SJSI
Syrian Journal for Science & Innovation

Volume 4 Issue 3
November 2026



https://journal.hcsr.gov.sy/

بسم الله الرحمن الرحيم



نحو آفاق جديدة من التكامل المعرفي والابتكار المستدام

أعزاءنا القراء، من باحثين ومبتكرين وصناع قرار، نرحب بكم في عدد جديد من **مجلة العلم والابتكار السورية**، العدد الذي يأتي متزامناً مع حراك بحثي استثنائي يسعى لتطويع التكنولوجيا والعلوم التطبيقية في خدمة المجتمع، وتقديم حلول مبتكرة لأعقد التحديات التقنية والبيئية والاقتصادية التي تواجهنا اليوم.

إنّ الفلسفة الأساسية التي تنطلق منها مجلتنا لا تقف عند حدود نشر المعرفة فحسب، بل تتعداها إلى إبراز **التكامل المعرفي بين مختلف القطاعات**، ودعم الحلول الهجينة التي تدمج وعي الإنسان بذكاء الآلة، وصولاً إلى تحقيق التنمية المستدامة وحماية البنى التحتية الحيوية. ويحمل هذا العدد في طياته نخبة من الدراسات المتميزة التي تجسد هذا التوجه بشكل عملي وتطبيقي.

في عالم رقمي متسارع، يفرض الأمن السيبراني نفسه كأولوية قصوى لحماية الأنظمة والمؤسسات. وفي هذا السياق، يسلط العدد الضوء على فجوة أمنية حرجة تتمثل في «العنصر البشري» عبر بحث نوعي يقدم **نظاماً هجيناً يجمع بين الأساليب القائمة على القواعد والتعلم الآلي لكشف هجمات الهندسة الاجتماعية**. يبرهن هذا البحث بكفاءة أن الحلول التقنية الصرفة لا يمكنها العمل بمعزل عن تمكين المستخدم ورفع وعيه الرقمي، مقدماً نموذجاً مبتكراً لحظر التهديدات في الوقت الفعلي بدقة إحصائية واعدة.

وامتداداً للابتكار الرقمي في حماية المنظومات المتصلة، تأخذنا دراسة أخرى إلى عمق الثورة الصناعية الرابعة، حيث يتم **تعزيز أمان إنترنت الأشياء (IoT) عبر معمارية هجينة مبتكرة تشترك فيها ثلاثة محاور**؛ تبدأ بتوليد مفاتيح تشفير ديناميكية مستوحاة من الهندسة المعقدة لشبكة العنكبوت، وتمر بضغط البيانات باستخدام تحليل المكونات الرئيسية (PCA)، وتتوج بتأمين سلامة البيانات عبر تقنية سلسلة الكتل (Blockchain) لمواجهة هجمات حجب الخدمة (DoS). ويمثل هذا البحث توازناً عبقرياً بين صلابة الحماية الأمنية وكفاءة استهلاك الطاقة في الشبكات اللاسلكية.

ولا تقتصر حدود الابتكار في مجلتنا على الفضاء الرقمي، بل تمتد لتلامس الأرض والبيئة والصحة المهنية. حيث يقدم هذا العدد دراسة هندسية وبيئية رائدة تقيم **تأثير استخدام الزيوليت الصناعي في إنتاج خلطات المجبول الإسفلتي الدافئ (WMA)** كبديل مستدام للمجبول الساخن التقليدي. يفتح هذا البحث آفاقاً تطبيقية هامة في قطاع البناء والتشييد المحلي عبر خفض درجات حرارة الإنتاج، مما ينعكس مباشرة كوفر اقتصادي في الطاقة، وتقليص ملموس للانبعاثات الغازية الضارة، حمايةً للبيئة ولصحة العمال على حد سواء.

وفي إطار الحلول البيئية المستدامة القائمة على الطبيعة، نختتم جولة هذا العدد بدراسة حيوية تطبيقية تبحث في **العلاقة بين صفات النمو الفيزيولوجية لنبات دوار الشمس وامتصاص المعادن الثقيلة في ظروف التلوث الحقلي**. تضع هذه الدراسة حجر أساس استرشادي غاية في الأهمية لبرامج «المعالجة النباتية» في الترب الملوثة شمال غرب سوريا، عبر تقديم نماذج تنبؤية بسيطة القياس وقليلة التكلفة تعتمد على مساحة الورقة وطول الساق لتقدير كفاءة التطهير الحيوي قبل الدخول في التحليلات المخبرية المعقدة.

أيها السيدات والسادة،

إنّ هذا التنوع الثري في أبحاث العدد– من خوارزميات الذكاء الاصطناعي والأمن السيبراني، إلى سلاسل الكتل وإنترنت الأشياء، وصولاً إلى المواد الإنشائية المستدامة والمعالجة الحيوية للبيئة– يؤكد مجدداً أنّ العقل العلمي السوري والمحلي قادراً على الابتكار وتقديم الحلول القابلة للتطبيق الميداني رغم كل التحديات.

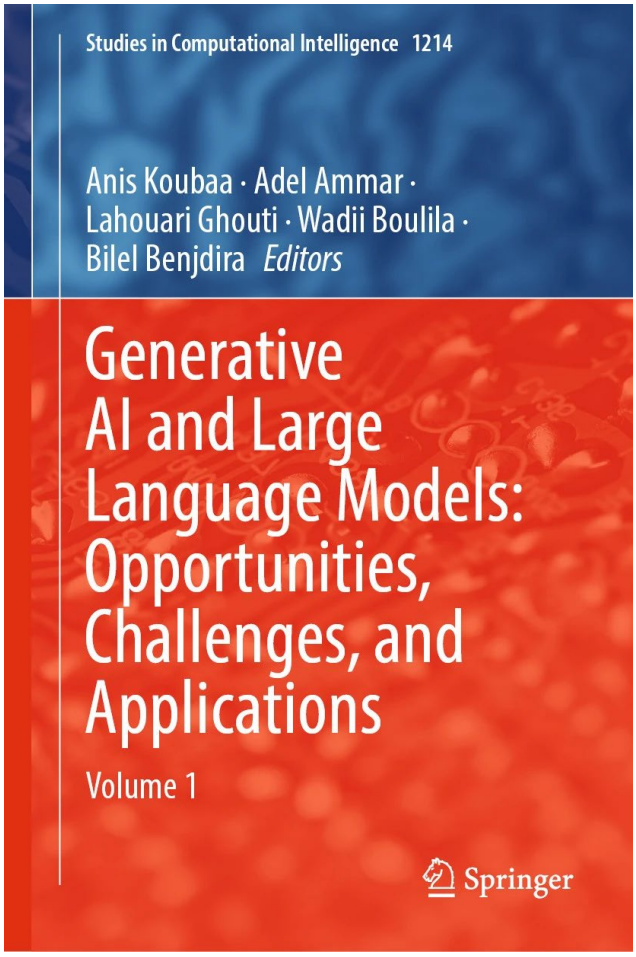
نأمل أن يجد القارئ في هذا العدد منبعاً للإلهام والمعرفة، ونعاهدكم في **مجلة العلم والابتكار السورية** أن نستمر في منصتنا هذه كجسر يربط بين الفكرة المبتكرة والتطبيق النافع.

نتمنى لكم قراءة ممتعة ومثمرة

والله وليّ التوفيق

رئيس التحرير

أ. د. نبيل نادر قوشجي



قراءة في كتاب: **(الذكاء الاصطناعي التوليدي والنماذج اللغوية الكبيرة: الفرص والتحديات والتطبيقات)**

المحررون: د. أنيس قوبعة، عادل عمار، الهواري الغوطي، وديع بليلة، بلال بن جديرة

صدر الكتاب في 2025: Springer Nature، ويقع في /310/ صفحة.

يشهد العالم اليوم تحولاً جذرياً في مجال الذكاء الاصطناعي بفضل التطور المتسارع للذكاء الاصطناعي التوليدي (Generative AI)والنماذج اللغوية الكبيرة(Large Language Models) ، التي أصبحت من أكثر التقنيات تأثيرًا في البحث العلمي والصناعة والتعليم والرعاية الصحية. وفي هذا السياق، يأتي كتاب **Generative AI and Large Language Models: Opportunities, Challenges, and Applications**ليقدم رؤية علمية متكاملة حول أحدث التطورات في هذا المجال، جامعاً بين الأسس النظرية والتطبيقات العملية والتحديات المستقبلية.

يضم الكتاب مجموعة من الفصول التي أعدها باحثون متخصصون، ويستعرض بصورة منهجية كيفية توظيف الذكاء الاصطناعي التوليدي في مجالات متعددة، من بينها التشخيص الطبي، ومعالجة اللغات الطبيعية، وتطوير البرمجيات، والاستدامة البيئية، إلى جانب تطبيقات الذكاء الاصطناعي متعددة الوسائط. كما يناقش البنية التقنية للنماذج اللغوية القائمة على معمارية **Transformer**، مع التركيز على التحديات المرتبطة بتدريب هذه النماذج وتحسين كفاءتها.

ومن أبرز الموضوعات التي يتناولها الكتاب اهتمامه باللغات منخفضة الموارد، إذ يعرض تجربة تطوير نموذج **ArabianGPT**بوصفه تكييفًا لنموذج GPT-2 لخدمة اللغة العربية. ويبرز هذا الفصل أهمية بناء نماذج لغوية تراعي الخصائص اللغوية والثقافية للغات التي لا تحظى بتمثيل

كافٍ في مجموعات البيانات العالمية، وهو توجه يكتسب أهمية متزايدة في ظل السعي إلى تطوير حلول ذكاء اصطناعي أكثر شمولًا وإنصافًا.

ولا يقتصر الكتاب على عرض فرص الذكاء الاصطناعي، بل يولي اهتمامًا واضحًا للتحديات التي ترافق انتشاره، مثل التحيز في البيانات، والهلوسة(Hallucination) ، والخصوصية، والأمن السيبراني، والملكية الفكرية، والمسؤولية الأخلاقية. ويؤكد أن نجاح تطبيقات الذكاء الاصطناعي لا يعتمد فقط على التقدم التقني، وإنما أيضًا على وجود أطر تنظيمية وأخلاقية تضمن الاستخدام المسؤول لهذه التقنيات، وهو ما يتوافق مع الاتجاهات العالمية الحالية في حوكمة الذكاء الاصطناعي.

ويمتاز الكتاب بأسلوب أكاديمي منظم يجمع بين الدراسات النظرية ودراسات الحالة، مع اعتماد مراجع حديثة وأمثلة تطبيقية، مما يجعله مرجعًا مناسبًا للباحثين وطلبة الدراسات العليا والمهندسين وصناع القرار المهتمين بتطبيقات الذكاء الاصطناعي. كما أن تنوع مؤلفي الفصول يضفي على الكتاب بعدًا متعدد التخصصات، ويتيح للقارئ الاطلاع على وجهات نظر علمية متنوعة حول مستقبل الذكاء الاصطناعي التوليدي.

تعريف بالمحررين:

- **أنيس قوبعة:** أستاذ في كلية الهندسة بجامعة الفيصل في الرياض.
- **عادل عمار:** باحث وأستاذ بجامعة الأمير سلطان في الرياض.
- **الهواري الغوطي:** خبير في الذكاء الاصطناعي وتعلم الآلة.
- **وديع بليلة:** باحث متخصص في التعلم العميق والذكاء الاصطناعي بجامعة الأمير سلطان.
- **بلال بن جديرة:** باحث في التعلم العميق والذكاء الاصطناعي بجامعة الأمير سلطان.

دراسة علاقات الارتباط بين صفات النمو الفيزيولوجية لنبات دوار الشمس وكمية امتصاص المعادن الثقيلة في ظروف التلوث الحقلي

Study of the Relationships Between Sunflower Growth Traits and Heavy Metal Uptake Under Field Pollution Conditions

عبد الحميد الخالد: وزارة التعليم العالي والبحث العلمي
عنوان التواصل: dep.educational@mohe.gov.sy
تاريخ التسليم: ١٩ تموز ٢٠٢٦ – **تاريخ القبول:** ١٠أيلول ٢٠٢٦

ABSTRACT

This study aimed to quantitatively evaluate the relationship between two physiological growth traits (stem height and leaf area) and the uptake of three heavy metals (chromium, cobalt, and nickel) from contaminated soils by sunflower (Helianthus annuus L.). A pot experiment was conducted using three sunflower cultivars (local, American, and Spanish) grown in soils collected from three locations in northwestern Syria (Harem, Sarmada, and Darat Izza) for 45 days. The experiment was arranged in a Randomized Complete Block Design (RCBD) with three replicates. Data were analyzed using Two-Way Analysis of Variance (Two-Way ANOVA), Pearson's correlation coefficient (r), and simple linear regression. The results revealed that leaf area was significantly and positively correlated with the uptake of all three heavy metals, with correlation coefficients of (r = ***0.783) for chromium, (r = ***0.691) for cobalt, and (r = ***0.825) for nickel ($p < 0.001$). Regression models based on leaf area explained between 47.8% and 68.0% of the variation in metal uptake ($R^2 = 0.478–0.680$). In contrast, stem height showed significant positive correlations only with chromium (r = **0.615, $p < 0.01$) and nickel (r = *0.470, $p < 0.05$), whereas no significant relationship was observed with cobalt. Although Harem soil contained the lowest total concentrations of the studied metals, it exhibited the highest uptake levels, indicating that metal bioavailability, rather than total soil metal concentration, is the principal determinant of metal uptake. These findings demonstrate the potential for developing preliminary predictive models that can be further refined and validated in future studies. Moreover, they highlight leaf area as a reliable proxy indicator for estimating phytoremediation efficiency prior to laboratory-based metal analysis.
Keywords: Sunflower (Helianthus annuus L.); heavy metals; metal uptake; leaf area; Pearson's correlation; simple linear regression; phytoremediation.

الملخص

تهدف هذه الدراسة إلى تقييم العلاقة الكمية بين صفتَي النمو الفيزيولوجيتين (طول الساق ومساحة الورقة) وكمية امتصاص ثلاثة عناصر ثقيلة (الكروم والكوبالت والنيكل) من التربة الملوثة في نبات دوار الشمس.(Helianthus annuus L.) أُجريت التجربة في أصص بثلاثة أصناف (المحلي والأمريكي والإسباني) على ترب مأخوذة من ثلاثة مواقع في شمال غرب سوريا (حارم وسرمدا ودارة عزة) لمدة 45 يوماً، وفق تصميم القطاعات الكاملة العشوائية (RCBD) بثلاثة مكررات. حُللت البيانات بتحليل التباين الثنائي (Two-Way ANOVA) ومعامل ارتباط بيرسون (r) والانحدار الخطي البسيط. أظهرت النتائج أن مساحة الورقة ارتبطت ارتباطاً طردياً دالاً إحصائياً مع كميات امتصاص العناصر الثلاثة جميعها(r=***0.783) للكروم، (r=***0.691)للكوبالت، (r=***0.825)للنيكل، ($p<0.001$)، وفسّرت نماذج الانحدار المبنية عليها ما بين 47.8% و68% من تباين كميات الامتصاص (R^2=0.478–0.680) أما طول الساق فارتبط ارتباطاً دالاً مع الكروم والنيكل فقط (r=**0.615)و(r=*0.470)، ($p<0.05$)، دون دلالة معنوية مع الكوبالت. وقد سجّل موقع حارم أعلى كميات امتصاص رغم امتلاكه أدنى التراكيز الكلية للعناصر، مما يؤكد أن الإتاحة الحيوية هي المحدد الفعلي للامتصاص لا التركيز الكلي للعنصر في النبات. أظهرت النتائج إمكانية بناء نماذج تنبؤية أولية يمكن تطويرها والتحقق من صلاحيتها في دراسات لاحقة، وتُعزز مساحة الورقة موقعها كمعيار استرشادي غير مباشر (Proxy indicator) لتقدير كفاءة المعالجة النباتية قبل إجراء التحليل المخبري.
الكلمات المفتاحية: دوار الشمس، المعادن الثقيلة، كمية الامتصاص، مساحة الورقة، الارتباط الخطي البسيط، الانحدار الخطي البسيط، المعالجة النباتية.

المقدمة

يُعدّ التلوث بالمعادن الثقيلة من أبرز التحديات البيئية التي تواجه القطاع الزراعي على المستويين المحلي والعالمي، إذ تتراكم هذه المعادن

في التربة نتيجة الأنشطة الصناعية والزراعية المكثفة والصراعات المسلحة، مُشكِّلةً خطراً متصاعداً على صحة الإنسان والنظم البيئية (Wuana & Okieimen, ٢٠١١). وعلى الرغم من توافر تقنيات معالجة كيميائية وفيزيائية متعددة، فإن تكلفتها المرتفعة وتأثيراتها السلبية على بنية التربة وميكروباتها دفعت الباحثين نحو المعالجة النباتية (Phytoremediation) كبديل مستدام يعتمد على استخدام النباتات واسعة الكتلة الحيوية لاستخلاص المعادن الثقيلة أو تثبيتها. ويتصدر نبات دوار الشمس (Helianthus annuus L.) قائمة النباتات المرشحة لهذه التقنية بفضل جملة من الخصائص الزراعية والفسيولوجية المتميزة؛ فهو سريع النمو، واسع الكتلة الحيوية، مقاوم لطيف واسع من الظروف المناخية والتربوية، وقادر على امتصاص عناصر ثقيلة متعددة كالرصاص والكروم والكادميوم والنيكل والكوبالت وتركيزها في أنسجته (Farid et al., ٢٠١٨; Stoikou et al., ٢٠١٧). وقد أكدت الأبحاث أن دوار الشمس يتميز بكتلة حيوية مرتفعة ونموٍّ سريع وقدرة على تخزين كميات ملحوظة من المعادن الثقيلة في أنسجته الدهنية (Huang et al., ٢٠٢٢)، مما يجعله موضع اهتمام واسع في برامج المعالجة النباتية. غير أن كفاءة الاستخلاص النباتي لا تحددها خصائص التربة فقط، وإنما تتأثر بشكل مباشر بالصفات الفسيولوجية للنبات ذاته. فالكتلة الحيوية الأكبر تعني بالضرورة مساحة امتصاص جذرية أوسع، وتدفقاً وعائياً أعلى عبر نظام الـ xylem يُسهم في نقل أيونات المعادن من الجذور نحو الأجزاء الهوائية Page & Feller (٢٠١٥) وقد أثبتت الدراسات الحديثة وجود ارتباط موجب دال إحصائياً بين الكتلة الحيوية الهوائية (AGB) ومعامل الانتقال (TF)، مما يُشير إلى أن النباتات الأكثر نمواً تُطوّر آليات أكثر كفاءة لنقل المعادن نحو المجموع الخضري(Huang et al., ٢٠٢٢) كما تُشكّل مساحة الورقة مؤشراً فسيولوجياً مزدوج الدلالة؛ إذ تعكس من جهة القدرة التمثيلية الضوئية وإنتاج المادة الجافة، ومن جهة أخرى السحب التبخيري– النتحي الذي يدفع تدفق المحلول الوعائي الحامل لأيونات المعادن الثقيلة صعوداً من الجذور نحو الأوراق. وعلى الرغم من توافر دراسات تقيّم كفاءة دوار الشمس في المعالجة النباتية عبر مؤشرات BAF وTF وكمية الامتصاص، فإن الدراسات التي تربط هذه المؤشرات ربطاً كمياً بصفات النمو الفسيولوجية المقيسة — كطول الساق ومساحة الورقة — تبقى محدودة في الأدبيات، وأكثر ندرةً في السياق السوري والإقليمي. وتأتي أهمية هذه الدراسة من كونها تُقدّم نماذج تنبؤية قابلة للتطبيق الميداني، تمكّن الباحثين والممارسين من تقدير كمية الامتصاص المتوقعة من صفات النمو البسيطة القياس قبل إجراء التحليل المخبري المكلف.

مبررات وأهمية البحث

على الصعيد البيئي، تمثّل الترب الزراعية في شمال غرب سوريا بيئةً ملوثة بالمعادن الثقيلة نتيجة تراكم النشاط الصناعي تاريخياً وتداعيات الصراعات المسلحة في السنوات الأخيرة، في ظل غياب شبه تام لبرامج المعالجة البيئية الممنهجة. ويُشكّل توصيف كفاءة الأصناف النباتية المتاحة في هذه الترب خطوةً لا غنى عنها لتصميم أي برنامج معالجة فعّال وذي جدوى اقتصادية. ومن الناحية المنهجية، تُقدّم هذه الدراسة نهجاً مغايراً للدراسات التقليدية التي تكتفي بقياس مؤشرات BAF وTF بوصفها قيماً مطلقة؛ إذ تسعى إلى ربطها بصفات النمو القابلة للقياس قبل الحصاد وبتكلفة زهيدة، مما يُتيح التنبؤ المبكر بالكفاءة المتوقعة قبل إجراء أي تحليل مخبري. أما تطبيقياً، تتيح نماذج الانحدار الخطي المُستخرجة تقدير الكمية المتوقع استخلاصها من المعادن الثقيلة بمجرد قياس مساحة الورقة أو طول الساق، دون الحاجة إلى تحليل أنسجة نباتية مكلف، مما يُخفّض تكلفة الرصد والتقييم البيئي الميداني بصورة ملموسة ويُيسّر تطبيقها على نطاق واسع.

ومن وجهة نظر أكاديمية، تُسدّ هذه الدراسة فجوةً بحثية في الأدبيات المتعلقة بالمعالجة النباتية في السياق السوري والإقليمي، إذ تربط بين نتائج ميدانية حقيقية ونماذج إحصائية تنبؤية في إطار متكامل يُغني قاعدة البيانات العلمية اللازمة لتطوير استراتيجيات المعالجة في الأراضي الجافة وشبه الجافة.

أهداف البحث

١.دراسة تأثير كل من الموقع والصنف النباتي في صفتَي طول الساق ومساحة الورقة لأصناف دوار الشمس المدروسة المزروعة في ترب ملوثة بالعناصر الثقيلة شمال غرب سوريا
٢.تحديد العلاقة الكمية بين كمية امتصاص المعادن الثقيلة (الكروم والكوبالت والنيكل) من جهة، وصفتَي طول الساق ومساحة الورقة من جهة أخرى، باستخدام معامل ارتباط بيرسون. (r)
٣.بناء نماذج انحدار خطي بسيط (Linear Regression) للأزواج الدالة إحصائياً، وتقييم قدرتها التفسيرية عبر معامل التحديد (R^2)، بهدف التنبؤ بكمية الامتصاص انطلاقاً من صفات النمو المقيسة.
٤.تحديد أي الصفتين (طول الساق أم مساحة الورقة) أقوى ارتباطاً بكفاءة امتصاص المعادن الثقيلة، لتوجيه برامج الانتخاب والتحسين الوراثي المستقبلية نحو الصفات الأكثر تأثيراً في كفاءة المعالجة النباتية.

الدراسة المرجعية

أولاً: تلوث التربة بالمعادن الثقيلة وأثره على صفات النمو النباتي

تتراكم المعادن الثقيلة في التربة عبر مصادر متعددة تشمل الصناعة والتعدين ومياه الصرف الصحي والأسمدة الكيميائية، وعند تجاوز تراكيزها العتبات الحرجة فإنها تُحدث تثبيطاً ملحوظاً في صفات النمو النباتي. فقد أظهرت دراسة أُجريت على ثلاثة أصناف من دوار الشمس في تربة ملوثة بالكروم أن أطوال الجذور والمجموع الخضري وأوزانها الرطبة والجافة انخفضت تدريجياً مع ارتفاع تراكيز الكروم في التربة. وبالمثل، أفادت دراسة أخرى بأن تطبيق ٢٠٠ mg/kg من الرصاص والكادميوم في التربة أدى إلى خفض الأوزان الرطبة للمجموع الخضري والجذور بنسبة تصل إلى ٧٦,٦٪ و٨٨,٥٪ على التوالي. هذا التثبيط في النمو له انعكاسات ذات بعدين على المعالجة النباتية: فمن ناحية يُقلل من الكتلة الحيوية المتاحة للامتصاص، أما ناحية أخرى قد يُشير إلى أن النبات يعيد تخصيص موارده الفسيولوجية للتعامل مع الإجهاد المعدني. (Farid et al. ٢٠١٨)
ثانياً: دوار الشمس في المعالجة النباتية يحتل دوار الشمس مكانةً بارزة في أبحاث المعالجة النباتية: فقد أظهرت دراسات مقارنة عدة أن مجموعه الخضري وجذوره سجّلا أعلى كميات امتصاص للكوبالت والكروم مقارنةً بالأنواع النباتية الأخرى المختبرة في الترب الملوثة ذاتها وفي دراسة حديثة أُجريت في ظروف البحر المتوسط وشبه الجافة، أكدت النتائج أن تراكيز النيكل في أنسجة دوار الشمس ارتفعت مع ارتفاع التركيز في التربة، مع تركّز أعلى في الجذور يليه السيقان والأوراق، وهو نمط يتوافق مع سلوك الاستبعاد المشابه لما رصده البحث الحالي. (Stoikou et al. ٢٠١٧) كما استُخدم دوار الشمس في تجارب التخليص المائي (Hydroponic Phytoremediation) لخمسة عناصر في آنٍ واحد. وأبدى قدرة على امتصاص الكادميوم والكروم والنيكل والزرنيخ والحديد بصورة متزامنة، مما يُظهر مرونته كمنظومة معالجة متعددة العناصر.

ثالثاً: الكتلة الحيوية كمؤشر لكفاءة الاستخلاص النباتي

يُجمع الباحثون على أن الكتلة الحيوية المرتفعة شرط أساسي لنجاح الاستخلاص النباتي. فقد أنتج (Nehnevajova et al. ٢٠٠٧) طفرات عملاقة من دوار الشمس أظهرت قدرة استخلاص للكادميوم والزنك والرصاص تفوق الشاهد بمقدار ٧,٥ و٩,٢ و٨,٢ مرة على التوالي مما يُثبت أن تضخيم الكتلة الحيوية يترجم مباشرةً إلى تضخيم الكمية الكلية للمعدن المُستخلَص. وقد استخدم باحثون نموذجاً رياضياً لتقدير تراكم المعادن الثقيلة في دوار الشمس

بوصفه دالةً للكتلة الحيوية عبر الزمن (Zhao et al., ٢٠٢٣)، مُقدِّمين بذلك شاهداً كمياً على الترابط بين نمو النبات وكمية الاستخلاص.ومن المنظور الفسيولوجي، كشفت الدراسات الحديثة أن الارتباط الموجب الدال بين الكتلة الحيوية الهوائية ومعامل الانتقال TF يعكس حقيقة أن النباتات الأكثر نمواً تمتلك طاقة أعلى لنقل المعادن من الجذور نحو المجموع الخضري عبر نظام الـ xylem Page & Feller (٢٠١٥) وهذا يعني أن اختيار صنف نباتي بكتلة حيوية أعلى يُحقق زيادة المساحة الامتصاصية الجذرية، وتعزيز كفاءة النقل الوعائي نحو الأجزاء القابلة للحصاد.

رابعاً: مساحة الورقة وعلاقتها بالنقل الوعائي للمعادن

تُمثل مساحة الورقة مؤشراً فسيولوجياً مركّباً يعكس حجم السطح الضوئي والسحب التبخيري–النتحي. ففي ظروف التربة الملوثة، يمتص النبات الماء من التربة عبر الجذور ويسحبه نحو الأوراق بفعل النتح، وهذا التدفق يحمل معه في محلول الـ xylem أيوناتٍ معدنية مذابة. وبذلك تصبح مساحة الورقة الأكبر مقترنةً بتدفق وعائي أعلى، مما يُفسّر قوة الارتباط بين هذه الصفة وكميات الامتصاص المسجّلة في البحث الحالي. وقد أشار (Deng et al., ٢٠٢١) إلى أن عنصر النيكل تحديداً يُظهر حركية مرتفعة داخل نظام الـ phloem و xylem، مما يجعله أكثر استجابةً للتدفق الوعائي المرتبط بمساحة الورقة مقارنةً بالكروم الأكثر احتجازاً في الجذور.

خامساً: الفجوة البحثية التي تُعالجها هذه الدراسة

على الرغم من غزارة الدراسات التي تُقيّم كفاءة دوار الشمس في المعالجة النباتية عبر مؤشرات BAF و TF، تظل الدراسات التي تربط ربطاً كمياً بين صفات النمو الفسيولوجية السهلة القياس — كطول الساق ومساحة الورقة — وكمية الامتصاص الفعلية للمعادن الثقيلة محدودةً في الأدبيات. وتتميز دراستنا الحالية بالعمل على ترب مُلوَّثة طبيعياً من مناطق شمال غرب سوريا، بثلاثة أصناف وطنية وأجنبية في ظروف بيئية فعلية، مما يجعل من نتائجها ذات قيمةٍ تطبيقية مباشرة لبرامج الانتخاب النباتي وتقييم الكفاءة الميدانية لمشاريع المعالجة النباتية في المنطقة.

المواد والطرائق

أولاً: المواد

١ .المادة النباتية:

اعتُمدت في هذه الدراسة ثلاثة أصناف من نبات دوار الشمس (Helianthus annuus L.) تتباين في أصولها الجينية:

صنف محلي (بلدي) خلطي التلقيح متكيف مع الظروف البيئية السائدة في المنطقة.

صنف أمريكي هجين، حُصل عليه من الإدارة العامة للبحوث العلمية الزراعية.

صنف إسباني، حُصل عليه من المصدر ذاته.

٢ .الأجهزة المستخدمة:

قِيست تراكيز المعادن الثقيلة في عينات التربة والأنسجة النباتية باستخدام جهاز مطيافية الانبعاث البصري البلازمي المقترن بالحث (ICP–OES)، موديل ICAP ٦٥٠٠ Duo من إنتاج شركة Thermo Scientific. وهُضمت العينات بجهاز هضم كلداهل موديل ١٢D–KDN. وشملت معدات التجربة أصصاً بارتفاع ٣٠ cm وقطر ٣٥cm، وموازين حساسة بدقة ٠,٠٠٠١ g لوزن الأجزاء الجافة و٠,٠١ g للأجزاء الطازجة، فضلاً عن مجففة هوائية ومطحنة للطحن الجاف.

ثانياً: طرائق البحث:

١ .وصف مواقع جمع العينات:

جُمعت عينات التربة المستخدمة في التجربة من ثلاثة مواقع في شمال غرب سوريا هي حارم وسرمدا ودارة عزة، وقد اعتُمد في الاختيار على توافر بيانات تلوث سابقة تُشير إلى تراكم ملحوظ للمعادن الثقيلة في هذه المناطق (Alhasan et al., ٢٠٢٣). ويُبين الجدول (١) الخصائص الفيزيائية والكيميائية للترب المدروسة قبل الزراعة، ويتضح منه تباينا في تراكيز المعادن الثقيلة بين المواقع رغم تقارب قيم pH التي تراوحت بين ٧,٧٧ و٧,٨٨، وهو مدى متعادل إلى قلوي خفيف.

الجدول (1) الخصائص الفيزيائية والكيميائية للترب المدروسة قبل الزراعة.

الخاصية	الواحدة	تربة حارم	تربة سرمدا	تربة دارة عزة
Ph	-	7.88	7.85	7.77
EC	$dS.m^{-1}$	0.492	0.462	0.295
O.M.	%	5.46	4.41	4.84
Total Cr	$mg.kg^{-1}$	112.48	149.17	172.83
Total Co	$mg.kg^{-1}$	20.601	50.727	29.57
Total Ni	$mg.kg^{-1}$	152.5	183.61	183.61

٢ .جمع عينات التربة وتحضيرها:

جُمعت التربة من عمق (٠–٥) cm عبر مسارات قطرية بطول ٥٠٠m انطلاقاً من نقطة مرجعية محددة في كل موقع، مع أخذ عينات فرعية من جانبي كل مسار على فواصل منتظمة لضمان التمثيل المكاني الكافي. دُمجت العينات الفرعية وجُففت في الهواء، ثم نُخلت عبر منخل ٢mm قبل الاستخدام. قِيست الخصائص الفيزيوكيميائية للتربة بما فيها pH والناقلية الكهربائية (EC) والمادة العضوية والتراكيز الكلية للمعادن الثقيلة قبل بدء التجربة.

٣ .تنفيذ التجربة الزراعية:

نُفِّذت التجربة في الحقل المفتوح في محطة بحوث كفريحمول التابعة لإدارة البحوث الزراعية في إدلب. زُرعت البذور في الأصص المملوءة بالتربة الملوثة بتاريخ ٢٠٢٣/٧/١٤، واعتُمد تصميم القطاعات الكاملة العشوائية (RCBD) بثلاثة مكررات لكل معاملة، ليبلغ المجموع الكلي للوحدات التجريبية ٢٧ وحدة. وُزِّعت الأصص عشوائياً بمسافات لا تقل عن ٣٥ cm بين كل نبات واعتُمدت نسبة ٧٠٪ من السعة الحقلية ريّاً محددة بالوزن مرتين أسبوعياً على مدار ٤٥ يوماً هي مدة التجربة.

٤ .قياس الصفات الفيزيولوجية:

عند الحصاد قِيس طول الساق بالسنتيمتر من قاعدة الساق حتى الزهرة، وقُدِّرت مساحة الورقة بالطريقة الوزنية النسبية (Gravimetric method) عبر رسم الورقة على ورقة مصنفرة موحدة الوزن واستنتاج مساحتها من نسبة الوزن. كما قِيس الوزن الرطب والجاف لكل جزء من أجزاء النبات (الجذر والساق والأوراق والأزهار) منفصلاً، وذلك بعد تجفيف العينات على درجة ٧٠°C لمدة ثلاث ساعات.

٥ .تحضير عينات النبات وتحليلها:

قُسِّمت النباتات بعد جمعها إلى أجزائها الأربعة (الجذر،

الساق، الأوراق، الأزهار)، وغُسلت بالماء منزوع الشوارد وجُففت وطُحنت. هُضمت العينات النباتية بطريقة الهضم حجماً/حجماً (HCl:HNO_3 = 3:1) الرطب بالماء الملكي على مرحلتين: هضم بارد لمدة 16 ساعة عند درجة الغرفة، يعقبه هضم ساخن عند 130°C لمدة ساعتين. من المزيج الحمضي، 10 ml من كل عينة بـ 1 g عولجت ، ثم(G4) ورُشِّحت المحاليل الناتجة بأقماع زجاجية مسامية دون تمديد إضافي لتعزيز ICP-OES حُلِّلت مباشرةً بجهاز ISO 15587 (ISO, 2002).حدود الكشف، وفقاً للمعيار أُرسلت العينات للتحليل في جامعة شولالونغكورن في بانكوك، تايلاند.

كما تم حساب كمية امتصاص المعدن (Metal uptake) اعتمادًا على تراكيز العناصر في النبات والكتلة الحيوية الجافة للنبات (Alaboudi et al., ٢٠١٨)، وباستخدام المعادلة التالية (Chaney et al., ١٩٩٧).

امتصاص المعدن بواسطة النبات: Metal Uptake
× Biomass Metal Uptake = C plant
(mg/kg) تركيز العنصر في النبات = C plant
Biomass = الكتلة الحيوية الجافة للنبات.

٦. *التحليل الإحصائي:*

حُلِّلت البيانات باستخدام برنامج GenStat–١٢ الإحصائي وفق تصميم القطاعات الكاملة العشوائية (RCBD)، وأُجري تحليل التباين الثنائي (Two–Way ANOVA) مع الموقع والصنف عاملَين مستقلَّين، وفُرِّق بين المتوسطات باستخدام اختبار الفروق المعنوية الأقل (LSD) عند مستوى ٠,٠١ كما حُسبت معاملات ارتباط بيرسون (r) ومعادلات الانحدار الخطي البسيط ومعاملات التحديد (R^2) لتحديد العلاقة الكمية بين صفات النمو الفيزيولوجية (طول الساق ومساحة الورقة) وكميات امتصاص المعادن الثقيلة.

النتائج

أولاً: صفتا طول الساق ومساحة الورقة:

الجدول (2): صفة طول الساق تبعاً للأصناف والمواقع.

المواقع / الأصناف	دارة عزة	حارم	سرمدا
الأمريكي	35.7333[a]	42.2333[a]	39.7333[a]
الإسباني	41.1333[a]	46.7667[a]	32.6333[a]
المحلي	—	46.4000[a]	42.7000[a]
Grand mean = 40.917	$L s d_{0.01}$ = 22.1335		C.V % = 22.7

الجدول (3): صفة مساحة الورقة تبعاً للأصناف والمواقع.

المواقع / الأصناف	دارة عزة	حارم	سرمدا
الأمريكي	795.0000[a]	1412.6667[a]	1212.6667[a]
الإسباني	684.0000[a]	1103.0000[a]	287.3333[a]
المحلي	—	1799.3333[a]	789.0000[a]
Grand mean = 1010.4	$L s d_{0.01}$ = 1073.2330		C.V % = 44.5

ثانياً: كمية امتصاص المعدن من قبل النبات: (MU)

الجدول (4): كمية الامتصاص لعنصر الكروم تبعاً للأصناف والمواقع.

