SHIHAB JOURNAL OF HUMANITIES
Volume (2), Issue (3) (Sept, 2026)
ISSN: 3136-2958

مجلة الشهاب للعلوم الإنسانية
المجلد (2) العدد 3 سبتمبر 2026
ردمد: 3136-2958

تقييم التصنيف الصفري للكشف عن الأخبار الكاذبة باللهجة الليبية باستخدام نموذج XLM-RoBERTa: دراسة استكشافية

وائل عثمان محمد النعاس

المعهد العالي للتقنيات الهندسية، الماجوري، بنغازي، ليبيا

Evaluating Zero-Shot Classification for Fake News Detection in the Libyan Dialect Using XLM-RoBERTa: An Exploratory Study

Wael Othman Mohammed Al-Naas
Higher Institute for Engineering Technologies, Al-Majouri, Benghazi, Libya
Waielothman83@gmail.com

تاريخ الاستلام: 2026/9/12 تاريخ المراج عة 9/15/ 2026 تاريخ القبول:2026/10/11- تاريخ النشر: 11/09/ 2026

الملخص

تهدف هذه الدراسة الاستكشافية إلى تقويم مخرجات نموذج XLM-RoBERTa-Large-XNLI في تصنيف الأخبار الحقيقية والكاذبة المكتوبة باللهجة الليبية دون تدريب إضافي. اعتمدت الدراسة ستة نصوص قصيرة أُسندت إليها وسوم مرجعية يدويا؛ نصان بالعربية الفصحى وأربعة نصوص باللهجة الليبية، ونُفذت التجربة باستخدام مكتبة Transformers على Google Colab. أظهر الجدول الأصلي توافقا بين تنبؤات النموذج والوسوم المرجعية في خمس حالات، مقابل اختلاف في حالة واحدة. وتفاوتت درجات الثقة المبلغ عنها، فبلغت 50.5% لأحد النصوص باللهجة الليبية و65.0% للتنبؤ غير المطابق للوسم المرجعي. ولا تمثل هذه الدرجات دقة التصنيف أو احتمالات معايرة للحقيقة. تشير الملاحظات إلى أهمية اختبار أثر اللهجة، لكنها لا تثبت وجود فجوة لهجية بسبب صغر العينة وغياب توثيق مستقل لصحة الأخبار. وتوصي الدراسة ببناء مدونة ليبية موثقة وتوسيع الاختبارات قبل تطوير نموذج متخصص.

الكلمات المفتاحية: الأخبار الكاذبة؛ اللهجة الليبية؛ التصنيف الصفري؛ XLM-RoBERTa؛ معالجة اللغة الطبيعية؛ تقييم النماذج.

Abstract

This exploratory study evaluates the outputs of XLM-RoBERTa-Large-XNLI for zero-shot classification of fake and genuine news in the Libyan dialect. Six short texts with manually assigned reference labels were examined: two in Modern Standard Arabic and four in Libyan Arabic. The experiment was reported as running with Transformers in Google Colab, without task-specific fine-tuning. The original results table shows agreement between model predictions and reference labels in five cases and disagreement in one. Reported model scores include 50.5% for one dialectal text and 65.0% for the misclassified case. These scores are classifier confidence values, not accuracy estimates or calibrated probabilities of truth. The sample is too small to establish a statistically reliable dialectal performance gap. Future work should verify source-level ground truth, compile a larger balanced Libyan dataset, and compare zero-shot and fine-tuned baselines.

Keywords: Fake News Detection; Libyan Dialect; Zero-Shot Classification; XLM-RoBERTa; Natural Language Processing; Model Evaluation.

1. المقدمة

أصبحت منصات التواصل الاجتماعي قناة مهمة لتداول المعلومات والأخبار، وهي بيئة تتداخل فيها المواد الصحيحة والمضللة وتتفاوت فيها موثوقية المصادر. وتتطلب معالجة الأخبار الكاذبة الجمع بين تحليل النص والتحقق من الأدلة والسياق والمصدر، لأن الخصائص اللغوية

وحدها لا تضمن صدق المحتوى (Shu et al., 2017). وفي السياق العربي، تزداد صعوبة التحليل الآلي بفعل التنوع الصرفي والإملائي واللهجي، فضلا عن محدودية مجموعات البيانات المشروحة لبعض الأصناف اللغوية.

