مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (10) العدد (11) 2026
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (10) No. (11) 2026

كشفُ نُعاسِ السَّائقِ باستخدامِ نموذجٍ هجينٍ قائمٍ على Vision Transformer وَالشَّبكاتِ العصبيّةِ المُتكرّرةِ

أ.د .م. يعرب ديوب *

د. م. فاطمة قداد **

م . آيه عباس ***

(تاريخ الإيداع 2026/3/16 . قُبِل للنشر في 2026/6/2)

□ ملخّص □

تُعدُّ حالةُ نُعاسِ السَّائقِ أثناءَ القيادةِ من العواملِ الرَّئيسةِ المُؤدّيةِ إلى الحوادثِ المُروريَّةِ، لما لها من تأثيرٍ مُباشرٍ على مُستوى الانتباهِ وزمنِ الاستجابةِ واتّخاذِ القرارِ. يهدفُ هذا البحثُ إلى تطويرِ نظامٍ ذكيٍّ لكشفِ نُعاسِ السَّائقِ بالاعتمادِ على تحليلِ تسلسلِ الإطاراتِ الوجهيَّةِ المُلتقطةِ من كاميرا داخلَ المركبةِ باستخدامِ تقنيَّاتِ التَّعلُّمِ العميقِ. يعتمدُ النِّظامُ المُقترَحُ على نموذجِ Vision Transformer (ViT) لاستخراجِ الميزاتِ المكانيّةِ العميقةِ من إطاراتِ الوجهِ، ثمّ دمجِ هذهِ الميزاتِ مع نماذجِ الشَّبكاتِ العصبيَّةِ المُتكرّرةِ، وهي Long Short-Term Memory (LSTM) و Gated Recurrent Unit (GRU) و Bidirectional LSTM (BiLSTM)، بهدفِ مُعالجةِ البُعدِ الزَّمنيّ وتمثيلِ تطوّرِ حالةِ السَّائقِ عبرَ الزَّمنِ . تمَّ تقييمُ أداءِ النِّظامِ باستخدامِ قاعدةِ البياناتِ العامَّةِ National Tsing Hua University – Driver Drowsiness Detection Dataset (NTHU-DDD)، مع تقسيمِ البياناتِ إلى مجموعاتِ تدريبٍ وتحقّقٍ واختبارٍ وفقَ بروتوكولٍ تجريبيٍّ واضحٍ. أظهرتِ النَّتائجُ التَّجريبيّةُ أنَّ نموذجَ ViT-LSTM حقَّقَ أعلى دقّةٍ في التَّصنيف، في حينٍ تميَّزَ نموذجُ ViT-GRU بزمنِ استجابةٍ أقلّ، ممّا يجعلهُ أكثرَ مُلاءمةً للتَّطبيقاتِ الآنيّة. تُؤكّدُ هذهِ النَّتائجُ فعاليَّةَ البُنيةِ المُقترَحةِ في تحقيقِ توازنٍ مُناسبٍ بينَ دقّةِ الكشفِ والكفاءةِ الزَّمنيّةِ، ممّا يعزّزُ إمكانيّةَ اعتمادِها ضمنَ أنظمةِ مُساعدةِ السَّائقِ الذَّكيَّةِ Advanced Driver Assistance Systems (ADAS).

الكلماتُ المفتاحيَّة: كشفُ نُعاسِ السَّائقِ، Vision Transformer، الشَّبكاتُ العصبيَّةُ المُتكرّرةُ، LSTM، GRU، BiLSTM، أنظمةُ مُساعدةِ السَّائقِ.

* أستاذٌ في جامعةِ طرطوس – كليَّةُ هندسةِ تكنولوجيا المعلوماتِ وَ الاتّصالاتِ – قسمُ هندسةِ تكنولوجيا المعلوماتِ (مشرف رئيسي).

** مُدَرِّسة في جامعةِ طرطوس – كليَّةُ هندسةِ تكنولوجيا المعلوماتِ وَ الاتّصالاتِ – قسمُ هندسةِ تكنولوجيا المعلوماتِ (مشرف مشارك).