المواقع / الأصناف	دارة عزة	حارم	سرمدا
الأمريكي	18.68	40.55	16.76
الإسباني	27.38	65.19	7.18
المحلي	—	73.99	20.80

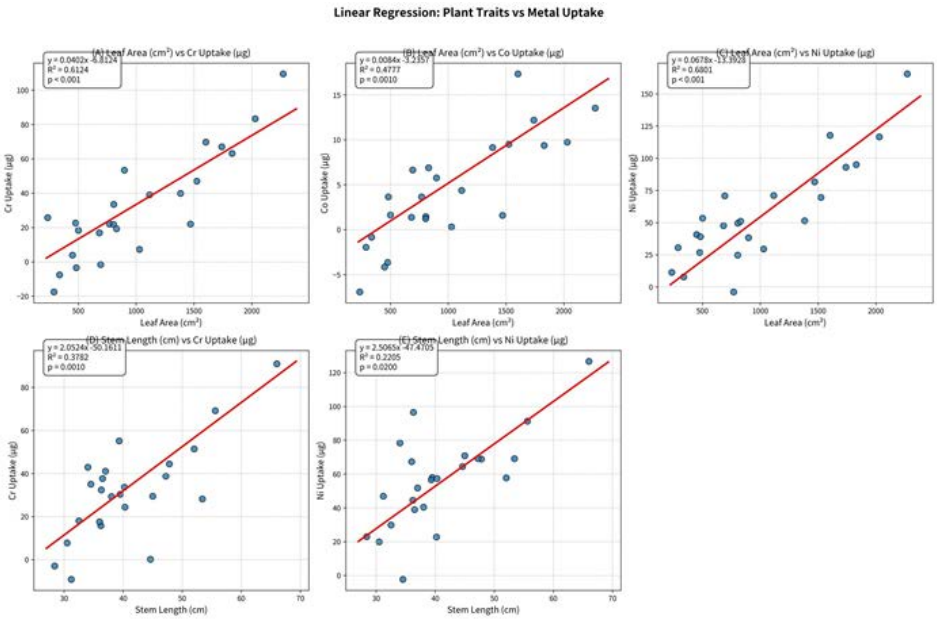
الجدول (5): كمية الامتصاص لعنصر الكوبالت تبعاً للأصناف والمواقع.			
المواقع / الأصناف	دارة عزة	حارم	سرمدا
الأمريكي	2.9995	1.883	2.647
الإسباني	2.13	6.91	1.69
المحلي	—	19.73	1.629

الجدول (6): كمية الامتصاص لعنصر النيكل تبعاً للأصناف والمواقع.			
المواقع / الأصناف	دارة عزة	حارم	سرمدا
الأمريكي	12.21	74.34	59.27
الإسباني	37.36	77.00	11.45
المحلي	—	145.77	23.29

ثالثاً: معامل ارتباط بيرسون والانحدار الخطي البسيط ومعامل التحديد

الجدول (7): معامل ارتباط بيرسون:(r)			
المتغير	الكروم	الكوبالت	النيكل
طول الساق	r= **0.615	r= 0.283 (NS)	r= *0.470
مساحة الورقة	r= ***0.783	r= ***0.691	r= ***0.825

الجدول (8): نماذج الانحدار الخطي				
المتغير	العنصر	معادلة الانحدار	R^2	P
مساحة الورقة	Cr	Cr_{uptake} (µg) = 0.0402 × LA (cm^2) – 6.812	0.612	0.001>
مساحة الورقة	Co	Co_{uptake} (µg) = 0.0084 × LA (cm^2) – 3.236	0.478	0.001>
مساحة الورقة	Ni	Ni_{uptake} (µg) = 0.0678 × LA (cm^2) – 13.393	0.680	0.001>
طول الساق	Cr	Cr_{uptake} (µg) = 2.052 × H (cm) – 50.161	0.378	0.01>
طول الساق	Ni	Ni_{uptake} (µg) = 2.507 × H (cm) – 47.471	0.221	0.05>



الشكل (١): مخططات التشتت

CC BY

المناقشة

أولاً: صفتا طول الساق ومساحة الورقة:

١.طول الساق: يُبين الجدول (٢) متوسطات طول الساق لأصناف دوار الشمس الثلاثة في المواقع المدروسة. بلغ المتوسط العام ٤٠,٩٢ cm بمعامل اختلاف CV=٢٢,٧٪، وبلغت قيمة LSD٠,٠١=٢٢,١٣ cm. ويُلاحظ أن هذه القيمة الكبيرة نسبياً لـ LSD تُشير إلى عدم وجود فروق معنوية دالة إحصائياً بين معظم المعاملات المدروسة في هذه الصفة، إذ تراوحت الأطوال بين ٣٢,٦٣ cm (الإسباني/سرمدا) و٤٦,٧٧ cm (الإسباني/حارم) دون أن يتجاوز الفرق بينهما حد الدلالة الإحصائية في أغلب المقارنات. ويُعزى هذا التباين المحدود إلى التجانس النسبي في ظروف الزراعة بين المواقع من حيث موعد الزراعة والري والمدة الزراعية الثابتة (٤٥ يوماً).

٢.مساحة الورقة: أما مساحة الورقة، الجدول (٣)، فقد سجّلت تبايناً واسعاً بين المعاملات إذ تراوحت بين ٢٨٧,٣٣ cm² (الإسباني/سرمدا) و١٧٩٩,٣٣ cm² (المحلي/حارم)، بمتوسط عام ١٠١٠,٤ cm² ومعامل اختلاف CV=٤٤,٥٪، وبلغت LSD٠,٠١=١٠٧٣,٢٣ cm². وبالنظر إلى قيمة LSD المرتفعة مقارنةً بالفروق المطلقة بين المعاملات، تبيّن أن الفروق لم تبلغ مستوى الدلالة الإحصائية في أغلب المقارنات، وإن أشارت القيم المطلقة إلى نمط واضح يتمثل في تفوّق موقع حارم في جميع الأصناف لا سيما الصنف المحلي. ويُعزى ارتفاع CV٪ في هذه الصفة إلى التباين الكبير في الوزن الجاف بين المكررات داخل المعاملة الواحدة، وهو نمط متوقع في تجارب التربة الملوثة حيث يختلف استجابة الأفراد النباتية للإجهاد المعدني (Farid et al., ٢٠١٨)

ثانياً: كمية امتصاص المعدن من قبل النبات: (MU)

١.عنصر الكروم: تراوحت كمية الامتصاص لعنصر الكروم بين (٧,١٨–٧٣,٩٩ µg/plant)، يُلاحظ أن جميع الأصناف في موقع حارم حققت أكبر كميات امتصاص وعلى وجه الخصوص الصنف المحلي الذي بلغت كميته ٧٣,٩٩ µg،أما في موقع سرمدا فقد تفوّق الصنف المحلي بكمية امتصاص بلغت ٢٠,٨٠ µg، في حين سجّل الصنف الإسباني في دارة عزة أكبر كمية في موقعه إذ بلغت ٢٧,٣٨ µg، بينما لم ينبت الصنف المحلي في هذا الموقع.

٢.عنصر الكوبالت: تراوحت كمية الامتصاص لعنصر الكوبالت بين (١,٦٢٩–١٩,٧٣ µg/plant)، الجدول (٥) تفوّق الصنف المحلي في موقع حارم تفوقاً لافتاً إذ بلغت كميته ١٩,٧٣ µg، يليه الصنف الإسباني في الموقع ذاته بـ ٦,٩١ µg أما الصنف الأمريكي في حارم فقد سجّل كمية امتصاص أقل ١,٨٨٣ µg مقارنةً بالصنف الأمريكي في دارة عزة ٢,٩٩٩٥ µg، بينما لم ينبت الصنف المحلي في موقع دارة عزة.

٣.عنصر النيكل: تراوحت كمية الامتصاص لعنصر النيكل بين (١١,٤٥–١٤٥,٧٧ µg/plant)، الجدول (٦) سجّل الصنف المحلي في موقع حارم أعلى قيمة في التجربة كلها إذ بلغت ١٤٥,٧٧ µg، مما يجعله الصنف الأكثر امتصاصاً للنيكل. وفي موقع سرمدا تفوّق الصنف الأمريكي بكمية ٥٩,٢٧ µg، في حين حقق الصنف الإسباني في دارة عزة أكبر كمية في موقعه بـ ٣٧,٣٦ µg بينما لم ينبت الصنف المحلي في موقع دارة عزة.

ثالثاً: معامل ارتباط بيرسون والانحدار الخطي البسيط ومعامل التحديد:

يوضح الجدول (٧) أن مساحة الورقة ارتبطت ارتباطاً موجباً ومعنوياً مع امتصاص العناصر الثلاثة، بينما كان ارتباط طول الساق معنوياً للكروم والنيكل فقط. كما يبين الجدول (٨) أن نماذج الانحدار المعتمدة على مساحة الورقة حققت أعلى قيم R^2 لدراسة طبيعة العلاقة بين صفات النمو الفيزيولوجية وكمية امتصاص المعادن الثقيلة، حُسب معامل ارتباط بيرسون (r) لصفتَي طول الساق ومساحة الورقة مع كميات امتصاص الكروم والكوبالت والنيكل عند مستوى دلالة ٠,٠١ أظهرت مساحة الورقة ارتباطات طردية دالة إحصائياً مع كميات امتصاص العناصر الثلاثة جميعها، إذ بلغت معاملات الارتباط (r=٠,٧٨٣*** ,p<٠,٠٠١) للكروم، و(r=٠,٦٩١*** ,p<٠,٠٠١) للكوبالت، و(r=٠,٨٢٥*** ,p<٠,٠٠١) للنيكل. ويعكس هذا الارتباط القوي العلاقةَ الفسيولوجية بين حجم السطح الهوائي الضوئي وكفاءة النقل الوعائي للعناصر؛ إذ تستلزم الورقة الأوسع تدفقاً مائياً أعلى عبر نظام الـ xylem يحمل معه كميات أكبر من أيونات المعادن الثقيلة من الجذور نحو الأجزاء الهوائية. أما طول الساق فقد أظهر ارتباطاً دالاً مع كمية امتصاص الكروم (r=٠,٦١٥** ,p<٠,٠١) والنيكل (r=٠,٤٧٠* ,p<٠,٠٥)، في حين لم يبلغ ارتباطه مع الكوبالت مستوى الدلالة الإحصائية (r=٠,٢٨٣ ,p=٠,١٨١)، وهو ما يُشير إلى أن طول الساق يرتبط بالامتصاص ارتباطاً غير مباشر، كونه يعكس حجم النبات الكلي لا آلية محددة لنقل عنصر بعينه. وقد جاءت جميع معاملات الارتباط الدالة موجبة الاتجاه، مما يؤكد أن تطور النمو الخضري لنبات دوار الشمس يُفضي إلى زيادة متوازية في كميات امتصاص المعادن الثقيلة، بمعزل عن تركيزها الكلي في التربة.

ثالثاً: الانحدار الخطي البسيط ومعامل التحديد (R^2)

بنيت نماذج تنبؤية للأزواج ذات الدلالة الإحصائية، وقد أعطت مساحة الورقة أفضل نماذج الانحدار للعناصر الثلاثة:

$$Cr_{uptake}\ (\mu g) = 0.0402 \times LA\ (cm^2) - 6.812$$
$$(R^2=0.612,\ p<0.001)$$

$$Co_{uptake}\ (\mu g) = 0.0084 \times LA\ (cm^2) - 3.236$$
$$(R^2=0.478,\ p<0.001)$$

$$Ni_{uptake}\ (\mu g) = 0.0678 \times LA\ (cm^2) - 13.393$$
$$(R^2=0.680,\ p<0.001)$$

تُشير معاملات التحديد إلى أن مساحة الورقة وحدها تُفسّر ٦١,٢٪ من تباين كمية امتصاص الكروم، و٤٧,٨٪ للكوبالت، و٦٨,٠٪ للنيكل. ويتصدر النيكل في هذه النماذج كونه الأكثر حركية داخل أنسجة النبات (Deng et al., ٢٠٢١)، مما يجعل حجم المجموع الخضري مؤشراً تفسيرياً أقوى لامتصاصه مقارنةً بالكروم الأكثر احتجازاً في الجذور. كما أعطى طول الساق نموذجاً دالاً للكروم والنيكل:

$$Cr_{uptake}\ (\mu g) = 2.052 \times H\ (cm) - 50.161\ (R^2=0.378,\ p<0.01)$$

$$Ni_{uptake}\ (\mu g) = 2.507 \times H\ (cm) - 47.471\ (R^2=0.221,\ p<0.05)$$

غير أن قيم R^2 لنماذج الطول أدنى بشكل ملحوظ من نظيراتها لمساحة الورقة، مما يُرجّح أن مساحة الورقة تُمثل المؤشر النباتي الأكثر ارتباطاً بكفاءة امتصاص المعادن الثقيلة في ظروف هذه الدراسة. وتُشير هذه النتائج مجتمعةً إلى أن صفات النمو الخضري — ولا سيما مساحة الورقة — نستطيع اعتبارها أداة تقدير أولية لكفاءة المعالجة النباتية، مما يفتح الباب أمام اختيار الأصناف النباتية بناءً على معاملات نموها قبل الزراعة في الترب الملوثة.

رابعاً: مخططات التشتت:

يظهر الشكل (١) تحليل الانحدار الخطي بين صفات النمو النباتية وكمية امتصاص المعادن الثقيلة (الكروم، الكوبالت، والنيكل) في أصناف دوار الشمس المزروعة في مواقع مختلفة.
١.العلاقة بين مساحة الورقة وامتصاص المعادن (A–C)
٢. العلاقة بين طول الساق وامتصاص المعادن. (D–E)
٣.كل نقطة تمثل مكرر فردي (n=٢٤) والخط الأحمر يمثل خط الانحدار الخطي.
٤. معادلات الانحدار، ومعامل التحديد (R^2)،

وقيم الدلالة الإحصائية موضحة داخل كل لوحة.

الاستنتاجات

١.أظهرت مساحة الورقة ارتباطاً طردياً دالاً إحصائياً مع كميات امتصاص العناصر الثقيلة الثلاثة جميعها (الكروم والكوبالت والنيكل)، مما يُقوي موقعها بوصفها أكثر صفة نباتية مرتبطة بكفاءة امتصاص المعادن الثقيلة في ظروف هذه الدراسة.

٢.فسّرت نماذج الانحدار المبنية على مساحة الورقة ما بين ٤٧,٨٪ و٦٨,٠٪ من تباين كميات الامتصاص (R^2=٠,٤٧٨– ٠,٦٨٠)، وهي نسب تفسير مرتفعة تُثبت إمكانية التنبؤ بكمية الامتصاص من صفة نباتية واحدة سهلة القياس.

٣.ارتبط طول الساق ارتباطاً دالاً مع امتصاص الكروم (r=٠,٦١٥ ,p<٠,٠١) والنيكل (r=٠,٤٧٠ ,p<٠,٠٥) فقط، دون دلالة معنوية مع الكوبالت، مما يُشير إلى أن هذه الصفة تعكس حجم النبات الكلي دون ارتباطها بآلية نقل عنصر بعينه.

٤.تصدّر النيكل نماذج الانحدار من حيث قوة العلاقة مع مساحة الورقة (R^2=٠,٦٨٠)، وهو ما يتوافق مع طبيعته الكيميائية الأكثر حركية داخل الأنسجة الوعائية مقارنةً بالكروم.

٥.تفوّق موقع حارم في كميات الامتصاص لجميع العناصر الثلاثة رغم امتلاكه أدنى التراكيز الكلية، مما يُؤكد أن الإتاحة الحيوية للعناصر في التربة – لا تركيزها الكلي – هي المحدد الفعلي لكمية الامتصاص.

٦.أعطى الصنف المحلي في موقع حارم أعلى كميات امتصاص للعناصر الثلاثة، ولا سيما للنيكل (١٤٥,٧٧ µg/ plant) والكروم (٧٣,٩٩ µg/plant)، مما يُشير إلى ملاءمته النسبية كصنف واعد في برامج المعالجة النباتية لترب المنطقة.

٧.تُثبت نتائج هذه الدراسة مجتمعةً أن صفات النمو الفيزيولوجية – وخاصة مساحة الورقة – يمكن اعتبارها مقياس مساعداً (Proxy indicators) للتنبؤ المبكر بكفاءة المعالجة النباتية، مما يفتح آفاقاً واسعة أمام برامج انتخاب الأصناف النباتية وتقييم كفاءتها بأساليب ميدانية بسيطة وغير مكلفة.

التوصيات:

١.اعتماد مساحة الورقة معياراً انتخابياً مبكراً في برامج تحسين أصناف دوار الشمس الموجهة للمعالجة النباتية، نظراً لقوة ارتباطها بكميات امتصاص المعادن الثقيلة الثلاثة المدروسة.

٢.إيلاء الصنف المحلي (البلدي) اهتماماً بحثياً خاصاً في الدراسات اللاحقة، كونه أظهر كميات امتصاص مرتفعة في أكثر من موقع، مع توصيف جيني ووراثي أعمق لفهم أسس هذا التفوق.

٣.إجراء دراسات مماثلة لأصناف أخرى من دوار الشمس والنباتات ذات الكتلة الحيوية المرتفعة، مع قياس صفات نمو إضافية كمساحة الجذر والوزن الجاف، في ترب مختلفة في درجة التلوث لبناء نماذج تنبؤية أكثر شمولاً وقدرةً تفسيرية.

٤.استخدام نماذج الانحدار الخطي المُستخرجة في هذه الدراسة كأداة تقييم ميدانية أولية في مشاريع المعالجة النباتية بشمال غرب سوريا، مع التحقق من صلاحيتها في مواقع وترب مختلفة قبل تعميمها.

٥.دراسة تأثير عوامل التحسين الكيميائي للتربة كـ (EDTAوحمض الستريك (على العلاقة بين مساحة الورقة وكمية الامتصاص، للتحقق من إمكانية تعزيز هذه العلاقة عبر رفع الإتاحة الحيوية للعناصر.

٦.تنفيذ تجارب متعددة المواسم لدراسة تطور العلاقة بين صفات النمو وكمية الامتصاص عبر الزمن، مما يُتيح تقدير الحاجة إلى عدد المواسم الزراعية اللازمة لتحقيق إزالة بيئية فعّالة.

المراجع

[1] Alaboudi, K. A., Ahmed, B., & Brodie, G. (2018). Phytoremediation of Pb and Cd contaminated soils by using sunflower (*Helianthus annuus*) plant. *Annals of Agricultural Sciences, 63*(1), 123-127. https://doi.org/10.1016/j.aoas.2018.05.007

[2]Alhasan, M., Lakmes, A., Alobaidy, M. G., AlHaeek, S., Assaf, M., Dawson, L., Pirrie, D., Abdeldayem, Z., & Bridge, J. (2023). A baseline survey of potentially toxic elements in the soil of north-west Syria following a decade of conflict. Environmental Science: Advances, 2(6), 886–897. https://doi.org/10.1039/D2VA00333C

[3]1. Chaney, R. L., Malik, M., Li, Y. M., Brown, S. L., Brewer, E. P., Angle, J. S., & Baker, A. J. M. (1997). Phytoremediation of soil metals. Current Opinion in Biotechnology, 8(3), 279–284. https://doi.org/10.1016/S0958-1669(97)80004-3

[4] Deng, T. H. B., van der Ent, A., Tang, Y. T., Sterckeman, T., Echevarria, G., Morel, J. L., & Qiu, R. L. (2021). Quantification of nickel and cobalt mobility and accumulation via the phloem in the hyperaccumulator Noccaea caerulescens.Metallomics,13(4),mfab012. https://doi.org/10.1093/mtomcs/mfab012

[5] Bhatti, M. S. H. (2018). Morphological, physiological and biochemical responses of different plant species to Cd stress. International Journal of Environmental Research and Public Health, 13(6), 1320. https://doi.org/10.3390/ijerph13060596

[6] Huang, Y., Zhao, L., Chen, X., Xu, X., & Li, T. (2022). Systematic evaluation of plant metals/metalloids accumulation efficiency: A global synthesis of bioaccumulation and translocation factors. Frontiers in Plant Science, 13, 906463. https://doi.org/10.3389/fpls.2022.906463

[7]ISO.(2002).Water quality — Dissolution for the determination of selected elements in water (ISO Standard No. 15587-1). International Organization for Standardization.

[8] Nehnevajova, E., Herzig, R., Federer, G., Erismann, K. H., & Schwitzguébel, J. P. (2007). Chemical mutagenesis — a promising technique to increase metal concentration and extraction in sunflowers. International Journal of Phytoremediation, 9(2), 149–165. https://doi.org/10.1080/15226510701232880

[9] Page, V., & Feller, U. (2015). Heavy Metals in Crop Plants: Transport and Redistribution Processes on the Whole Plant Level. Agronomy, 5(3), 447–463. https://doi.org/10.3390/agronomy5030447

[10]Stoikou, V., Andrianos, V., Stasinos, S., Kostakis, M. G., Attiti, S., Thomaidis, N. S., & Zabetakis, I. (2017). Metal uptake by sunflower (Helianthus annuus) irrigated with water polluted with chromium and nickel. Foods, 6(7), 51. https://doi.org/10.3390/foods6070051

[11] Zhao, H., Xia, B., Fan, C., Zhao, P., & Shen, S. (2023). Modelling heavy-metal phytoextraction capacities of Helianthus annuus L. and Brassica napus L. Chemosphere, 338, 139500. https://doi.org/10.1016/j.chemosphere.2023.139500

[12]Wuana, R. A., & Okieimen, F. E. (2011). Heavy metals in contaminated soils: A review of sources, chemistry, risks and best available strategies for remediation. ISRN Ecology, 2011, 402647.https://doi.org/10.5402/2011/402647

مساهمات المؤلفين:أ.د. عبد الحميد الخالد

تضارب المصالح:يعلن المؤلفون أنه ليس لديهم أي مصالح متضاربة.

توافر البيانات والمواد:جميع البيانات متوفرة في النص الرئيس عدا الأشكال والمراجع.

Developing a Hybrid Algorithm that Combine Rule-based and Machine Learning Methods to Detect Social Engineering Attacks

تطوير خوارزمية نظام هجين يجمع بين الأساليب القائمة على القواعد والتعلُّم الآلي لكشف هجمات الهندسة الاجتماعية

دانا نجيب لجامعة الافتراضية السورية

باسل الخطيب : وزارة التعليم العالي، الجامعة الافتراضية السورية، جامعة دمشق_سوريا.

عنوان التواصل: dana_najeeb_300318@svuonline.org – t_balkhatib@svuonline.org

تاريخ التسليم: ٢٠ تموز ٢٠٢٦ – **تاريخ القبول:** ١٥ تشرين الاول ٢٠٢٦

ABSTRACT

Social engineering is a science that utilizes social interaction to persuade an individual or organization to comply with an attacker's request via a computer-related entity. Recently, social engineering attacks have grown rapidly, exploiting human naivety to gain unauthorized access to systems and information. Many models have been developed to address this problem. The hybrid algorithm proposed in this study is based on automating most cases of the optimized model for manual detection of social engineering attacks, while applying a dynamic blacklist algorithm to verify each line of text generated by a potential attacker, matching with dynamic topic blacklist, and employing natural language processing and machine learning, the proposed hybrid system successfully automated the detection of social engineering attacks. This achieved substantial operational efficiency, demonstrating the efficacy of the proposed algorithm as a defensive tool. The dataset used consists of 50,000 emails, collected from the public repository (Kaggle), balanced samples were integrated into a unified test file to feed the proposed hybrid system and measure its security response. The accuracy reached 0.9740 despite ignoring other features, which is considered excellent.

Keywords: Cybersecurity, Social Engineering Attacks, Attack Detection Techniques, Machine Learning, Artificial Intelligence Technologies.

الملخص

الهندسة الاجتماعية علم يعتمد التفاعل الاجتماعي وسيلةً لإقناع فرد أو منظمة بالامتثال لطلب مهاجم عبر كيان مرتبط بالحاسوب. برزت هجمات الهندسة الاجتماعية مؤخراً كتهديد متزايد للأمن السيبراني. استغل المهاجمون نقاط الضعف البشرية للوصول غير المصرح به للأنظمة والمعلومات الحساسة. تم تطوير نماذج متعددة لمعالجة هذه المشكلة. خوارزمية النظام الهجين المقترحة في هذه الدراسة قائمة على أتمتة معظم حالات النموذج المُحسَّن للكشف اليدوي عن هجمات الهندسة الاجتماعية، وتطبيق خوارزمية القوائم السوداء الديناميكية بالمطابقة مع نص أنشأه مهاجم محتمل، وباستخدام معالجة اللغة الطبيعية والتعلم الآلي، تمكن النظام الهجين المقترح من اكتشاف هجمات الهندسة الاجتماعية آليا، وحقق كفاءة تشغيلية ملموسة أثبتت كفاءة الخوارزمية المقترحة كأداة دفاعية لكشف هجمات الهندسة الاجتماعية. مجموعة البيانات المستخدمة (50.000) رسالة بريد الكتروني. جُمعت من مستودعات (Kaggle) العامة. دمجت العينات المتوازنة بملف اختبار موحد مغذي للنظام الهجين المقترح لقياس استجابته الأمنية. بلغت الدقة 0.9740، وتُعتبر جيدة، بسبب الاكتفاء بالمقارنة النصية لنص البريد في هذا الاختبار.

الكلمات المفتاحية: الأمن السيبراني، هجمات الهندسة الاجتماعية، تقنيات الكشف عن الهجمات، التعلُّم الآلي، تقنيات الذكاء الاصطناعي.

CC BY

المقدمة

شهد القرن الحادي والعشرين اندماجاً كاملاً لتكنولوجيا المعلومات (IT) في بنية المجتمعات المعاصرة، وركيزة أساسية لإدارة القطاعات الحيوية الحكومية والمالية والصحة والتعليم والاتصالات وغيرها. هذا الاعتماد الرقمي الشامل فتح افاقاً جديدة للتهديدات السيبرانية، وحول التكنولوجيا نفسها الى أداة لإلحاق الضرر بالمؤسسات والأفراد. برز أمن المعلومات (IS) كخدمة استراتيجية لحماية البيانات أثناء حفظها واستخدامها ونقلها من التهديدات الداخلية والخارجية. ورغم التطور الهائل للتقنيات الدفاعية للبرمجيات وأجهزة مكافحة الثغرات الرقمية، الا أن المهاجمين وجدوا في العنصر البشري منفذاً أكثر فاعلية عبر ما يعرف بالهندسة الاجتماعية (SE) التي تعتمد استغلال الخطأ البشري، والتلاعب النفسي والثقة لخداع المستخدمين واستدراجهم لكشف بيانات حساسة (كلمات مرور، حسابات بنكية) وتسهيل اختراق الأنظمة المحظورة. تؤكد الإحصاءات خطورة هذا التهديد، وتشير التقارير الأمنية لعام 2026 الى دور العنصر البشري في أغلب خروقات البيانات العالمية، وتُستخدم تقنيات «القرصنة البشرية» بشكل شائع في الجرائم الإلكترونية لخداع المستخدمين لكشف معلومات محمية، أو نشر إصابات ببرامج ضارة، أو منح الوصول إلى أنظمة محظورة. هذه التقنيات تستغل عدم وعي الانسان ليؤدي عملاً يفتح ثغرة أمنية أو يفضي بمعلومات سرية. وقد تستخدم في عمليات احتيال عبر الإنترنت أو على أرض الواقع، وقد تمكّن المهندس الاجتماعي (المهاجم) من اختراق المباني أو الأنظمة أو الحسابات لتحقيق منافع معنوية أو مادية أو أمنية، دون الحاجة إلى مهارات تقنية عالية كما هي الحال في «الهاكرز» و«الكراكرز» الذين يتميزون بمهارات برمجية ومعرفة حاسوبية واسعة بالعتاد وأنظمة التشغيل والشبكات وثغراتها الأمنية. ما يحتاجه المهاجم هنا فقط مهارات اجتماعية ومهارات تسويق، لإقناع الضحية وأخذ ما يريده دون إثارة شبهات. نعرّف الهندسة الاجتماعية (SE) أنها تقنية يستخدمها المهاجمون للتلاعب بالأفراد لإفشاء معلومات حساسة أو القيام بأعمال يمكن أن تعرض أمن المؤسسة للخطر، ويعد الاحتيال الالكتروني (Phishing) التطبيق التقني الأكثر انتشاراً لها. تتزايد هجمات الهندسة الاجتماعية (SE) بسبب نقص المعلومات حول الموضوع [1]. هناك حاجة للمزيد من البحث في هذا القطاع لرفع مستوى الوعي بشأن هجمات الهندسة الاجتماعية (SE). أصبحت هجمات الهندسة الاجتماعية (SE) معقدة بشكل يصعب اكتشافه. يتم تطوير نماذج قائمة على القواعد (RBS) والتعلم الآلي (ML) والشبكات العصبونية الاصطناعية (ANN) ومعالجة اللغة الطبيعية (NLP) لمعالجة هذه المشكلة باستمرار، لكنها تغفل دور المستخدم. تهدف هذه المقالة البحثية الى تطوير خوارزمية نظام هجين يجمع بين الأساليب القائمة على القواعد (الذكاء الاصطناعي الرمزي) والتعلم الآلي والطريقة اليدوية لتمكين المستخدم، وكشف هجمات الهندسة الاجتماعية (SE). الشق الأول هو تطوير نموذج للكشف عن هجمات الهندسة الاجتماعية بالطريقة اليدوية الأقل تكلفة، معتمد على الإنسان، يمكن استخدامه لتعليم الناس كيفية حماية أنفسهم من بعض أنواع الهجمات. الشق الثاني الكشف بالطريقة الآلية الأكثر تكلفةً وتعقيداً، ومعتمد على النظام. وفي هذا السياق، يقترح البحث أسلوبًا يستخدم خوارزمية القوائم السوداء الديناميكية للموضوعات (Topic Blacklist–TBL) للتحقق من موضوعات النقاش لكل سطر من نص أنشأه مهاجم محتمل. وباستخدام تقنيات معالجة اللغة الطبيعية والتعلم الآلي، يمكن للنظام اكتشاف هجمات الهندسة الاجتماعية (SE) ألياً. النموذج الهجين المقترح مبني على أساس أتمتة أجزاء من خوارزمية الكشف اليدوي، للوصول الى تكامل معرفي يدمج اليات الكشف الالي ضمن حالات خوارزمية الكشف اليدوي.

الجدول1 قائمة الاختصارات

IT	Information Technology	تقنية المعلومات
IS	Information Security	أمن المعلومات
SE	Social Engineering	الهندسة الاجتماعية
ML	machine learning	التعلم الالي
AI	Artificial intelligence	الذكاء الاصطناعي
DL	Deep Learning	التعلم العميق
NLP	Natural Language Processing	معالجة اللغة الطبيعية
ANN	Artificial Neurons Networks	اشبكة العصبونية الاصطناعية
SEADM	Social Engineering Attacks Detection Model	نموذج كشف هجمات الهندسة الاجتماعية
TBL	Topic BlackList	القائمة السوداء الديناميكية للموضوعات
SAI	Symbolic AI	الذكاء الاصطناعي الرمزي
RBS	Rule- based systems	الأنظمة المعتمدة على القواعد
URL	Uniform Resource Locator	محدد موقع الموارد المحدد (رابط ويب)

الدراسات المرجعية

أُجريت العديد من الدراسات البحثية لتعزيز فاعلية تقنيات الكشف عن هجمات الهندسة الاجتماعية (SE)، ومنعها. ومع ذلك، لم تتمكن هذه الدراسات من القضاء على عيوب هذه التقنيات بالكامل. تم تصنيف أكثر الطرق شيوعاً للكشف عن هجمات الهندسة الاجتماعية (SE) إلى ثلاثة أنواع: اجتماعية، وتقنية، وهجينة (Hybrid) [2]. تعتمد الطرق الاجتماعية للكشف اليدوي على الإنسان لكشف الهجوم [3،4،5]، أما الطرق التقنية الآلية فتعتمد على النظام لكشف الهجوم باستخدام التقنيات وأدوات الذكاء الاصطناعي (AI). جميع استراتيجيات الكشف المستخدمة حاليًا هي خيارات قابلة للتطبيق لاكتشاف هجمات الهندسة الاجتماعية (SE) أو منعها. لكنها تحوي نقاط ضعف تستدعي التحديث بانتظام لمواكبة التغيرات السريعة. أيضا معظمها تعتمد على بنية تحتية مكلفة بدلاً من حلول موجهة للمستخدم (هدف الهجوم)، لذلك حتى الشركات التكنولوجية الضخمة التي تنفق الكثير على الأمن عرضة لهجمات الهندسة الاجتماعية (SE). ان اغفال دمج وعي المستخدم لتعزيز الطرق التقنية المتاحة هو الفجوة المعرفية التي تهدف هذه الدراسة لمعالجتها كحجر أساس للنموذج المقترح.

CC BY



الشكل رقم 1 تصنيف تقنيات الكشف عن هجمات الهندسة الاجتماعية (SE)

الجدول٢ تصنيف التقنيات المستخدمة في الكشف عن هجمات الهندسة الاجتماعية (SE)

التقنية أو النهج	الحلول	الإيجابيات والسلبيات	المراجع
النهج الاجتماعي	تثقيف وتدريب المستخدمين أثناء وبعد التوظيف على الأمن السيبراني ونقاط ضعفه	يُعتبر النهج الاجتماعي تدخلاً مُركّزاً على الإنسان. تُركّز حلول الكشف عن الاحتيال الإلكتروني المعتمد على الهندسة الاجتماعية (SE) والوقاية منه على آليات الحماية من تلاعبات المستخدمين بدلاً من التلاعب المباشر بالنظام. يُتيح النهج الاجتماعي للمستخدمين سبل حماية أنفسهم والعواقب المترتبة على عدم حمايتهم. على الرغم من دقة وفعالية هذه التقنية الاجتماعية، لا يزال المتسللون يجدون طريقهم للتلاعب بالمستخدمين نظرًا للطبيعة المتطورة لهجمات الاحتيال الالكتروني والحالة النفسية للمستخدم وقت الهجوم	[4،5،3]
التعلم الآلي	استخدام تقنيات الذكاء الاصطناعي (AI) (التعلم الآلي، الشبكات العصبية الاصطناعية (ANN)، التعلم العميق، معالجة اللغة الطبيعية، الشبكات العصبية التلافيفية، الغابة العشوائية)	يُعد هذا النهج الأكثر شيوعًا في الكشف الفني عن الاحتيال الإلكتروني المعتمد على الهندسة الاجتماعية (SE)، لمرونته واعتماده خوارزميات تصنيف البيانات. لا يُعد حلاً مثالياً لمكافحة هذه الهجمات، لأنه يغفل العامل البشري. تتراوح دقة هذه الخوارزميات بين 92% و99.68%. فالخوارزميات لا تستطيع محاكاة سلوك الإنسان بدقة تامة	[6,7,8]
قوائم الحظر Blacklist Whitelist Blocklist Phishtank	قاعدة بيانات حالات الاحتيال الالكتروني المبلغ عنها والكلمات والمواضيع الدالة على الهجوم	تشتهر هذه المواقع الإلكترونية بدقة معلوماتها عن حالات الاحتيال الإلكتروني المبلغ عنها، الا ان المشكلة تكمن في تشتت البيانات التي توفرها هذه المواقع. فمثلاً، قد يصنف أحد المواقع موقعاً ما على أنه موقع احتيالي، بينما لا يحتوي موقع آخر أي معلومات عنه. مما يصعب على المستخدم اكتشاف هذه المواقع الاحتيالية.	[9]
هيكلية/ محتوى الموقع الإلكتروني	تطابق المحتوى الموجود في صفحات الويب مع محتوى الصفحات الأخرى	يعمل تحليل محتوى المواقع الإلكترونية وطريقة مقارنة التشابه بشكل جيد في مكافحة عمليات الاحتيال التي تهدف الى تقليد صفحات الإنترنت. إلا أن هذه الطريقة قد لا تتمكن من اكتشاف عمليات الاحتيال الالكتروني التي تستخدم محتوى ونظام تصميم مشابهين لمواقع الإنترنت الأصلية.	[10,11]
تشابه عناوين (URL)	مقارنة عناوين (URL) للصفحات الإلكترونية المشبوهة بعناوين الصفحات الإلكترونية الموثوقة	يعمل هذا النهج بشكل جيد في مكافحة عمليات الاحتيال الإلكتروني التي تستخدم عناوين مواقع ويب مزيفة عبر تقليد عناوين المواقع الأصلية. إلا أن هذا النهج لا يُمكنه الكشف عن عمليات الاحتيال الالكتروني التي تستخدم عناوين فريدة وغير مكررة.	[12,13]
سجل عمليات الوصول إلى الموقع الالكتروني	تطابق عناوين (URL) لصفحات الويب مع سجل الوصول الذي يوفره الخادم.	هذا النهج عملي لمستخدمين على نفس الخادم الذي تم إنشاء سجل الوصول إليه. وليكون النظام فعالاً، يجب أن يتضمن سجل الوصول بيانات من جميع خوادم الويب.	[14,8]
متعدد المستويات	جمع المؤلفون بين اثنين أو أكثر من تدخلات النظام وبين تدخل المستخدم.	يبدو أن هذا النهج هو الأنسب للكشف عن عمليات الاحتيال الإلكتروني المعتمد على الهندسة الاجتماعية (SE)، يجمع بين التدخلات التقنية وتدخلات المستخدمين. الا أن التحدي يكمن في تحديد أفضل الاستراتيجيات لتحقيق أفضل النتائج	[2,15,16]
النهج الهجين Hybrid) (Approach	جمع المؤلفون بين نهجين أو أكثر	يعتمد هذا النهج دمج خوارزميتين أو أكثر من خوارزميات التعلم الآلي، بالإضافة إلى منهجية للحد من مخاطر الاحتيال الالكتروني، بهدف تعزيز دقة تصنيف الالكتروني. وتقتصر تلك الدراسات السابقة في هذا النهج على عمليات الاحتيال حلول تقنية صرفة دون دمج آليات تفاعلية لحماية وتوعية المستخدمين النهائيين.	[2,15,17]

يوضح الجدول 2 الدراسات الحديثة مصنفةً حسب الاجتماعية (SE)، والايجابيات والسلبيات المتعلقة بها. التقنيات المستخدمة في الكشف عن هجمات الهندسة

المواد والطرائق

تتمثل منهجية النظام المقترح الهجين من خلال دمج تقنية النهج الاجتماعي بخوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) والتعلم الآلي لكشف هجمات الهندسة الاجتماعية (SE). بناءً على ذلك، تم تطوير خوارزمية النموذج المقترح المخصص للكشف اليدوي عن هجمات الهندسة الاجتماعية (SE) أولا، وبعد تحليل كافة العوامل السيكولوجية المتوفرة في الدراسات المرجعية. ولاختباره بشكل فعال، تم استخدام سيناريو بريد ذريعة (احتيال الكتروني) ثم استبيان موافق لأسئلة النموذج اليدوي التي تعتبر بوابات النموذج. في قسم آخر من الدراسة تم استعراض الإطار النظري لأنواع من هجمات الهندسة الاجتماعية (SE) والتقنيات المستندة للتعلم الآلي في كشفها (النهج التقني) والمتوفرة في الدراسات المرجعية. ثم تم الاستعانة ببرمجية مفتوحة المصدر على (GitHub) [18] وقاعدة بيانات متوازنة مكونة من (11.000) سطر من الروابط الخبيثة المتوفرة على (Data Mendeley) [19]، لدراسة وتحليل واختبار فاعلية التعلم الآلي [20] في اكتشاف أهم ميزات الروابط الاحتيالية (URL) المستخدمة في هجمات الاحتيال الالكتروني بالاعتماد على الذكاء الاصطناعي (AI). أيضا تمت دراسة فاعلية القوائم السوداء الديناميكية للموضوعات (TBL) لتعزيز أليات الفحص الذكي، كما تم تطوير أداة توعوية كنموذج أولي (prototype) لتوضيح فكرة القوائم السوداء الديناميكية للموضوعات (TBL) [21] ، وكيفية تفاعل المستخدم مع القوائم بالإضافة والحذف عبر تطبيق جافا سكريبت كواجهة تفاعلية حية. في النهاية تم استخدام مجموعة بيانات ضخمة متوازنة من نصوص البريد الالكتروني [24]، لتدريب نموذج التعلم الآلي. وأخيرا تم الحصول على نموذج هجين يجمع بين خوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) القابلة لتدخل المستخدم مع خوارزميات التعلم الآلي لكشف رسائل الاحتيال الالكتروني، وتجربة النموذج الهجين المقترح على تطبيق بايثون سحابي [22].