تناولت الدراسات السابقة كشف الأخبار الكاذبة باللغة العربية باستخدام التحليل النصي والتعلم العميق، وأتاحت قواعد بيانات لتصنيف مصداقية الأخبار (Himdi et al., 2022; Khalil et al., 2022; Najadat et al., 2022). وفي مجال تصنيف النصوص العربية عموما، اختبر Garash (2025) خوارزميات k-NN وSVM وNaive Bayes على نصوص إخبارية بالفصحى، بينما درس AlZawam and Ahmad (2026) تحسين تصنيف النص العربي عبر ضبط SVM بخوارزمية Harris Hawks Optimization. وعلى الرغم من أهمية هذين العملين في تقييم خوارزميات التصنيف، لا يمكن اعتبار نتائجهما دليلا مباشرا على كفاءة كشف الأخبار الكاذبة باللهجة الليبية، لاختلاف المهمة والبيانات وبيئة التقييم.

وفي ليبيا، تتداول بعض المنشورات الإخبارية تعبيرات محلية تختلف عن الفصحى في المفردات والتراكيب، مثل «توا» و«واجد» و«شنو». ومن ثم تبرز الحاجة إلى اختبار ما إذا كانت نماذج التصنيف متعددة اللغات قادرة على التعامل مع هذه الصياغات دون ضبط متخصص. ويركز هذا البحث على فحص مخرجات نموذج XLM-RoBERTa-Large-XNLI في عينة صغيرة من النصوص، على أن تكون النتائج مؤشرات أولية لتوليد الفرضيات، لا حكما عاما على أداء النموذج أو على اللهجة الليبية.

2. مشكلة البحث وأسئلته

تتمثل مشكلة البحث في محدودية الأدلة التجريبية المتاحة داخل الدراسة بشأن ملاءمة التصنيف الصفري للأخبار القصيرة المكتوبة باللهجة الليبية، مقارنة بنصوص عربية فصيحة، مع صعوبة التمييز بين انخفاض درجة ثقة النموذج ووقوع خطأ فعلي في التصنيف. ويزداد هذا الإشكال عند غياب مصادر مستقلة توثق صحة الأخبار والوسوم المستخدمة مرجعا للتقييم.

السؤال الرئيس: ما الأنماط الأولية التي تظهر في مخرجات نموذج XLM-RoBERTa-Large-XNLI عند تصنيف النصوص الإخبارية باللهجة الليبية مقارنة بالنصوص الفصيحة ضمن العينة المختبرة؟

1. ما مدى توافق تنبؤات النموذج مع الوسوم المرجعية اليدوية في الحالات الست؟
2. كيف تتفاوت درجات الثقة المبلغ عنها بين الحالات، وما حدود تفسيرها؟
3. ما المتطلبات المنهجية اللازمة للانتقال من الاستكشاف الأولي إلى اختبار قابل للتعميم؟

3. أهداف البحث وأهميته

يسعى البحث إلى وصف أداء نموذج التصنيف الصفري على عينة أولية، وتوثيق نتائج كل حالة بصورة قابلة للمراجعة، وتحليل حدود الثقة في المخرجات، واقتراح مسار لتطوير مجموعة بيانات ليبية واختبارات أكثر صرامة. وتكمن أهميته التطبيقية في توجيه الجهود اللاحقة نحو تقييم أنظمة مساعدة للتحقق الإخباري تراعي الخصوصية اللغوية المحلية من دون الادعاء بأن النموذج الآلي بديل عن التحقق المستقل من الوقائع.