*** طالبةُ ماجستير في جامعةِ طرطوس– كليَّةُ هندسةِ تكنولوجيا المعلوماتِ والاتّصالاتِ – قسمُ هندسةِ تكنولوجيا المعلوماتِ.

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (10) العدد (11) 2026
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (10) No. (11) 2026

Driver Drowsiness Detection Using a Hybrid Vision Transformer and Recurrent Neural Network Model

Prof .Eng. Yaroub Dayoub [*]
Dr.Eng. Fatima Kadad [**]
Eng. Aya Abbas [***]

(Received 16/3/2026. Accepted 2/6/2026)

□ ABSTRACT □

Driver drowsiness during driving is considered one of the major factors contributing to traffic accidents, as it directly affects attention level, reaction time, and decision-making ability. This paper aims to develop an intelligent driver drowsiness detection system based on the analysis of facial frame sequences captured by an in-vehicle camera using deep learning techniques. The proposed system relies on the Vision Transformer (ViT) model to extract deep spatial features from facial images, which are then integrated with recurrent neural network models, namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM), to model temporal dependencies and represent the evolution of the driver's state over time. The system performance was evaluated using the public National Tsing Hua University - Driver Drowsiness Detection Dataset (NTHU-DDD), with the data divided into training, validation, and testing sets following a clear experimental protocol. Experimental results indicate that the ViT-LSTM model achieves the highest classification accuracy, while the ViT-GRU model demonstrates lower response time, making it more suitable for real-time applications. These findings confirm the effectiveness of the proposed architecture in achieving a balance between detection accuracy and temporal efficiency, highlighting its potential applicability within intelligent Advanced Driver Assistance Systems (ADAS).

Keywords: Driver drowsiness detection, Vision Transformer, recurrent neural networks, LSTM, GRU, BiLSTM, Advanced Driver Assistance Systems.

[*] Professor in Information and Communication Technology Engineering Faculty, Information Technology Department (Main Supervisor).
[**] Lecturer in Information and Communication Technology Engineering Faculty, Information Technology Department (Co-supervisor).
[***] Master Student in Information and Communication Technology Engineering Faculty, Information Technology Department.

1. مُقدّمة :