خوارزمية النموذج المقترح المخصص للكشف اليدوي عن هجمات الهندسة الاجتماعية (SE)

يهتم النموذج المقترح بتعديل الإصدار الأول والثاني لخوارزمية (SEADM)، ويستخدم من قبل الأفراد لتثقيف أنفسهم حول أساليب الخداع والهجوم. يعتمد النموذج المقترح على القواعد (RBS) والمحاكمة المنطقية لبعض المعايير كالاستعجال، الخوف والفضول، وإضافة بعض القواعد كالثقة والصلاحيات، المعاملة بالمثل، الالحاح، والتحيز في الإدراك، بهدف رفع وعي المستخدم لكشف هجمات الهندسة الاجتماعية (SE). يساعد النموذج المحسن المقترح الأفراد في طرح أسئلة تقييم ذاتية لحماية أنفسهم من أنواع مختلفة من هجمات الهندسة الاجتماعية (SE). يعتمد النموذج المقترح شجرة قرار متعددة المستويات لتكون عملية اتخاذ القرار أكثر مرونة وقابلية للإدارة. يعرض النموذج تدفق الإجراءات وكيف يجب على «المستلم» التعامل مع أشكال الطلب الواردة. تم تصميم حالات النموذج للوصول إلى إحدى الحالتين الرئيسيتين، إيقاف الطلب (SF) أو تنفيذ الطلب (SS). يعزى هذا الاختزال الاجرائي للمخرجات لغرض الدراسة واعتبار الرسائل المؤجلة مدخلات جديدة، تعود الى بداية شجرة القرار للتحقق، ليكون المخرج الحتمي النهائي ثنائي التصنيف ينحصر في الإيقاف التام او التنفيذ الامن.

–الحالة SF: أوقف الطلب: هذا هو شرط النتيجة السلبية (اكتشاف هجوم هندسة اجتماعية (SE)).

–الحالة SS: تنفيذ الطلب: هذه هي حالة النتيجة الإيجابية للنموذج (لا يعتبر هجوما).

–الحالة S1: فهم الطلب: تقوم هذه الحالة بتقييم ما إذا كان مستلم الطلب يفهم الطلب تمامًا.

–الحالة S2: طلب معلومات لفهم الطلب أكثر.

–الحالة S3: هل يحاول مقدم الطلب خلق شعور بالإلحاح؟

–الحالة S4: هل يمكن للمستلم أن يبطئ الطلب ويتأكد من شرعيته؟

–الحالة S5: هل يحث مقدم الطلب على فتح مرفق أو النقر فوق رابط ويب (URL)؟

–الحالة S6: هل يستطيع المستلم التحقق من المرفق أو رابط الويب (URL) قبل النقر عليه؟

–الحالة S7: هل الطلب وصل بشكل غير متوقع أو يجيب على سؤال لم يطرح ويطلب معلومات شخصية؟

–الحالة S8: هل يتمتع المستلم بسلطة تنفيذ الطلب؟

–الحالة S9: هل الإجراء أو المعلومات المطلوبة متاحة للعامة؟

–الحالة S10: هل يمكن التحقق من هوية مقدم الطلب؟

–الحالة S11: هل يتمتع مقدم الطلب بالسلطة اللازمة لطلب الإجراء أو المعلومات؟

تتكامل هذه الحالات (S1–S11) لتشكل معاً بناءً منهجياً ينقسم لأربع بوابات أمنية تتوزع على مستويات خدمة استراتيجية الدفاع العميق (Defense–in–Depth):

–المستوى الادراكي: تقييم وضوح الطلب الوارد ويشمل الحالات S1, S2)).

–المستوى النفسي السلوكي: رصد الضغوط السيكولوجية ويشمل الحالات S3, S4)).

–المستوى الفني (التقني): كشف المؤشرات التقنية كالروابط والمرفقات الخبيثة ويشمل الحالات S5, S6,S7)).

–المستوى الإداري التنظيمي: يفحص صلاحيات وهوية مقدم الطلب وشرعية تداول البيانات مؤسساتياً ويشمل الحالات S8,S9,S10,S11)).

يفيد التقسيم المنهجي في تتبع السلوك الأمني لتحديد المستوى الأمني الذي تم رصد هجوم هندسة اجتماعية (SE) فيه أو فشل اكتشافه. بناءً على تحليل البيانات وتوصيات خبراء الهندسة الاجتماعية (SE)، تم تطوير الإرشادات المنهجية التالية:

القاعدة 1: كن حذرًا إذا طلب مقدم الطلب معلومات حساسة –كمعلوماتك الشخصية أو بيانات غير عامة.

–القاعدة 2: كن متشككًا إذا حاول مقدم الطلب خلق شعور إلحاح، كالتهديد بتجميد حسابك إذا أهملت المعلومات.

–القاعدة 3: كن حذرًا إذا حثك مقدم الطلب على فتح مرفق أو النقر على رابط (URL) بدافع الخوف أو الفضول.

–القاعدة 4: كن حذرًا إذا كان مقدم الطلب يستخدم اسمًا أو عنوان بريد إلكتروني مزيف.

–القاعدة 5: كن حذرًا إذا هددك مقدم الطلب أو حاول تخويفك.

المجتمع والعينة وأدوات جمع البيانات

محاكاة الاحتيال الالكتروني (Phishing Simulation) هي تقنية تدريب لاختبار قدرة الناس وجاهزيتهم للتعرف على هجمات الاحتيال الالكتروني والإبلاغ عنها، وتعد وسيلة فعالة لتحسين الوعي الأمني وتقليل مخاطر هجمات الاحتيال الالكتروني. تم استخدام سيناريو بريد الذريعة كأحد أمثلة الهجمات الواقعية للاحتيال الالكتروني، والذي قد يمثل هجوم هندسة اجتماعية أو مجرد رسالة الكترونية حقيقية. في هذه الدراسة تتمثل الذريعة بدعوة الطلاب للمشاركة بندوة علمية ونشر أبحاث حول دراستهم وتحفزهم بأن المشاركة سترفع درجتهم النهائية. بعد موافقة مدير برنامج الدراسات العليا في علوم الويب/ الجامعة الافتراضية السورية، ودكتور مادة أمن المعلومات/ الجامعة الدولية الخاصة، تم ارسال النص التالي عبر بريد الكتروني صمم خصيصاً للهدف البحثي باسم مشابه لبريد الجامعة الافتراضية السورية الالكتروني (svu.symp.adm@gmail.com)، ليوحي

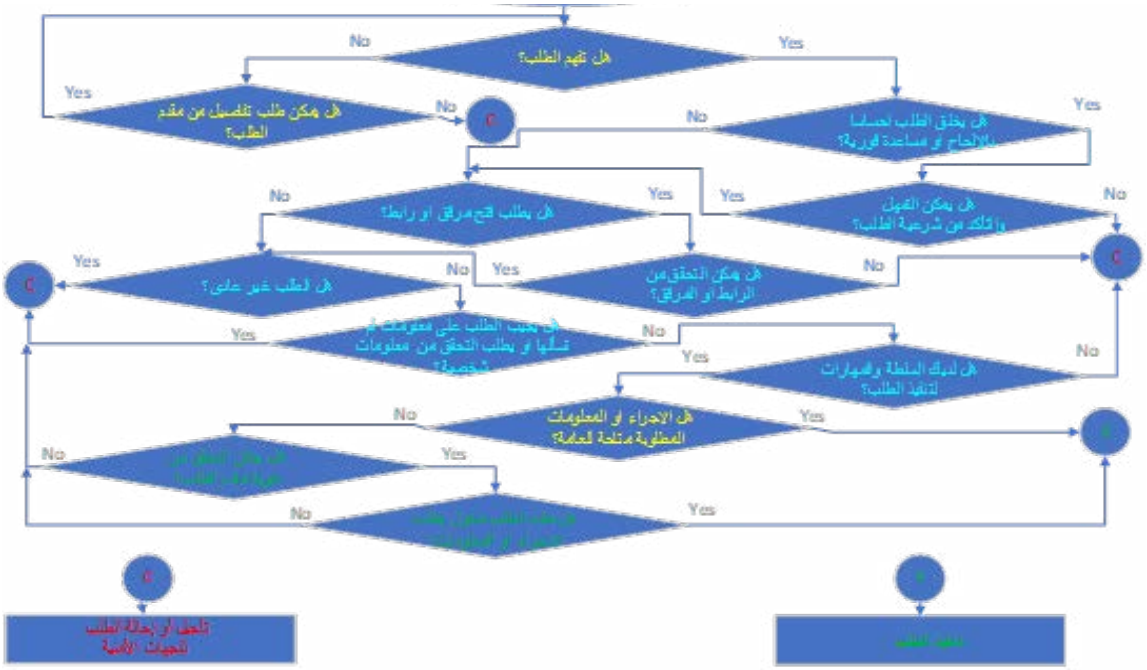
CC BY

the Use of Artificial Intelligence Technologies
- Cybersecurity
- Social Engineering
- Cloud Computing: Containers and Virtual Servers
- Sub-solutions for AI Technologies in Web Applications
- Data Storage, Sharing, and Retrieval and Big Data Processing on the Web
- Other topics.

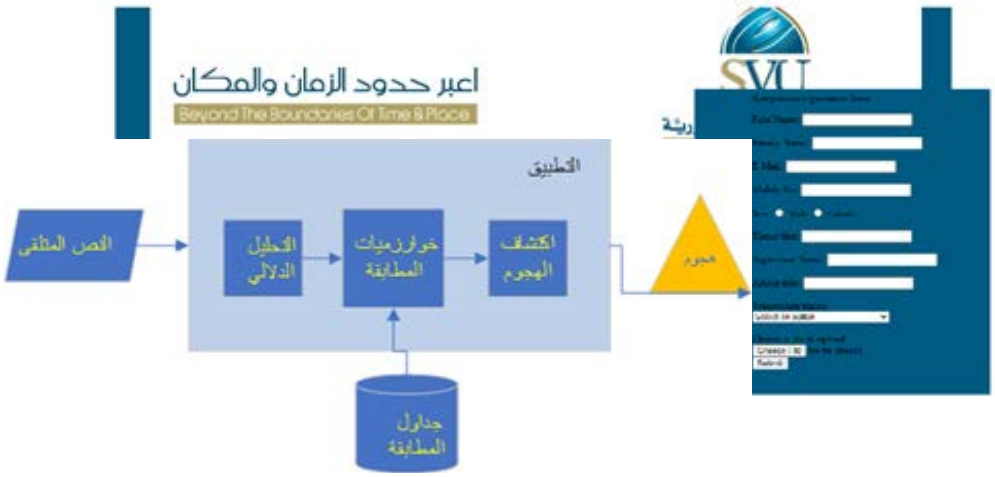
To register, please fill out the student's information and the research to be submitted via the following link: (www.svu-srf.kesug.com) username: admin password: 1234

للمستلم أنه من الجهة الناظمة للندوة العلمية. بالطبع الهدف إحصاء عدد الطلاب الذين استخدموا رابط الويب (URL) للولوج الى الموقع الوهمي البسيط والذي لا يحوي أي نوع من أنواع الهجمات اطلاقاً [23]، ولا يحتفظ بأي معلومة يتم إدخالها من قبل المشاركين. نص رسالة بريد الذريعة:

We invite you to participate in the Scientific Symposium on various topics in web science, scheduled to be held at the Syrian Virtual University on Thursday, December 18 2025 Participating research and articles should reflect summaries or the current portion of each student's research, based on the specific thesis topic. The symposium will cover the following topics:
- Web Application Development and



الشكل رقم 2: الخوارزمية المقترحة للكشف اليدوي عن هجمات الهندسة الاجتماعية (SE)



الشكل رقم 3: صفحة الرابط المرفق

خوارزميات القوائم السوداء الديناميكية للموضوعات (TBL–Top–ic–blacklists)

تنتمي هذه الخوارزميات الى النهج التقني لكشف الاحتيال الالكتروني المعتمد على الهندسة الاجتماعية (SE)، يستخدم النموذج الهجين المقترح الخوارزميات التقنية التالية للكشف عن الهجوم:

خوارزمية الكشف عن الهندسة الاجتماعية (SE) بالاعتماد على (TBL)

حلقة خارجية تمسح كل سطر في نص أرسله مهاجم محتمل (TEXT)، تتكرر الحلقة الداخلية خلال كل إدخال (n في TBL). يستدعي السطر 3 خوارزمية تحدد إذا كان موضوع الإدخال (n) موجودًا في السطر الأول. عند المطابقة يتم اكتشاف هجوم وعرض رسالة تحذير.

1. For each l Є TEXT.
2. For each n Є TBL.
3. If Match Topic (l,n).
4. ATTACK_DETECTED.

خوارزمية تطابق الموضوع بالاعتماد على (TBL)

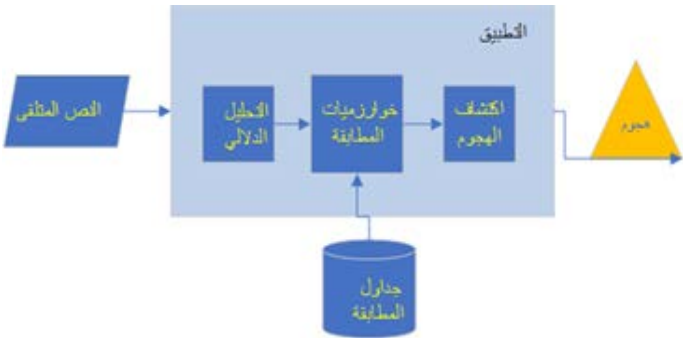
الغرض من الحلقة الخارجية (السطر 1) هو فحص كل رمز مميز في السطر ومقارنته بـ (n.action)، الفعل المرتبط بإدخال (n في TBL) . يتم إجراء المقارنة مع (n.action) في (السطر 2) . إذا تم العثور على الفعل، فسيتم الدخول للحلقة الداخلية (السطر 3) لمسح الرموز المميزة المتبقية على السطر ومقارنتها بـ (n.resource)، المورد الموافق. إذا تم العثور على كل من الفعل والمورد، فسترجع الدالة (TRUE) في (السطر 5). تقوم دالة المقارنة التي يتم استدعاؤها في (السطرين 2 و4) بإجراء مقارنة سلسلة، ولكنها تقارن أيضًا الوسيطة الأولى بأي مرادفات للوسيطة الثانية

1. For i = 0 to | token in l |
2. If Compare (l[i], n.action).
3. For j = I + 1 to |token in l|.
4. If Compare (l[j], n.resource).
5. Return TRUE.
6. Return FALSE.

يتفاعل المستخدم مع القوائم السوداء الديناميكية (TBL) المضمنة داخل نظامه بالإضافة والحذف، مما يزيد مناعة نظامه ويرفع وعيه ومسؤوليته الرقمية. القوائم المقترحة محلياً ضمن (TBL) هي جداول من أفعال ومصادر تنتمي دلالياً وسياقيا الى عوامل الخوارزمية الهجينة المقترحة بعد الاستفادة من خوارزمية (SEADM) الإصدار الأول والثاني، واضافة عوامل أخرى تؤشر الى هجمات بالهندسة الاجتماعية (SE)، كالاستعجال وطلب المال وتجاوز الصلاحيات وانتحال السلطة وغيرها. تقترح الدراسة دمج مجموعة من خوارزميات معالجة اللغة الطبيعية (NLP) لتقطيع النص الى أسطر ثم الى كلمات، ثم التنظيف والتجذير، وخوارزمية مطابقة الكلمات مع القوائم السوداء الديناميكية للموضوعات (TBL)، ثم إطلاق التحذيرات الملائمة في الوقت الفعلي، بالطبع الفكرة مرنة قابلة للتطبيق على مختلف نواقل الهجوم (اتصال احادي او غير مباشر كنص بريد الكتروني، أو ثنائي الاتجاه كالدردشات). تستخدم عادةً برامج جدر الحماية (Firewall) وأيضاً مصفاة البريد الالكتروني الأمن (Email Filter) وغيرها، جميعها تستخدم خوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) يتم تحديثها باستمرار بالمواقع والخوادم والروابط الخبيثة وبصمات الملفات المكتشفة دوريا وغيرها الكثير، مثل قواعد بيانات ارقام الهواتف الاحتيالية. تبنى هذه القوائم والمكتبات وتتحدث باستمرار من قبل أهم شركات الأمن السيبراني في العالم، مثالها (Phish Tank, Open Phish , Virus Total). يقوم النظام الهجين المقترح على أتمتة أغلب حالات النموذج المحسن للكشف اليدوي (الاجتماعي) عن الهجمات بالهندسة الاجتماعية (SE)، عبر المطابقة مع القوائم السوداء الديناميكية للموضوعات (TBL) المنتمية دلالياً الى حالات النموذج المحسن للكشف اليدوي (الاجتماعي)، بالطبع يمكن تحميل قوائم سوداء ديناميكية للموضوعات (TBL) ومكتبات محلياً ومجاناً داخل النظام البرمجي. أو التواصل مع خوادم الفحص السحابية العالمية، أو استعمال تقنية (Bloom Filter). يمكن استخدام بعض أو جميع ما ذكر بحسب أهمية المؤسسة، مع مراعاة خصوصية وسرية وفيدرالية البيانات، والأخذ بعين الاعتبار أن الامن مقدم على التأخير. برمجياً كل هذا متاح ليخدم الهدف البحثي للدراسة الحالية. تم تصميم تطبيق جافا سكريبت أولي كمنصة تفاعلية لشرح الفكرة.

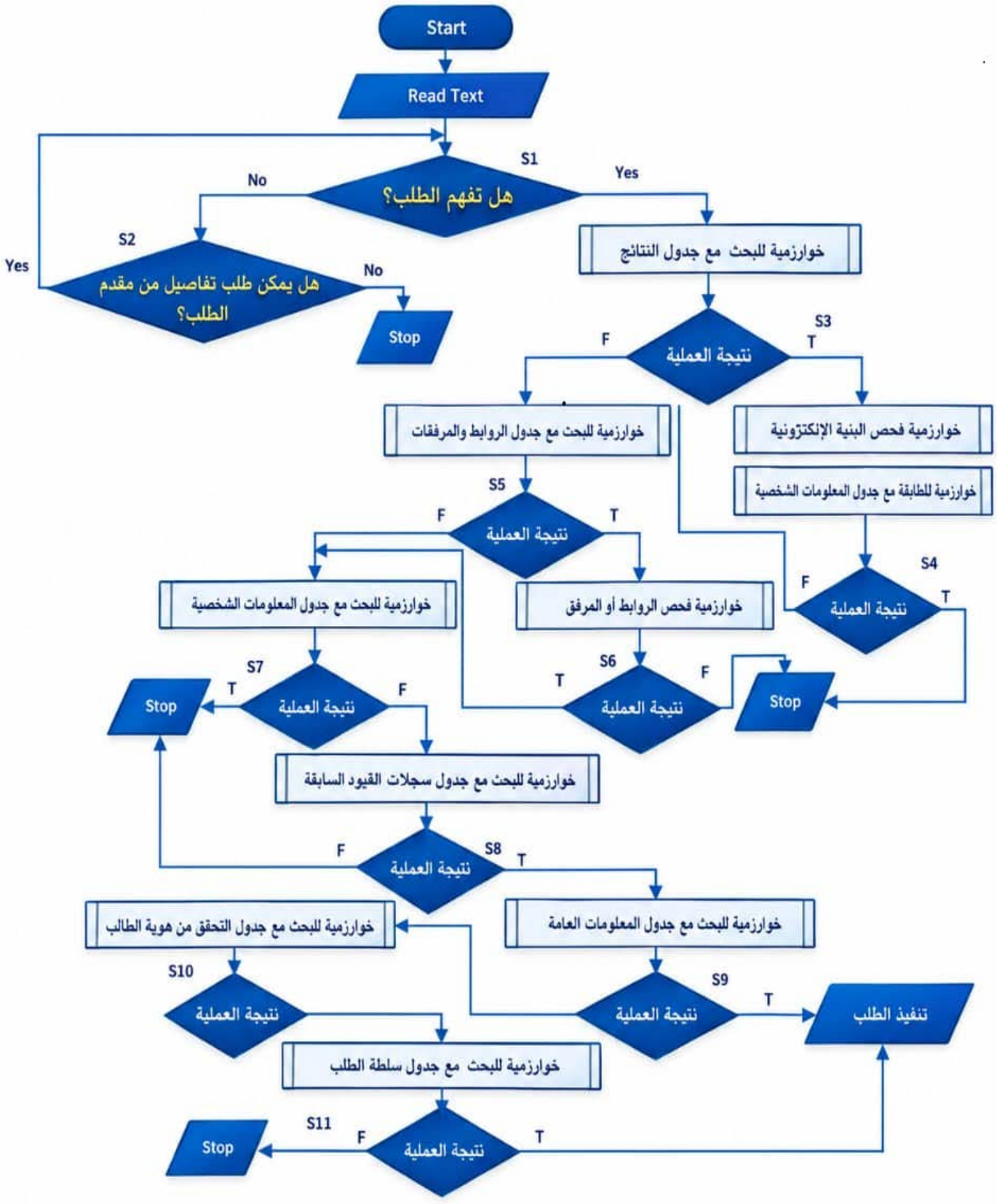
النظام الهجين المقترح

تعتمد الأنظمة المعتمدةعلى القواعد (RBS) والذكاء الاصطناعي الرمزي (SAI) على مجموعة قواعد محددة مسبقًا لتحديد هجمات الهندسة الاجتماعية (SE). تستند هذه القواعد إلى أنماط الهجوم والسلوكيات المعروفة. وتعتمد حظر تلقائي مؤقت لكلمات ومواضيع معينة في الوقت الفعلي، بناءً على سياق الأحداث أو سلوك المستخدمين. من خلال هذا المنحى، يقترح البحث نموذجاً قائماً على القواعد والتعلم الآلي ومعالجة اللغة الطبيعية، محسن ومخصص وقائم على المحاكمة المنطقية لبعض القواعد والمعايير مثل: قلة الوعي، الاستعجال، الثقة، الخوف والفضول، وتعديل خوارزمية(SEADM) بإضافة بعض القواعد عليها، كالثقة والصلاحيات، المعاملة بالمثل والالحاح، التحيز في الإدراك، بهدف رفع وعي المستخدم لكشف هجمات الهندسة الاجتماعية (SE). يقوم النظام الهجين المقترح بتطبيق خوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) للمطابقة مع جداول الحظر، حسب الحالة في النموذج اليدوي، حيث تمثل كل حالة (جدول) قائمة موضوعات مستقلة ضمن (TBL)، للتحقق من موضوعات المناقشة لكل سطر من نص أنشأه مهاجم محتمل، واستخدام معالجة اللغة الطبيعية والتعلم الآلي، لاكتشاف هجمات الهندسة الاجتماعية (SE) آليا. يقترح البحث طريقة لاكتشاف هجمات الهندسة الاجتماعية (SE) عبر إجراء تحليل دلالي لجميع النصوص المرسلة إلى الضحية لتخمين المواضيع والكلمات المحظورة التي تتم مناقشتها في كل سطر ومقارنتها مع الكلمات المخزنة في القوائم حسب الحالة. ثم التحقق والتأكد من المعالجة والإجابة الآلية لاتخاذ القرار المناسب والانتقال للحالة التالية حسب الخوارزمية الهجينة المقترحة، إذا تم اكتشاف هجوم، فسيتم إنشاء رسالة تحذير للضحية.



الشكل رقم 4: المخطط العام لنظام الكشف الهجين المقترح

خوارزمية النظام الهجين المقترحة



جداول الحالات

تبنى الجداول الديناميكية على أهداف هجوم الهندسة الاجتماعية (SE) الشائعة. وترتبط كل كلمة جذر في بمجموعة مرادفات.

الجدول 3 القائمة السوداء للموضوع	
الفعل (الطلب)	الموارد
إرسال	الأموال
إرسال	البيانات الحساسة
اتصل	بالرقم
زيارة	الموقع

الجدول 4 الإلحاح	
#	الكلمات الدالة على الالحاح
1	فورا
2	عاجل
3	سريع
4	التصرف الآن
5	موعد نهائي
6	حد زمني
7	مدة محددة

الجدول 5 الأفعال المرتبطة بالإلحاح	
الموارد	الفعل (الطلب)
الحساب	تجميد
الحساب	تعطيل
الحساب	إيقاف
انذار	توجيه
الخدمات	انهاء
الراتب	خصم
الترقية	تأخير

الجدول ٦ المعلومات الشخصية	
#	الكلمات الدالة على المعلومات الشخصية
1	كلمة المرور
2	رقم بطاقة الائتمان
3	عنوان السكن
4	عدد افراد الأسرة
5	مدارس الأولاد
6	الراتب الشهري
7	الحالة الاجتماعية
8	البريد الالكتروني الشخصي
9	الحساب البنكي
10	الجنس

الجدول 7الروابط والمرفقات (URL)	
الفعل (الطلب)	الموارد
افتح	المرفق
افتح	الملف
افتح	الرابط
افتح	صفحة الويب
انقر	الرابط
حمل	المرفق
نزل	المرفق
انتقل	للصفحة
أملئ	الاستبيان
أملئ	النموذج

الجدول8 صلاحيات التنفيذ (للمستلم)	
رقم الموظف	
المنشور على العام S9	
الاذن ١	
الاذن ٢	

الجدول 9 المعلومات المتاحة للعامة	
يشمل هذا الجدول قائمة عناوين المواضيع والمعلومات المتاحة للعامة، والفقرات المنشورة على الموقع الالكتروني للمؤسسة.	
#	البند
١	من نحن
٢	اهدافنا
٣	رؤيتنا
٤	مشاريعنا
٥	منتجاتنا
٦	عروضنا
٧	الاتصال بنا

الجدول ١٠ التحقق من هوية مقدم الطلب	
الرقم	
الاسم	
الهاتف	
البريد الالكتروني	
صورة الهوية	

الجدول ١١ سلطات مقدم الطلب	
الرقم	
سلطة مستوى ١	
سلطة مستوى ٢	
سلطة مستوى ٣	

CC BY

لتقييم خوارزمية الكشف اليدوي المقترحة في القسم 3.1. تم ارسال بريد الذريعة لعدد 140 طالب، 60 طالب منهم ماجستير علوم الويب في الجامعة الافتراضية السورية. تم جذب 25 ضحية من أصل 60 طالب، أي بنسبة بلغت 41,67 % من اجمالي عينة المجموعة الأولى، وعند ارسال الرسالة الى عدد 80 طالب سنة رابعة وخامسة هندسة معلوماتية / الجامعة الدولية الخاصة، تم جذب 50 ضحية أي بنسبة بلغت 62.50 % من اجمالي عينة المجموعة الثانية، مما يؤكد زيادة احتمال نجاح الهجوم وجذب ضحايا أكثر في مجتمعات أقل خبرةً في حال تغيير نص الذريعة ليناسب مجتمعات متنوعة بحسب اهتماماتها، ويثبت مرونة ناقل الهجوم حيث لا يقتصر بالضرورة على رسائل البريد الإلكتروني، بل يمتد ليشمل النصوص المنشورة على منصات التواصل الاجتماعي المختلفة. لاحقاً، تم اجراء استبيان لمعرفة مدى استفادة الطلاب من التجربة البحثية، ولمعرفة في أي مرحلة من حالات النموذج تبين للمستلم انه كان بريداً وهمياً، حيث أن أسئلة الاستبيان تمثل الحالات المنهجية للنموذج المقترح (S1–S11). لدينا 75 حالة هجوم محتمل في العالم الحقيقي بين مجموعة من الاختصاصيين. النتيجة المرجوة من هذا الاختبار هي توجيه المستلمين لاتخاذ قرار إيقاف الطلب (SF) لجميع الحالات الخمس والسبعين، مما يرفع الوعي الأمني للمستلم لاتخاذ القرار المناسب. يعرض الجدول التالي أسئلة الاستبيان:

الجدول 12 أسئلة الاستبيان

رقم القاعدة	القاعدة	الطلب المؤجل		تنفيذ الطلب
		نعم	لا	
S1	هل تفهم ما هو المطلوب؟	66	9	لا
S2	هل يمكن طلب تقديم تفاصيل أكثر عن الطلب من مقدم الطلب؟	57	18	
S3	هل يحاول مقدم الطلب خلق شعور إلحاح؟	29	46	
S4	هل يمكن للمستلم أن يبطئ ويتأكد من شرعية الطلب؟	58	17	
S5	هل يحثك مقدم الطلب على فتح مرفق أو النقر على رابط (URL)؟	42	33	
S6	هل تستطيع التحقق من المرفق أو الرابط (URL) قبل الضغط عليه؟	53	22	
S7	هل الطلب غير عادي ووصل بشكل غير متوقع أو يجيب على سؤال لم يطرح ويطلب معلومات شخصية؟	26	49	
S8	هل لدى المستلم صلاحية تنفيذ الطلب؟	53	22	
S9	هل الإجراء أو المعلومات المطلوبة متاحة للعامة؟	41	34	
S10	هل يمكن التحقق من هوية مقدم الطالب؟	50	25	
S11	هل يمتلك مقدم الطلب السلطة اللازمة لطلب الإجراء أو المعلومات؟	42	33	

المناقشة

نجح تسلسل أسئلة الاستبيان في تصفية المخاطر تدريجياً، ورفع مستوى وعي المشاركين لرفض الطلب في النهاية، لأن الأسئلة مصممة بالفعل لتثير الشك بوجود ثغرات قد تؤدي لهجوم حقيقي. سلط النموذج الضوء على سلوكيات التلاعب السيكولوجي بالمستخدمين بدلاً من التلاعب بالنظام، ورفع وعي المستخدمين بدلاً من التركيز التقليدي على التقنيات والأجهزة المكلفة. وبناءً عليه يمكننا استخدام النموذج السلوكي المحسن المقترح للدفاع ضد هجمات الهندسة الاجتماعية (SE) كأداة لرفع وعي المستخدمين، ورفع أداء التقنيات السيبرانية لمواجهة الهجمات. ان القيمة الابتكارية المضافة تكمن في دمج النهج السلوكي مع خوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) والتعلم الالي (ML) كفكرة رائدة لإشراك وعي المستخدم في رفع كفاءة التقنيات لمواجهة الهجمات، تنجح هذه البنية الهجينة في سد الفجوة البحثية، حيث اكتفت الدراسات السابقة بالحل التقني او دمج مجموعة تقنيات أو دراسة نظرية للعوامل الاجتماعية والسيكولوجية المسببة للهجوم دون حلول إجرائية. لاختبار الكفاءة الإجرائية وصحة المنطق البرمجي صمم تطبيق جافا سكريبت يحاكي ألية الكشف باستخدام خوارزمية القوائم السوداء الديناميكية للموضوعات (TBL) الممثلة للأنظمة المعتمدة على القواعد (RBS) والذكاء الاصطناعي الرمزي (SAI). تتسم بنية النموذج الهجين المقترح بقدرتها الاستيعابية على دمج خوارزميات التعلم الآلي (ML)، لمعالجة المشكلة على نطاق أوسع، ومرونتها التصميمية البرمجية. لذلك، تم بناء مجموعة بيانات متوازنة مؤلفة من (50.000) رسالة بريد الكتروني مرمزة ثنائياً (0) رسالة شرعية، (1) رسالة احتيال الكتروني. تم توزيعها مناصفةً (50:50) لضمان توازن البيانات. جُمعت عينات مجموعة البيانات الكلية (Dataset) من المستودعات القياسية العامة كالتالي: (25.000) رسالة بريد إلكتروني شرعية من بيانات شركة إنرون القياسية Enron (Email Corpus) وهي أشهر قاعدة بيانات عالمية تحوي مئات الألاف من رسائل البريد الالكتروني الحقيقية، و(25.000) رسالة احتيال الكتروني من بيانات نازاريو القياسيةNazario Phishing) (Corpus التي تحوي مئات ألاف رسائل الهندسة الاجتماعية (SE) الحقيقية التي استهدفت الضحايا بأساليب الخداع النفسي، والروابط الخبيثة وطلب البيانات. دمجت العينات وتم اكمال النقص بتوازن من (Spam Assassin Public Corpus) وعينات (Kaggle) العامة لتكوين ملف بيانات موحد ومغذٍ للتطبيق البرمجي المطور، لقياس استجابته الأمنية. وبعد اجراء المعالجة اللغوية للنصوص كتجريد الرموز وتوحيد حالات الأحرف، تم تطبيق تقنيةTF–IDF) (Vectorizerلاستخلاص الميزات النصية مع وضع قيود معيارية على أقصى حجم للميزات لمنع ظواهر الافراط في التخصيص (overfitting) والتسريب المعرفي وفرط التدريب، قُسّمت مجموعة البيانات الاجمالية عشوائياً بنسبة 80% (40,000 رسالة) لغرض التدريب و20% (10,000 رسالة) المتبقية لغرض الاختبار، وتم اخضاع خوارزميات التصنيف المتنافسة للتحليل المعمق عبر مصفوفة الارتباك (confusion matrices) الخاصة بمجموعة الاختبار المعزولة والبالغة (10,000) رسالة. تعتمد الخوارزمية المقترحة آلية تكامل هجين، أولاً، تتم مقاطعة الكلمات(Tokens)

CC BY

عبر استعلامات مباشرة مع الجدول المعرفية الديناميكية (الفحص القائم على القواعد (RBS–Matching)) لرصد أفعال الطلب ومواردها ومرادفاتها المرتبطة بالإلحاح أو المعلومات الشخصية والروابط الحساسة ، التي تشير مباشرةً الى هجوم هندسة اجتماعية، حيث تستخلص هذه الميزات بشكل مستقل لكل رسالة لتجنب تسرب البيانات (Data Leakage) . ثانياً، تشغيل مصفوفة التردد العكسي لاستخلاص الأوزان الإحصائية للكلمات والأنماط الدلالية للرسالة. تُربط مخرجات المسارين معاً عبر الية التكديس الافقي لتوليد متجه ميزات هجين(Hybrid Feature Vector) ، ليغذي هذا المتجه الهجين المتكامل خوارزميات التعلم الآلي الكلاسيكي المقارنة، حيث يتم تدريبها واختبار كفاءتها على المجموعات المعزولة مسبقاً لضمان دقة التقييم وقدرة النماذج على التعميم. يوضح الجدول13 أداء أبرز الخوارزميات في حالتي التدريب والاختبار.