4. الدراسات السابقة والإطار النظري

4.1 كشف الأخبار الكاذبة والتصنيف الصفري

قدم Shu et al. (2017) عرضا لخصائص الأخبار الكاذبة على وسائل التواصل وطرائق كشفها، وأشاروا إلى أهمية المعلومات السياقية والتفاعلات الاجتماعية إلى جانب النص. وفي المقابل، صاغ Yin et al. (2019) التصنيف الصفري بوصفه ربط النص بوسوم لم يسبق تدريب النموذج عليها في المهمة نفسها، وأوضحوا توظيف الاستدلال النصي ضمن هذا الإطار. وهذان العملان يحددان أساس المشكلة والمنهج؛ الأول يتعلق بموثوقية المعلومات، والثاني بآلية تصنيف نصوص دون عينات تدريب خاصة بالمهمة.

4.2 النماذج العربية ومجموعات بيانات التضليل

قدم Antoun et al. (2020) نموذج AraBERT لمعالجة اللغة العربية، بينما تناول Conneau et al. (2020) التعلم التمثيلي متعدد اللغات الذي تستند إليه عائلة XLM-R. وطرح Abdul-Mageed et al. (2021) نموذجي ARBERT وMARBERT لدعم فهم العربية ولهجاتها. كما قدم Haouari et al. (2021) مجموعة ArCOV19-Rumors للتحقق من الادعاءات المتداولة على تويتر. وتشترك هذه المساهمات في معالجة مشكلات التمثيل اللغوي والبيانات، لكنها لا توفر بمفردها اختبارا مباشرا لمهمة الأخبار الليبية القصيرة المعتمدة هنا.

ركز Himdi et al. (2022) على كشف الأخبار الكاذبة عربيا عبر التحليل النصي؛ وقدم Khalil et al. (2022) مجموعة AFND واسعة النطاق لمصداقية الأخبار، مع وسم يعتمد على تقدير موثوقية المصدر، وهو أمر يختلف عن التحقق المستقل من صحة كل خبر. ودرس Najadat et al. (2022) التعلم العميق باستخدام العلاقة بين العنوان والمقال. وتظهر هذه الأعمال تنوع تعريف المهمة، بين مصداقية المصدر والتحقق من الادعاء والتوافق بين العنوان والمضمون، بما يستلزم تحديد مهمة الدراسة الحالية بدقة.

4.3 الدراسات الإنجليزية المنشورة في المجلات الليبية المطلوبة

اختبر Garash (2025)، في مجلة العلوم الشاملة، خوارزميات تصنيف تقليدية على مجموعة من 700 مقال إخباري مكتوب بالعربية الفصحى ضمن سبع فئات موضوعية؛ وذكر أن SVM حققت دقة 89.3% في إعداد تجربته. وتفيد هذه النتيجة في إبراز أهمية وجود بيانات موسومة ومقاييس تقييم واضحة، إلا أنها لا تصلح للمقارنة الرقمية المباشرة بنتائج البحث الراهن لاختلاف حجم العينة ومهمة التصنيف وطريقة بناء البيانات.

أما AlZawam and Ahmad (2026)، في مجلة الفاروق للعلوم، فقد تناولا تحسين تصنيف النصوص العربية من خلال ضبط معلمات SVM واختيار السمات بخوارزمية Harris Hawks Optimization. ويسهم عملهما في توضيح أثر تمثيل النص وضبط الخوارزميات في أداء التصنيف، في حين تتميز الدراسة الحالية بتركيزها على تصنيف صفري ثنائي الوسم لنصوص قصيرة ذات تعبيرات ليبية. وتتمثل الفجوة البحثية المحددة في نقص التقييمات الموثقة للتصنيف الصفري على عينات ليبية ذات وسم خبري محقق، لا في غياب جميع أعمال معالجة العربية.

5. منهجية الدراسة

5.1 تصميم البحث والعينة

اعتمد البحث تصميم دراسة حالة استكشافية وصفية ذات عنصر تجريبي محدود، ويحلل البيانات الرقمية الواردة في النسخة الأصلية للتجربة. تضمنت العينة ستة نصوص قصيرة؛ نصين بالفصحى وأربعة صنفتها الدراسة الأصلية ضمن اللهجة الليبية. أُسنِدت إلى النصوص وسوم ثنائية يدويا (حقيقي/كاذب)، من غير إرفاق روابط المنشورات أو سجلات تحقق مستقلة لكل حالة؛ لذا تُسمى هذه الوسوم «وسوما مرجعية مبلغ عنها»، ولا تُعامل بوصفها حقائق محققة خارجيا.

