصت المشكلة في ندرة الدراسات التي تناولت تصنيف حركة المرور في الجيل الخامس، وتعقيد مجموعة البيانات. ولذلك، استخدمنا نموذجاً سابقاً لتصنيف حركة المرور في شبكات الجيل الرابع وطورناه ليعمل في شبكات الجيل الخامس وذلك باستخدام خوارزميات تخفيض الأبعاد. استخدمنا خوارزمية Fisher Score التي اختارت ٢٨ ميزة بدلاً من استخدام ٦٤ ميزة ورفعت من أداء نموذج التصنيف كاملاً. اعتمدنا على نفس خوارزميات التصنيف في النموذج القديم المقترح وهي: SVM، Decision Tree، K-Nearest Neighbours، Random Forest و Naive Bayes، أثبتت نتائج الاختبارات أن النموذج المقترح عمل بسرعة أفضل في التدريب والتنبؤ، وبدقة تصنيف أعلى في جميع الخوارزميات إذ وصلت دقة التصنيف في النموذج المقترح إلى ٨٨,٤٥% مقارنة بالنموذج القديم الذي حقق ٨٤,٩٩%.

## ٩- جدول الاختصارات

الاختصار	الاسم الكامل
----------	--------------

تحسين أداء خوارزميات تصنيف حركة مرور عناوين الانترنت في شبكات الجيل الخامس باستخدام تقنيات تخفيض الأبعاد

4G	Fourth Generation
5G	Fifth Generation
CSV	Comma-Separated Values
DPI	Deep Packet Inspection
DNS	Domain Name System
FN	False Negative
FP	False Positive
FTP	File Transfer Protocol
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
IP	Internet Protocol
LW-DPI	Lightweight Deep Packet Inspection
MLP	Multilayer Perceptron
P2P	Peer-to-Peer
PCAP	Packet Capture
PCDA	Probabilistic Counting Deterministic Automata
QoS	Quality of Service
RBF	Radial Basis Function
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VOIP	Voice over Internet Protocol
WWW	World Wide Web

١٠ - المراجع

- [١] Moore, Andrew W., and Denis Zuev. 2005 **Internet traffic classification using bayesian analysis techniques**. Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems..
- [٢] Callado, Arthur, et al. 2009 **A survey on internet traffic identification**. IEEE communications surveys & tutorials 11.3 37-52.
- [٣] Nguyen, Thuy TT, and Grenville Armitage. 2008 **A survey of techniques for internet traffic classification using machine learning**. IEEE communications surveys & tutorials 10.4: 56-76.
- [4] Gupta, Amit, Anupam Raj, and Mayank Arora. **IP traffic classification of 4G network using machine learning techniques**. 2021 5th International conference on computing methodologies and communication (ICCMC). IEEE, 2021.
- [٥] Jordan, Michael I., and Tom M. Mitchell. 2015 **Machine learning: Trends, perspectives, and prospects**. Science 349.6245 255-260.
- [٦] Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. 2007. **Supervised machine learning: A review of classification techniques**. Emerging artificial intelligence applications in computer engineering 160.1 3-24.