الجدول ١٣ نتيجة دقة التدريب والاختبار للمصنفات الأربعة للنموذج الهجين المقترح

نموذج التعلم الآلي	دقة التدريب على 40000 رسالة	دقة الاختبار على 10000 رسالة
Support Vector Machines (SVM)	**0.9850**	**0.9740**
Naive Bayes	**0.9780**	**0.9630**
K-Nearest	**0.9620**	**0.9410**
Decision Tree	**0.9450**	**0.9120**

تظهر القراءة التحليلية للنتائج تفوقاً واضحاً للنماذج الخطية في التصنيف النصي ثنائي الفئة. فقد سجلت خوارزمية (SVM) أعلى دقة اختبار بلغت 97.40% متفوقةً بفارق ضئيل على مصنف (Naive Bayes) الذي حقق 96.30%. في المقابل، سجل مصنف (K–Nearest) دقة اختبار متزنة بلغت 94.10%، متقدماً على نموذج (Decision Tree) الذي استقر عند دقة اختبار واقعية بلغت 91.20%. إن تراجع أداء شجرة القرار يُعد سلوكاً منطقياً متسقاً مع المراجع العلمية[25] ، حيث يميل التصنيف المعتمد على شجرة القرار إلى ارتكاب أخطاء تصنيفية عندما تكون الحدود الفاصلة بين الكلمات متداخلة وكثيفة. يلاحظ من مراجعة قيم الفوارق الإحصائية بين دقة التدريب ودقة الاختبار ثبات مؤشر القدرة التعميمية لجميع النماذج، حيث انحصر الفارق في نطاق آمن منضبط (بين 1.1% إلى 2.1%)، مما يقلل بشكل كبير من احتمال حدوث إفراط في التخصيص(Overfitting) . خضعت مخرجات الخوارزميات التصنيفية المتنافسة لتحليل معمق عبر مصفوفة الارتباك (confusion matrices) على مستوى عينة الاختبار المستقلة البالغة 10,000 رسالة. تُلاحظ كفاءة خوارزمية(SVM) التي تربع أداؤها على قمة التصنيف، في تقليص الإنذارات الخاطئة وفشل الكشف، حيث نجحت في تحديد 4,870 رسالة تصيد بشكل صحيح من أصل 5,000 عينة (True Positives)، وانحصر فشل الكشف في 130 رسالة فقط تسللت عبر النظام (False Negatives)، وأظهرت مرونة أمنية عالية برصد 4,870 رسالة عمل شرعية (True Negatives)، ولم تسجل سوى 130 إنذار خاطئ تسبب في حجز بريد شرعي (False Positives)، مما يمنح النموذج وثوقية عالية للعمل كخط دفاع أول دون تعطيل أعمال الموظفين. في المرتبة الثانية، أظهر مصنف (Naive Bayes) سلوكاً قريباً جداً من (SVM)، وانحصرت الأخطاء التصنيفية لديه في 185 عينة خطأ لكل فئة من أصل عينات فحص الاختبار المستهدفة. يُعزى هذا الاستقرار والمعدل المنخفض للإنذارات الخاطئة لدى بايز الساذج إلى حساسيته الاحتمالية العالية لتكرار الكلمات، حيث تمكن بكفاءة من تصفية الميزات اللغوية التي يستغلها المهاجمون لإثارة الضغط النفسي[26] (عناصر الإلحاح). في المقابل، أظهر نموذج الجار الأقرب ميلاً واضحاً لارتكاب أخطاء من النوع الثاني (فشل كشف الاحتيال الالكتروني)، وأخفق في كشف ما يقارب 300 رسالة احتيالية لاعتماده على المسافات الإقليدية المتداخلة في الفضاءات الرقمية متعددة الأبعاد[27]. أما مصنف شجرة القرار، فقد عانى من تشتت إحصائي انعكس صراحةً في مصفوفة الارتباك الخاصة به، حيث سجل أعلى معدل أخطاء إجمالي في التجربة، تمثل في وقوع حوالي 440 عينة خطأ موزعة بين إنذارات خاطئة وفشل كشف. هذا القصور يبرهن عملياً على أن أشجار القرار المنفردة تميل إلى اتخاذ قرارات حادة وجامدة عند مواجهة نصوص تحتوي على تكرار وتداخل سياقي متعمد من قبل المهاجمين، مما يجعلها أقل مرونة من النماذج الخطية والاحتمالية عند التعامل مع تعقيد وتطور هجمات الهندسة الاجتماعية (SE) المعاصرة[28] . يعرض الجدول14 صياغة نموذجية لنتائج النماذج الأفضل في الدراسة. القراءة التحليلية للنتائج المدرجة في الجدول 41 تثبت علمياً أن الخوارزميات الكلاسيكية، عند ضبط معاملاتها الفائقة وتغذيتها ببيانات متوازنة بدقة، تمتلك قدرة تنافسية عالية تضاهي الأنظمة الأمنية الحديثة في التمييز القاطع بين أنماط رسائل العمل الشرعية ورسائل الهندسة الاجتماعية(ES) الموجهة، يلاحظ أيضاً تطابق قيم المؤشرات الإحصائية ، هذا التساوي الرياضي يعزى لتوازن توزيع فئات عينة الاختبار، وتناظر الأخطاء التصنيفية المستخرجة من مصفوفة الارتباك، مما يعكس استقراراً هيكلياً في أداء المصنفات. بناءً على تحليل البيانات والنتائج، والمقارنة مع الدراسات المرجعية، نجح النموذج الهجين المقترح في كشف هجمات الهندسة الاجتماعية(ES) على عدة مستويات، وسد الفجوة البحثية المتروكة في الدراسات المرجعية، وحقق الهدف المعرفي للدراسة بالوصول الى نموذج كشف هجين يدمج بين الجداول المعرفية الديناميكية و أليات الكشف الآلي المتطورة ضمن حالات خوارزمية الكشف اليدوي، وساهم في اشراك المستخدمين ورفع وعيهم لكشف هجمات الهندسة الاجتماعية(ES).

جدول 14: نتائج المقاييس الإحصائية على عينة الاختبار (10,000) رسالة للنماذج الأفضل في الدراسة

النموذج	حجم ومصدر مجموعة البيانات	Accuracy الدقة	Precision الضبط	F1-Score التوازن
(SVM)	50,000 رسالة (Enron+Nazario+Combined)	0.9740	0.9740	0.9740
Naive Bayes	50,000 رسالة (Enron+Nazario+Combined)	0.9630	0.9630	0.9630

الاستنتاجات والتوصيات

يبرهن هذا البحث أن الحلول التقنيّة الصرفة تبقى قاصرة في مواجهة هجمات الهندسة الاجتماعية (SE)، مالم تراع التحديث البرمجي الدائم والتدريب والتوعية لتعزيز دفاعات المستخدمين، نظراً لتزايد انتشار هجمات الهندسة الاجتماعية (SE). هناك حاجة ملحة لدعم الأشخاص بحسب خياراتهم ومهنتهم وقدراتهم. نجح النظام الهجين المقترح بسد هذه الفجوة عبر أتمتة أغلب حالات النموذج المحسن للكشف اليدوي عن هجمات الهندسة الاجتماعية (SE)، وتطبيق خوارزمية القوائم السوداء الديناميكية (TBL) على نص أنشأه مهاجم محتمل، وباستخدام تقنيات معالجة اللغة الطبيعية (NLP) والتعلم الآلي (ML)، تمكن النظام الهجين المقترح من اكتشاف هجمات الهندسة الاجتماعية (SE) آليا، وحقق كفاءة تشغيلية ملموسة، مما يثبت كفاءة الخوارزمية المقترحة كأداة دفاعية في كشف هجمات الهندسة الاجتماعية (SE). ولتوسيع نطاق الاستفادة من النموذج الهجين المقترح، توصي هذه الدراسة مستقبلا بتوسيع الطاقة الاستيعابية للنموذج بدراسته على منصات التراسل الفوري وشبكات التواصل الاجتماعي لضمان مرونة التصنيف في بيئات تواصل متنوعة. كما يوصي البحث بدعم الأنظمة متعددة اللغات، والتحول نحو نماذج اللغة الضخمة (LLMs)، وتضمين خوارزميات التعلم العميق (DL) لمعالجة النوايا ورصد الحالة العاطفية للمستخدم عند تطوير النموذج الهجين المقترح مستقبلاً، أيضاً توصي الدراسة بالبحث في معالجة تحدي الهواتف النقالة، حيث توفرت معظم المؤشرات الأمنية لشاشات الحواسب، بينما ما يزال حجم شاشة الهواتف النقالة لا يسمح بذلك.

المراجع

[1] Nagaraj K, Bhattacharjee B, Sridhar A, Sharvani GS. Detection of Phishing Websites Using a Novel Twofold Ensemble Model. Journal of Systems and Information Technology. 2018 Oct;20(4):392-411. Available from: https://doi.org/10.1108/JSIT-09-2017-0074.

[2] Aonzo S, Merlo A, Tavella G, Fratantonio Y. Phishing Attacks on Modern Android. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security; 2018 Oct 15-19; Toronto, Canada. New York: ACM; 2018. p. 1793-1806. Available from: https://doi.org/10.1145/3243734.3243778.

[3] Lastdrager EE. Achieving a consensual definition of phishing based on a systematic review of the literature. Crime Sci. 2014 Sep 3;3:9. Available from: https://doi.org/10.1186/s40163-014-0009-y.

[4] Alsufyani AA, Alzahrani SM. Social Engineering Attack Detection Using Machine Learning: Text Phishing Attack. Indian J Comput Sci Eng. 2021 Jun;12(3):743-751. Available from: https://doi.org/10.21817/indjcse/2021/v12i3/211203298.

[5] Iuga C, Nurse JRC, Erola A. Baiting the hook: factors impacting susceptibility to phishing attacks. Hum Cent Comput Inf Sci. 2016 Dec;6:8. Available from: https://doi.org/10.1186/s13673-016-0065-2.

[6] Adu-Manu K, Ahiable R. Detecting Phishing Using a Multi-Layered Social Engineering Framework. J Comput Sci. 2023;5:13-32. Available from: ProQuest; https://www.proquest.com/openview/08a4c605b57abbcae828fbb328f25bc0/1?pq-origsite=gscholar&cbl=4585457 [Accessed 2026 Jul 6]. doi:10.32604/jcs.2023.043359.

[7] Waelchli S, Walter Y. Reducing the Risk of Social Engineering Attacks Using SOAR Measures in a Real World Environment: A Case Study. Comput Secur. 2025 Jan;148:104137. Available from: https://doi.org/10.1016/j.cose.2024.104137.

[8] Bell S, Komisarczuk P. An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank. In: Proceedings of the Australasian Computer Science Week Multiconference; 2020 Feb 4-6; Melbourne, Australia. New York: ACM; 2020. p. 1-11. Available from: https://doi.org/10.1145/3373017.3373020.

[9] Venkatesha S, Reddy KR, Chandavarkar BR. Social Engineering Attacks During the COVID-19 Pandemic. SN Comput Sci. 2021 Mar;2(2):78. Available from: https://doi.org/10.1007/s42979-020-00443-1.

[10] Hijji M, Alam G. A Multivocal Literature Review on Growing Social Engineering Based Cyber-Attacks/Threats During the COVID-19 Pandemic: Challenges and Prospective Solutions. IEEE Access. 2021;9:7152-7169. Available from: https://doi.org [11] "How Experts Detect Phishing Scam Emails." https://doi.org/10.1145/3415231.

[11] Pan Y, Ding X. Phishing page detection via learning classifiers from page layout features. EURASIP J Wirel Commun Netw. 2019 Feb 20;2019(1):33. Available from: doi:10.1186/s13638-019-1361-0

[12] Malloy T, Ferreira MJ, Fang F, Gonzalez C. Training Users Against Human and GPT-4 Generated Social Engineering Attacks. arXiv [Preprint]. 2025 Feb 3 [cited 2026 Oct 5]:arXiv:2502.01764 [cs.HC]. Available from: https://doi.org/10.48550/arXiv.2502.01764.

[13] Sahingoz OK, Buber E, Demir O, Diri B. Machine Learning Based Phishing Detection from URLs. Expert Syst Appl. 2019 Mar;117:345-357. Available from: doi:10.1016/j.eswa.2018.09.029

[14] Adebowale MA, Lwin KT, Hossain MA. Intelligent phishing detection scheme algorithms using deep learning. J

الاختصارات:

الجدول رقم 1

اساهمات المؤلفين الإطار المفاهيمي: باسل الخطيب– دانا نجيب

المنهجية: دانا نجيب

التحقيقات: باسل الخطيب– دانا نجيب

إدارة المشروع: باسل الخطيب

الإشراف: باسل الخطيب

الكتابة – المسودة الأصلية: دانا نجيب

الكتابة – المراجعة والتحرير :باسل الخطيب– دانا نجيب

تضارب المصالح:يعلن المؤلفون أنه ليس لديهم أي مصالح متضاربة.

توافر البيانات والمواد:جميع البيانات متوفرة في النص الرئيس عدا الأشكال والمراجع.

المواد التكميلية: قائمة المراجع – قائمة الأشكال – قائمة الجداول (المدرجة في النص).

التمويل: لا يوجد

Enterprise Inf Manag. 2020 May 20;34(3):849-868. Available from: https://doi.org/10.1108/JEIM-01-2020-0036
[15] Zamir A, Khan H, Iqbal T, Yousaf N, Aslam F, Anjum A, et al. Phishing web site detection using diverse machine learning algorithms. Electron Libr. 2020 Jun 25;38(3):543-564. Available from: https://doi.org/10.1108/EL-05-2019-0118
[16] Soyemi J, Mudasiru H. An enhanced authentication scheme for preventing phishing attacks on WhatsApp accounts. In: Proceedings of the 2nd International Conference on Science, Technology, Engineering and Management (ICSTEM); 2020 Nov 10-11; Ilaro, Nigeria. FPI; 2020. p. 45-53. Available from: https://erepository.federalpolyilaro.edu.ng/handle/123456789/1627
[17] Roy SS, Thota P, Naragam KV, Nilizadeh S. From Chatbots to Phishbots?: Phishing Scam Generation in Commercial Large Language Models. In: Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP); 2024 May 19-23; San Francisco, CA, USA. Piscataway: IEEE; 2024. p. 36-54. Available from: https://doi.org/10.1109/SP54263.2024.00182
[18] Najeeb D. Forked Repository [Software on the Internet]. Zenodo; 2026. Available from: https://doi.org/10.5281/zenodo.23168960
[19] Hannousse A, Yahiouche S. Web page phishing detection [Dataset]. Mendeley Data; 2021. Available from: Mendeley Data; 2021. doi:10.17632/c2gw7fy2j4.3.
[20] Najeeb D. Phishing Features Detection and Analysis Code [Computer Program on the Internet]. Google Colab; 2026 [cited 2026 Oct 5]. Available from: https://colab.research.google.com/drive/1CSNK2zKcS7CxYTi2Yi_B5HLMhKApkM0O?usp=sharing
[21] Najeeb D. JavaScript Application [Software on the Internet]. Zenodo; 2026. Available from: https://doi.org/10.5281/zenodo.23168848
[22] Najeeb D. Win [Software on the Internet]. Zenodo; 2026. Available from: https://doi.org/10.5281/zenodo.23168574
Najeeb D. Social Engineering Detection and Feature Engineering Code [Computer Program on the Internet]. Google Colab; 2026 [cited 2026 Oct 5]. Available from: https://colab.research.google.com/drive/13kySjOAxcyvoCaNzOQ-WiXolLsu5c0Qt?usp=sharing
[23] Najeeb D. Web Page [Internet]. Zenodo; 2026. Available from: https://doi.org/10.5281/zenodo.23169057
[24] Alam NA. Phishing Email Dataset [Dataset]. Kaggle; 2024 [cited 2026 Oct 5]. Available from: https://www.kaggle.com/datasets/naserabdullahalam/phishing-email-dataset
[25] Napierala K, Stefanowski J. Types of minority class examples and their influence on learning classifiers. J Intell Inf Syst. 2016;46(3):563-597. doi: 10.1007/s10844-015-0368
[26] Salloum SA, Al-Emran M, Monem AA, Shaalan K. A survey of text mining in social media: highlights and trends. Int J Inf Manag Data Insights. 2021;1(2):100028. doi:10.1016/j.jjimei.2021.100028
[27] Cvetko T. The Curse Of Dimensionality in KNN Classifiers. Towards AI. 2023;12(4):102-115. doi:10.21384/tai.knn.2023.04
[28] Trevisan T. Machine Learning and Text Classification: Stability and Overfitting of Single Decision Trees. Politesi. 2026;14(2):112-126. doi:10.1057/pol.dt.2026.03.

Enhancing IoT Security Using Spider Web-Based Dynamic Keys and Blockchain Encryption)

(تعزيز أمان إنترنت الأشياء باستخدام مفاتيح ديناميكية مستوحاة من شبكة العنكبوت وتشفير سلسلة الكتل

Razan Rezk: Syrian Virtual University _ Damascus _Syria
Wasim Al-Juneidi: Syrian Virtual University _ Damascus _Syria
Corresponding: razanrezk96@gmail.com
Submitted date: 23 August2026 - **Accepted date:** 1 October 2026.

ABSTRACT

The present investigation proposes a novel hybrid security architecture that reconciles robust defense against Denial-of-Service (DoS) attacks with resource-efficient operation within Internet of Things-based wireless sensor networks (IoT-WSNs). The framework synergistically integrates three principal techniques: Spider-Web-based Dynamic Key (SWDK) generation for lightweight cryptographic operations, Principal Component Analysis (PCA) for data dimensionality reduction, and blockchain technology for immutable, tamper-resistant authentication. This model underwent comprehensive evaluation via the NS-3 network simulator, employing a cluster-based topology consisting of one base station and four sink nodes acting as cluster heads, with network densities varying from 3 to 9 sensors per cluster under static node and open-channel assumptions. Performance was assessed under two flooding attack configurations: a distributed, concurrent assault transmitting at 500 packets per second, and a concentrated, high-intensity assault utilizing larger packet sizes (1024 bytes) and escalated rates of up to 1500 packets per second. Empirical findings attest to the pronounced efficacy of the integrated framework; Under distributed attacks, the model curbed the malicious packet delivery ratio to below 14% while attaining a classification accuracy surpassing 93%, thereby reducing the overall attack success rate by 73% to 85%. In the context of concentrated, high-density attacks, although the blockchain layer independently demonstrated only modest resilience, the complete triple-integrated system yielded exceptional latency stabilization, achieving an average reduction of approximately 70% (with observed spanning improvements of 55% to 82%). This security enhancement inherently introduced marginal performance trade-offs, evidenced by a mean throughput decline of roughly 25% and a concomitant reduction in the legitimate Packet Delivery Ratio (PDR). Nevertheless, given the stringent security demands of mission-critical domains such as healthcare and smart infrastructure, this compromise is justifiable. Subsequent research directions should prioritize the optimization of queuing mechanisms, the implementation of adaptive compression protocols, and the periodic renewal of cryptographic key parameters, alongside exploratory efforts toward hardware acceleration and topological scalability.

Keywords: Wireless Sensor Networks (WSNs), Internet of Things (IOT), Data Security Spider Web Dynamic Key, Blockchain, Data Compression, Secure Data Aggregation.

الملخص

تهدف هذه الدراسة إلى تطوير نموذج أمني هجين مبتكر يوازن بين الحماية القوية ضد هجمات حجب الخدمة (DoS) وكفاءة استخدام الموارد في شبكات الاستشعار اللاسلكية لإنترنت الأشياء (IoT–WSN). يدمج الإطار المقترح ثلاث تقنيات رئيسية: توليد المفاتيح الديناميكي القائم على شبكة العنكبوت (SWDK) للتشفير الخفيف، وتحليل مكونات المفاتيح (PCA) لضغط البيانات، وتقنية سلسلة الكتل (Blockchain) للمصادقة المقاومة للتلاعب. تم تقييم النموذج باستخدام محاكي شبكة عالي الدقة (NS–3) ببنية عنقودية تتألف من محطة أساسية مركزية وأربع عقد طرفية (مستقبلات) كرؤوس للعناقيد، مع كثافات شبكية متدرجة (من 3 إلى 9 مستشعرات لكل عنقود) بافتراض عقد ثابتة ونماذج قنوات مفتوحة. تم اختبار النموذج في ظل سيناريوهين لهجوم الفيضان: هجوم موزع متزامن بمعدل 500 حزمة/ثانية، وهجوم مركز بحجم حزمة أكبر (1024 بايت) ومعدل يصل إلى 1500 حزمة/ثانية. أظهرت النتائج كفاءة النموذج الهجين العالية. في بيئة موزعة، نجح النظام في خفض نسبة تسليم الحزم الخبيثة إلى أقل من 14%، وحقق دقة تصنيف تتجاوز 93%، مما أدى إلى انخفاض معدل نجاح الهجمات الإجمالي بنسبة تتراوح بين 73% و85%. أما في الهجمات المركزية عالية الكثافة، فرغم أن أداء تقنية البلوك تشين وحده لم يكن حاسمًا، إلا أن التكامل الثلاثي حقق استقرارًا ممتازًا في زمن الاستجابة، حيث خفض متوسط زمن الاستجابة بنسبة 70% (بين 55% و82%). وعلى الرغم من أن المفاضلة الأمنية أدت إلى انخفاض طفيف في الإنتاجية الإجمالية (حوالي 25%) وانخفاض في معدل تسليم الحزم المشروعة، إلا أن هذا حل وسط مقبول نظرًا لطبيعة التطبيقات الحيوية كالرعاية الصحية والبنية التحتية الذكية. ويوصى بإجراء تحسينات مستقبلية على إدارة قوائم الانتظار، والضغط التكيفي، والتجديد الدوري للمعاملات المشفرة، مع إمكانية تسريعها عبر الأجهزة وزيادة قابليتها للتوسع..

الكلمات المفتاحية:شبكات المستشعرات اللاسلكية (WSNs)، إنترنت الأشياء (IoT)، البنية النجمية (Star Topology)، أمان البيانات، المفتاح الديناميكي لشبكة العنكبوت (SWDK)، سلسلة الكتل (Blockchain)، ضغط البيانات (PCA)، التجميع الآمن للبيانات.

INTRODUCTION

The proliferation of Internet of Things (IoT) technologies, coupled with advances in network speed and data management, has accelerated progress in domains requiring continuous remote monitoring, including smart homes, industrial control, and healthcare services [1]. However, this evolution introduces significant cybersecurity challenges, particularly at the network edge, where sensors and actuators interact directly with the physical environment [2]. With the global number of connected devices projected to surpass 30.9 billion [3], the frequency and sophistication of cyberattacks are escalating, necessitating robust yet lightweight security mechanisms. Traditional encryption and authentication schemes, designed for stable and homogeneous computing environments, are ill-suited to the heterogeneous, resource-constrained nature of modern IoT deployments [4]. This challenge is especially pronounced in Wireless Sensor Networks (WSNs), a prevalent IoT paradigm comprising numerous heterogeneous nodes with limited energy and processing capabilities. The urgency of defending against resource-draining threats, notably Denial-of-Service (DoS) attacks, is paramount. However, excessively complex cryptographic operations accelerate battery depletion and exacerbate communication delays due to packet overhead, underscoring the need for lightweight security strategies to ensure network longevity [5][6][7][8]. To address this gap, the present work proposes a secure architectural framework that integrates lightweight encryption, data compression, and blockchain technology. Specifically, a Spider-Web-inspired Dynamic Key (SWDK) enhances device identification and security, Principal Component Analysis (PCA) reduces transmission overhead by compressing data, and blockchain-based encryption consolidates data integrity and future-proofs the network against adversaries, all while preserving energy efficiency. Scientific Contributions Mutual Triple Integration: The synergistic combination of SWDK, PCA, and Blockchain provides comprehensive protection without imposing excessive resource strain. Attack-Type Sensitivity: The discernible variance in performance between distributed and concentrated attacks provides critical insights for designing adaptive defenses. Practical Viability: The ~70% latency reduction under concentrated attacks meets the stringent requirements of real-time Internet of Things (IoT) applications [9].

Previous studies

About denial of service in wireless sensor networks, Demirci et al [1] proposed a kind of security mechanism meant to spot nodes that are compromised already. Their method looks at packet loss ratios, and also checks both the incoming and outgoing data volumes, while using an improved search routine, TDFS, to trace the attack back from the base station all the way to where it started. After they manage to identify the misbehaving node, its communications are rerouted toward a honeypot, so it can be contained there. Their simulations showed the idea is better than EVO and MULTI, especially in how fast it detects the threat, reaching a detection probability of 0.8 even when the likelihood of a malicious node is only 0.05. Similarly, Al Ebrahim et al. [2] built a machine learning driven framework aimed at identifying and mitigating DoS and DDoS assaults in real time inside software-defined WSNs, meaning SDWSN. Instead of focusing on one single signal, they used traffic related characteristics like packet types, packet sizes, sending rates, source and destination addresses, plus some statistical indicators. Using these features, they trained a classifier to separate normal data flows from adversarial ones. The reported outcomes were striking: the packet delivery fraction rose from 1.45% up to 72.74%, the delay dropped significantly from 12,016 ms down to 297 ms, the detection precision reached 100%, and the approach removed false alarms entirely. As cyberattacks keep getting more sophisticated, a combination of various tactics into one single coherent framework has become way more important, mainly to reduce computational overhead. In that direction, Prasad et al. [10] propose a hybrid plan for WSN DoS detection that uses a joint feature selection optimization idea. It blends the Grasshopper Optimization Algorithm (GOA) and a Genetic Algorithm (GA)—they call it GOA-GA. The whole thing is designed to shrink the feature set you actually need, and then do traffic classification with four separate ML models, namely MLP, KNN, CART, and LR. The goal is to label traffic as either normal or belonging to one of four attack categories, and it's tested on a particular WSN benchmark. In their results, the GOA-GA plus KNN demonstrated the best performance, achieving the highest accuracy of 95.51% with the shortest computation duration too. At the same time, Elumalai et al. [4] came up with a more general intrusion detection framework for WSNs, where they combined a temporal anomaly module—an Autoencoder teamed with a GRU—with a lightweight blockchain. That blockchain uses PBFT and smart contracts, aiming to isolate malicious nodes and update routing tables when needed. Their experiments reported 97.4% accuracy, 2.1% false positives, a 61 ms response time, and 2.4 mj energy consumption per operation. Overall, the reported outcomes look better than approaches that rely only on deep learning, or only on blockchain, or even sys-

tems built purely on learning components without that hybrid balance. Additionally, A. Jawahar et al. [5] put together a combined security architecture, mixing blockchain and ML, for dealing with DDoS attacks in IoT setups. They used customized Random Forest plus SVM models, to do closer traffic pattern scrutiny, then they designed a dedicated blockchain for recording security incidents, and finally they linked both parts with an intelligent integration layer. That layer basically routes the analytic outputs into smart contracts so the system can make quick isolation decisions. With all that, they reported around 94% detection rate, with only about a 3.2% false-positive ratio. Mohamed Anwar and Pavalarajan [6] place extra weight on keeping data intact and private, so they designed a rather advanced encryption setup for IoT networks. In their approach, the key is generated on the fly and it refreshes every time a piece of data is sent, which helps the scheme stay less predictable and more robust, even when someone is trying to sneak in or breach things. On another track, Shakor et al. [7] came up with a security model that works in two phases for cloud storage. Phase one is basically about producing dynamic AES keys, so each individual file ends up being encrypted with a key that keeps shifting. Phase two then leans on blockchain, not just to store those keys but to archive them together with their related metadata too, and at the same time they use ECC protection for the public keys in order to strengthen what happens during both sending and storage. Overall, their plan gives strong cipher coverage, enables decentralized control over the cryptographic keys, blocks unauthorized access, and still stays adaptable, meaning it can grow with the newer and more demanding needs of modern cloud security requirements.Blockchain really does shine as a leading way to plant trust, security, and decentralization inside wireless sensor networks, even if at first glance it looks kind of heavy. Draz et al. [8] introduced a blockchain-backed protocol that helps with energy exchange between nodes in those resource scarce WSNs. Instead of going all in on raw computation, they hinge the whole method on smart contracts plus a lighter consensus approach called Proof-of-Stake (PoS), so the protocol tends to reduce both computational load and energy consumption. Their simulations showed it lowers energy waste and also makes the network more adaptable, extending the service time by about 20% when you compare it with standard solutions. Then, Arachchige et al. [9] presented an IoT security design that uses blockchain specifically to counter DDoS attacks, by deploying a distributed protective mechanism that detects intrusions early and then isolates the affected nodes. In their setup, they didn't use Proof-of-Work, they swapped it out for the more energy conscious Proof-of-Authority. Meanwhile, Rai et al. [11] looked at how blockchain and IoT can come together to strengthen data privacy and security in energy domains, and they especially zoomed in on nuclear power. Their proposal describes a secure information governance structure that ties encryption, integrity checks, and one shared communication framework together. They argue that blockchains built in immutability and transparency, provide solid protection for data overall. Taken as a whole, these investigations kind of converge on the notion that the upcoming security paradigms for IoT and WSN ecosystems will be anchored in hybrid frameworks that somehow synergize cryptographic robustness with blockchain dependability [15]. Crucially, they also keep underscoring the ongoing imperative to minimize computational burdens, ensuring that it stays aligned with the tightly constrained resource environments, typical of these networks [12][13][14].

MATERIALS AND METHODS

The suggested approach puts together a unified framework that integrates dynamic encryption strategies with blockchain driven security. Initially, the data collected from sensor nodes gets protected using a dynamic key created via the Spider Web inspired Dynamic Key (SWDK) method. Even if this key boosts the system's overall safety, it also tends to artificially expand the transmitted packet size, so communication overhead goes up, a bit more than expected. To calm that downside, Principal Component Analysis (PCA) is integrated specifically for data compression, before the information is relayed to the aggregation unit. Then the compressed and encrypted streams coming from several sensors finally meet at a central aggregator, where they are processed together, in a group sense. After that, an extra blockchain encryption layer is added, to further back up data dependability and integrity, so it does not get easily tampered with. The network works with a mesh kind of topology including N nodes and L links. From those, we pick a subset M ($M \in N$) as data aggregation units, using the ones with the highest remaining energy levels. Overall, we can write this as a connected graph, G=(N,L),and it contains three different kinds of nodes, not really more. ***Sensors***: these are the terminal nodes in charge of sensing the surroundings and then forwarding the collected data to the nearest aggregation point, more or less. ***Sinks***: these act like cluster heads, they collect the data coming from the nearby sensors, do some processing and aggregation, and then they relay the combined information to-

CC BY

ward the base station. ***Base station***: it is located at the heart of the network, and it acts as the ultimate sink that receives every aggregated report from the four sink nodes, all of them. The simulation deployment comprises one base station and four distinct clusters, each containing a sink node along with a variable number of sensors ranging from 3 to 9 per cluster. To assess the influence of network size on overall performance and to gauge the proposed framework's adaptability across different densities, the sensor count is adjusted in each simulation trial—covering a spectrum typical of small- to mid-scale wireless sensor networks. The spatial arrangement of nodes remains consistent throughout: the base station is fixed at coordinate (50, 50), the four sinks are located at a uniform distance of 25 meters from the base, and the sensors are evenly distributed in circular patterns with a radius of 10 meters surrounding each sink. While this configuration is somewhat idealized and does not incorporate physical obstructions, it remains sufficiently representative for practical experimental validation.

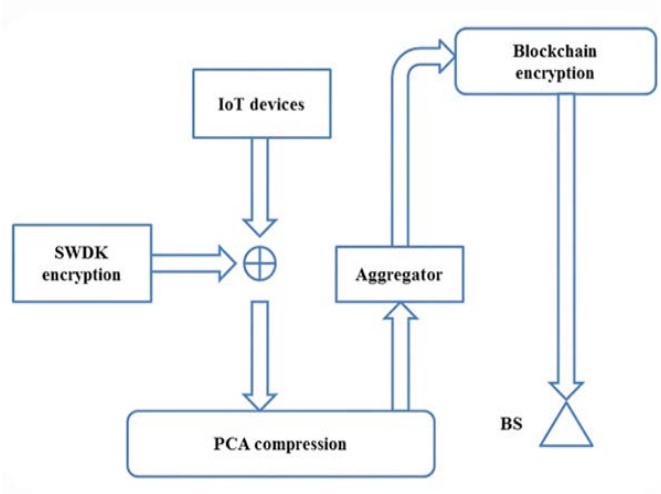


Figure 1: The proposed model

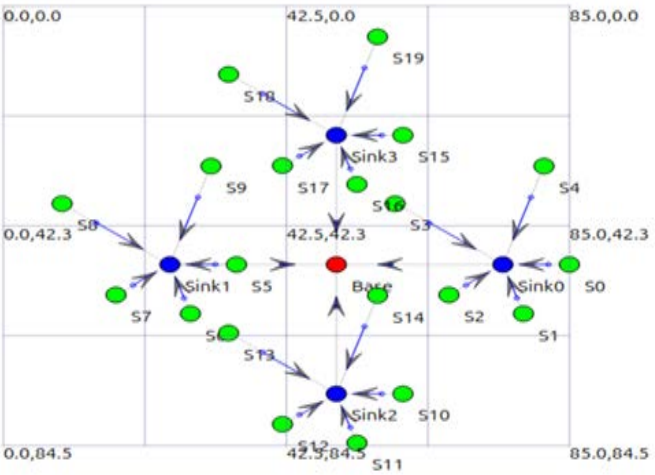


Figure 2: Sensor network topology under investigation.

Dynamic Key Generation (SWDK): The SWDK is derived from the geometrical principles of spiderwebs, yielding an irregular and non uniform structure that offers strong resistance against cryptanalysis and ensures a high security level. The web is composed of radial and spiral strands placed at variable intervals, with their exact offsets governed by configurable parameters—these may incorporate node specific identifiers (such as serial codes or physical positions) to generate diversity across keys. The intersections of these strands define the underlying cryptographic foundation; hence, restoring the key demands thorough familiarity with the parameter set of the Spiderweb Simulation Component (SWC), a model carefully designed to emulate the natural web weaving process of spiders.The key selection process is governed by a set of criteria, namely: a)Data sensitivity and confidentiality: selecting encryption keys appropriate to the sensitivity level of the transmitted data—using stronger keys for highly sensitive data and less robust keys for data with lower sensitivity. b)Susceptibility to attacks: employing dynamic key rotation based on detected attacks. c)Node resource constraints: matching the key to the node's capabilities to ensure no delays occur during data transmission. d)Data category: each data block is classified into a specific category, and a distinct security key is dynamically allocated based on its category and criticality. e)Threat exposure and acceptable risk margins: key selection is based on anticipated threat exposure and the acceptable risk level for each data type.

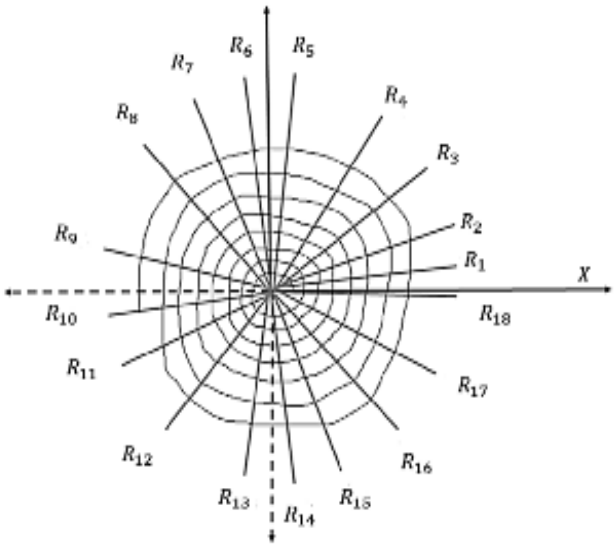


Figure 3:The spiderweb network

The radial lines can be defined by the relation:

$$ad_0 = ad + \Delta ad$$
$$ad_{next} = ad_0 + \Delta ad$$

The spiral lines can be defined by the relation:

$$d_1 = d_{init} + \Delta d$$
$$d_{next} = d_1 + \Delta d$$

The encryption function can be defined by the relation:

$$X'_n = X_n \oplus K_n$$

Principal Component Analysis (PCA): Within the proposed framework, Principal Component Analysis (PCA) fulfills a twofold function: it is deployed either on the raw sensor data matrices at the local node prior to the SWDK-based dynamic encryption, or on the aggregated data matrix structure prior to the key-matching stage:

$$X'_n = X_n \oplus K_n$$

This paradigm leverages PCA's intrinsic capacity to reduce linear dependencies and minimize packet dimensionality, resulting in a marked decrease in network transmission load. Critically, the compression mechanism preserves the integrity of the security model, as the dynamic key's randomness and the ciphertext's decipherability remain entirely intact. The compression operation is formally expressed as:

$$CX'_n = f_c(X'_n) = [{}^{X'_n}_{c}]$$

wherein the ciphertext is supplemented by a compression coefficient. This formulation ensures lossless reconstruction at aggregation nodes while simultaneously achieving enhanced transmission efficiency through reduced in-transit data volumes, without compromising the stipulated security constraints. **Blockchain**: The integration of blockchain technology with Internet of Things (IoT) frameworks constitutes a robust mechanism for safeguarding data integrity and ensuring tamper resistance, as it relies on a distributed ledger architecture that renders any unauthorized modification attempts computationally expensive. Building upon this principle, a private/permissioned blockchain, specifically tailored for clustered wireless sensor network environments, has been adopted. Within this architecture, a lightweight consensus mechanism grounded in Proof-of-Authority (PoA) is distributed among the cluster head nodes (sink nodes) and the base station, with the aim of supporting energy efficiency and significantly reducing response latency compared to conventional protocols such as PBFT or Proof-of-Work. The construction of the blockchain proceeds through a structured multi-stage procedure, which is detailed as follows: Data Segmentation: The aggregated dataset (CX) at the sink node is initially partitioned into j discrete blocks, formally denoted as

$$b_1(CX), b_2(CX), \ldots, b_j(CX)$$

Category-Based Dynamic Encryption: Each data block b_i is subsequently classified into a specific category belonging to the set

$$(Cb_1, Cb_2, \ldots, Cb_n)$$

A distinct security key is then dynamically allocated to each block, contingent upon its designated category and the relative criticality of its contents. This category-aware encryption strategy ensures that stronger cryptographic primitives are assigned to high-sensitivity data segments, whereas shorter key lengths—and consequently, smaller data payloads—are utilized for less critical information. The key selection process is governed by a set of predefined criteria, including data confidentiality requirements, anticipated threat exposure, and acceptable risk margins. Following the assignment and application of the appropriate key, the encryption of the block is finalized. **Blockchain Assembly and Structuring:** For each encrypted block, a final data structure FD_i is constructed, comprising the following four essential components:

Cb_i : The encrypted data payload.

K : The security key employed for encryption.

P : A cryptographic pointer to the preceding block in the chain.

H : The cryptographic hash digest of the current block.

Collectively, these structures form the complete blockchain sequence FD_1,FD_2,...,FD_n, which securely persisted in the database prior to its transmission to the base station. Under this architectural framework, successful reconstruction of the original plain text by an adversarial entity is contingent upon prior access to the dynamically distributed keys, thereby ensuring robust data confidentiality and resilience against unauthorized decryption. **The simulation framework** employed in this study was the NS-3 platform, a high-fidelity environment capable of modeling wireless sensor network behavior with a degree of realism that closely approximates practical deployment scenarios. This platform

CC BY

is particularly well-suited for assessing network robustness against adversarial cyber threats. Two distinct attack classes were emulated through the introduction of malicious nodes configured to inject high-rate packets with spoofed origins. This strategy was designed to exhaust available bandwidth resources and induce queue saturation at both intermediate downstream nodes and the base station (BS). Quantitative performance evaluation under these attack conditions was conducted using the following metrics: **Total Throughput**: Defined as the aggregate volume of data successfully delivered to the destinations, normalized by the total simulation duration, and quantified in kilobits per second (kbps). **Average Delay:** Computed as the mean temporal interval between packet transmission at the source and its successful arrival at the destination. This calculation relies upon a TimestampTag mechanism that records both transmission and reception instants; the difference is subsequently averaged over the total number of successfully received packets. This metric comprehensively encapsulates wireless propagation latency, queuing delays, and processing overhead at the receiver. **Packet Delivery Ratio (PDR):** Determined as the ratio of the number of packets successfully received at their destinations to the total number of packets originating from the sources. **Legitimate Packet Delivery Ratio (Legitimate PDR):** A specific indicator that quantifies the network's capacity to forward authentic data flows in the presence of adversarial activity. **Illegitimate Packet Delivery Ratio (Attack_PDR):** A metric that measures the proportion of maliciously injected packets that successfully traverse the network and reach their intended endpoints. These metrics collectively facilitate a comprehensive and systematic evaluation of the proposed framework's efficacy in mitigating DoS and flooding attacks while preserving reliable communication in resource-constrained sensor network environments. Two distinct attack scenarios were designed to evaluate the resilience of the proposed framework: **Type 1: Concurrent Attack on the Base Station and All Downstream Nodes** In this scenario, three malicious nodes were deployed within a medium sized network a configuration deemed sufficient to assess the attack's impact, while acknowledging that real world deployments may involve varying numbers of adversaries. The attack traffic was generated at a fixed rate of 500 packets per second, distributed uniformly among the base station and the four downstream nodes, corresponding to 100 packets per second directed at each target. This transmission rate, equivalent to approximately 0.5 Mbps, can induce a measurable degradation in Wi Fi network performance. Both legitimate and attack packets were configured with a fixed size of 512 bytes; this standardization was intentionally adopted to isolate the effects of encryption and compression by eliminating packet size variability. **Type 2: Targeted Attack on the Base Station and a Single Sink Node** In the second scenario, three attacking nodes were employed, each transmitting at a rate of 1,000 packets per second to a single sink node and 1,500 packets per second to the base station. These elevated rates are indicative of a high intensity attack. Additionally, the attack packet size was set to 1,024 bytes twice the size of legitimate packets (512 bytes) thereby artificially amplifying the network load.