5.2 أداة التصنيف وبيئة التنفيذ

استخدمت الدراسة، بحسب الملف الأصلي، النموذج الجاهز joeddav/xlm-roberta-large-xnli عبر مهمة zero-shot-classification في مكتبة Transformers على منصة Google Colab، من دون ضبط لاحق باستخدام بيانات ليبية. وقد عُرضت فئتا التصنيف بالعربية: «خبر حقيقي» و«خبر كاذب». ويجب في أي إعادة تنفيذ توثيق إصدار النموذج والمكتبة وقالب فرضية الاستدلال (hypothesis_template)، لأن تغيير الصياغة قد يؤدي إلى اختلاف درجات النموذج.

5.3 طريقة قراءة المخرجات

رُصد لكل نص الوسم المرجعي والتنبؤ والفئة المختارة ودرجة الثقة المرتبطة بها وفقا للجدول المرفق في المادة الأصلية. ويُقصد بدرجة الثقة هنا درجة المخرجات التي يعرضها المصنف للوسم المختار، ولا تساوي بالضرورة احتمال صحة الخبر في الواقع. ويمكن عد حالات الاتفاق

مع الوسم المرجعي وصفيا، ولكن لا يجوز الاستدلال من ست حالات فقط على أداء مستقر أو فرق ذي دلالة إحصائية بين الفصحى واللهجة الليبية.

5.4 الضوابط والقيود

لم تتضمن المادة الأصلية وصفا قابلا للتحقق لأسلوب اختيار المنشورات أو تواريخ جمعها أو آلية التحكيم بين المقيّمين أو روابط التحقق من صحة المحتوى. كما أن الاختبار لا يتضمن تقسيم تدريب/اختبار ذا معنى إحصائي بسبب استخدام التصنيف الصفري والعينة الصغيرة. لذلك اقتصر التحليل على الوصف الدقيق لما هو موثق، مع تجنب استنتاج علاقة سببية بين المفردة اللهجية وقرار النموذج.

6. النتائج

يبين الجدول (1) النتائج الست كما وردت في الملف الأصلي بعد إعادة تنظيمها تحريريا، دون إعادة تشغيل النموذج أو تعديل الدرجات المسجلة. وتشير النسبة بين قوسين إلى درجة ثقة المخرجات بالوسم المتوقع، لا إلى دقة النظام.

رقم	نص الخبر	الوسم المرجعي	تنبؤ النموذج (الدرجة)	المطابقة
1	الحكومة تعلن توزيع سيارات مجانا لكل المواطنين	كاذب	كاذب (60.0%)	مطابق
2	وزارة التعليم تعلن موعد الامتحانات النهائية	حقيقي	حقيقي (57.9%)	مطابق
3	توا الحكومة بتوزع فلوس واجد بكل لكل الشباب	كاذب	كاذب (50.5%)	مطابق
4	شنو صار في بنغازي اليوم؟ البلدية قالت فيه صيانة للضي توا	حقيقي	كاذب (65.0%)	غير مطابق
5	علاش ما قالولناش؟ لقو بير ذهب في سبها وبيقسموه علينا	كاذب	كاذب (74.8%)	مطابق
6	بكل تأكيد، وزارة الصحة تعلن وصول شحنة أدوية جديدة	حقيقي	حقيقي (86.7%)	مطابق

جدول (1). نتائج تصنيف النصوص الستة الواردة في المادة الأصلية؛ الدرجة هي درجة ثقة النموذج وليست دقة التصنيف.

6.1 الوصف الكمي للنتائج

اتفق التنبؤ مع الوسم المرجعي المبلغ عنه في خمس حالات من أصل ست، ووقع اختلاف في الحالة الرابعة. وفي النصين المصنفين ضمن الفصحى ظهر اتفاق في الحالتين، مقابل اتفاق في ثلاث من الحالات الأربع المصنفة ضمن اللهجة الليبية. وهذه أعداد وصفية تخص عينة صغيرة غير احتمالية، ولا تصلح لتقدير معدل نجاح النموذج في المجتمع المستهدف أو للمقارنة الاستدلالية بين المجموعتين.