تُعدّ حوادثُ السّيرِ من أبرزِ أسبابِ الوفياتِ والإصاباتِ عالميّاً، إذ تشيرُ تقاريرُ مُنظّمةِ الصّحّةِ العالميّةِ إلى أنّ ما يقارب 1.19 مليون شخصٍ يفقدون حياتهم سنويّاً نتيجةَ الحوادثِ المُروريّةِ، إضافةً إلى عشراتِ الملايينِ من الإصاباتِ غيرِ المُميتةِ (World Health Organization,2023). ويُعدّ العاملُ البشريّ السَّببَ الرَّئيسَ لمُعظمِ هذهِ الحوادثِ، ويأتي نُعاسُ السَّائقِ ضمنَ أهمّ هذهِ العواملِ لما لهُ من تأثيرٍ مُباشرٍ على الانتباهِ وزمنِ الاستجابةِ واتّخاذِ القرارِ أثناءَ القيادةِ(National Highway Traffic Safety Administration,2023). تشيرُ الدِّراساتُ إلى أنّ القيادة في حالةِ النُّعاسِ تؤدّي إلى تدهورٍ كبيرٍ في الأداءِ الإدراكيّ والحركيّ، مُماثلٍ للتَّأثيراتِ النَّاتجةِ عن القيادةِ تحتَ تأثيرِ الكحولِ، كما تزيدُ بشكلٍ ملحوظٍ من خطرِ الحوادثِ المُروريّةِ (Bioulac et al.,2022). ولذلكَ يُعدُّ الكشفُ المُبكّرُ عن النُّعاسِ مكوّناً أساسيّاً في أنظمةِ السَّلامةِ المُروريّةِ الحديثةِ وأنظمةِ مُساعدةِ السَّائقِ المُتقدّمةِ. يندرجُ كشفُ نُعاسِ السَّائق ضمنَ مجالاتِ الرُّؤيةِ الحاسوبيّةِ والتَّعلّمِ العميقِ، حيثُ تعتمدُ الأنظمةُ الحديثةُ على تحليلِ مُؤشّراتٍ بصريّةٍ مثلُ حركاتِ العينِ، ومُعدّلِ الرَّمشِ، ووضعيَّةِ الرَّأسِ، وتعابيرِ الوجهِ لتقديرِ حالةِ اليقظةِ لدى السَّائق وقد أسهمَ التَّطوّرُ الكبيرُ في تقنيَّاتِ التَّعلّمِ العميقِ في تحقيقِ تحسيناتٍ ملحوظةٍ مُقارنةً بالطَّرائقِ التَّقليديَّةِ المُعتمدةِ على الميزاتِ اليدويّةِ .في هذا السّياقِ، استُخدمتِ الشَّبكاتُ العصبيّةُ الالتفافيّةُ Convolutional Neural Networks (CNN) على نطاقٍ واسعٍ لاستخراجِ الميزاتِ المكانيّةِ من الصُّورِ، بينما استُخدمتِ الشَّبكاتُ العصبيّةُ المُتكرّرةُ(RNN) Recurrent Neural Networks ونماذجُها المُتقدّمةُ مثل Long Short-Term Memory (LSTM) و Gated Recurrent Unit (GRU) و Bidirectional LSTM (BiLSTM) لتحليلِ العلاقاتِ الزَّمنيّةِ في تسلسلاتِ الفيديو. ومع ذلكَ، فإنَّ اعتمادَ هذهِ النَّماذجِ على البُنيةِ الالتفافيّة التَّقليديّة قد يحدُّ من قُدرتها على تمثيلِ العلاقاتِ البصريّةِ طويلةِ المدى بدقّةٍ عاليةٍ. ومع ظهورِ نماذجِ مُحوّلاتِ الرُّؤية المُعتمدةِ على آليّةِ الانتباه الذَّاتي، أصبحَ بالإمكانِ استخراجُ تمثيلاتٍ بصريّةٍ أكثرَ قُدرةٍ على التقاطِ العلاقاتِ المكانيّةِ المُعقّدةِ مُقارنةً بالشَّبكات الالتفافيّة التَّقليديّة (Dosovitskiy et al.,2021). ورغمَ ذلكَ، فإنَّ توظيفَ هذهِ النَّماذج في كشفِ نُعاسِ السَّائقِ، ولا سيّما عندَ دمجها مع نماذجِ التَّحليلِ الزَّمنيّ، ما يزالُ محدوداً نسبيّاً في الأدبيّاتِ العلميّةِ. انطلاقاً من ذلكَ، يهدفُ هذا البحثُ إلى تطويرِ نظامٍ ذكيٍّ لكشفِ نُعاسِ السَّائقِ يعتمدُ على دمجِ نموذجِ Vision Transformer (ViT) لاستخراجِ الميزاتِ المكانيّةِ مع نماذجِ الشَّبكاتِ العصبيّةِ المُتكرّرةِ لمُعالجةِ البُعدِ الزَّمنيّ، بهدفِ تحسينِ دقّةِ الكشفِ ودعمِ أنظمةِ مُساعدةِ السَّائقِ الذَّكيّةِ في البيئاتِ الواقعيّةِ .

وعلى الرَّغمِ من التَّقدُّمِ المَلحوظِ في تقنيَّاتِ كشفِ نُعاسِ السَّائقِ، لا تزالُ العديدُ من الدِّراساتِ تُعاني من أحدِ اتِّجاهين: إمّا التَّركيزُ على التَّمثيلِ المكانيّ دونَ استثمارٍ كافٍ للبُعدِ الزَّمنيّ، أو اعتمادُ نماذجَ زمنيَّةٍ مع تمثيلٍ بصريٍّ محدودٍ، إضافةً إلى ارتفاعِ التَّعقيدِ الحسابيّ في بعضِ المُقارباتِ الحديثةِ.