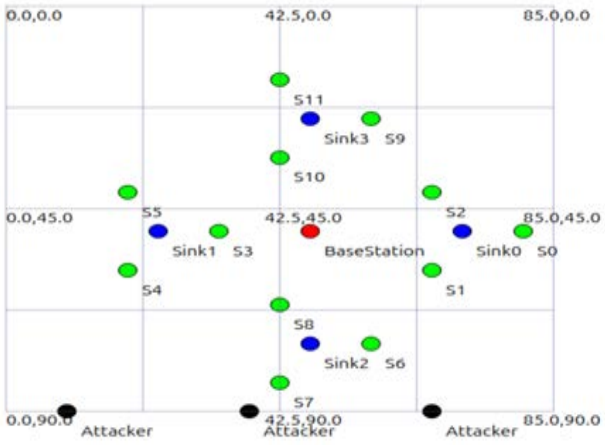


Figure 4: Network nodes and malicious nodes.

RESULTS

Impact of the Attack

An evaluation of the attack's influence on network behavior reveals a pronounced deterioration in Quality of Service (QoS) for legitimate traffic. The attack induces a substantial elevation in end-to-end delay alongside a notable decline in the Packet Delivery Ratio (PDR) for authentic packets. Aggregate throughput registers an increase, attributed to its composition, which accounts for both malicious packets arriving at interfaces and legitimate data. A comparative analysis of the two attack configurations indicates that a focused assault on a single sink induces significantly worse performance in terms of both latency and PDR than a distributed assault across multiple sinks. This is attributable to the complete saturation of the targeted node's ingress port, which generates severe localized congestion.

Performance of Blockchain-Based Protection

Distributed Attack Scenario: Under the distributed attack scenario, the blockchain mechanism demonstrated remarkable effectiveness. It successfully curtailed the proportion of malicious packets reach-

CC BY

ing network interfaces to below 7% and attained classification accuracy exceeding 93%, thereby neutralizing the overwhelming majority of malicious traffic while sustaining relatively modest latency levels. Focused Attack Scenario: Under the concentrated attack scenario, blockchain performance proved less conclusive. Although it reduced the attack success rate relative to an unprotected network configuration, between 25% and 56% of malicious packets still reached their intended targets, with the exact proportion depending on the number of active sensor nodes. This outcome suggests that blockchain in isolation is sufficient for repelling high volume, targeted attacks concentrated on individual ports. Nevertheless, a discernible improvement in end to end latency was observed with blockchain deployment compared to the baseline network under attack. However, latency remained elevated during concentrated attacks, approaching approximately 0.3 seconds.

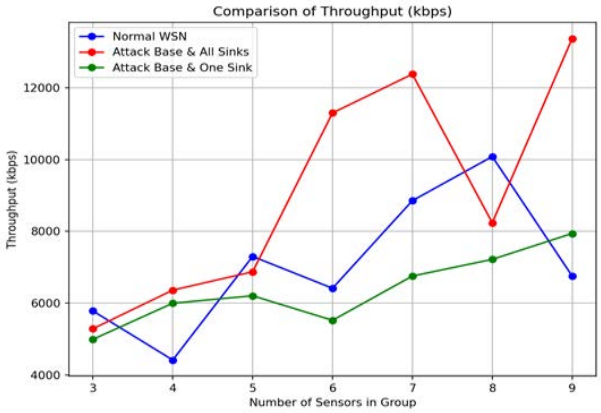


Figure 5: Impact of the attack on throughput.

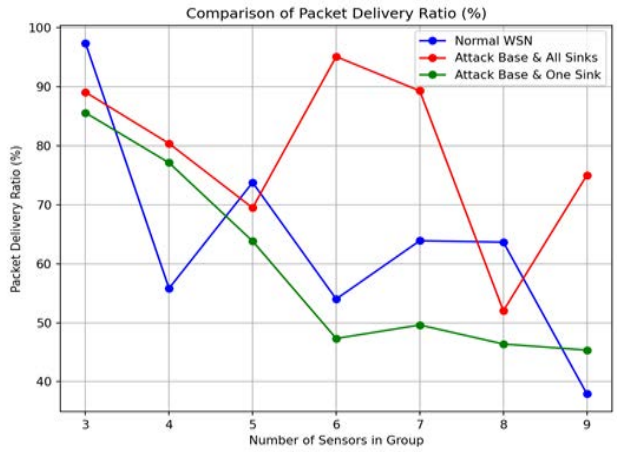


Figure 6: Impact of the attack on packet delivery ratio.

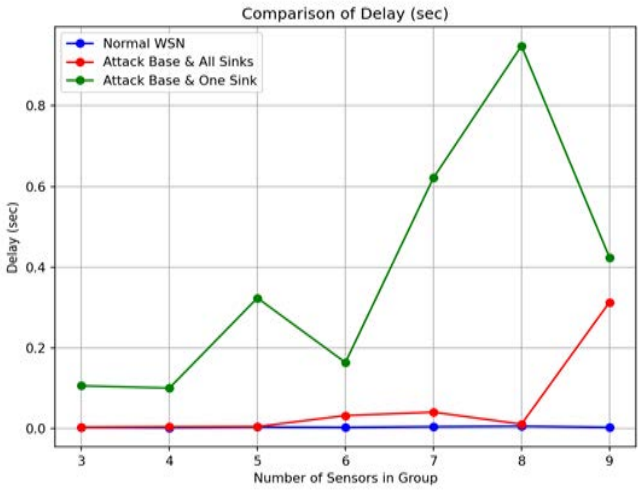


Figure 7: Impact of the attack on packet delivery latency.

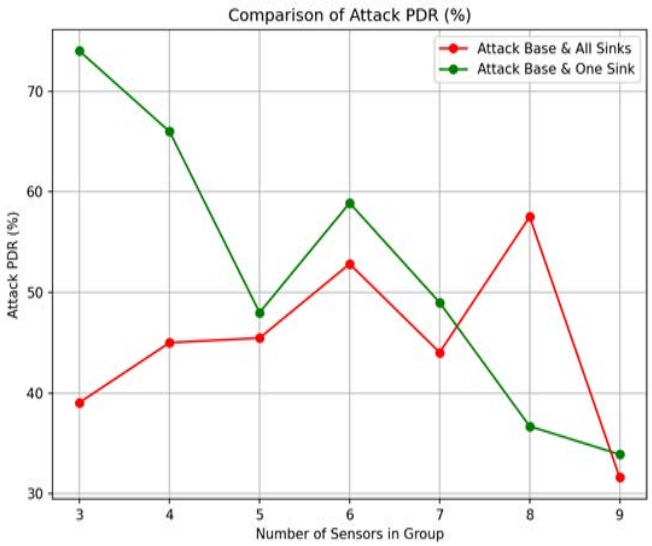


Figure 8:Comparison of packet delivery ratios for the two attack types.

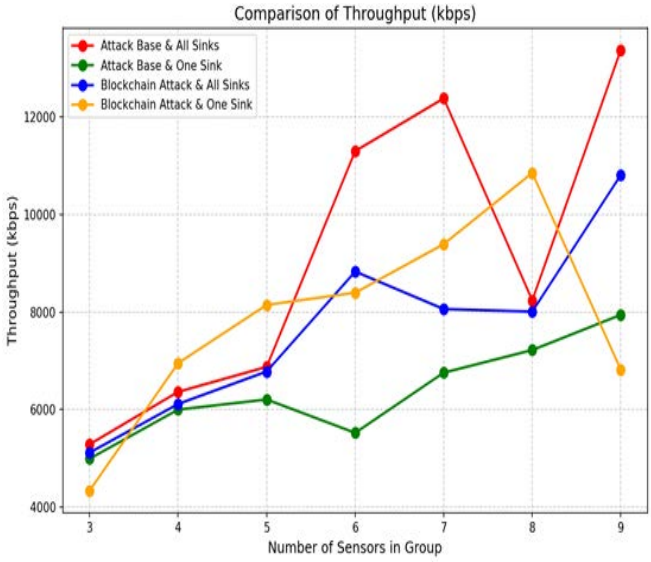


Figure 9: Impact of blockchain on throughput.

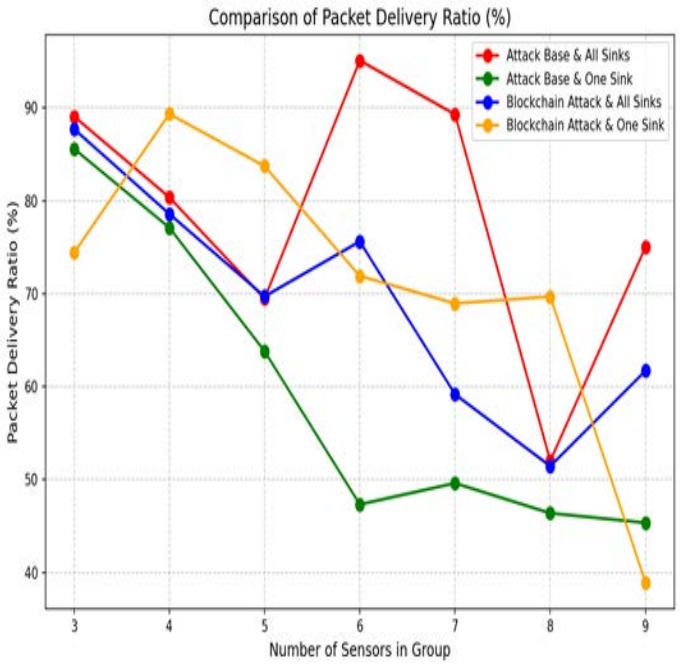


Figure 10: Impact of blockchain on packet delivery ratio.

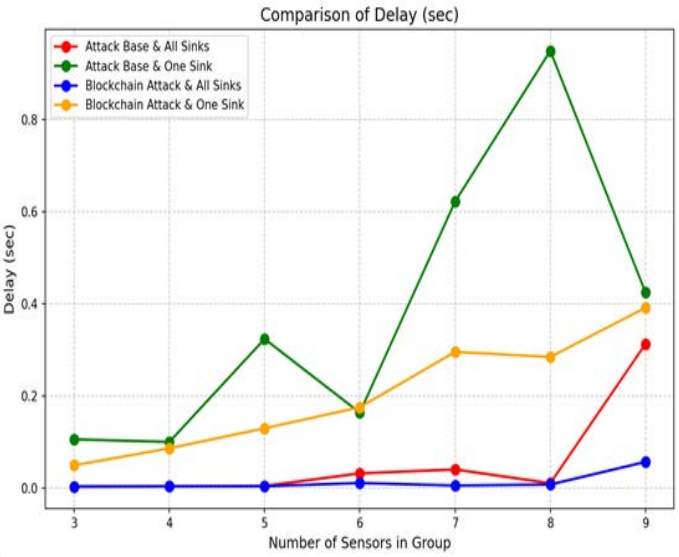


Figure 11: Impact of blockchain on packet delivery delay.

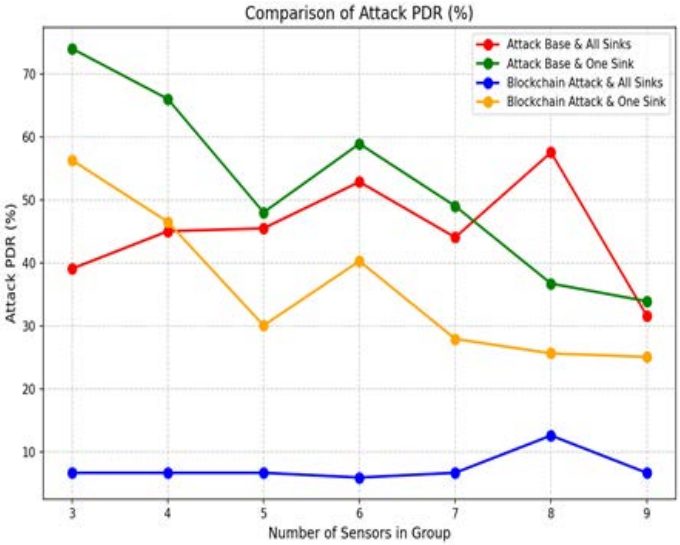


Figure 12: Impact of blockchain on malicious packet delivery.

Impact of SWDK Encryption

Distributed Attack Scenario: The deployment of SWDK encryption under distributed attack conditions resulted in a measurable throughput decline, ranging from 2% to 45.5% with a mean reduction of approximately 21%, relative to the unprotected state. A corresponding deterioration was observed in the Packet Delivery Ratio (PDR), which decreased by 8% to 45.8%. Latency behavior under SWDK was heterogeneous across the network. While most sensor nodes experienced reduced delays, counterintuitive improvements were recorded at specific nodes (3, 4, 5, and 9). Notably, the proportion of illegitimate packets delivered to the network interfaces declined dramatically, falling between 66% and 84%. Focused Attack Scenario: Under the concentrated attack model, the outcomes exhibited greater variability. Throughput improved across the majority of sensor configurations. The Packet Delivery Ratio (PDR) demonstrated significant improvement, with gains ranging from 21% to 44%. A and pronounced latency reduction—between 65% and 90%—was also observed. Moreover, the reception rate of illegitimate packets was curtailed by 19% to 47%. Implementing encryption resulted in an additional operational overhead caused by the increase in packet size following the addition of the encryption key leading to a 21% reduction in throughput compared to the unencrypted scenario.

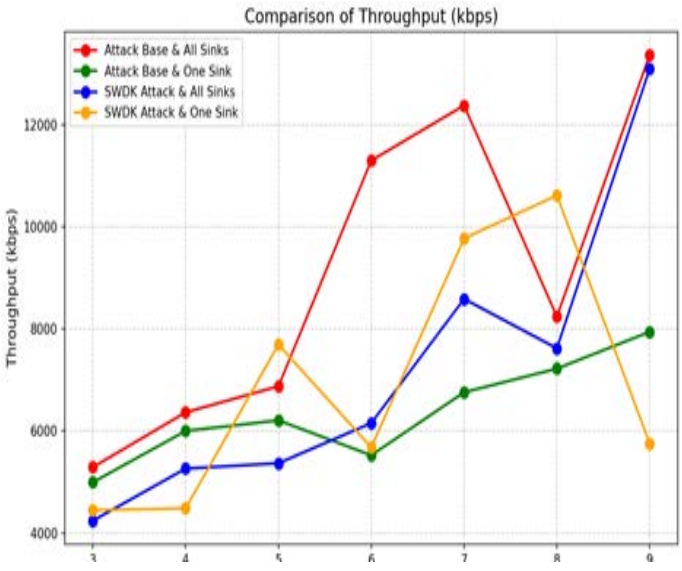


Figure 13: Effect of SWDK on throughput.

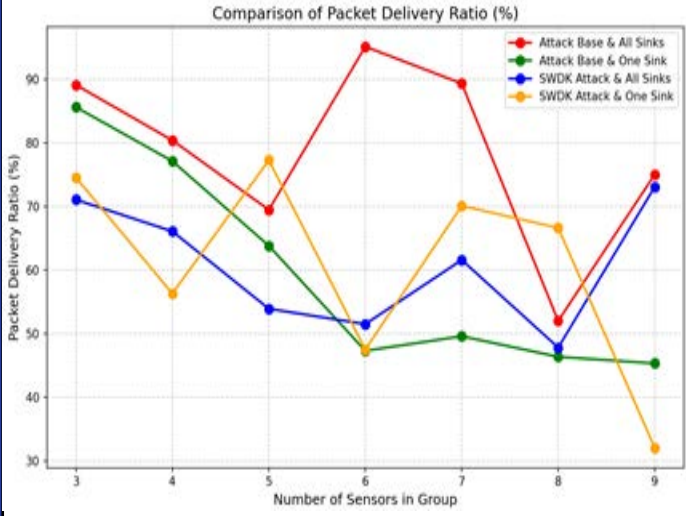


Figure 14: Effect of SWDK on packet delivery ratio.

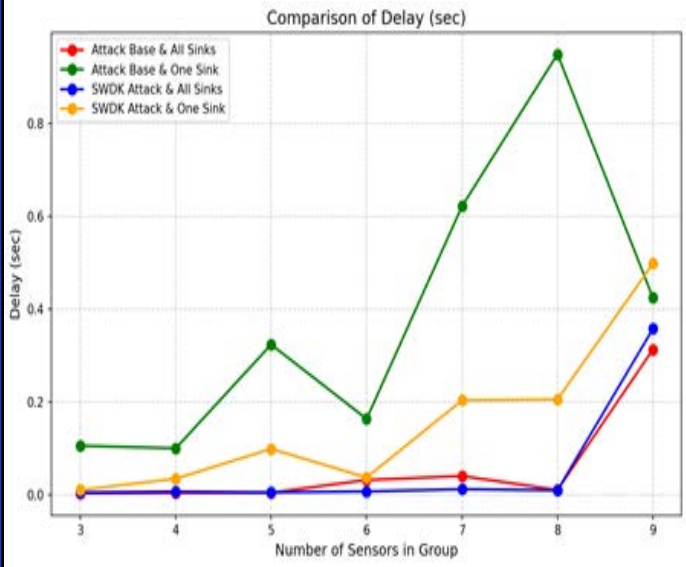


Figure 15: Effect of SWDK on packet delivery delay.

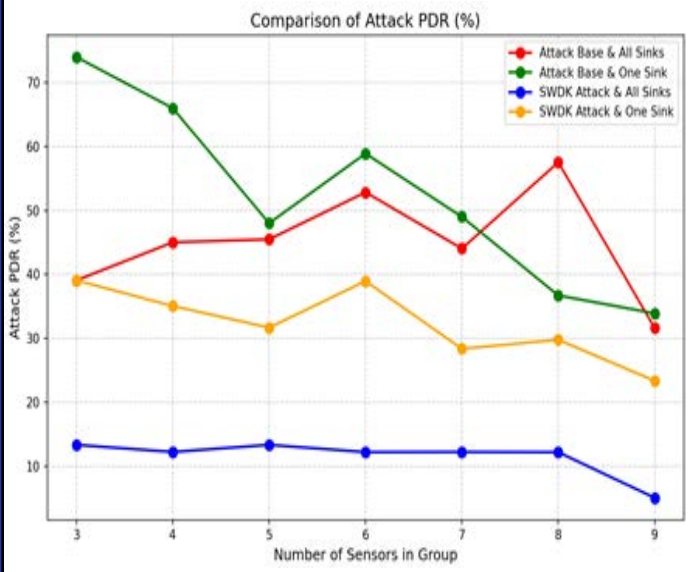


Figure 16: Effect of SWDK on malicious packet delivery.

Impact of the proposed model (Blockchain + PCA + SWDK)

Distributed Attack Scenario: In the context of distributed attacks, the proposed integrated framework fulfilled its core defensive mandate with exceptional efficacy. The overall attack success rate was curtailed by 73% to 85%, as the Attack_PDR declined precipitously from a concerning baseline of 31%–57% to a minimal residual level of actually 6%–14%. This substantial security enhancement was achieved alongside certain operational trade-offs: a mean throughput reduction of approximately 25% and a corresponding decline in the Packet Delivery Ratio (PDR). Latency outcomes exhibited great variability across scenarios; While some configurations recorded notable improvements, others experienced significant increases ranging from 83% to 92%. Focused Attack Scenario: Conversely, under the concentrated attack model, the integrated framework exhibited markedly improved performance. Through registered significant gains, and PDR demonstrated significant improvement with improvements ranging between 12% and 34%. Latency was reduced across all test cases by 55% to 82%, achieving an average latency reduction of approximately 70% in single-sink environments. The reduction in Attack_PDR ranged from 12% to 45% with an average of roughly 33%, as the metric shifted from a baseline of 34%–74% down to 23%–40%.

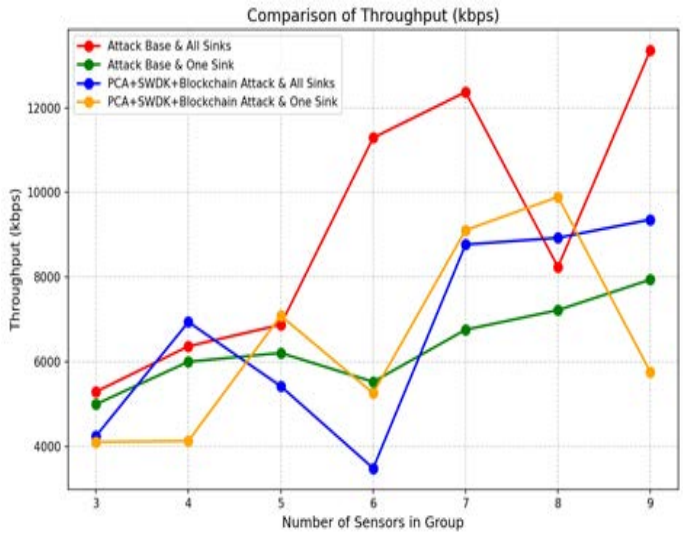


Figure 17: Impact of the proposed model on throughput.



Figure 18: Impact of the proposed model on packet delivery ratio.

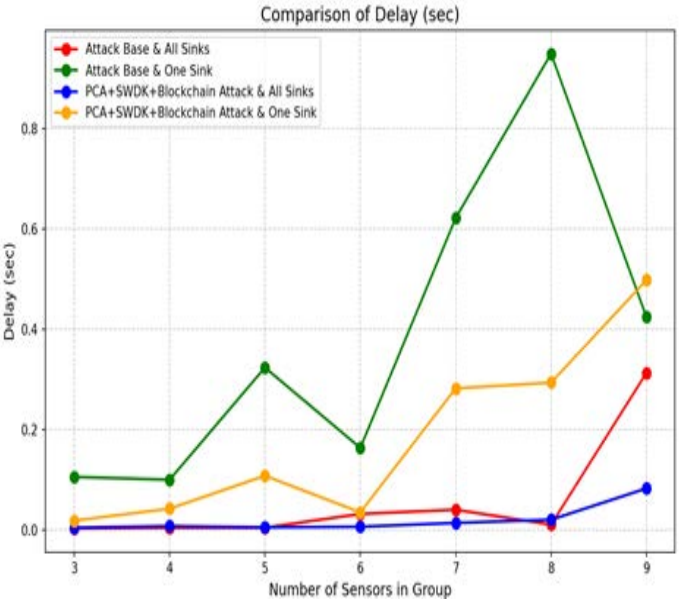


Figure 19: Impact of the proposed model on packet delivery delay.

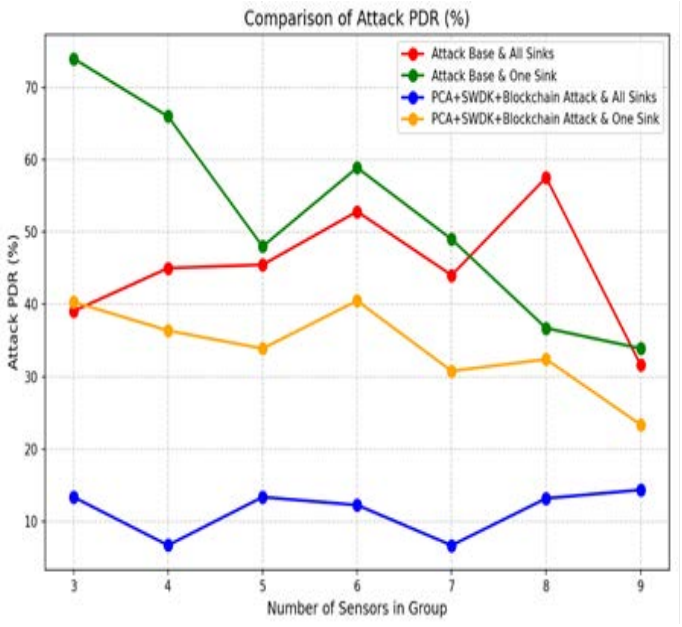


Figure 20: Impact of the proposed model on malicious packet delivery.

DISCUSSION

Blockchain

The blockchain layer, when deployed in isolation, demonstrated decisive performance against distributed attacks (malicious packet delivery <7%, accuracy >93%), which is consistent with the findings of Arachchige et al. [9]. However, it exhibited a notable shortcoming under concentrated attacks targeting a single sink node, permitting 25%–56% of malicious packets to pass through. This contrasts with the uniform traffic scenarios evaluated by Elumalai et al., who reported 97.4% accuracy and 2.1% false positives. Furthermore, the latency observed with blockchain alone reached approximately 0.3 seconds, compared to 61 ms reported by Elumalai et al., a discrepancy attributed to differences in the consensus mechanism (PBFT) and network scale. This observation supports the conclusions of Draz et al. [8] regarding the trade-off between computational cost and enhanced network adaptability.

SWDK

The SWDK encryption, which is derived from the geometry of a spiderweb, achieved a 66%–84% reduction in illegitimate packets, with a concurrent throughput reduction of 21%. This performance is comparable to the dynamic key system proposed by Mohamed Anwar & Pavalarajan [6]. SWDK excels in concentrated attack scenarios, improving the Packet Delivery Ratio (PDR) by 21%–44% and reducing latency by 65%–90%, primarily due to the rapid discarding of malicious packets and the alleviation of queue congestion—a dimension not highlighted by Shakor et al. [7]. Moreover, the unexpected latency reduction observed at specific nodes (3, 4, 5, and 9) is attributable to the prevention of retransmission for damaged packets.

Efficacy of the Integrated Hybrid Model (SWDK + PCA + Blockchain)

The complete model surpassed both its individual components and prior hybrid frameworks, as evidenced by the following: Attack Success Rate: Decreased by 73%–85% in distributed attacks, surpassing the detection rate of Jawahar et al. [5] (94% with 3.2% false positives). Latency Stability: Achieved an average reduction of ~70% in concentrated attacks, outperforming the 297 ms delay reported by Al Ebrahim et al. [2], while simultaneously achieving a lower malicious packet delivery rate (6%–14%). Trade-offs: Throughput decreased by 25%, which is consistent with the computational costs inherent in hybrid solutions such as those presented by Prasad et al. [10] (accuracy of 95.51%). However, the current model provides a more favorable and justifiable balance for sensitive applications, achieving a classification accuracy of 93%. Scalability: Multi-sink distributions demonstrated superior efficienviding quantitative

cy, effectively extending the principles of Rai et al. [11] for mitigating Denial-of-Service (DoS) attacks. The model exhibits performance fluctuations under high-density distributed attacks (latency increase of 83%–92%), attributed to the complex interplay between processing overhead, compression ratios, and queue management. This observation echoes the warnings of Abughazaleh et al. Concerning network saturation. Furthermore, the assumptions of static nodes and open-channel models constrain the generalizability of the results compared to dynamic or mobile environments (Halgamuge & Niyato [12]).

Comparison with Related Studies

In comparison with Reference [6], with which the present study shares a unified architectural philosophy, both approaches employ a Spider-Web-based Dynamic Key (SWDK) for lightweight encryption, Principal Component Analysis (PCA) for data compression, and a blockchain layer for tamper-resistant data aggregation. Both also adopt a two-stage security model—from sensor to aggregator and from aggregator to the base station—along with a category-aware encryption strategy that assigns varying key strengths according to data sensitivity. Nevertheless, the two studies diverge in several fundamental respects. First, whereas Reference [6] focuses primarily on reducing energy consumption, transmission cost, and storage cost within a 50×50 network simulated in MATLAB, the present study targets resilience against Denial-of-Service (DoS) and flooding attacks in an NS-3 environment, using a cluster-based topology comprising one base station and four sink nodes. Second, Reference [6] incorporates a routing mechanism based on neural networks optimized by the Bacterial Foraging Optimization (BFA) algorithm, while the proposed model relies on a private (permissioned) blockchain operating under a Proof-of-Authority (PoA) consensus mechanism, specifically tailored for cluster-based Wireless Sensor Networks (WSNs). Third, and most importantly, Reference [6] does not evaluate its scheme under adversarial attack scenarios, nor does it report attack-related metrics such as the Attack Packet Delivery Ratio (Attack_PDR) or the attack success rate. In contrast, the proposed model demonstrates a 66%–84% reduction in illegitimate packet delivery under distributed attacks, a 73%–85% reduction in the overall attack success rate, and an approximately 70% reduction in average latency under concentrated attacks, albeit accompanied by a trade-off of roughly 25% in throughput. Thus, the SWDK layer in the present study extends the findings of Reference [6] by shifting the evaluation from a purely performance-oriented assessment to a security-oriented, attack-aware one, providing quantitative evidence that SWDK is not only resource-efficient but also effective in countering DoS and flooding threats. Unlike the system in Reference [3]—which relied on a Proof-of-Authority (PoA) framework with a distributed detection mechanism, achieving high detection accuracy (exceeding 97%) and a latency of 61 ms, yet neglecting data compression—the proposed model integrates encryption to ensure security and data confidentiality alongside a compression algorithm to reduce data size prior to transmission. Furthermore, it employs blockchain technology to enhance reliability, mitigate flooding attacks, and detect malicious packets before they reach their targets. Consequently, this tripartite model successfully blocked attack packets at rates ranging from 73% to 85% in distributed attack scenarios (involving the base station and four sinks) and achieved a 70% response rate in concentrated attack scenarios (involving the base station and a single sink), albeit at the expense of a 25% reduction in throughput.

Table 1: Comparative summary of evaluated scenarios.

Scenario	Average Throughput (kbps)	Legitimate PDR (%)	Average Delay (s)	Attack_ PDR (%)	Security-Performance Balance	notice
No Protection (Under Attack)	7374–7741	56.00–66.31	0.0714–0.3934	48.00–60.28	None	—
Blockchain Only	8497–8787	69.50–73.49	0.0386–0.1770	3.79–44.00	Moderate	Best throughput (single-sink attack)
SWDK Encryption Only	7636–8354	64.00–71.85	0.0799–0.2180	8.75–38.50	Good	Balanced performance (single-sink attack)
Proposed Model	7157–7732	61.70–66.36	0.0380–0.1660	9.67–40.70	Excellent	

CONCLUSIONS AND RECOMMENDAIIONS

Based on what this study found, a few core recommendations are suggested to push the integrated framework further, especially in terms of performance, security, and scalability: First, the queue handling at the aggregation node needs to be upgraded by adding smarter scheduling algorithms. Such mechanisms are essential to prevent latency degradtion and even risk a sort of collapse once the active sensors get to around 5 or 6 nodes, which showed up as a critical turning point in the experiments. Second, adaptive compression is worth adopting, meaning the compression factor inside the Principal Component Analysis (PCA) approach should be adjustable on the fly. In practice, the compression ratio should rise automatically as the network gets close to its saturation zone, so congestion is reduced and communication reliability stays intact, even when traffic is heavy. Third, there's a clear point about accepting engineering trade-offs. A small drop in throughput might be allowed on purpose, if that leads to big improvements in detection accuracy and better temporal stability. Still, these trade-offs should be tuned carefully to match the needs of each particular use case, so you end up with a good equilibrium between security and performance, not just one side getting all the attention. Fourth, it is recommended to do periodic renewal of the spiderweb network parameters. This work should be handled through a secure protocol, so the confidentiality stays intact and robustness remains solid for the dynamic keys, and so you can avoid any likely key compromise over longer operational stretches. Finally, future work should concentrate on scaling up and hardware acceleration. It would be good to test the proposed system on larger scale network topologies, and also run it against a wider spectrum of attack types. In addition, doing a hardware level implementation of the SWDK encryption layer is encouraged, to sharply cut down computational overhead, which in turn should improve overall efficiency and responsiveness when you deploy it in real-world situations.

REFERENCES

1. Demirci S, et al. Depth-First Search-Based Malicious Node Detection with Honeypot Technology in Wireless Sensor Networks. Mathematics. 2026;14(6):1050.

2. Al Ebrahim A, Badawieh A. Machine Learning-Based Detection and Prevention of DoS and DDoS Attacks in SDWSN. Iraqi Journal for Electrical and Electronic Engineering. 2026;22(1):114-127.

3. Vailshery LS. Global IoT and Non-IoT connections 2010–2025 [Internet]. Statista; 2021 [cited 2026 Aug 23]. Available from: https://www.statista.com/statistics/1101442/iot-number-of-connected-devicesworldwide/.

4. Elumalai G, et al. A blockchain-driven intrusion detection model for secure communication in IoT-WSN mesh architectures. Peer-to-Peer Networking and Applications. 2026;19(2):60.

5. Jawahar A, et al. DDoS mitigation using blockchain and machine learning techniques. Multimedia Tools and Applications. 2024;83(21):60265-60278.

6. Mohamed Anwar A, Pavalarajan S. Spider web-based dynamic key for secured transmission and data-aware blockchain encryption for the Internet of things. IETE Journal of Research. 2024;70(1):499-514.

7. Shakor MY, et al. Dynamic AES encryption and blockchain key management: a novel solution for cloud data security. IEEE Access. 2024; 12:26334-26343

8. Draz U, et al. Decentralized energy swapping for sustainable wireless sensor networks using blockchain technology. Mathematics. 2025;13(3):395.

9. Arachchige KG, et al. Blockchain-enabled mitigation strategies for distributed denial of service attacks in IoT sensor networks: An experimental approach. Computers, Materials & Continua. 2024;81(3):3679-3705.

10. Prasad A, Arockiasamy K, Gunasekaran K. A hybrid metaheuristic algorithm with machine learning for detecting denial-of-service attacks in wireless sensor networks. Frontiers in Artificial Intelligence. 2026; 9:1738152.

11. Rai HM, et al. Enhancing data security and privacy in energy applications: Integrating IoT and blockchain technologies. Heliyon. 2024;10(19): e38100.

12. Halgamuge MN, Niyato D. Adaptive edge security framework for dynamic IoT security policies in diverse environments. Computers & Security. 2025; 148:104128.

13. Saheed YK, Abdulganiyu OH, Ait Tchakoucht T. Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the internet of things networks with edge capabilities. Applied Soft Computing. 2024; 155:111434.

14. Abughazaleh N, et al. DoS attacks in IoT systems and proposed solutions. International Journal of Computer Applications. 2020;176(33):16-19.

15. Ren B, et al. A multiagent deep reinforcement learning autonomous security management approach for Internet of Things. IEEE Internet of Things Journal. 2024;11(15):25600-25612.

Author contributions

Conceptualization: Razan Rezk

Methodology: Razan Rezk

Investigation: Razan Rezk – Dr. Wasim Al-Juneidi

Project Administration: Razan Rezk

Supervision: Dr. Wasim Al-Juneidi

Writing – Original Draft: Razan Rezk

Writing – Review & Editing: Razan Rezk – Dr. Wasim Al-Juneidi

Competing interests: No conflict of interest exists.

Data and materials availability: All data are available within the main text and supplementary materials

Abbreviations

abbreviation	Expression
IoT	Internet of Things
WSN	Wireless Sensor Network
WSNs	Wireless Sensor Networks
DoS/DDoS	Denial of Service / Distributed Denial of Service
SWDK	Spider Web Dynamic Key
SWC	Spider Web Construction
PCA	Principal Component Analysis
PDR	Packet Delivery Ratio
Legitimate PDR	Legitimate Packet Delivery Ratio
Attack PDR	Attack Packet Delivery Ratio
NS-3 / ns3	Network Simulator 3
QoS	Quality of Service
SDWSN	Software-Defined Wireless Sensor Network
TDFS	Tree-based Depth-First Search
ML	Machine Learning
GOA	Grasshopper Optimization Algorithm
GA	Genetic Algorithm
MLP	Multilayer Perceptron
KNN	K-Nearest Neighbors
CART	Classification and Regression Trees
LR	Logistic Regression
GRU	Gated Recurrent Unit
PBFT	Practical Byzantine Fault Tolerance
SVM	Support Vector Machine
AES	Advanced Encryption Standard
ECC	Elliptic Curve Cryptography
PoS	Proof-of-Stake

Conversational Explanations for Bayesian Networks: An Interactive LLM-Driven Framework for Explainable AI

التفسيرات الحوارية للشبكات البايزية: إطار عمل تفاعلي مدعوم بنماذج اللغة الكبيرة للذكاء الاصطناعي القابل للتفسير

Lamis Olaiky :Syrian Virtual University, Damascus-Syria.
Corresponding: lamis.olaiky@gmail.com
Submitted date: 27 August 2026 - **Accepted date:** 29 October 2026.