تراوح مقدار ثقة الوسم المتوقع في الجدول بين 50.5% و86.7%. وتبين الحالة الثالثة قرب درجتي الفئتين المحتملتين من التعادل في تلك الحالة وحدها، لكنها لا تثبت أن النموذج «لم يفهم» الكلمات اللهجية. كما تبين الحالة الرابعة أن التنبؤ غير المطابق جاء بدرجة 65.0%، وهي درجة ثقة للتصنيف الخاطئ لا «نسبة خطأ» مقدارها 65%.

6.2 تفسير النتائج

تسمح النتائج بطرح فرضية تستحق الاختبار بشأن أثر التعبير اللهجي في مخرجات النموذج، لكنها لا تعزل هذا الأثر عن صياغة النص أو طوله أو غموض مضمونه أو التوزيع غير المتوازن للحالات. وبوجه خاص، لا يجوز تعميم القول بأن الكلمات «توا» و«واجد» و«علاش» علامات إخبارية كاذبة؛ إذ يمكن ظهورها في أخبار صحيحة أيضا. ويظل التحقق من الخبر متعلقا بالأدلة والمصادر لا بنمط العبارة وحده.

7. المناقشة ومقارنة النتائج بالدراسات السابقة

تشترك الدراسة الحالية مع أعمال كشف الأخبار الكاذبة العربية في اهتمامها بمعالجة المحتوى الإخباري، لكنها تختلف من حيث نوع المدخلات وحجم البيانات وإجراء التقييم. فقد درست أعمال Himdi et al. (2022) و Najadat et al. (2022) مشكلات كشف التضليل على بيانات أكثر تنظيما، في حين اعتمدت التجربة الحالية نصوصا قصيرة قليلة العدد وتصنيفا صفريا دون تدريب خاص. لذلك لا يصح إجراء مقارنة مباشرة بين درجات الثقة المعروضة هنا ومقاييس الدقة أو F1 المنشورة في تلك الأعمال.

ويبرز الاختلاف نفسه عند المقارنة مع Garash (2025)، الذي درس تصنيفا موضوعيا متعدد الفئات لمقالات بالفصحى، ومع AlZawam and Ahmad (2026)، اللذين بحثا تحسين تصنيف النصوص العربية باستخدام SVM وخوارزمية تحسين. وتساند الدراستان أهمية تمثيل النص وتحديد المهمة وتوحيد بروتوكول التقييم، ولا تقدم أي منهما معيارا عدديا مكافئا لتقييم صحة أخبار ليبية في وضع zero-shot. وبذلك تتحدد الإضافة المناسبة للبحث الراهن في توثيق حالة تطبيق محلية أولية واقتراح تجربة لاحقة قابلة للتكرار والاختبار، لا في إثبات تفوق نموذج أو إخفاقه على مستوى اللهجة.

من ناحية المنهج، تستلزم المقارنة اللاحقة جمع منشورات حقيقية موثقة المصدر، وتعيين حالة الصدق اعتمادا على تحقق مستقل، ثم إعداد عينات متكافئة بحسب طول النص ونوع الموضوع والفترة الزمنية، واختبار أكثر من صياغة لوسوم التصنيف. وبعد ذلك يمكن حساب الدقة والاسترجاع و F1 ومصفوفة الالتباس وفواصل الثقة، ومقارنة النموذج الصفري بمرجع بسيط ونموذج عربي مضبوط على بيانات محلية.

8. حدود الدراسة

1. صغر حجم العينة (ست حالات فقط) وعدم إمكان إجراء استدلال إحصائي موثوق منها.
2. عدم توفير روابط المصادر وسجلات تحقق مستقلة للأخبار، ما يحد من يقين الوسوم المرجعية.
3. عدم توثيق إصدار البرمجيات وتهيئة فرضيات التصنيف وسجلات المخرجات القابلة لإعادة الإنتاج بالكامل.
4. وجود تفاوت بين بعض القيم الرقمية المعروضة في مواضع مختلفة من المادة الأصلية؛ ولذلك اعتمد الجدول (1) بوصفه السجل الوحيد المنظم للحالات، ويلزم التحقق منه قبل الإرسال النهائي للمجلة.
5. عدم توافر خط أساس مقارن أو تصميم يضبط أثر اللهجة بصورة مستقلة عن أثر المحتوى.