وانطلاقاً من هذهِ الفجوةِ، تتمثّلُ القيمةُ المُضافةُ لهذا البحثِ في:

- اقتراحِ بُنيةٍ هجينةٍ مُتكاملةٍ تجمعُ بينَ نموذجِ Vision Transformer والنَّماذجِ الزَّمنيَّةِ (LSTM، GRU، BiLSTM) ضمنَ إطارٍ مُوحَّدٍ لتحليلِ تسلسلاتِ الفيديو.
- إجراءِ مُقارنةٍ مَنهجيَّةٍ بينَ عدّةِ نماذجٍ زمنيَّةٍ تحتَ نفسِ ظروفِ التَّجربةِ لتحديدِ النَّموذجِ الأكثرِ توازناً من حيثِ الدِّقّةِ والكفاءةِ.

- تحقيقِ توازنٍ فعّالٍ بينَ دقّةِ الكشفِ وزمنِ الاستجابةِ بما يدعمُ تطبيقاتِ الزَّمنِ الحقيقيّ داخلَ المركباتِ.
- اعتمادِ بروتوكولِ تقييمٍ صارمٍ قائمٍ على بياناتِ اختبارٍ غيرِ مرئيَّةٍ (Unseen Test Set) مع فصلٍ تامٍّ بينَ التَّدريبِ والاختبارِ لضمانِ قياسٍ واقعيٍّ لقُدرةِ التَّعميمِ.

1-1 أهميَّةُ البحثِ وَ أهدافِهِ:

تندرجُ أنظمةُ كشفِ نُعاسِ السَّائقِ ضمنَ مجالِ أنظمةِ مُساعدةِ السَّائقِ المُتقدّمةِ ADAS، والَّتي تهدفُ إلى تقليلِ مُعدّلاتِ الحوادثِ المُروريّةِ وتحسينِ مُستوى السَّلامةِ على الطُّرقِ من خلالِ مُراقبةِ حالةِ السَّائقِ بشكلٍ مُستمرٍّ واتّخاذِ إجراءاتٍ تحذيريّةٍ عندَ اكتشافِ مُؤشّراتِ الخطرِ (European Commission,2020) . وتُعدّ مهمَّةُ الكشفِ المُبكّرِ عن النُّعاسِ من أكثرِ المهامّ تعقيداً في هذا المجالِ نظراً للطَّبيعةِ الدّيناميكيّةِ لسُلوكِ السَّائقِ وتغيُّرِ حالتهِ الفيزيولوجيَّةِ والمعرفيّةِ تدريجيّاً عبرَ الزَّمنِ (Bioulac et al.,2022). تكمنُ صعوبةُ تصميمِ هذهِ الأنظمةِ في ضرورةِ تحقيقِ توازنٍ دقيقٍ بينَ دقّةِ الكشفِ وزمنِ الاستجابةِ، إذ إنَّ التَّأخرَ في اكتشافِ حالةِ النُّعاسِ ولو لثوانٍ قليلةٍ قد يؤدّي إلى زيادةٍ كبيرةٍ في خطرِ الحوادثِ المُروريّةِ (National Highway Traffic Safety Administration,2023). لذلكَ فإنَّ الأنظمةَ المُعتمدةَ على تحليلِ إطارٍ مُنفردٍ غالباً ما تكونُ غيرَ كافيةٍ لتمثيلِ التَّطوُّرِ التَّدريجيّ لحالةِ السَّائقِ، ممَّا يستدعي الاعتمادَ على نماذجٍ قادرةٍ على تحليلِ التَّسلسلِ الزَّمنيّ للإشاراتِ البصريّةِ المُستخرجةِ من الفيديو.