ABSTRACT

Explainable Artificial Intelligence (XAI) is important for improving the interpretability of probabilistic models in decision-support applications. Although Bayesian Networks (BNs) provide probabilistic reasoning through Directed Acyclic Graphs (DAGs) and Conditional Probability Tables (CPTs), their numerical outputs can be difficult for non-technical decision-makers to interpret. Traditional explanation approaches may provide feature importance or attribution measures but do not necessarily provide contextual, natural-language explanations of probabilistic outcomes. To address this gap, this paper introduces an architectural framework that combines Bayesian Networks with Large Language Models (LLMs) to generate contextual and interactive explanations. The framework was evaluated on 300 held-out credit scoring cases, representing 30% of the dataset, after training the Bayesian Network on the remaining 70%. The Bayesian Network achieved an accuracy of 82.67%, precision of 81.04%, recall of 99.54%, and F1-score of 89.34%. The generated explanations achieved average ROUGE-1, ROUGE-2, ROUGE-L, and BLEU scores of 0.4499, 0.1685, 0.3338, and 0.0417, respectively. An LLM-as-a-Judge evaluation using Qwen 2.5 yielded an overall mean quality score of 3.11/5.00 (Faithfulness: 2.98, Clarity: 3.92, Completeness: 2.42), with an average LLM latency of 15.75 seconds per case. These results demonstrate the potential of combining probabilistic inference with constrained natural-language generation to improve the accessibility and interpretability of Bayesian network-based decision support.
Keywords: Explainable AI (XAI), Bayesian Networks, Large Language Models (LLMs), Natural Language Explanation, Decision Support Systems, Interactive Explanation

الملخص

يُعد الذكاء الاصطناعي القابل للتفسير (XAI) مهماً لتحسين قابلية تفسير النماذج الاحتمالية في تطبيقات دعم اتخاذ القرار. وعلى الرغم من أن الشبكات البايزية (BNs) توفر استدلالاً احتمالياً من خلال الرسوم البيانية الموجهة الخالية من الحلقات (DAGs) وجداول الاحتمالات الشرطية (CPTs)، فإن مخرجاتها الرقمية قد تكون صعبة التفسير بالنسبة لمتخذي القرار غير المختصين. وقد توفر أساليب التفسير التقليدية مقاييس لأهمية الخصائص أو مساهمتها، لكنها لا تقدم بالضرورة تفسيرات سياقية باللغة الطبيعية للمخرجات الاحتمالية. ولردم هذه الفجوة، تقدم هذه الورقة إطاراً معمارياً يجمع بين الشبكات البايزية ونماذج اللغة الكبيرة (LLMs) لتوليد تفسيرات سياقية وتفاعلية. يستخرج الإطار الأدلة الاحتمالية والعلاقات ذات الصلة من الشبكة البايزية ويمررها إلى نموذج اللغة من خلال قوالب أوامر منظمة ومقيدة، مما يتيح توليد تفسيرات باللغة الطبيعية تستند إلى نتائج الاستدلال البايزي. وعند تقييم الإطار ضمن حالة دراسية لتقييم الجدارة الائتمانية تشمل 300 حالة اختبار، حققت الشبكة البايزية دقة تصنيف بلغت 82.67% (بمقياس F1 بلغ 89.34%)، وبلغ متوسط مقياس ROUGE-1 مقدار 0.4499، إضافة إلى درجة جودة كلية بلغت 3.11 / 5.00 من خلال التقييم باستخدام نموذج لغوي كحَكَم (Qwen 2.5 LLM-as-a-Judge) وبمتوسط زمن استجابة بلغ 15.75 ثانية لكل حالة. وتوضح هذه النتائج إمكانية الجمع بين الاستدلال الاحتمالي وتوليد اللغة الطبيعية المقيد لتحسين سهولة الوصول إلى تفسيرات قرارات الشبكات البايزية وقابليتها للفهم
الكلمات المفتاحية: الذكاء الاصطناعي القابل للتفسير، الشبكات البايزية، نماذج اللغات الكبيرة، التفسير باللغة الطبيعية، أنظمة دعم القرار، التفسير التفاعلي

INTRODUCTION

Artificial intelligence (AI) models are increasingly used in high-stakes decision-making environments, including financial risk assessment and enterprise management [11, 22]. Bayesian Networks (BNs) provide a formal framework for probabilistic reasoning in these domains by representing variables and their conditional dependencies through Directed Acyclic Graphs (DAGs) and Conditional Probability Tables (CPTs) [10, 31, 35]. However, interpreting BN structures and probabilistic outputs can be challenging for users without specialized knowledge of probabilistic modeling. Although methods for ex-

plaining Bayesian network inference have been developed [17, 20, 36], translating model outputs into accessible explanations that support interactive exploration remains an important challenge. This challenge is particularly relevant in high-stakes settings, where users need to understand the basis and limitations of model predictions to support informed decision-making [15, 23, 33]. To address this challenge, this study investigates a conversational Explainable AI (XAI) framework that combines Bayesian probabilistic inference with large language model (LLM)-based explanation generation [18, 45]. In this architecture, the BN performs the probabilistic inference, while the LLM generates natural-language explanations based on structured applicant information, BN outputs, and instructions governing the interpretation of the network's structure. The framework also supports follow-up questions, allowing users to explore individual credit-assessment cases through conversational interaction [13, 29, 37]. The study focuses on the consistency of generated explanations with the underlying probabilistic results and network structure, as well as the limitations of using prompt constraints to guide the language model. Llama 3.2 is used as the language-generation component in the implemented system; the study does not assume that the proposed framework depends on this specific model. To guide this investigation, the study addresses the following research questions:

•RQ1: How can Bayesian Network (BN) inference outputs and structural information be used to generate natural-language explanations through a large language model?

•RQ2: To what extent are the generated explanations consistent with the numerical outputs and specified structure of the underlying Bayesian Network?

•RQ3: What limitations arise when using a large language model to provide interactive explanations and recommendations for individual credit-assessment cases?

The study investigates these questions through the implementation and evaluation of a conversational framework that separates probabilistic inference from natural-language generation. The evaluation examines the consistency of generated explanations with the BN outputs and structure, while the qualitative analysis of case scenarios explores the behavior and limitations of the conversational interface

Related Work and Background

Bayesian Networks and Interpretability Challenges:

Probabilistic graphical models, particularly Bayesian Networks (BNs), were established by Pearl [31] to model causal relationships and conditional dependencies under uncertainty using Directed Acyclic Graphs (DAGs). Subsequent surveys by Heckerman et al. [14] highlighted the efficacy of BNs across medical diagnosis, financial risk, engineering, and robotics. However, BNs suffer from exponential computational complexity and severe interpretability bottlenecks [2, 14]. Barber [2] attributed this opacity to derived variables and noisy Conditional Probability Tables (CPTs). To systematically address this cognitive load, Lacave and Díez [20] proposed a framework categorizing BN explanations by granularity (micro vs. macro), reasoning nature (causal vs. non-causal), and user personalization, demonstrating that Maximum A Posteriori (MAP) inference yields more concise human-readable explanations than full state representations. Early natural language and narrative attempts focused either on logical argumentation via support graphs [11] or scenario-based explanations in legal domain applications [40]. Nevertheless, these classical approaches remained constrained by rigid rule-based templates, high computational overhead, and limited adaptability to quantitative financial domains [40].

Explainable AI (XAI) Frameworks and Post-Hoc Interpretability:

To formalize model interpretability, Doshi-Velez and Kim [8] established three evaluation tiers: application-grounded, human-grounded, and functionally-grounded metrics. Lipton [23] delineated the philosophical dimensions of XAI, distinguishing intrinsic transparency (simulatability, decomposability, algorithmic transparency) from post-hoc explanations (textual, visual, local examples), noting that local post-hoc explanations are often the most practical solution for complex decision systems under regulatory mandates [41].

While post-hoc methods aim to illuminate deep learning models [34], they face critical structural limitations, including explanation inconsistency [19, 34], a lack of standardized evaluation benchmarks [30, 34], and the inherent trade-off between predictive accuracy and interpretability [33]. These challenges spurred a shift toward leveraging natural language generation and Large Language Models (LLMs) to communicate model behavior interactively [5, 37].

Conversational AI and LLM-Grounded Explanation Generation:

The emergence of generative AI and LLM-driven interfaces transformed data interaction [13, 46] by utilizing prompt engineering and Chain-of-Thought (CoT) reasoning to decompose complex inference into human-understandable steps [42, 46]. However, ungrounded LLMs remain prone to hallucinations

CC BY

and statistical inaccuracies [13, 16, 22], especially when decoupled from exact calculation engines [13, 42]. Specialized conversational agents have been developed to explain black-box models, such as Deep Reinforcement Learning (DRL) in healthcare and finance [24, 28], though they struggle to balance scientific rigor with user accessibility.

To improve grounding, structured text-generation approaches have emerged at the intersection of XAI and LLMs

•BN-to-Text Baselines: Keppens [17] introduced pre-LLM Natural Language Generation (NLG) for BN query results via support graphs. Sevilla et al. [36] advanced this by extracting factor arguments (sub-graph reasoning paths) into natural language, validated through human user studies in medicine.

•LLM-XAI Integration: Recent advances in explainable AI highlight the potential of leveraging LLMs to naturalize complex decision outputs, though unconstrained models remain prone to generating plausible yet unfaithful narratives [26]. To address this in financial decision-making, Dyomin et al. [9] combined traditional scoring classifiers with feature attribution tools, using an LLM strictly as a grounded translator for model-derived risk indicators.

Research Gap and Main Contributions

Previous research has addressed the generation of natural-language explanations for Bayesian Network (BN) inference through approaches that explicitly represent probabilistic reasoning. Keppens [17], for example, proposed a natural-language generation approach based on qualified support graphs to identify and communicate how observations affect a query result within a BN. Sevilla et al. [36] developed an explanation approach based on factor arguments, representing the flow of evidence toward a target variable, and evaluated its usefulness in the medical domain. These studies demonstrate that structured, model-grounded explanations of Bayesian reasoning can be generated without relying on an LLM to perform the underlying probabilistic inference. However, their primary focus is on constructing explanations of BN reasoning, rather than on the particular combination of a separate BN inference component, an LLM-based conversational interface, and an evaluation of the numerical consistency of the resulting dialogue.

A different line of research uses LLMs to communicate predictions and explanation evidence from machine-learning models. For example, Dyomin et al. [9] combine Logistic Regression and LightGBM with SHAP-based feature attribution and an LLM that translates feature contributions into natural-language explanations. This approach addresses the communication of feature-attribution results in credit-risk assessment, whereas the present study uses a BN to represent conditional probabilistic dependencies and obtain the credit-approval prediction. These approaches therefore differ in their underlying reasoning and the type of model information supplied to the language-generation component. More broadly, the use of LLMs as explanation interfaces raises the need to evaluate whether generated narratives preserve the supplied numerical results and model relationships rather than introducing unsupported claims.

Against this background, this study investigates a hybrid approach that combines BN-based probabilistic inference with LLM-generated conversational explanations for credit-assessment cases. Its focus is not on claiming that either BN explanation or LLM-based explanation is new in itself, but on implementing and evaluating their integration within a single decision-support interface, with explicit prompting constraints and quantitative and qualitative assessment of generated explanations.

The main contributions of this research are as follows:

1.A Decoupled Hybrid XAI Architecture:

BN performs probabilistic inference independently of the LLM, which generates natural-language explanations from structured applicant information, BN outputs, and explicit instructions governing the interpretation of the network's structure

2.A Constraint-Guided Prompting Protocol:

Development and evaluation of a prompt design intended to guide the local LLM in preserving the numerical results and specified relationships of the BN. The system achieved a Numerical Faithfulness Rate of 91.33% under the evaluation procedure used in this study. This result measures numerical consistency according to the adopted metric; it does not establish the elimination of hallucinations in general.

3.A Multi-Metric Evaluation Pipeline:

Evaluation of generated explanations using lexical-overlap metrics, BERTScore, numerical faithfulness, decision alignment, and LLM-based judging with Qwen 2.5. The pipeline compares generated explanations with manually constructed reference texts and assesses complementary aspects of explanation quality.

4.An Interactive Conversational Decision-Support Implementation:

Implementation of an interface that maintains applicant inputs and BN evaluation results across follow-up questions, allowing users to explore expla-

nations of individual credit-assessment cases. The implementation was evaluated using 300 test cases.

MATERIALS AND METHODS

Framework Overview and Strategic Objectives

The proposed framework establishes a structured, three-tiered architecture that integrates Bayesian Network (BN)-based probabilistic reasoning with Large Language Model (LLM)-based natural language generation for decision-support applications. Rather than relying on the LLM as an autonomous decision-maker, the architecture strictly separates the quantitative inference process from the narrative explanation layer, positioning the Bayesian Network as the primary computational engine. As illustrated in Figure 1, the operational pipeline spans three sequential functional layers: ***Data Ingestion Layer:*** Preprocesses, discretizes, and standardizes raw applicant attributes into discrete evidence states suitable for probabilistic propagation. ***Probabilistic Inference Layer:*** Serves as the Bayesian core, responsible for factorizing the directed acyclic graph (DAG), updating conditional probability tables (CPTs), and computing exact posterior marginals for the target outcome (e.g., loan approval status). ***Explanation and Natural Language Layer:*** Maps Maps structured probabilistic outputs and intermediate variable dependencies into an accessible, interactive narrative using an LLM, with prompt constraints intended to support consistency with the underlying BN structure and inference results

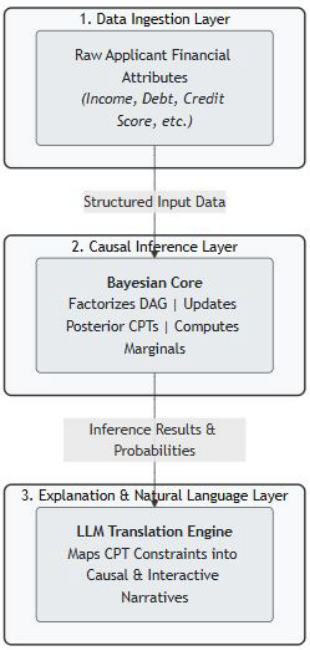


Figure 1System Architecture Diagram

Empirical Case Study: Bank Loan Approval Application

To validate and evaluate the practical implementation of this architecture, a bank loan approval decision system was adopted as an empirical case study. Loan underwriting presents an ideal domain due to its strict regulatory requirements, the necessity for auditability, and the presence of complex conditional dependencies among financial metrics. Although the framework is evaluated here in a credit-assessment context, its general architecture could be investigated in other domains, such as medical diagnostics and insurance underwriting. However, its suitability for these applications remains to be evaluated. The framework was evaluated using a publicly available credit risk dataset comprising 1,000 loan applicant records with financial and demographic attributes, where the target variable is binary loan approval (Loan). To suit the discrete requirements of Bayesian Networks, continuous features (e.g., annual income and credit score) were discretized into uniform state brackets (e.g., low, medium, high, very high) using quantile-based discretization (pd.qcut). Categorical variables were label-encoded to preserve structural dependencies while mitigating state-space explosion. The loan evaluation model uses a three-tiered hierarchical BN structure developed through a combination of data-driven feature selection and domain-informed topological constraints: ***Primary Evidence Nodes (Layer 1):***

Observed input attributes including annual_income, credit_score, existing_loans_count, employment_status, and num_dependents

Intermediate Derived Nodes (Layer 2):

Domain-expert intermediate variables—specifically financial_capacity (derived from income, employment, and dependents) and credit_worthiness (derived from credit score and active loans). These intermediate nodes act as structural abstraction layers, capturing higher-level financial concepts prior to final risk evaluation.

Target Decision Outcome Node (Layer 3):

The final classification node representing Loan Approval Status (Loan). Probabilistic Inference Engine: Bayesian Network Architecture and Pipeline The final classification node representing Loan Approval Status (Loan). **Probabilistic Inference Engine: Bayesian Network Architecture and Pipeline**

Data Preparation and Feature Discretization

This layer serves as the primary entry interface for applicant profiles. It standardizes raw continuous and categorical financial parameters-including income level, credit score, active debt burden, number of dependents, and employment status-into discrete evidence states suitable for probabilistic propagation. Continuous numerical variables (e.g.,

CC BY

credit_score and annual_income) are discretized into four uniform quantile-based state brackets using pd.qcut (e.g., categorizing credit scores into discrete evaluation tiers) to facilitate exact case classification and Bayesian Network construction. Categorical variables and state labels are encoded to preserve structural dependencies while assigning human-readable representations that support seamless integration with the subsequent natural-language explanation layer.

Domain-Engineered Intermediate Variable Construction

To incorporate domain knowledge into the Bayesian Network architecture, two derived intermediate nodes, financial_capacity and credit_worthiness, were constructed using predefined rules based on observed applicant attributes. Unlike statistically latent variables, these nodes are explicitly computed from available input features. They are subsequently incorporated into the Bayesian Network as intermediate variables, providing a hierarchical representation of financial capacity and creditworthiness within the probabilistic model. The financial_capacity node is derived from annual_income, employment_status, and num_dependents, whereas the credit_worthiness node is derived from credit_score and existing_loans_count.

Financial Capacity (Pillar_1):

Synthesized from primary evidence attributes including annual_income, employment_status, and num_dependents

$$Fin\ Capacity\ Score = Income\ bin + Employment\ bin - \mathbb{I}(Dependents\ bin \geq 2)$$

$$financial\ capacity\ score = \begin{cases} 0\ (Low\ Capacity), if\ fin\ score \leq 0 \\ 1\ (Medium\ Capacity), \quad if\ 0 < fin\ score \leq 2 \\ 2\ (High\ Capacity), \quad if\ fin\ score > 2 \end{cases}$$

1. **Credit Worthiness ($Pillar_2$):** Synthesized from `credit_score` and `existing_loans_count`.

$$Credit\ Worthiness\ Score = credit\ score\ bin - existing\ loans\ bin$$

$$credit\ worthiness = \begin{cases} 0\ (Poor\ Worthiness), if\ cred\ score \leq 0 \\ 1(Fair\ Worthiness), if\ cred\ score = 1 \\ 2(Excellent\ Worthiness), if\ cred\ score \geq 2 \end{cases}$$

Dependents Penalty Threshold Logic

The dependents condition I(dep_bin ≥ 2) enforces a symmetrical bipartition of the discretized state space {0, 1, 2, 3}. Partitioning the set into Low Family Burden ({0, 1}) and High Family Burden ({2, 3}) maintains statistical class balance while capturing a recognized financial inflection point. In retail underwriting, one dependent falls within baseline cost-of-living expectations, whereas two or more dependents significantly increase fixed household liabilities, directly impairing disposable cash flow and Debt-Service Coverage Ratios (DSCR).

Hierarchical Topology and Hybrid Structure Learning

The statistical foundation consists of a Directed Acyclic Graph G = (V, E), where vertices V represent domain variables and edges E denote direct conditional dependencies. The joint probability distribution over V factorizes according to the chain rule for Bayesian Networks:

$$P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P\left(V_i \mid \text{Parents}(V_i)\right)$$

Hybrid Feature Selection

To prevent exponential state-space explosion in Conditional Probability Tables (CPTs) and preserve exact inference tractability, a hybrid feature selection mechanism was executed prior to topology construction. Feature importance was quantified using a weighted composite score: Mutual Information (٪۷۰ weight) measured direct entropy reduction relative to the binary target variable (loan_approved), aligning with the information-theoretic foundations of Bayesian joint distributions. Decision Tree importance (٪۳۰ weight) evaluated non-linear conditional interactions via Gini impurity reduction. Based on this composite ranking, the top five primary predictor features were retained, while secondary and demographic attributes (age, loan_amount, gender) exhibited negligible

explanatory power and were pruned (Table 1).

$$Score = 0.70 \times Mutual\ Information + 0.30 \times Decision\ Tree\ Importance$$

Table-1 - Hybrid Feature Selection and Importance Rankings

	tree_ importance	mutual_ information	final_ score
num_ dependents	0.2347	0.0915	0.1344
credit_score	0.2369	0.0725	0.1218
existing_ loans_count	0.2238	0.0705	0.1165
employment_ status	0.2353	0.0573	0.1107
annual_ income	0.0666	0.0209	0.0346
marital_ status	0.0027	0.0020	0.0022
age	0.0000	0.0002	0.0002
loan_amount	0.0000	0.0002	0.0001
gender	0.0000	0.0002	0.0001

Structural Optimization

Purely data-driven algorithms (e.g., Hill-Climb Search with K2 score) led to spurious correlations and high parent cardinality. To preserve exact inference tractability, a three-tiered hierarchical DAG was established, where primary inputs (Tier_1) flow exclusively into the derived pillars(Tier_2), which in turn dictate the target decision outcome (loan_approved,(Tier_3))

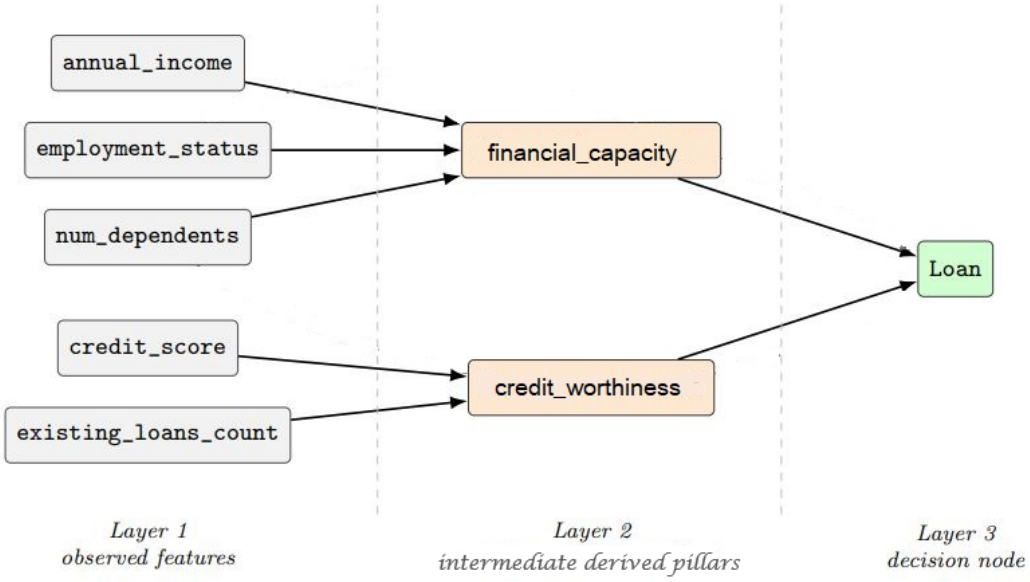


fig 2. Bayesian Network architecture

Parameter Learning and CPT Smoothing

To prevent data leakage, the feature discretization sequence (such as binning continuous variables using quantile-based boundaries via pd.qcut) and structural parameter estimation for the Conditional Probability Tables (CPTs) were executed exclusively on the 70% training partition (700 records). The resulting discretization boundaries and learned parameters were subsequently locked and applied to the independent 30% test partition (300 records). Parameter learning used a Bayesian Estimator with a Dirichlet uniform prior (BDeu) and an Equivalent Sample Size (ESS = 10) to handle sparse parent configurations. Network structural integrity and conditional sum constraints

were formally validated using check_model(). The compiled network artifact was then serialized using joblib for zero-overhead runtime loading.

$$\sum_k P(V_i = k | Parents(V_i)) = 1.0$$

Representative Conditional Probability Tables (CPTs)

To demonstrate the mathematical integrity of the parameter learning phase, representative excerpts of the learned CPTs for both intermediate mediators and the final target decision node are detailed below

Target Decision Node (loan): As shown in Table 2, the decision logic exhibits strong convergence. When both intermediate mediators (financial_capacity and credit_worthiness) are at optimal levels (High), the probability of loan approval reaches 0.9880. Conversely, critical vulnerabilities in creditworthiness (Low) shift the outcome towards rejection ($P \geq 0.50$).

Table2- Sample CPT Excerpt for Target Decision Node

Target Outcome (Loan)	Financial capacity	Credit worthiness	Conditional Probability
Rejected (0)	Low (0)	Low (0)	0.9954
Rejected (0)	Medium (1)	Low (0)	0.5
Rejected (0)	High (2)	Low (0)	0.5
Rejected (0)	Low (0)	Medium (1)	0.6239
Rejected (0)	High (2)	Medium (1)	0.0023
Approved (1)	Low (0)	Low (0)	0.0046
Approved (1)	High (2)	Low (0)	0.5
Approved (1)	Medium (1)	Low (0)	0.586
Approved (1)	Medium (1)	Medium (1)	0.9982
Approved (1)	High (2)	High (2)	0.988

Intermediate Mediators (credit_worthiness & financial_capacity):

Tables 3 and 4 illustrate how primary evidence features interact to define intermediate risk states. The BDeu smoothing estimator prevents deterministic zero-probabilities (e.g., maintaining residual probabilities such as 0.003 or 0.004), guaranteeing model adaptability during runtime inference.

Table 3 Sample CPT Excerpt for credit_worthiness Node

Derived State (credit_ worthiness)	Credit Score	Active Loans (existing_ loans_ count)	Conditional Probability
Low (0)	Low (Bin 0)	0	0.9931
Low (0)	Very High (Bin 3)	3	0.0043
Medium (1)	Low (Bin 0)	1	0.0028
Medium (1)	Medium (Bin 1)	0	0.9925
Medium (1)	Medium (Bin 1)	3	0.0029
High (2)	Low (Bin 0)	0	0.0034
High (2)	Medium (Bin 1)	1	0.003
High (2)	Very High (Bin 3)	0	0.9925
High (2)	Very High (Bin 3)	4	0.0025

Table4 - Sample CPT Excerpt for financial_capacity Node

Derived State (financial_ capacity)	Annual Income	Employment Status	Dependents (num_ dependents)	Probability
Low (0)	Low (Bin 0)	Employed	0	0.9939
Low (0)	Low (Bin 0)	Employed	3	0.9931
Low (0)	Medium (Bin 1)	Employed	0	0.0042
Low (0)	High (Bin 2)	Unemployed	1	0.0028
Low (0)	Very High (Bin 3)	Employed	1	0.0042
Medium (1)	Low (Bin 0)	Unemployed	1	0.0032
Medium (1)	Medium (Bin 1)	Employed	3	0.9927
High (2)	Medium (Bin 1)	Employed	3	0.0037
High (2)	High (Bin 2)	Employed	1	0.9909
High (2)	Very High (Bin 3)	Unemployed	1	0.9945

Operational Runtime Inference Pipeline

To bridge offline parameter learning with live execution, the system operationalizes inference through a two-stage pipeline:Offline Parameter Estimation: Prior to application runtime, deterministic aggregation rules are applied across the historical dataset (N = 1,000) to generate the intermediate pillar states, after which conditional probability distributions (CPTs) are estimated and serialized.

Online Runtime Execution (evaluate.py): During live operation, when raw applicant attributes are submitted, the application layer dynamically computes the exact pillar states. These computed pillar values are combined with the primary discretized inputs into a unified evidence dictionary E, which is injected into the VariableElimination inference engine to calculate marginal posterior probabilities.

Model Validation Setup and Evaluation Metrics

To evaluate the predictive performance of the Bayesian Network prior to integrating the LLM explanation layer, the dataset (N = 1,000) was split into a 70% training set (700 records) and a 30% test set (300 records).

The network parameters were learned using the training subset, after which the compiled model was executed on the remaining 300 test cases to evaluate its overall classification performance. Model performance is quantified using standard classification metrics: Accuracy, Precision, Recall, and F_1Score. Evaluating these probabilistic outputs verifies the network›s classification consistency before passing its inference outputs to the downstream explanation engine. (Quantitative results are presented in next Section)

Interactive Explanation Engine: LLM Integration and Conversational Framework

To translate probabilistic inference outcomes into accessible natural-language explanations while avoiding dependence on cloud-based language-model services and using prompt constraints to reduce unsupported claims [16, 44], the framework integrates a locally deployed Llama 3.2 instance orchestrated via the Ollama execution engine. Rather than allowing unconstrained natural language generation, the explanation layer enforces strict numerical grounding and constraint-guided prompt engineering to ensure robust alignment with the underlying Bayesian Network (BN).

Grounded Explanation and Translation Architecture

The runtime architecture operates locally to guarantee eliminates cloud transmission latency and strict data privacy. To balance structural determinism with interactive flexibility, the engine executes Llama 3.2 across two distinct operational pipelines:

Automated Evaluation Mode (evaluate.py): Designed for immediate, highly deterministic structured reporting following BN inference.

Interactive Conversational Mode (chat.py): Tailored for multi-turn, context-aware interrogation by risk officers.

To enforce strict numerical grounding and prevent hardware instability during local CPU execution, spe-

cific hyperparameters were tailored per operational mode, as detailed in Table 6.

table 6: Local Llama 3.2 Execution Hyperparameters Across Operational Modes

Hyperparameter / Setting	Evaluation Mode (evaluate.py)	Interactive Chat Mode (chat.py)	Operational Rationale
Model Runtime	Llama 3.2	Llama 3.2	Locally hosted via Ollama framework
CPU Thread Limit (num_thread)	4	4	Constrains thread concurrency to prevent hardware thermal throttling
Temperature (temperature)	0.1	0.0	Near-zero sampling variability to promote more consistent outputs, although this setting does not guarantee factual accuracy or eliminate hallucinations.
Top-P Sampling (top_p)	0.9	0.1	Restricted nucleus sampling for highly deterministic token selection
Maximum Tokens (num_predict)	140	250	Strict output length cap enforcing concise 3-bullet or 3–4 sentence outputs
Repetition Penalty (repeat_penalty)	Default (1.0)	1.2	Penalizes token looping during multi-turn conversational querying

Data Humanization and Decision Governance Boundaries

Prior to prompt construction, raw discrete evidence states and Bayesian posterior probabilities are converted into intuitive domain representations using a humanization translation dictionary (human_labels). This translation spans both input features and derived hierarchical pillars:

Primary Evidence Mapping: Raw discretized categories are mapped to qualitative descriptors (e.g., credit_score → {Very Poor, Poor, Good, Excellent}; employment_status → {Unemployed, Employed (Stable)})

Intermediate Derived Pillars: Hierarchical intermediate states are humanized into standardized risk capacity tiers (financial_capacity → {Low Capacity, Medium Capacity, High Capacity}; credit_worthiness → {Poor Worthiness, Fair Worthiness, Excellent Worthiness})Furthermore, the calculated posterior approval probability P(Approval) is mapped into explicit operational decision categories:

$$Deceision = \begin{cases} Likely\ Approved, if\ P(Approval) \geq 60\% \\ Borderline - Requires\ Manual\ Human\ Review, if\ 40\% < P(Approval) < 60\% \\ Likely\ Rejected, if\ P(Approval) < 40\% \end{cases}$$

Unlike traditional single-threshold classifiers (P = 0.50) that exhibit high decision entropy near the boundary, this framework adopts a dual-threshold mechanism (P ≥ 0.60 for approval, P<0.40 for rejection) establishing an explicit Uncertainty Buffer Zone (0.40 < P < 0.60):

Automated Processing (P ≥ 0.60 or P < 0.40): High-confidence profiles with decisive probabilistic evidence are processed automatically.

Human-in-the-Loop (HITL) Governance (0.40 <P< 0.60): Borderline cases with conflicting evidence are systematically flagged for manual credit officer review.

This 20-percentage-point uncertainty interval is a design choice intended to flag borderline cases for human review. It does not, by itself, establish reduced financial exposure or compliance with regulatory requirements, including those of the EU AI Act.

Constraint-Guided Prompt Engineering and

The humanized profile JSON and calculated posterior probability metrics are injected into a system prompt bounded by strict deterministic constraints: Professional B2B Tone Directive: Mandates an institutional credit analyst persona. Conversational greetings, introductory fluff, and pleasantries are strictly prohibited, enforcing an objective third-person perspective (referring strictly to "the applicant's profile"). Mathematical Alignment and Anomaly Detection: The LLM evaluates logical agreement between the Bayesian posterior probability and the underlying humanized metrics. If P(Approval) falls within the critical range between 40% and 60% while coinciding with high-risk metrics, the model is instructed to output an explicit anomaly flag. Otherwise, outputs are tagged as Aligned with banking logic. Mitigation of Numerical Hallucinations: The LLM is strictly prohibited from inventing, calculating, or introducing financial figures (e.g., dollar amounts, interest rates, or loan terms). The only numerical figure

CC BY

permitted in the generated explanation is the exact posterior probability calculated by the BN core. Causal Reasoning Alignment: Prompts force Llama 3.2 to trace causal drivers along the network's Directed Acyclic Graph (DAG) topology: moving from primary inputs through intermediate derived pillars (financial_capacity and credit_worthiness) to the final decision state. 4.4 Output Format Specifications In initial reporting mode (evaluate.py), the generated explanation is constrained to a deterministic three-bullet structure: Decision Status: Declares the decision outcome, the exact BN posterior probability percentage (P%), and the alignment/anomaly verification tag. Risk Evaluation: Highlights key negative drivers threatening repayment (limited to a maximum of two sentences) or declares a standardized non-risk statement. Mitigating Strengths: Details offsetting high-performing profile attributes (maximum two sentences) or suggests qualitative credit structuring adjustments. (The complete, raw System Prompt Templates for both evaluate.py and chat.py are provided in Appendix A). Interactive Conversational Engine (chat.py) To enable dynamic post-inference analysis for credit risk officers, the interactive conversational engine (chat.py) establishes a stateful session linked directly to the primary Bayesian Network evaluation output (last_evaluation). To prevent arbitrary or ungrounded queries, the endpoint enforces a strict evaluation dependency, requiring an active applicant profile prior to initializing context-aware interrogation. Upon receiving a query, the conversational pipeline executes a keyword-driven intent classification mechanism to dynamically route the inquiry into target instruction templates while maintaining the base system constraints: Causal Justification Mode: Triggered by causal indicators (e.g., why, reject, cause), this path prompts Llama 3.2 to articulate the specific structural bottlenecks occurring within the intermediate derived pillars (financial_capacity vs. credit_worthiness) that mathematically drive the posterior approval probability. Qualitative Risk Mitigation Mode: Activated by optimization queries (e.g., improve, mitigate, fix), this path structures actionable recommendations into two distinct domains: Internal Variable Shifting: Controllable client-side behavior modifications aimed at improving specific weak intermediate pillars. Loan Structuring Adjustments: Institutional policy adaptations (e.g., requiring collateral or reducing requested capital) to absorb credit exposure. General Inquiry Mode: Directly addresses technical questions regarding profile attributes while strictly maintaining mathematical alignment with the calculated Bayesian outputs. To ensure reproducible and grounded explanations during local execution, runtime inference options are strictly bounded to meet three practical requirements: Zero Temperature (Temperature = 0.0): Minimizes generative randomness to produce consistent, stable explanations for identical profiles and reduce potential hallucinations. Constrained Nucleus Sampling (Top_p = 0.1): Filters out out-of-context or informal words, restricting output selection to relevant, professional banking terminology. Controlled Output Length (Num Predict = 250, Repeat Penalty = 1.2): Limits responses to concise 3–4 sentence summaries, preventing redundant text loops and reducing compute latency on local hardware.

Rule-Based Ground-Truth Baseline for LLM Evaluation

To systematically evaluate the faithfulness, semantic equivalence, and numerical consistency of the explanations generated by Llama 3.2, an automated, deterministic rule-based baseline engine was developed in Python. Operating independently of generative language models (Zero-LLM configuration), this engine serves as a deterministic reference benchmark. The evaluation pipeline constructs this baseline across the N=300 test profiles through a programmatic, two-stage rule-mapping mechanism: Rule-Based Factor Analysis: The baseline script interfaces directly with the compiled Bayesian Network inference core (pgmpy) via VariableElimination. For any given evidence profile E, it executes custom condition-mapping functions that evaluate discrete categorical states (such as credit score tiers, income levels, and loan counts) to programmatically separate inputs into positive contributing factors and negative risk drivers. Deterministic Text Instantiation: The exact posterior approval probability P(Approval | E) and the extracted factors are then programmatically formatted into structured, predefined textual templates. This ensures that the reference text strictly reflects the underlying probabilistic inference without model-driven variation. By establishing this baseline across all N = 300 unique credit scenarios, we created a standardized evaluation benchmark. This enables automated, scalable measurement of the downstream LLM's syntactic precision, contextual semantic alignment (via BERTScore), and adherence to numerical outputs.

Technical Stack and Software Implementation

The proposed framework is implemented using a modular, fully open-source software stack: Dataset Source & Preprocessing: The empirical loan evaluation dataset used for structure learning and CPT parameterization is publicly available . Data standardization, continuous feature discretization,

CC BY

and categorical encoding were implemented using Pandas, NumPy, and scikit-learn (LabelEncoder). Probabilistic Modeling & Inference Core: Structural topology learning and conditional probability estimation were executed via pgmpy using a hybrid approach (HillClimbSearch constrained by domain expertise) and a BayesianEstimator with a BDeu prior. Exact marginal probability queries were processed using the VariableElimination inference algorithm, with trained model artifacts serialized via joblib. Backend API & Data Validation: Backend REST services were developed using FastAPI, enforcing strict request payload validation and data integrity via Pydantic schemas (BaseModel). Local Language Model Execution: Natural language translation and conversational explanations were orchestrated via requests interfacing with a locally hosted Ollama engine running the Llama 3.2 model. Presentation Layer: The user interface was developed using React.js, providing an interactive dashboard for real-time risk reporting, counterfactual scenario exploration, and multi-turn querying

Results

Bayesian Network Performance Evaluation Predictive Performance of the Bayesian Core To ensure methodological consistency across all architectural layers, the empirical dataset was partitioned using a standard 70% training and 30% testing split (N = 1,000 total records; N = 300 held-out test profiles). The Bayesian Network (BN) core was parameterized using the 70% training subset and evaluated on the held-out test set. For each test profile, input variables were processed using standard discretization and encoding procedures, while domain-expert-derived intermediate variables (Financial Capacity and Credit Worthiness) were computed using predefined deterministic rules and incorporated into the inference evidence. The state with the highest posterior probability for the Loan decision node was selected as the predicted outcome and evaluated against ground-truth loan statuses. To contextualize the Bayesian Network's performance against standard statistical benchmarks, two baseline models were evaluated on the same independent test split (N = 300) using the exact same five primary predictor features: a Majority-Class Baseline (which deterministically predicts the majority 'Approved' status) and a standard Logistic Regression model. As detailed in Table 5, the Majority-Class Baseline achieves an accuracy of 73.00% with a 100.00% Recall, directly reflecting the empirical class imbalance of the underlying credit dataset (where 729 out of 1,000 cases, or 72.90%, represent loan approvals). The proposed Bayesian Network core demonstrates a clear predictive improvement over this naive baseline, reaching an overall accuracy of 82.67%, a Precision of 81.04%, a Recall of 99.54%, and an F1-score of 89.34%. While standard Logistic Regression achieves a higher accuracy of 88.33%, it functions purely as a discriminative 'black-box' classifier without explicit conditional dependency representations. The Bayesian Network effectively trades off a marginal difference in classification accuracy for complete structural explainability, transparent Conditional Probability Tables (CPTs), and a DAG topology that enables interactive conversational reasoning.