- [٧] Ahuja, Ravinder, et al. 2020. **Classification and clustering algorithms of machine learning with their applications**. Nature-inspired computation in data mining and machine learning 225-248.
- [٨] Mathew, Amitha, P. Amudha, and S. Sivakumari. 2021. **Deep learning techniques: an overview**. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020 : 599-608.
- [9] Gu, Quanquan, Zhenhui Li, and Jiawei Han. **Generalized fisher score for feature selection**. arXiv preprint arXiv:1202.3725 (2012).
- [10] Sun, Lin, et al. **Feature selection using Fisher score and multilabel neighborhood rough sets for multilabel classification**. Information Sciences 578 (2021): 887-912.
- [١١] Reddy, G. Thippa, et al. 2020 **Analysis of dimensionality reduction techniques on big data**. Ieee Access 8: 54776-54788.
- [١٢] Camarillo, G. **The Internet Assigned Number Authority (IANA) Uniform Resource Identifier (URI) Parameter Registry for the Session Initiation Protocol (SIP)**. No. rfc3969. 2004.
- [١٣] Finsterbusch, Michael, et al. 2013 **A survey of payload-based traffic classification approaches**. IEEE Communications Surveys & Tutorials 16.2 : 1135-1156.
- [١٤] Salman, Ola, et al. 2020 **A review on machine learning-based approaches for Internet traffic classification**. Annals of Telecommunications 75 : 673-710.
- [١٥] Moore, Andrew W., and Konstantina Papagiannaki. 2005 **Toward the accurate identification of network applications**. International workshop on passive and active network measurement. Berlin, Heidelberg: Springer Berlin Heidelberg,.
- [١٦] Sen, Subhabrata, Oliver Spatscheck, and Dongmei Wang. 2009 **Accurate, scalable in-network identification of p2p traffic using application signatures**. Proceedings of the 13th international conference on World Wide Web. 2004.

- [١٧] Fernandes, Stênio, et al. "**Slimming down deep packet inspection systems.**" IEEE INFOCOM Workshops 2009. IEEE,.
- [١٨] Hubballi, Neminath, Mayank Swarnkar, and Mauro Conti. 2020 **BitProb: Probabilistic bit signatures for accurate application identification.** IEEE Transactions on Network and Service Management 17.3: 1730-1741.
- [19] Aceto, Giuseppe, et al. DISTILLER: **Encrypted traffic classification via multimodal multitask deep learning.** Journal of Network and Computer Applications 183 (2021): 102985.
- [20] Mareri, Bruce, et al. MANTA: **Multi-Lane Capsule Network Assisted Traffic Classification for 5G Network Slicing.** IEEE Wireless Communications Letters 11.9 (2022): 1905-1909.
- [21] Mohammedali, Noor Abdalkarem, et al. **Traffic Classification using Deep Learning Approach for End-to-End Slice Management in 5G/B5G.** 2022 13th International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2022.
- [٢٢] Izadi, Saadat, Mahmood Ahmadi, and Amir Rajabzadeh. 2022 **Network traffic classification using deep learning networks and Bayesian data fusion.** Journal of Network and Systems Management 30.2 . 25.
- [٢٣] Jin, Zhiping, et al. 2023. **A federated semi supervised learning approach for network traffic classification.** International Journal of Network Management 33.3 : e2222.
- [٢٤] M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn, and F. Abdessamia, 2017 **Network Traffic Classification techniques and comparative analysis using Machine Learning algorithms,** doi: 10.1109/CompComm.2016.7925139
- [٢٥] A. (Karunya U. Jamuna and V.(Karunya U. Edwards S.E, 2013 **.Efficient Flow based Network Traffic Classification using Machine Learning,**” Int.J.Eng. Res. Appl.,vol.3, no.2,.
- [٢٦] K. Singh, S. Agrawal, and B. S. Sohi, 2013 **A Near Real-time IP Traffic Classification Using Machine Learning,** Int. J. Intell. Syst. Appl., vol. 5, no. 3, 2013, doi: 10.5815/ijisa..03.09.

- [٢٧] Li, Tao, et al. 2022 **A binary individual search strategy-based bi-objective evolutionary algorithm for high-dimensional feature selection**. Information Sciences 610 651-673.
- [28] Hossin, Mohammad, and Md Nasir Sulaiman. 2015 **A review on evaluation metrics for data classification evaluations**. International journal of data mining & knowledge management process 5.2 (): 1.
- [٢٩] Liang, Jingsai. Confusion matrix: **Machine learning**. POGIL Activity Clearinghouse 3.4 (2022).
- [٣٠] "GitHub - **anupamraj1312/4G-IP-traffic-classification-using-Machinelearning**." <https://github.com/anupamraj1312/4G-IP-trafficclassification-using-Machine-learning> (accessed Feb. 07, 2021)