9. الخاتمة والتوصيات

تؤكد هذه الدراسة الاستكشافية إمكانية تشغيل نموذج XLM-RoBERTa-Large-XNLI لتصنيف نصوص إخبارية عربية، بما فيها نصوص ذات تعبيرات ليبية، دون ضبط إضافي. وقد توافقت تنبؤاته مع الوسوم اليدوية المبلغ عنها في خمس حالات من ست، مع تسجيل حالة اختلاف. غير أن حجم العينة وطبيعة التوسيم وعدم توثيق التحقق من الوقائع تحد من القيمة الاستدلالية للنتائج، وتمنع إثبات فجوة لهجية أو إطلاق حكم نهائي بشأن كفاءة النموذج.

توصي الدراسة ببناء مجموعة بيانات ليبية متنوعة ومتوازنة ذات مصادر معلنة ووسوم تم التثبت منها، وتوثيق إعدادات التشغيل والبذور والإصدارات، واختبار حساسية النموذج لصياغة الملصقات، ثم مقارنة تصنيفه الصفري بطرائق مدربة على بيانات محلية. ويمكن النظر في تطوير نموذج متخصص للهجة الليبية لاحقا إذا أثبتت دراسات أكبر الحاجة إلى ذلك وجدواه، مع إبقاء التحقق البشري والرجوع إلى المصادر الأصلية جزءا أساسيا من العمل الإخباري.

المراجع

Abdul-Mageed, M., Elmadany, A., & Nagoudi, E. M. B. (2021). ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In Proceedings of the 59th Annual Meeting of the Association for Computational

Linguistics (pp. 7088–7105). Association for Computational Linguistics. https://aclanthology.org/2021.acl-long.551/

AlZawam, H. A., & Ahmad, T. M. A. (2026). Strengthening support vector machine capability through Harris Hawks optimization for Arabic text categorization. *Al-Farooq Journal of Sciences*, 2(4), 675–685. https://doi.org/10.65405/nsmnf959

Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT: Transformer-based model for Arabic language understanding. In Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools (pp. 9–15). European Language Resources Association. https://aclanthology.org/2020.osact-1.2/

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 8440–8451). Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.747/

Garash, S. S. (2025). A comparative study of Arabic text classification using k-NN, SVM, and Naive Bayes. *Comprehensive Journal of Science*, 10(37), 867–879. https://doi.org/10.65405/.v10i37.326

Haouari, F., Hasanain, M., Suwaileh, R., & Elsayed, T. (2021). ArCOV19-Rumors: Arabic COVID-19 Twitter dataset for misinformation detection. In Proceedings of the Sixth Arabic Natural Language Processing Workshop (pp. 72–83). Association for Computational Linguistics. https://aclanthology.org/2021.wanlp-1.8/

Himdi, H., Weir, G., Assiri, F., & Al-Barhamtoshy, H. (2022). Arabic fake news detection based on textual analysis. Arabian Journal for Science and Engineering, 47(8), 10453–10469. https://doi.org/10.1007/s13369-021-06449-y

Khalil, A., Jarrah, M., Aldwairi, M., & Jaradat, M. (2022). AFND: Arabic fake news dataset for the detection and classification of articles credibility. Data in Brief, 42, 108141. https://doi.org/10.1016/j.dib.2022.108141

Najadat, H., Tawalbeh, M., & Awawdeh, R. (2022). Fake news detection for Arabic headlines-articles news data using deep learning. International Journal of Electrical and Computer Engineering, 12(4), 3951–3959. https://doi.org/10.11591/ijece.v12i4.pp3951-3959

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36. https://doi.org/10.1145/3137597.3137600

Yin, W., Hay, J., & Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (pp. 3914–3923). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1404