Rule-Based Ground-Truth Baseline Construction

Evaluating N = 300 test cases through manual human expert annotation introduces subjective bias, inter-annotator variance, and scalability constraints. To establish a rigorous reference benchmark for evaluating the faithfulness of Llama 3.2 to the Bayesian Network (BN) outputs, an automated rule-based explanation engine was implemented. This engine dynamically extracts posterior probability vectors P(Y | E) and conditional probability table (CPT) parent-child dependencies directly from the BN core, inserting these inference outcomes into structured, deterministic text templates. The construction of this rule-based baseline serves four primary analytical functions within the evaluation pipeline: Standardization and Consistency: It enforces uniform evaluation criteria across all 300 test profiles, eliminating human scoring variability. Numerical and Factual Precision: It directly anchors reference text to exact BN probability calculations, ensuring zero gold-standard error. Experimental Reproducibility: It provides a deterministic baseline that guarantees complete replication under identical BN configurations. Feasibility and Scalability: It offers an automated, verifiable benchmark for evaluating LLM text alignment against underlying BN outputs at scale. Operational Note: This rule-based reference engine does not function as an independent manually constructed reference texts (Ground Truth) for evaluating the predictive capability of the BN itself. Instead, it defines the precise information expected to be verbalized by the downstream LLM. Consequently, the resulting evaluation metrics measure the syntactic, semantic, and quantitative faithfulness of the generated explanations relative to the probabilistic BN core.

Automated NLG Evaluation Results

To evaluate the natural language explanations generated by the Llama 3.2 model, a deterministic rule-based explanation engine was constructed as a reference baseline. This baseline maps the probabilistic outputs (P(Y | E)) and conditional probabil-

CC BY

ity dependencies from the Bayesian Network into structured, predefined textual templates. These reference texts were used to assess the lexical overlap, semantic similarity, and selected aspects of numerical and decision-level alignment of the generated explanations across 300 held-out test profiles.

table 5: Performance Benchmarking of the Bayesian Network Core against Baselines (N = 300)				
Model / Benchmark	Accuracy	Precision	Recall	F1-Score
Majority-Class Baseline	73.00%	73.00%	100.00%	84.39%
Proposed Bayesian Network (BN)	82.67%	81.04%	99.54%	89.34%
Logistic Regression	88.33%	90.71%	93.61%	92.13%

Table 7 - Summary of Automatic Evaluation Metrics for Generated Explanations (N = 300)		
Evaluation Category	Metric	Mean Score (%)
Lexical Overlap	BLEU	4.17%
	ROUGE-1	44.99%
	ROUGE-2	16.85%
	ROUGE-L	33.38%
Semantic Similarity	BERTScore Precision	83.34%
	BERTScore Recall	83.89%
	BERTScore F1	83.59%
Explanation Faithfulness	Numerical Faithfulness Rate	91.33%
Decision Alignment	Direct Decision Alignment Accuracy	28.67%

Summary of Observed Results

Lexical Overlap: As shown in Table 7, the BLEU score was 4.17%, while ROUGE-1, ROUGE-2, and ROUGE-L reached 44.99%, 16.85%, and 33.38%, respectively. These metrics quantify different forms of lexical overlap between the generated explanations and the manually constructed reference texts. Differences in wording and explanatory detail may affect these scores. Semantic Similarity: The generated explanations achieved a BERTScore Precision of 83.34%, a Recall of 83.89%, and an F1 score of 83.59%. These results indicate semantic similarity between the generated explanations and the reference texts according to BERTScore. However, because the references contain only the key input values and final decisions rather than complete explanatory narratives, these scores should not be interpreted as direct evidence of factual correctness or complete semantic equivalence. Numerical Faithfulness and Decision Alignment: The Numerical Faithfulness Rate was 91.33%, indicating consistency with the numerical reference criteria under the implemented evaluation procedure. Direct Decision Alignment Accuracy was 28.67%, based on regex-based and exact keyword matching of the decision status. The latter result reflects explicit decision-label matching under the implemented parsing procedure and does not, by itself, establish the overall correctness of the generated explanations. These metrics assess specific aspects of the generated explanations relative to the constructed references. They provide evidence about alignment with the recorded inputs and decisions but do not independently establish the real-world validity of the Bayesian Network's predictions or quantify all forms of hallucination.

LLM-as-a-Judge & Semantic Quality Analysis

To complement automatic text-matching metrics, Qwen 2.5 (7B), running locally via Ollama, was used as an LLM-based judge to evaluate the generated explanations against manually constructed reference texts (Ground Truth). For each test case, these reference texts contained the key input values provided to the Bayesian Network and its corresponding final decision, without additional explanatory content. The evaluation covered 300 test cases and assessed three dimensions: faithfulness, clarity, and completeness, using a five-point rating scale. Qwen 2.5 achieved an overall mean score of 3.11 out of 5.0. The mean scores were 2.98 for faithfulness, 3.92 for clarity, and 2.42 for completeness. Clarity received the highest rating, whereas completeness received the lowest. These results indicate that the generated explanations were rated more favorably in terms of clarity than faithfulness and completeness, suggesting room for improvement in their alignment with the reference information and their coverage of relevant details The judge-model evaluation complements the automatic metrics by providing an additional assessment of the generated explanations relative to the constructed references. However, these scores represent model-based judgments and

should not be interpreted as independent proof of factual correctness, comprehensive hallucination detection, or real-world decision validity.

Table 8 - LLM-Based Evaluation of Generated Explanations Using Qwen 2.5 (7B) (N = 300)

Evaluation Dimension	Mean Score (out of 5)
Faithfulness	2.98
Clarity	3.92
Completeness	2.42
Overall Mean	3.11

Operational Validation: Live Interactive Case Scenarios

To rigorously validate the hybrid Hierarchical Bayesian Network and LLM-based explanation framework, we implemented a multi-faceted evaluation pipeline across a test sample of 300 comprehensive scenarios. The aggregate evaluation provided a quantitative assessment across metrics, including a Bayesian Network Accuracy of 82.67% (Precision: 0.8104, Recall: 0.9954, F1-Score: 0.8934), lexical alignment (ROUGE-1 Average: 0.4499, ROUGE-2 Average: 0.1685, ROUGE-L Average: 0.3338, BLEU Average: 0.0417), and a qualitative evaluation by the Qwen 2.5 judge with an Overall Mean Score of 3.11 / 5.0 (Clarity: 3.92, Faithfulness: 2.98, Completeness: 2.42) at an average latency of 15.75 seconds per case. The aggregate evaluation metrics provide an overview of the predictive performance of the Bayesian Network and the quality of the generated explanations. To complement these quantitative results, three illustrative scenarios are examined, representing high-approval, high-rejection, and borderline probability cases. These examples demonstrate the framework's explanation and interaction capabilities while also highlighting potential limitations in numerical interpretation, structural reasoning, and decision consistency. The interactive responses are presented as qualitative examples and were not included in the automated single-shot explanation evaluation: Low-Risk Profile (Deterministic Approval) Input Evidence Vector (E): Credit Score = 770, Annual Income = $147,905, Existing Loans = 0, Dependents = 1, Employment = Employed. BN Posterior Inference: P(Approval | E) = 99.90% (High Confidence Tier). Generated Narrative Output: Decision Status: Approved | 99.90% | Aligned with banking logic (Strong Profile). Risk Evaluation: The model's approval of a subprime profile despite a good credit score and employed status is questionable. Further review is necessary to understand the decision-making process. Mitigating Strengths: High annual income combined with zero existing debt and stable employment provides maximum financial stability and repayment capacity. Automated Explanation Evaluation: Automated Lexical Metrics: Evaluated against the reference texts, achieving strong alignment (ROUGE-1 = 0.2149, BLEU = 0.0393). LLM-Based Judge Assessment: Qwen 2.5 evaluated the explanation with an overall score of 3.33/5.0, indicating satisfactory alignment with the Bayesian posterior probability and professional readability. Interactive Sub-Query Tracing: The following conversation demonstrates the interactive reasoning of the system when probed with specific exploratory queries: Q1: What are the primary drivers for this high approval probability? "The Final Model Approval Probability of 99.90% is primarily driven by the applicant's profile having a Raw Credit Score Category of "Excellent" and Financial Capacity Assessment of "High Financial Capacity", which are strongly indicative of low credit risk. Additionally, the existing debt burden being categorized as "No Debt Burden" further reduces the likelihood of default. These factors collectively contribute to an extremely high probability of approval." Q2: Given this excellent risk evaluation, what qualitative loan structuring terms or preferential options should the bank consider offering this applicant? "The Final Model Approval Probability of 99.90% indicates a strong confidence in The Risk Profile's creditworthiness and financial capacity. Given these excellent assessments, the bank should consider offering preferential loan terms such as a lower interest rate or extended repayment period to further reduce the risk burden on The Applicant. This decision is grounded in the high probability that The Risk Profile will meet their loan obligations without defaulting (99.90% confidence). Offering flexible loan structures aligns with this assessment, ensuring The Bank's exposure to potential losses remains minimal while still providing a competitive offering to The Applicant" High-Risk Profile (Deterministic Rejection) Input Evidence Vector (E): Credit Score = 480, Annual Income = $24,135, Existing Loans = 2, Dependents = 4, Employment = Employed. BN Posterior Inference: P(Approval | E) = 0.21% (High Confidence Rejection Tier). Generated Narrative Output: Decision Status: Likely Rejected | 0.21% | Aligned with banking logic. Risk Evaluation: The low credit score and relatively low income may indicate a higher risk profile, but the model's rejection aligns with banking logic due to its high baseline probability. This decision is not indicative of a critical anomaly.

Mitigating Strengths: High employment status mitigates the overall risk, while having more dependents reduces the debt-to-income ratio. Automated Explanation Evaluation: Automated Lexical Metrics: Evaluated against the reference texts, achieving consistent alignment (ROUGE-1 = 0.512, BLEU = 0.0256). LLM-Based Judge Assessment: Qwen 2.5 evaluated the explanation with an overall score of 3.0/5.0, reflecting moderate alignment with the Bayesian posterior probability and standard readability. Interactive Sub-Query Tracing: The following conversation demonstrates the interactive reasoning of the system when probed with specific exploratory queries: Q1: What are the primary drivers for this high rejection probability? "The Financial Capacity Assessment and Credit Worthiness Evaluation are interconnected yet distinct components of the risk profile. The applicant's Low Financial Capacity is primarily driven by their 2 Active Loans, which indicate a substantial debt burden. This bottleneck in financial capacity constrains the ability to absorb additional credit or manage existing expenses. Conversely, the Poor Credit Score Category and High Risk associated with the Credit Worthiness Evaluation are largely influenced by the same underlying factors: the applicant's Very Low Income and 2 Active Loans. The stable employment status mitigates some of these concerns; however, it does not entirely offset the risk posed by their debt burden. The combination of these factors creates a compensating trade-off between Financial Capacity Assessment and Credit Worthiness Evaluation, resulting in an overall high rejection probability due to the cumulative impact of debt burden on both assessments." Q2: Given this high-risk evaluation, what specific corrective measures or restructuring steps would this applicant need to undertake to shift their profile from a rejected status to an acceptable risk threshold? "The applicant's Financial Capacity Assessment is constrained by their existing debt burden and family dependents, which limits their ability to take on additional loans. This bottleneck in Financial Capacity Assessment negatively impacts the overall risk profile. In contrast, while Credit Worthiness Evaluation indicates a high-risk category due to poor credit scores, historical network patterns suggest that stable employment with low income can mitigate this risk when combined with other positive factors (e.g., employed status). The 0.21% approval probability reflects the net effect of these competing assessments: the applicant's limited financial capacity outweighs their relatively good employment and family stability. To address the bottleneck, the officer should consider alternative loan options that do not exacerbate debt burden or explore income-enhancing strategies to improve Financial Capacity Assessment before pursuing further credit approvals." Borderline Decision Evaluation & Uncertainty Zone Input Evidence Vector (E): Credit Score = 490, Annual Income = $149,295, Existing Loans = 3, Dependents = 2, Employment = Employed. BN Posterior Inference: P(Approval | E) = 50.00% (Critical Threshold / Manual Review Zone). Generated Narrative Output: Decision Status: Approved (Borderline) | 50.00% | Aligned with banking logic (Risk Correctly Captured). Risk Evaluation: The subprime profile was not flagged despite a low credit score, indicating inadequate risk assessment. Insufficient weight may have been placed on the credit score metric. Mitigating Strengths: High income and stable employment mitigate some risk but are insufficient to offset the adverse credit profile. Automated Explanation Evaluation: Automated Lexical Metrics: Evaluated against the reference texts, achieving strong alignment (ROUGE-1 = 0.3571, BLEU = 0.0114). LLM-Based Judge Assessment: Qwen 2.5 awarded an overall score of 3.0/5.0, reflecting a satisfactory alignment in articulating the nuances of the borderline risk zone and maintaining model fidelity under high uncertainty. Interactive Sub-Query Tracing: To further investigate this borderline evaluation, we initiated the following dialogue: User Inquiry: How can we improve this client's profile to achieve approval? System Response: "The applicant's profile exhibits a contradictory mix of high income and poor credit score, indicating potential for financial capacity but elevated credit risk due to existing debt burden. **INTERNAL VARIABLE SHIFTING** To improve this client's profile, consider the following: Encourage The applicant to prioritize debt repayment by allocating more funds towards loan payments. Suggest a temporary reduction in discretionary spending to reduce the overall debt-to-income ratio. Recommend regular financial planning and budgeting sessions with the applicant. LOAN STRUCTURING** To mitigate risk, consider adjusting bank policies: Offer collateral options that better align with the applicant's credit profile (e.g., secured loans). Implement a more stringent loan approval process for high-risk borrowers to ensure adequate capital buffers. Consider offering alternative repayment terms or interest rate adjustments to reduce the overall cost of borrowing. By implementing these strategies, it is possible to improve the risk profile and in-

crease the likelihood of successful loan outcomes"

Discussion:

1. Interpretation of Bayesian Network Predictive Performance

The empirical evaluation of the Bayesian Network (BN) core achieved a classification accuracy of 82.67% and an F1-score of 89.34% on the held-out test subset (N = 300). These results validate the capacity of the probabilistic graph to capture underlying decision patterns embedded within the historical credit dataset. Key technical and methodological insights regarding these results include:

1. Class Distribution and Efficacy of Structural Feature Decomposition: The introduction of intermediate variable structures—specifically Financial Capacity and Credit Worthiness—was designed using analytical domain heuristics to structure raw input features into discrete probabilistic dependencies. Evaluating the test partition (N = 300) yields an overall Bayesian Network accuracy of 82.67%, with Precision at 81.04%, Recall at 99.54%, and an F1-score of 89.34%. To investigate the elevated Recall score, a class distribution analysis of the baseline credit dataset (N = 1,000) was conducted. The empirical dataset exhibits an inherent class imbalance, comprising 729 Approved cases (72.90%) against 271 Rejected cases (27.10%). Consequently, the Bayesian Network learned this empirical prior distribution, resulting in a prediction boundary aligned with the majority class prior rather than an artificially balanced decision threshold. This confirms that the classification behavior directly reflects the underlying dataset distribution while maintaining exact probabilistic inference tractability.

2. Reflecting Empirical Data Patterns and Boundaries: The remaining ~18% classification variance highlights two operational factors: boundary discretization effects in continuous variables, and inherent inconsistencies or noise within historical human credit decisions. Because the BN reflects the empirical probability distributions of the underlying dataset, its predictive output mirrors real-world decision patterns rather than idealized assumptions.

3. Validation Alignment and Scope Considerations: The current performance was evaluated using an independent 30% held-out test split, ensuring strict separation between parameter learning and performance estimation. While evaluating the model against external multi-institutional datasets or conducting blind human-expert panel scoring remains an objective for future work, the current held-out evaluation provides a statistically valid and rigorous benchmark for downstream explanation generation.

2. Critical Analysis of Natural Language Generation Performance

1. Lexical vs. Semantic Divergence: The low BLEU and ROUGE-2 scores (4.17% and 16.85%, respectively), alongside a BERTScore F1 of 83.59%, reflect differences in surface-level wording alongside semantic similarity between the generated explanations and the template-based reference texts. These metrics capture complementary aspects of textual similarity and should not be interpreted as direct evidence of factual correctness or complete semantic equivalence.

2. Mitigation of Numerical Hallucinations: The high Numerical Faithfulness Score (91.33%) indicates that the generated explanations were numerically consistent with the underlying BN probabilities in 91.33% of cases under the implemented evaluation procedure. This result suggests that the proposed integration strategy can support numerical consistency between generated narratives and BN outputs. However, it does not independently establish that structured evidence injection and constraint-guided prompt engineering reduce numerical hallucinations, as this would require a separate evaluation.

3. Methodological Constraints of Programmatic Decision Matching: The Direct Decision-Matching Rate of 28.67% reflects the operational limitations of strict regex-based parsing rather than a systemic failure of LLM semantic decision alignment. Because programmatic matching required exact keyword matches (e.g., "Approved"), descriptive or alternative domain formulations -such as "The applicant displays a highly favorable profile for credit extension"- were categorized as non-matches despite conveying the correct decision context. This limitation underscores the necessity of incorporating LLM-as-a-Judge semantic evaluation to assess decision-support quality beyond lexical matching.

3. Methodological Analysis of Judge Performances

1. Assessment of Explanation Quality: The Qwen 2.5 evaluation provided a complementary assessment of the generated explanations beyond automatic text-matching metrics. The mean clarity score of 3.92 out of 5.0 was the highest among the evaluated dimensions, followed by faithfulness at 2.98 and completeness at 2.42.

2. Interpretation of Judge Scores: The overall mean score of 3.11 out of 5.0 indicates a moderate evaluation outcome, with clarity rated more favorably than faithfulness and completeness. These results suggest that, although the explanations were generally considered relatively clear, their alignment with the reference information and their coverage of relevant details require further improvement.

3. Methodological Limitations: The Qwen-based assessment provides an additional model-based eval-

CC BY

uation of the generated explanations against the manually constructed reference texts. However, these scores should be interpreted as supplementary judgments rather than independent verification of factual correctness or comprehensive detection of hallucinations. The evaluation is also limited by the scope and content of the constructed references.

4. Qualitative Insights from Live Case Scenarios and Interactive Risk-Mitigation Recommendations

The three representative case scenarios provide qualitative insights into the explanatory capabilities and limitations of the proposed framework across low-risk, high-risk, and borderline credit profiles. The analysis distinguishes the probabilistic predictions produced by the Bayesian Network (BN) from the natural-language explanations generated by Llama 3.2, with particular attention to structural consistency, probabilistic interpretation, and the reliability of interactive recommendations.

4.1. Evaluation Scope and Single-Shot Explanation Performance

The quantitative Natural Language Generation (NLG) metrics (ROUGE and BLEU) and the Qwen 2.5-based evaluation were applied exclusively to the initial single-shot explanations, consisting of three components: Decision Status, Risk Evaluation, and Mitigating Strengths. These explanations were compared against deterministic rule-based reference templates generated from the applicant attributes and Bayesian Network outputs. The results indicate that linguistic coherence does not necessarily ensure accurate or complete representation of the underlying probabilistic model. Accordingly, the following case analysis examines whether the generated explanations preserve the meaning of the predicted outcome, correctly attribute input features to the corresponding intermediate assessments, and distinguish model-supported evidence from plausible but unverified financial interpretations.

4.2 Case-Based Analysis of Explanatory Limitations

Low-Risk Profile: Overinterpretation of the Predicted Outcome

In the low-risk scenario, the interactive explanation associated the 99.90% approval probability with favorable financial capacity and creditworthiness indicators. However, it also interpreted this probability as indicating an equivalent likelihood of fulfilling loan obligations without default. This interpretation exceeds the scope of the BN, which predicts loan approval rather than repayment or default. The explanation also described the applicant as having "maximum financial stability and repayment capacity." Although the favorable indicators support a strong approval assessment within the implemented model, this absolute wording is not established by the reported probability alone. These observations illustrate two distinct limitations: extending the predicted outcome to an unmodeled financial event and overstating the strength of the assessment through imprecise language. Neither observation, by itself, establishes an error in the BN's numerical inference.

High-Risk Profile: Structural Attribution and Unsupported Financial Claims

In the high-risk scenario, the initial explanation described the financial burden associated with having multiple dependents as a reduction in the debt-to-income ratio. However, the implemented framework incorporates the number of dependents into a rule-based Financial Capacity Assessment through penalties applied to higher dependent-burden categories; it does not directly calculate the conventional debt-to-income ratio. The explanation therefore introduced a financial interpretation that was not supported by the implemented mechanism. The interactive responses revealed additional concerns. When asked about the primary drivers of rejection, the conversational model associated existing loans with both Financial Capacity Assessment and Credit Worthiness Evaluation, although existing_loans_count is structurally assigned to Credit Worthiness Evaluation in the implemented BN. The response also referred to "historical network patterns" to explain the mitigating effect of stable employment under low-income conditions, without presenting explicit conditional probability evidence supporting that claim. These observations concern the fidelity of the generated explanations rather than necessarily the correctness of the final approval probability. They demonstrate that a plausible financial narrative can misrepresent the network's structure or present an unsupported relationship as an established model finding.

Borderline Profile: Interpretation of Conflicting Indicators and Human Review

In the borderline scenario, the BN assigned a 50.00% approval probability and reported an Approved classification. We clarify that exact probability values of 50.00% in certain Conditional Probability Table (CPT) entries arise from the Dirichlet uniform smoothing prior applied during parameter learning to handle sparsely populated or unobserved parent configurations; thus, this outcome reflects data sparsity and structural neutrality rather than a true conflict of empirical evidence. The initial narrative questioned whether the applicant's low credit score had received sufficient consideration, while the interactive response acknowledged the combination of high income, poor creditworthiness indicators, and existing debt burden.

CC BY

The interactive response proposed debt repayment reduced discretionary spending, financial planning, and alternative loan structures. However, it did not recalculate the posterior approval probability following any proposed change to the applicant's attributes. These suggestions should therefore be treated as qualitative risk-mitigation recommendations rather than validated counterfactual explanations. The framework also applies a separate human-review advisory to cases with approval probabilities between 40% and 60%, inclusive. Under this policy, the 50.00% case retains the BN's Approved classification while receiving an additional recommendation for review by a credit officer. The advisory does not alter the network's prediction or constitute a new classification; it is a supplementary decision-support mechanism intended to draw attention to cases requiring further assessment. When examining the raw narrative outputs generated across the representative case scenarios, certain conceptual or phrasing discrepancies become apparent. For instance, in Case 1, despite an excellent credit score of 770, the model's generated narrative unsuitably utilized the term 'subprime' while simultaneously classifying the profile as aligned with banking logic. Similarly, minor semantic or structural variations appear in subsequent scenarios. To mitigate such issues, the prompt engineering protocol was carefully structured to enforce strict operational constraints, limit numerical hallucination, and guide the language model along the Bayesian network's causal pathways. However, despite these structured prompting safeguards, lightweight locally hosted models inherently exhibit limitations when handling multi-constraint reasoning tasks. The purpose of integrating this local setup was to test the practical feasibility of the decoupled hybrid architecture within an offline, resource-accessible environment. Consequently, the minor variations, inappropriate terminology (such as subprime for high-scoring profiles), or subtle semantic discrepancies occasionally found in the raw outputs are an expected characteristic of models with limited parameter sizes when navigating complex inference constraints. It is important to emphasize that this prototype was intentionally designed as an exploratory proof-of-concept. While these raw outputs successfully demonstrate the functional integration of probabilistic inference and natural-language generation, they also highlight the boundaries of edge-deployed language models. Transitioning this framework into an operational, production-grade financial decision-support system would certainly necessitate scaling up to advanced, high-capacity language models to guarantee absolute narrative fidelity and eliminate such explanatory discrepancies.

4.3 Architectural Constraints and Feature-Space Limitations

The framework's explanatory capabilities are constrained by the variables and probabilistic relationships represented in the BN and supplied to the language model through the structured prompt context. To limit Conditional Probability Table (CPT) dimensionality and mitigate data sparsity, the implemented BN retains five primary input variables and two domain-engineered intermediate nodes. Although this compact structure reduces model complexity, it also limits the information available for probabilistic inference and subsequent explanation generation. For example, the exclusion of loan_amount prevents the framework from explicitly incorporating the requested loan size into its assessment. An applicant may therefore receive a high approval probability despite a relatively low credit score, while the framework remains unable to account for the potential effect of the requested amount. Such a result may reflect a limitation of the available feature space rather than an error in the network's inference, and should be interpreted in the context of the variables represented in the model. Overall, the case scenarios demonstrate that evaluating conversational Bayesian explanations requires more than verifying the reported approval probability. The explanations must also preserve the meaning of the predicted target, correctly represent the network's structural dependencies, avoid unsupported financial claims, and distinguish qualitative recommendations from changes to the model's calculated outcome. The current framework should therefore be regarded as a proof of concept for conversational explanations of Bayesian predictions rather than a comprehensive credit-risk assessment system. Future work may incorporate additional financial variables, more explicit extraction of probabilistic evidence, and systematic evaluation of interactive responses to improve explanatory reliability.

5. Literature Benchmarking & Comparative Position

To position this study within the broader Explainable AI (XAI) literature, the proposed integration of Llama 3.2 with a Bayesian Network (BN) is discussed in relation to established approaches to Bayesian explanation, natural language generation (NLG), and post-hoc feature attribution.

Positioning Against BN Explanation Baselines

Previous research has investigated methods for explaining Bayesian network inference through natural language generation and structured explanatory representations [17, 36]. Such approaches provide mechanisms for translating probabilistic reasoning

CC BY

into human-readable explanations. The proposed framework adopts a complementary architecture by combining a BN-based probabilistic inference core with an LLM-based language layer. This design supports flexible, applicant-specific explanations and conversational interaction, while retaining the BN as the source of the underlying predictions. Unlike predefined templates, an LLM-based language layer offers greater flexibility in phrasing and interaction; however, the factual accuracy of the generated explanations depends on how faithfully the language model represents the BN's structure and outputs. Although prior studies have addressed natural-language explanations of Bayesian network inference, the proposed framework differs in its application domain, language-generation architecture, and interaction setting. Keppens [17] presents an NLG approach based on qualified support graphs for generating explanations of Bayesian inference, whereas Sevilla et al. [36] investigate natural-language explanations based on factor arguments in the medical domain. In contrast, this study applies Bayesian inference to credit approval assessment and uses Llama 3.2 as a conversational language layer to explain applicant-specific BN outputs and respond to follow-up questions. The framework also examines the consistency of generated explanations with the underlying network structure and inference results. Thus, its contribution lies not in introducing natural-language explanation of Bayesian networks itself, but in integrating BN-based credit assessment with LLM-generated conversational explanations and evaluating the resulting explanations in credit-related scenarios. Alignment with LLM-Based XAI Frameworks Recent research has explored the use of large language models to generate natural-language explanations of structured model outputs and AI decisions [9, 26]. In line with this research direction, the proposed framework supplies Llama 3.2 with applicant information, BN inference results, and explicit instructions governing how the network's structure should be interpreted. These constraints are intended to reduce unsupported interpretations and improve consistency between the generated explanations and the underlying model. Nevertheless, the qualitative analysis of case scenarios highlights the importance of evaluating generated explanations against the model's actual structure and outputs, since the use of explicit prompting constraints alone does not guarantee complete faithfulness. Probabilistic Reasoning and Post-Hoc Feature Attribution (SHAP/LIME) The proposed framework uses a Bayesian Network as its probabilistic reasoning core rather than relying exclusively on post-hoc feature attribution methods such as SHAP (SHapley Additive exPlanations) [27] or LIME (Local Interpretable Model-agnostic Explanations) [32]. These methods provide ways to explain model predictions through feature contributions or locally interpretable approximations, whereas a BN explicitly represents conditional probability distributions and dependencies among variables. The approaches therefore provide different forms of explanatory information and should not be considered interchangeable. In the proposed framework, the BN represents the probabilistic relationships between applicant attributes, intermediate assessments, and the final loan approval outcome. Llama 3.2 uses the supplied model information to produce natural-language explanations of the BN's predictions. This architecture is intended to provide explanations informed by the network's probabilistic structure rather than relying solely on feature-attribution scores. However, a BN's representation of conditional dependencies does not, by itself, establish causal relationships, and the use of an LLM does not automatically guarantee that these dependencies will be accurately conveyed in natural language. Accordingly, the proposed framework is best characterized as a conversational approach to explaining Bayesian predictions, rather than as a demonstrated method of causal explanation. Ethical Considerations and Regulatory Relevance The use of conversational decision-support systems in financial applications raises ethical and regulatory considerations, particularly regarding transparency, human oversight, and the interpretation of model outputs. The proposed framework is designed as an explanatory interface that communicates Bayesian Network (BN) predictions in natural language, rather than as an autonomous decision-maker or a bias-mitigation mechanism. From an ethical perspective, the use of socioeconomic attributes, such as the number of dependents, may raise fairness concerns if these attributes contribute to systematically unfavorable outcomes for particular groups. The language model provides explanations of the BN's outputs but does not independently assess or mitigate potential biases in the underlying model. Accordingly, algorithmic fairness and potential disparities across demographic subgroups remain outside the scope of this study and require dedicated evaluation. From a regulatory perspective, the framework addresses aspects relevant to transparency and human oversight in automated decision-support systems. By translating BN outputs into natural-language explanations and providing a human-review advisory for cases with intermediate approval probabilities, the framework aims to support credit officers in inter-

CC BY

preting model predictions. These capabilities may be relevant to discussions of transparency and human oversight under regulatory frameworks such as the General Data Protection Regulation (GDPR) and the European Union Artificial Intelligence Act (EU AI Act). However, the framework has not been subjected to a formal regulatory compliance assessment, and its implementation alone does not establish compliance with these requirements.

Table 9 - Qualitative Comparison of the Proposed Framework with Representative XAI and BN Explanation Approaches

Baseline / Methodology	Representative Literature	Primary Mechanism	Distinguishing Capability of the Proposed Framework
BN Natural Language Generation	Keppens (2019) [١٧]	Generation of natural-language explanations using a structured model of Bayesian inference and support graphs	Combines BN-based probabilistic inference with an LLM-based language layer to support flexible, context-sensitive explanations and interactive dialogue.
User-Evaluated BN Explanations	Sevilla et al. (2024) [٣٦]	Natural-language explanations of BN reasoning evaluated in a user study	Extends the explanation setting to conversational interaction and applicant-specific credit assessments.
LLM-Based Explanation of Credit-Risk Predictions	Dyomin et al. (2026) [٩]	Conversion of structured machine-learning feature contributions into natural-language explanations	Uses a BN as the probabilistic inference core and supplies its outputs and structural information to an LLM for explanation generation.
Post-hoc Feature Attribution (SHAP/LIME)	Bücker et al. (2020) [٤٧]	Feature-attribution methods and other approaches to model transparency and explainability in credit scoring	Explicitly represents conditional probabilistic dependencies in the BN rather than relying exclusively on post-hoc feature-attribution outputs.

Conclusions and recommendations

Concluding Remarks

This study presented an initial proof of concept for generating natural-language explanations of Bayesian Network (BN) inferences using a locally deployed Large Language Model (Llama 3.2) in a credit decision-support context. The proposed implementation separates probabilistic inference performed by the BN from natural-language generation performed by the LLM, which receives structured applicant information and BN outputs to produce explanations of the resulting predictions. Evaluation of the BN on an independent test set yielded a classification accuracy of 82.67%, with a precision of 0.8104, recall of 0.9954, and F1-score of 0.8934. These results provide a quantitative baseline for the implemented credit-classification model. However, the observed performance should be interpreted in light of the dataset, class distribution, and evaluation setup, rather than as evidence of general suitability for real-world credit decisions. For natural-language generation, the system achieved a Numerical Faithfulness Rate of 91.33% and a BERTScore F1 of 83.59%. An additional evaluation using Qwen 2.5 as a judge model produced a mean overall score of 3.11 out of 5.0, with scores of 2.98 for faithfulness, 3.92 for clarity, and 2.42 for completeness. These results indicate that the generated explanations were generally clearer than they were complete or consistently faithful, highlighting the distinction between linguistic readability and the accurate communication of model outputs. The separation of BN inference from language generation is a central design characteristic of the implementation: the LLM is used to communicate the results of probabilistic inference rather than to calculate the underlying probabilities independently. Nevertheless, this architectural separation does not guarantee that every generated explanation accurately preserves the numerical results or the relationships represented in the network. The observed evaluation results and qualitative limitations therefore support treating the system as an exploratory decision-support prototype rather than a val-

idated solution for operational credit assessment.

Acknowledged Limitations

Several technical and methodological limitations should be considered when interpreting the findings.

1. Dependence on a Simplified BN Topology: The implemented BN contains five applicant input variables and two intermediate assessment nodes. Although this structure provides a manageable setting for investigating the integration of probabilistic inference and natural-language generation, it represents only a simplified model of credit assessment. More complex financial risk settings may involve additional variables, dependencies, and sources of uncertainty that are not represented in the current topology.

2. Evaluation Focused on Consistency with BN Outputs: The reference explanations were constructed programmatically using the BN's probability outputs. Consequently, the evaluation measures how closely the generated explanations correspond to the model-derived references; it does not independently establish the real-world validity of the BN's predictions, the appropriateness of its assumptions, or the fairness of the resulting decisions. High consistency with a reference derived from the BN should therefore not be interpreted as evidence that the underlying credit assessment is objectively correct.

3. Limitations of Programmatic Decision Matching: The direct decision-matching rate was 28.67% under the exact parsing procedure used in the evaluation. This result indicates that the implemented matching procedure did not reliably identify the intended decision in many generated responses. Because exact or regex-based matching can be sensitive to phrasing and paraphrasing, the result may reflect limitations in both the generated responses and the evaluation procedure. Further analysis is required to distinguish semantic decision errors from parsing failures.

4. Residual Risk of Unsupported or Misleading Explanations: Despite a Numerical Faithfulness Rate of 91.33%, some generated explanations may contain numerical inconsistencies, unsupported interpretations, or misleading descriptions of relationships among variables. Prompt constraints reduce the scope for unsupported interpretations but do not eliminate these risks. Moreover, if the underlying BN reflects limitations or biases in its data, assumptions, or structure, a fluent natural-language explanation does not resolve those limitations and may communicate the resulting prediction persuasively without adequately conveying its uncertainty. Human review remains important, particularly in high-stakes credit-assessment settings.

Interpretive Reliability of the Language Model

The qualitative case analysis illustrates both the potential and the limitations of LLM-generated explanations. In borderline cases, the language model can describe competing considerations in an applicant's profile, such as a low credit score alongside a high annual income. However, a plausible narrative does not necessarily demonstrate that the model has correctly represented the BN's probability distributions or the roles of individual variables. For example, descriptions of financial burden or the relative importance of credit-related factors may introduce interpretations that are not directly supported by the specified network structure or its outputs. These observations underscore the need to distinguish linguistic plausibility from model-grounded accuracy. The current implementation should therefore be understood as a means of making BN outputs more accessible through conversational interaction, not as an independently validated authority on credit risk. Further refinement and evaluation with domain experts are needed before the approach can be considered for operational use in high-stakes decision-making.

Recommendations for Future Research

Future work should address the limitations identified in this study through the following directions:

1. Expanding BN Complexity: Evaluate the framework using larger and more complex Bayesian Networks with additional variables, intermediate nodes, and higher-dimensional Conditional Probability Tables (CPTs). This would help assess whether the prompting approach remains effective as the amount and complexity of the information supplied to the LLM increase.

2. Improving Semantic Decision Evaluation: Replace or supplement exact regex-based decision matching with a more robust evaluation procedure that distinguishes genuine decision mismatches from differences in wording. The revised procedure should be validated against manually reviewed examples to establish whether generated responses communicate the intended BN decision accurately.

3. Strengthening Explanation Faithfulness and Completeness: Refine the prompting protocol and evaluation criteria to identify unsupported claims, incorrect descriptions of variable relationships, omissions of relevant evidence, and confusion between approval probability and other outcomes, such as repayment or default. Numerical consistency should be evaluated separately from structural consistency and overall explanatory quality.

4. Human-in-the-Loop Validation: Conduct structured evaluations with credit-risk professionals to assess the factual accuracy, completeness, interpretability, and practical usefulness of the generated explanations. Such studies could investigate whether conversational explanations help users understand

model outputs more effectively than conventional numerical summaries, without assuming in advance that they improve decision quality or user trust.

5. Assessing Robustness and Fairness: Investigate how changes in applicant characteristics, BN assumptions, and the distribution of evaluation data affect both the underlying predictions and their natural-language explanations. Dedicated fairness analyses would also be required to determine whether the credit model produces systematically different outcomes across relevant applicant groups.

Overall, this study provides an initial implementation and empirical evaluation of a conversational interface for explaining Bayesian Network outputs in credit assessment. Its findings demonstrate the feasibility of separating probabilistic inference from natural-language generation while also revealing limitations in numerical consistency, completeness, and automated decision evaluation. Addressing these limitations through more comprehensive testing, improved evaluation methods, and expert validation is an important next step toward establishing the reliability and practical value of this approach.

References

1. Absar, N. et al. (2021). Explaining Bayesian networks in natural language: A review. Journal of Artificial Intelligence Research, 71, 1101–1145.
2. Barber, D. (2011). Bayesian reasoning and machine learning. Cambridge University Press.
3. Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of Machine Learning Research, 3, 1137–1155.
4. Berger, J. O. (1985). Statistical decision theory and Bayesian analysis. Springer.
5. Zytek, A., Pidò, S., & Veeramachaneni, K. (2024). LLMs for XAI: Future directions for explaining explanations. arXiv preprint arXiv:2405.06064.
6. Bollen, K. A. (1989). Structural equations with latent variables. Wiley.
7. Darwiche, A. (2009). Modeling and reasoning with Bayesian networks. Cambridge University Press.
8. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
9. Dyomin, A., & Li, M. (2026). Interpretable decision support for credit risk assessment using machine learning and LLM-based explanations. Proceedings of the International Conference on Advanced Credit Risk Modeling and Explainable AI, 115–128.
10. Fenton, N., & Neil, M. (2018). Risk assessment and decision analysis with Bayesian networks. CRC Press.
11. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI Magazine, 38(3), 50–57.
12. Gunning, D. et al. (2019). DARPA's Explainable Artificial Intelligence (XAI) program. AI Magazine, 40(2), 44–58.
13. Hassani, H., & Silva, E. S. (2023). The role of ChatGPT in data science: How AI-assisted conversational interfaces are revolutionizing the field. Data, 8(7), 112.
14. Heckerman, D., Mamdani, A., & Wellman, M. P. (1995). Real-world applications of Bayesian networks. Communications of the ACM, 38(3), 24–26.
15. Jacovi, A. et al. (2021). Formalizing trust in artificial intelligence: Prerequisites, definitions and measurement scales. In Proc. ACM FAccT (pp. 639–649).
16. Ji, Z. et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
17. Keppens, J. (2019). Explainable Bayesian network query results via natural language generation systems. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law (ICAIL '19, pp. 42–51).
18. Koller, D., & Friedman, N. (2009). Probabilistic graphical models: Principles and techniques. MIT Press.
19. Krishna, S. et al. (2023). The disagreement problem in explainable machine learning: An empirical study. Neurocomputing, 558, 126720.
20. Lacave, C., & Díez, F. J. (2002). A review of explanation methods for Bayesian networks. The Knowledge Engineering Review, 17(2), 107–127.
21. Lakkaraju, H. et al. (2022). Rethinking explainability: Towards a human-centered approach for explainable AI. Communications of the ACM, 65(1), 24–32.
22. Lin, P. (2015). Why ethics matters for autonomous cars. In Autonomous Driving (pp. 69–85). Springer.
23. Lipton, Z. C. (2018). The mythos of model interpretability. ACM Queue, 16(3), 31–57.
24. Liu, P. et al. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1–35.
25. Liu, Y. et al. (2023). Trustworthy LLMs: A survey on large language models' reliability and explainability. arXiv preprint arXiv:2308.05374.
26. Turpin, M., Michael, J., Perez, E., & Bowman, S. R. (2024). Language models often offer plausible but misleading reasons for their predictions. Advances in Neural Information Processing Systems, 36, 42100–42115.
27. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
28. Metzger, A., Bartel, J., & Laufer, J. (2023). An AI chatbot for explaining deep reinforcement learning decisions of service-oriented systems. In International Conference on Service-Oriented Computing (pp. 115–130). Springer.
29. Metzger, M. et al. (2023). Interactive chatbots for explaining complex AI decisions. Journal of Artificial Intelligence Research, 78, 411–445.
30. Nauta, M. et al. (2023). From qualities to criteria: A survey on evaluation metrics for explainable AI. IEEE Transactions on

Pattern Analysis and Machine Intelligence, 45(7), 8848–8869.
31. Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann.
32. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proc. ACM SIGKDD (pp. 1135–1144).
33. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.
34. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K.-R. (Eds.). (2019). Explainable AI: Interpreting, explaining and visualizing deep learning. Springer.
35. Scutari, M., & Denis, J.-B. (2021). Bayesian networks: With examples in R. CRC Press.
36. Sevilla, J. et al. (2024). Explaining Bayesian networks reasoning to the general public. CEUR Workshop Proceedings, 4061, Paper 3.
37. Slack, D. et al. (2023). Explaining machine learning models with interactive natural language conversations. Nature Machine Intelligence, 5(8), 892–901.
38. Timmer, S. T. et al. (2016). A two-phase method for extracting explanatory arguments from Bayesian networks. International Journal of Approximate Reasoning, 80, 475–494.
39. Touvron, H. et al. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
40. Vlek, C. S., Prakken, H., Renooij, S., & Verheij, B. (2016). A method for explaining Bayesian networks for legal evidence with scenarios. Artificial Intelligence and Law, 24(3), 263–290.
41. Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.
42. Wei, J. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824–24837.
43. White, J. et al. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382.
44. Zhang, Y. et al. (2023). Siren's song in the AI ocean: A survey on hallucination in large language models. arXiv preprint arXiv:2309.01219.
45. Zhao, H. et al. (2024). Explainable AI meets LLMs: A survey on explainable artificial intelligence in the era of large language models. arXiv preprint arXiv:2402.04616.
46. Zhao, W. X. et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.
47. Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2020). Transparency, auditability and explainability of machine learning models in credit scoring. arXiv preprint arXiv:2009.13384.
48. Calabrese, R., & Martin-Barragan, B. (2023). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(2), 755–769

Author contributions
Conceptualization: Lamis Olaiky
Methodology: Lamis Olaiky
Software / Pipeline Implementation: Lamis Olaiky
Validation & Investigation: Lamis Olaiky
Formal Analysis: Lamis Olaiky
Writing – Original Draft: Lamis Olaiky
Competing interests: The author declares that there are no competing interests
Data and materials availability: All data and materials used in the analysis are available within the main text or the supplementary materials

Abbreviations
BN: Bayesian Network
CPT: Conditional Probability Table
DAG: Directed Acyclic Graph
LLM: Large Language Model
XAI: Explainable Artificial Intelligence
SHAP: SHapley Additive exPlanations
LIME: Local Interpretable Model-agnostic Explanations
API: Application Programming Interface

Studying the Effect of Synthetic Zeolite on Properties of Warm Mix Asphalt

دراسة تأثير الزيوليت الصناعي في خواص المجبول الإسفلتي الدافئ

Mahmoud abdullh : Faculty of Civil Engineering- Homs university.
Corresponding: mahmoudaliabdullh@gmail.com
Submitted date: 10 September 2026 - **Accepted date:** 25 October 2026.

ABSTRACT

Warm mix asphalt (WMA) technology allows for lower production Hot Mix Asphalt (HMA) production temperatures by)20-40(°C and achieves economic and environmental benefits through energy savings and reduced emissions of carbon dioxide (CO2) and other gases that pollute the environment and are harmful to the health of workers and employees in this field. The purpose of this research is to evaluate the use of Synthetic zeolite in the manufacture of warm asphalt mixes and studying the changes in properties of warm mix asphalt compared to the properties of hot mix asphalt. The synthetic zeolite, bituminous binder, and aggregates used in this study were characterized and described. Two types of asphalt mixtures were done according to the Marshall method at varying asphalt binder contents (4-7%): Hot mixtures and warm mixtures containing 0.3% of synthetic zeolite by the total weight of the asphalt mix, all asphalt mixtures have been prepared from two asphalt grades (60/70, 85/100) and single aggregate source. the changes in the properties of warm asphalt mixtures modified with Synthetic zeolite were investigated and compared with reference mixtures with indicators such as Density, Marshall stability, Flow and Air voids. Results of Marshall tests indicate the success of the Synthetic zeolite in production of WMA temperatures at 135C°, in which the physical and mechanical properties of the WMA were close to the properties of the hot mix asphalt.

Keywords: Hot mix asphalt, Warm mix asphalt, Synthetic Zeolite, Marshall test.

الملخص

تسمح تكنولوجيا المجبول الإسفلتي الدافئ بتخفيض درجات حرارة انتاج المجبول الساخن بمقدار (20–40)°C، بالإضافة إلى تحقيق مجموعة كبيرة من الفوائد الاقتصادية والصحية والبيئية من خلال الوفر الحاصل بالطاقة وغيرها من الغازات الملوثة للبيئة والمؤذية لصحة العمال والعاملينCo2المستهلكة وتقليل انبعاث غاز ثاني أوكسيد الكربون في هذا المجال. يهدف هذا البحث إلى دراسة إمكانية استخدام الزيوليت الصناعي في إنتاج المجبول الإسفلتي الدافئ، ودراسة تم في هذه الدراسة تحديد خواص الزيوليت .التغيرات الحاصلة في خواصه بالمقارنة مع خواص المجبول الإسفلتي الساخن عند نسب رابط الصناعي والروابط الإسفلتية والحصويات المستخدمة، وتصميم نوعين من الخلطات الإسفلتية وفق طريقة مارشال اسفلتي متغيرة (4-7%): الخلطات الإسفلتية الساخنة المرجعية، والخلطات الدافئة المعدلة بالزيوليت الصناعي عند نسبة وزنية واحدة (0.3%) من الوزن الكلي للمجبول الإسفلتي، حيث تم تصنيع جميع الخلطات الإسفلتية باستخدام نوعين رابط بيتومين (60/70، 85/100) ونوع واحد من الحصويات الكلسية. تم تمت دراسة التغيرات الحاصلة في خواص الخلطات الإسفلتية الدافئة المنتجة باستخدام بالزيوليت الصناعي ومقارنتها مع الخلطات الإسفلتية الساخنة المرجعية باستخدام بارامترات مارشال المعروفة تُشير نتائج اختبارات مارشال إلى نجاح الزيوليت الصناعي في إنتاج المجبول .من ثبات وانسياب وكثافة وقيم الفراغات الهوائية الإسفلتي الدافئ عند درجات حرارة إنتاج 135°C، وبخواص فيزيائية وميكانيكية جيدة وقريبة من خواص المجبول الساخن.

الكلمات المفتاحية: المجبول الإسفلتي الساخن، المجبول الإسفلتي الدافئ، الزيوليت الصناعي، اختبار مارشال.

INTRODUCTION

Warm mix asphalt has been used over the past few decades as an alternative to hot mix asphalt in order to achieve a wide range of environmental, economic, and engineering benefits. These include reducing the temperatures required for the production and compaction of asphalt mixtures, lowering the energy needed for production, and decreasing emissions of gases and fumes. It also improves working conditions in asphalt plants and paving sites, which in turn has a positive impact on workers' health. Additionally, it results in lower energy consumption, enhanced safety, slower aging of the asphalt binder, and improved properties of the asphalt mixture. It allows the mixture to be transported over longer distances while maintaining its workability, enables placement in thicker layers,

facilities opening to traffic within a short period and supports the use of higher percentages of recycled paving materials [1,-4].The mixing temperatures of warm mix asphalt mixtures ranged from 100°C to 140°C compared to the mixing temperatures of 150-180°C for conventional hot mix asphalt. WMA technologies are commonly classified as those that use foaming techniques, organic or wax additives, and chemical additives [5-7].Synthetic zeolite is considered a hydro-thermally mineral that contains about 21% by weight of water, which is released from the zeolite structure as the temperature increases. It is produced by chemical reactions between sodium silicate (Na2SiO3), sodium aluminate (NaAlO2), and various silicate and aluminate-bearing materials such as fly ash, clay minerals, and perlite, under specific conditions of varying temperatures, pressures, and reaction times. The addition of zeolite into the asphalt mix causes the release of the crystalline water and forming a micro-foaming effect in the mix, which leads to the expansion of the binder volume and reduced viscosity. It facilitates better coatings of the asphalt on aggregates at lower temperatures during the production process, transport and the laying of the asphalt mix [8-12]. In 2005, Hurley and Prowell conducted the first laboratory study on the use of synthetic zeolite in production of WMA. Their study included the use of two types of aggregates (granite, and limestone), two types of asphalt binders (PG 64-22 and PG 58-28) and Aspha-Min additives. WMA samples were compacted at temperatures (149, 129, 110, 88)°C, and the mixing temperatures were twenty degrees higher than the compaction temperature, Laboratory tests indicated an improvement in the compaction of WMA samples involving synthetic zeolite regardless of the type of asphalt, the type of aggregates and the compaction temperature. On the other hand, the lower compaction temperature used when producing warm mix asphalt with Aspha-min® may increase the potential for moisture damage. Hence, the addition of 1.5% hydrated lime has resulted in acceptable performance [12]. Mohamed et al. (2025) evaluated the performance of half warm mix asphalt(HWMA) prepared with Synthetic zeolite and addressed the gap in comparative analysis between HWMA and HMA by conducting comprehensive tests, including indirect tensile strength (ITS), loss of stability and scanning electronic microscope (SEM). This paper concluded that HWMA modified with 10% Synthetic zeolite, 5% epoxy and 10% stearic acid achieved higher ITS values and reduced susceptibility compared to HMA. It was also concluded that HWMA's potential as a sustainable and high performance alternative to HMA without compromising mechanical properties [13].Sağlık et al. (2016) used Synthetic zeolite (Aspha-Min) in production WMA and evaluated the properties of warm mix asphalt. They used bitumen 50/70 and added Synthetic zeolite at 0.3% by total weigh of asphalt mixtures. The results indicated that the performance of WMA was comparable to hot mix asphalt in terms of rutting and fatigue resistance, while a slight reduction in the moisture resistance of the warm mixtures [14].Bichajło et al. (2025) used zeolites to enable reconciliation of technological requirements of mastic asphalt and asphalt roof isolation. the mixtures containing 0 and 5% of two types of zeolites and asphalt binders 35/50 or elastomer-SBS-modified asphalt binder PMB 25/55-60 were used in their study. Laboratory results revealed that the addition of a 5% amount of zeolite by asphalt mass makes it possible to decrease the mastic asphalt laying temperature by 30 °C, which could be very important to achieve environmental and economic goals and pavement durability [15].Some studies found that the moisture resistance of WMA formulated with zeolite decreased compared to hot asphalt mixes when the production temperature of the asphalt mix decreased [16,17]. While other studies indicated an improvement in the indirect tensile strength modulus (ITSR) of warm asphalt mixes containing zeolite as they decreased in production temperature [18].

objectives:

WMA is a new topic in Syria, and there aren't a lot of studies in the Arab world of production WMA concrete using Synthetic zeolite. Therefore, the aims of this work are:

1) Investigate and evaluate the influence of Synthetic zeolite additives on physical and mechanical properties of warm mix asphalt.
2) Achieving economic goals through cost savings in asphalt mix production due to the reduced energy and fuel required.
3) Production environmentally friendly asphalt mix by reducing the emission of gases and odors from asphalt plants, in addition to preserving the health of workers in this field.

MATERIAL CHARACTERISTICS

1.Asphalt binder:

Two Asphalt binder grades were used in this study (60/70, 85/100) obtained from Homs and Baniyas Refinery Company. The Physical Characteristics of Asphalt binder were tested according to standard specifications are listed in table 1.

Table 1: Physical Characteristics of asphalt binder				
Test	Test Result		ASTM Specifications	
			60/70	85/100
Penetration (ASTM D.5)	61.5	98	60-70	85-100
Ductility (ASTM D.113), cm	150	150	100> cm	
Softening Point (ASTM D.36), °C	47.8	45.8	54-46	43-51

Aggregates:

The coarse and fine aggregates were supplied from Hsia City. The gradation of the test specimens was performed in accordance with ASTM of surface course, Fig.1 shows gradation of these aggregates. It has been selected and incorporated in preparing all asphalt concrete mixes used in this study. The mechanical and physical characteristics of used aggregates have been tabulated in table 2.

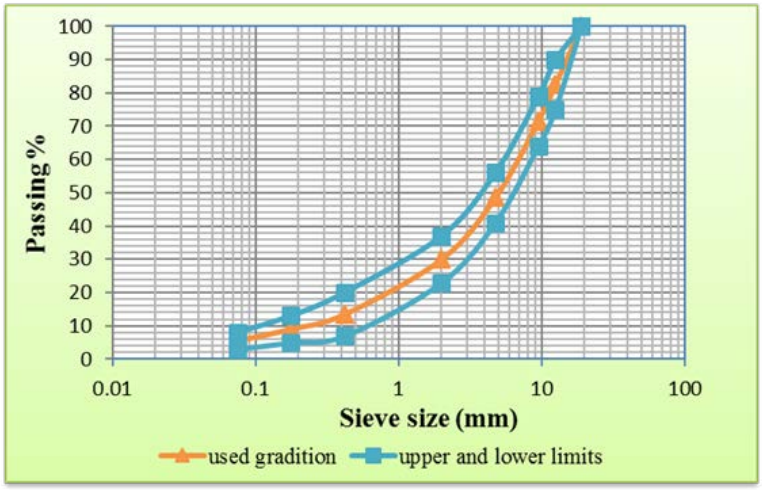


Fig 1: Aggregates gradation of asphalt concrete

Table 2: Mechanical and physical characteristics of aggregates	
Test	Test Result
Resistance of Abrasion (ASTM C.535)	18.7%
Sand Equivalent (ASTM D.2419)	74.4%
Effective specific gravity of aggregate	2.638

3.Synthetic zeolite:

•Synthetic zeolite samples used in this study imported from the People›s Republic of China. Table. 3 presents the chemical analysis of the synthetic zeolite samples, indicating that synthetic zeolite consists mainly composed of silica (SiO2) and alumina (Al2O3) that constitute the zeolitic structure in addition to its high crystalline water content.

Table 3: Chemical composition of Synthetic zeolite							
Oxide	Sio_2	Al_2o_3	Cao	Tio2	K2o	Na2o	H2o
Wt%	41.03	28.13	0.05	9.53	0.21	0.05	21

•Experimental Methods:

There are no local specifications to production WMA mixtures and a possible reduction in WMA mixing and compaction temperatures depends on trial studies, type of additives and recommendations of companies producing it. However, zeolite additives allow reduction in the production placement temperatures of (20-30)ºC with the HMA[19-

CC BY

[22]. The possibility of using synthetic zeolite to produce WMA mixtures has been investigated. WMA samples were manufactured according to the Marshall test (ASTM.D 1559) at a mixing temperature of 135ºC and a compaction temperature of 125ºC. synthetic zeolite additives were crushed with grain size of 0-0.075mm (passing from No.200) and added to aggregates at the rate of 0.3% of the mix at the same time with asphalt and mixed.



Fig. 2: Production of WMA samples

RESULTS

Effect of synthetic zeolite on asphalt mixtures characteristics:

•Figures (3-6) show the design results of reference hot asphalt mixes and warm mixes containing 0.3% synthetic zeolite at varying asphalt contents (4-7%), where Marshall curves were drawn and the ideal asphalt ratio was determined on each curve, achieving maximum density, maximum stability, and average technical requirements for flow and air voids. Table (4) shows the properties of hot and warm asphalt mixes at the optimum asphalt ratios.

•The changes in the properties of warm asphalt mixtures modified with synthetic zeolite were investigated and compared with reference mixtures. Marshall Properties at optimum asphalt content results has been tabulated as shown in Table. 4.

Table. 4: Marshall properties of HMA and WMA mixtures at optimum asphalt content

Mix type	Optimum asphalt content (%)	Marshall Stability (Kg)	Marshall flow (mm)	Density (g/cm3)	Voids in total mix (%)	Marshall quotient (KN/mm)
$HMA_{60/70}$	5.4	1420.67	3.49	2.346	3.6	4.07
$HMA_{85/100}$	5.1	1306	3.70	2.343	4	3.52
$WMA_{60/70}$	5.8	1262	3.64	2.319	4.2	3.05
$WMA_{85/100}$	5.6	1101.33	3.79	2.320	4.3	2.9

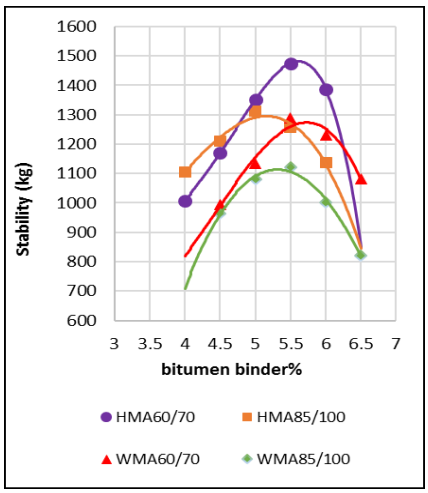


Fig. 3: stability values for asphalt mixture types

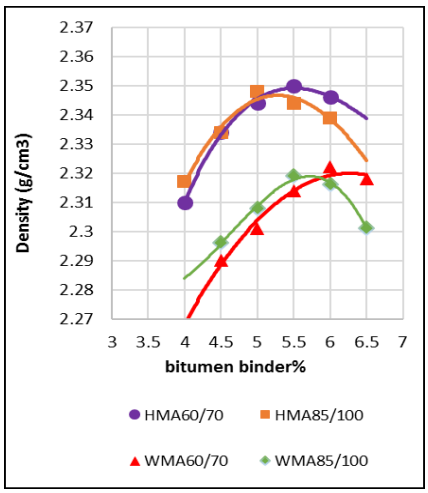


Fig. 4: Density values for asphalt mixture types

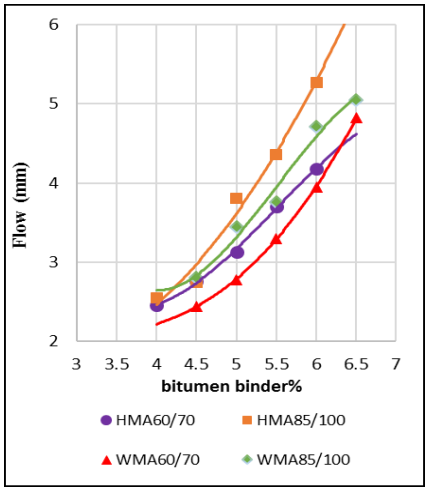


Fig. 5: flow values for asphalt mixture types

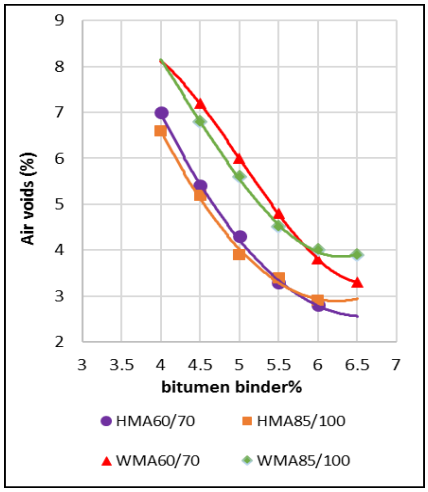


Fig. 6: Air voids values for asphalt mixture types

Fig. 7 shows the test results of Marshall stability for all mixtures. The results show that the Marshall stability for all mixtures were acceptable as compared with the specification limits (Min. 800kg) [ASTM].

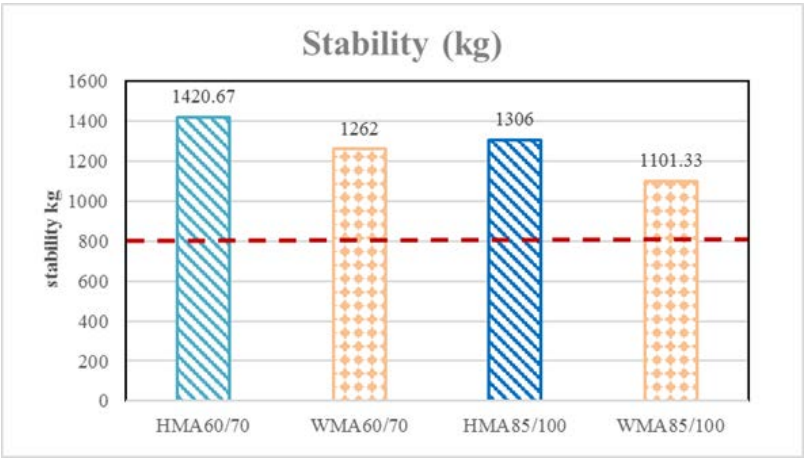


Fig. 7: Marshall stability values of different asphalt mixture types

Marshall stability values indicate a decrease in the stability values of warm asphalt mixes designed using synthetic zeolite. Marshall stability values decreased by 11% for mixes designed using asphalt 60/70 and by 16% for mixes designed using asphalt 85/100. This decrease in stability values is attributed to the reduction in heat energy used during its production. Consequently, the role of synthetic zeolite comes into play in maintaining good stability values within the limits of the acceptance specifications (<800kg). The good stability values can be attributed to the role of synthetic zeolite in creating foam in the asphalt mix and improving its workability. Figure. 8 shows the flow of warm and hot asphalt mixes and Marshall quotient coefficients at optimum asphalt binders, where the results

were within the acceptable specifications limits of (2-4)mm.These changes were accompanied by a decrease in the resistance to permanent deformation of the warm mixes, expressed by a decrease in Marshall quotient coefficients, while remaining within the acceptable specifications limits.

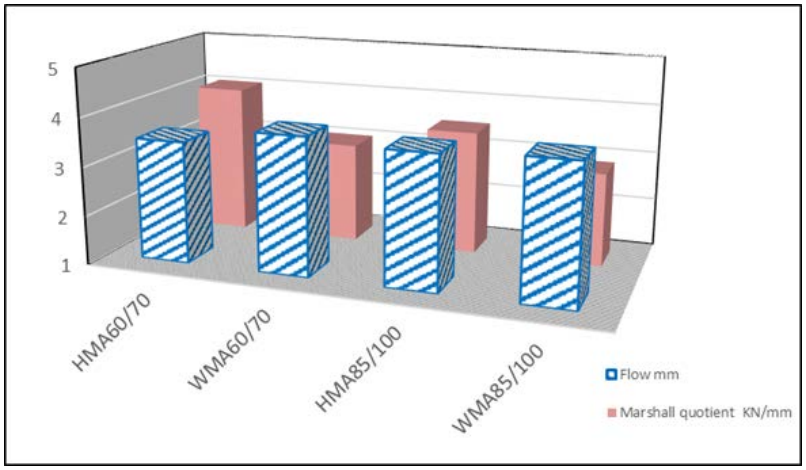


Fig. 8: Flow and Marshall quotient values of different asphalt mixture types

Figure. 9 shows the changes in the values of air voids for warm asphalt mixes where it is noted that the values of air voids for warm mixes increase compared to hot mixes, while remaining within the limits of the acceptable specifications (3-5)%.

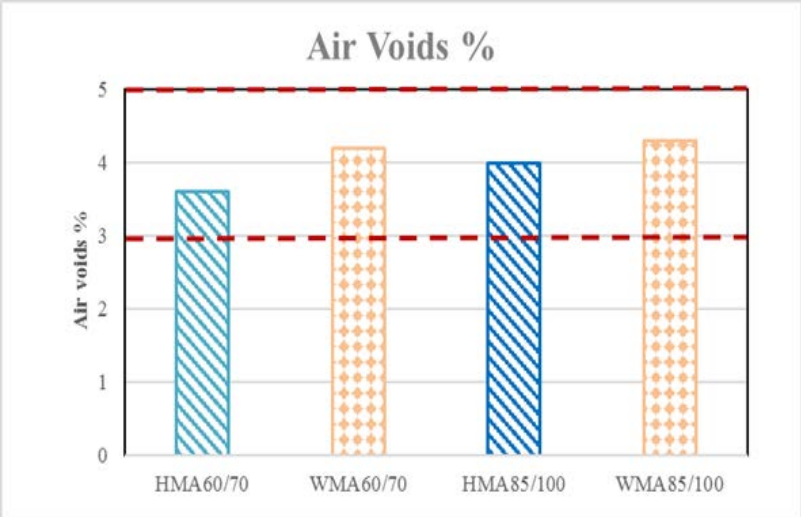


Fig. 9: Air voids values of different asphalt mixture types

Warm asphalt mixes showed a significant increase in the required asphalt binder ratios, as shown in Figure. 10.

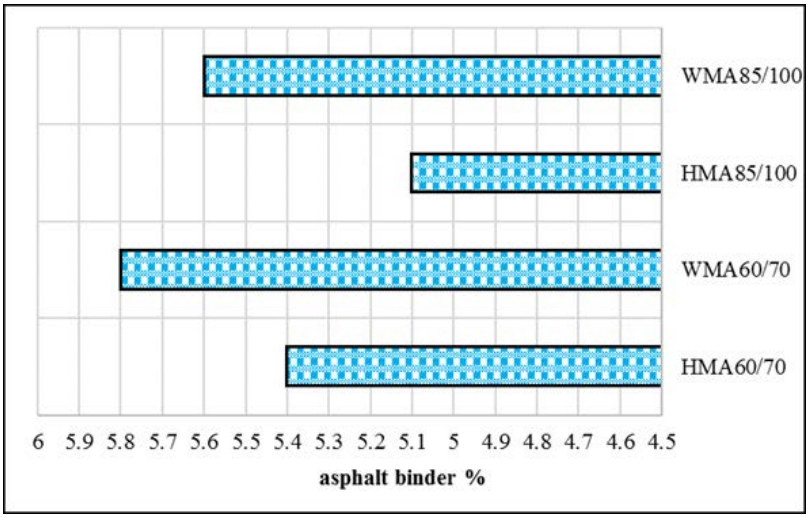


Fig. 10: Optimum asphalt content of different asphalt mixture types

DISCUSSION

The results of Marshall tests indicate to decrease in stability and density and increase in flow and air vales of warm mix asphalt. Marshall stability values decreased by 11% for mixes designed using asphalt 60/70 and by 16% for mixes designed using asphalt 85/100. it could be due to the lower thermal energy used during production. Consequently the role of synthetic zeolite becomes important in maintaining acceptable stability values within specification limits (800>kg). The relatively good stability values can be attributed to the ability of synthetic zeolite as a foaming additives in improve workability of asphalt concrete. These results are consistent with previous studies related to synthetic zeolite additives [23-25].The Marshall quotient (MQ) expresses the ratio between stability and flow, and is used to determine the resistance of asphalt mixtures to permanent deformation (rutting). The higher the value of this quotient, the greater the resistance of asphalt mixtures to permanent deformation. The result showed a decrease in the permanent deformation resistance of the warm mixes, expressed as a decrease in Marshall stiffness indices, while remaining within the acceptable specification [26,27].Warm asphalt mixes showed an increase in the required asphalt binder ratios. The required binder ratio for $WMA_{60/70}$ warm mixes increased by 7%, and by 10% for $WMA_{85/100}$ mixes. These changes are consistent with the results of previous studies regarding the increased need for binder with decreasing production temperatures [23-25]. The changes in the properties of warm asphalt mixtures at optimum asphalt content was investigated. The results showed good physical and mechanical properties of WMA compared with HMA. However, the stability decreased by 16% while maintaining flow values and air voids within the limits of the specifications (flow: 2-4mm, air voids 3-5%) [ASTM]. The changes in void values demonstrate the role of synthetic zeolite in maintaining the compactability of warm asphalt mixes. This result is consistent with the results of previous studies [12,28].

CONCLUSIONS AND RECOMMENDATIONS

Conclusions:

Based on the results of studying the effect of synthetic zeolite in asphalt concrete mixtures, it can be deduced:

1.The successful use of synthetic zeolite in production warm mix asphalt with good physical and mechanical properties within specification limits.

2.An increase in the optimum asphalt binder content was observed in certain WMA mixtures, depending on the type of additive and aggregate characteristics, which can be attributed to reduced binder absorption by aggregates at lower production temperatures

3.Results of Marshall tests showed decreasing in stability values of warm mix asphalt compared to hot mix asphalt, while flow values and air voids decreased.

RECOMMENDATIONS

1.Study the performance of the warm mix asphalt involving synthetic zeolite such as rutting potential and fatigue life.

2.Study the effect of other additives (organic and chemical additives) on the characteristics of asphalt binder and WMA mixtures.

REFERENCES

1. **Sihankar P, Jhunarani O, RIMA S. Study On Rutting Performance Of Synthetic Zeolite Based Warm Mix Asphalt Using Railway Carriage Workshop Incineration Ash As Filler. Innovative Infrastructure Solutions, 2024, 9(11).**
2. **Bocci M, Virgili A, Grilli A, Pannunzio V. Experimental Analysis Of The Potentialities And Limits Of Asphalt Mixtures By Adding Zeolite, 4th International Siiv Congress – Palermo (ITALY), 12-14, 2007.**
3. **Caputo P, Abe A, Loise V, Porto M, Calandra O P, Angelico R, Rossi, C. The Role of Additives in Warm Mix Asphalt Technology: An Insight into Their Mechanisms of Improving an Emerging Technology. Nanomaterials (Basel). 2020 Jun 19;10(6):1202.**
4. **Diab A, Sangiorgi C, Ghabchi R, Zaman M, Wahaballa A. Warm Mix Asphalt (WMA) technologies: Benefits and drawbacks—a literature review. Functional Pavement Design – Erkens et al. (Eds) © 2016 Taylor & Francis Group, London, 2016.**
5. **Wasilewska M, PacholakX R, Gierasimiuk P, Gardziejczyk W, Woszuk A, Bichajlo L, Siwowski, T. The Effect of a Zeolite Addition to Modified Bitumen on the Properties of Stone Matrix Asphalt Lärmarmer Mixtures Produced as Warm Mix Asphalt. Materials (Basel). 2024.**
6. **Mahmood H, AL-Hadidy A. Mechanical behavior of Synthetic zeolite-warm mix asphalt containing sulfur waste as a sustainable filler. International journal of pavement research and technology, 2025.**
7. **MO L, LI X, FANG X, HUURMAN M, WU S. Laboratory investigation of compaction characteristics and performance of warm mix asphalt containing chemical additives. Construction and Building Materials, VOL. 37, 239–247, 2012.**
8. **Liu N, Liu L, Li M, Sun L. Effects of zeolite on rheological properties of asphalt materials and asphalt – filler interactionability. Construction and building materials, 2023.**
9. **Mattaparthi S, Vootkuri N, Biligiri K. Zeolite-Like Warm-Mix Asphalt Additive: Dosage Optimization Based on Asphalt Binder Characterization. Journal of Testing and Evaluation, 2023.**
10. **Belc A, Ciutina A, Buzatu R, Belc F, Costescu C. Environmental Impact Assessment of different Warm Mix Asphalts. Sustainability, 2021, 13(12), 11869.**
11. **Franus M, Wdowin M, Bandura L, Franus W. Removal of environmental pollutions using zeolites from fly ash: A review. Fresenius Environmental Bulletin. VOL. 24, NO3a, 854–866, 2015.**
12. **Hurley C, Prowell B. D. Evaluation of Aspha-Min**

zeolite for use in warm mix asphalt National Center for Asphalt Technology, NCAT Auburn University, 2005.
13. Mohamed H, Jhunarani O, Elsayed A, Abed el-latif M, Mostafa, A. Performance evaluation of half warm asphalt mixture prepared with natural and nano Synthetic zeolite, chemical additives and polymers. JES. Journal of Engineering sciences, 2025.
14. Sağlik M, Demir B, Oral H, Ozturk T, Güngör A. Development and performance of new additives for warm mix asphalt technologies. 6th Eurasphalt & Eurobitume Congress, Prague, Czech Republic, 2016.
15. Bichajło L, Gardziejczyk W, Gierasimiuk P, Kołodziej K, Kowalski K, Malinowski S, Siwowski T, Wasilewska M. Effect of the Addition of Zeolites on the Resistance to Permanent Deformations of Mastic Asphalt Bridge Pavement. Materials (Basel), 2025.
16. Sanchez-Alonso E, Vega-Zamanillo A, Castro-Fresno D, DelRio-Prat M. Evaluation of compactability and mechanical properties of bituminous mixes with warm additives. Constr. Build. Mater, 2011.
17. Woszuk A, Franus W. Properties of the Warm Mix Asphalt involving clinoptilolite and Na-P1 zeolite additives. Constr. Build. Mater, 2016.
18. Judycki J, Stienss M. Research on Warm mix asphalt at lower Temperature, Gdansk University of Technology, Gda´ nsk, Poland, 2011.
19. Mansfeld R, Barth R, Beer F, Breitbach P, Gogolin D, Pass F, Radenberg M, Riebesehl G, Sadzulewsky S, Wölfle H. Leitfaden Temperatur abgesenkte Asphalte, Deutscher Asphaltverband e.V, Bonn, Germany, 2009.
20. Gandhi T, Xiao F, Amirkhanian SN, Estimating indirect tensile strength of mixtures containing anti-Stripping agents using an artificial neural network approach. Int J Pavement Res Technol, 2009, 2(1):1–12.
21. Kristjansdottir O, Muench S, Michael L, Burke G, Assessing potential for warm-mix asphalt technology adoption. J Transport Res Board, 2007.
22. Xiao F, Zhao W, Gandhi T, Amirkhanian SN, Influence of Anti-Stripping Additives on Moisture Susceptibility of Warm Mix Asphalt Mixtures. J Mater Civil Eng. 2010.
23. Al-Jumaili M, Al-Jameel H, Influence of Selected Additives on Warm Asphalt Mixtures Performance. Kufa Journal of Engineering, 2015.
24. ALONSO A, TEJEDA E, MORENO F, RUBIO M, MEDEL E. A comparative study of natural zeolite and synthetic zeolite as an additive in warm asphalt mixes, Materiales de Construcción, VOL. 63, 310, 195-217.
25. Sentosa L, Subagio B, Rahma H, Yamin R, Warm mix asphalt mixture using modified asbuton semi extraction modify and synthetic zeolite additive. MATEC Web of Conferences, 276, 03003 (2019).
26. Massara A, Bulgis, Setiawan A, The Marshall Characteristics of Mixed Asphalt Concrete-Wearing Course Using Kudo Gum Additive. Civil Engineering and Architecture 12(1): 109-120, 2024.
27. Almeida A, Sergio M, Evaluation of the Potential of Sasobit REDUX Additive to Lower Warm-Mix Asphalt Production Temperature. Materials, Portugal, 2019.
28. Yang L, Tan Y, Compacting Properties of Zeolite Based Warm Asphalt Mixture. Advanced Engineering Forum, 2012, Vol. 5, 299-304.
29. ASTM. Standard Specification for Hot-Mixed, Hot-Laid Bituminous Paving Mixtures. Annual Book of ASTM Standard, D3515-01, Vol. 04.03, 2003.

Author contributions: Mahmoud abdullh

Ethical Approval: Oral informed consent was obtained from parents/guardians and assent from adolescents, in accordance with the Declaration of Helsinki and CIOMS guidelines for research in conflict settings. Consent to publish Not applicable (the manuscript contains no individual person's identifiable data).

Data Availability Statement

All data are available in the main text or the supplementary materials.

Competing interests: Authors declare that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.