تأتي أهميَّةُ هذا البحثِ من كونهِ يقترحُ بُنيةً هجينةً تجمعُ بينَ التَّمثيلِ المكانيّ العميقِ باستخدامِ نموذجِ ViTالقائمِ على آليّةِ الانتباهِ الذَّاتيّ، والتَّحليلِ الزَّمنيّ باستخدامِ نماذجِ الشَّبكاتِ العصبيّةِ المُتكرّرةِ RNN، ممَّا يتيحُ نمذجةَ تطوُّرِ حالةِ السَّائقِ عبرَ الزَّمنِ بدقّةٍ أعلى مُقارنةً بالأنظمةِ التَّقليديّةِ المُعتمدةِ على الشَّبكاتِ الالتفافيّةِ CNNأو على تحليلِ الإطاراتِ بشكلٍ مستقلٍّ. كما يوفّرُ هذا الدَّمجُ تمثيلاً غنيّاً للعلاقاتِ المكانيّةِ داخلَ الصُّورةِ مع القُدرةِ على تتبّعِ الأنماطِ الزَّمنيّةِ طويلةِ الأمدِ المُرتبطةِ بحالةِ النُّعاسِ (Dosovitskiy et al.,2021).

يهدفُ هذا البحثُ إلى تحقيقِ مجموعةٍ من الأهدافِ العلميّةِ والتَّطبيقيّةِ تتمثَّلُ فيما يلي:

- تطويرُ نظامٍ ذكيّ لِكشفِ نُعاسِ السَّائقِ اعتماداً على تحليلِ تسلسلِ الإطاراتِ الوجهيّةِ المُلتقطةِ من كاميرا داخلَ المركبةِ.
- الاستفادةُ من قُدراتِ نموذجِ ViTفي استخراجِ ميزاتٍ مكانيّةٍ عاليةِ المستوى من إطاراتِ الوجهِ.
- دراسةُ تأثيرِ نماذجِ الشَّبكاتِ العصبيَّةِ المُتكرّرةِ (LSTM، GRU، BiLSTM) على دقّةِ الكشفِ وزمنِ الاستجابةِ في بيئاتِ القيادةِ الواقعيّةِ.
- تقييمُ كفاءةِ النّظامِ المُقترحِ باستخدامِ المقاييسِ الإحصائيَّةِ المُعتمدةِ في مجالِ التَّعلّمِ العميقِ والرُّؤيةِ الحاسوبيّةِ.
- اقتراحُ نموذجٍ قابلٍ للتّطبيقِ ضمنَ أنظمةِ مُراقبةِ السَّائقِ الحديثةِ وأنظمةِ السَّلامةِ المُروريَّةِ الذَّكيّةِ في الزَّمنِ الحقيقيّ.

ومن المتوقّع أن تُسهمَ هذهِ المُخرجاتُ في تحسينِ موثوقيّةِ أنظمةِ كشفِ النُّعاسِ وتعزيزِ قابليّتها للتَّطبيقِ في البيئاتِ الواقعيَّةِ ضمنَ أنظمةِ مُساعدةِ السَّائقِ الذَّكيّةِ.

2-1 الدِّراساتُ السَّابقةُ:

شهدَتْ أنظمةُ كشفِ نُعاسِ السَّائقِ تطوُّراً ملحوظاً في السَّنواتِ الأخيرةِ، ويمكنُ تصنيفُ هذهِ الدّراساتِ وفقَ المنهجيّةِ المُتبّعةِ إلى عدّةِ اتّجاهاتٍ رئيسيّةٍ.

أولّاً: الأساليبُ التّقليديّةُ القائمةُ على استخراجِ الميزاتِ اليدويّةِ (Feature-based Methods): اعتمدَتِ الدّراساتُ المُبكّرةُ على تحليلِ مُؤشّراتٍ هندسيّةٍ مُستخرجةٍ من ملامحِ الوجهِ، مثلُ (Eye Aspect Ratio) EAR و (Mouth Aspect Ratio) MAR و PERCLOS.حيثُ قامَتْ دراسةٌ بتطويرِ نظامٍ للكشفِ الفوريّ عن النُّعاسِ يعتمدُ على انحناءِ الجفنِ، مُحقّقاً دقّةَ 95% (Arora et al., 2024) . كما قدَّمتْ دراسةٌ أُخرى نظاماً قائماً على تحليلِ نمطِ العينِ باستخدامِ EAR، مُحقّقةً دقّةَ 94.9% (Maior et al., 2020). تميَّزتْ هذهِ الأساليبُ ببساطتها وسرعتها، لكنَّها أظهرَتْ حساسيّةً عاليةً للتّغيّراتِ في ظروفِ الإضاءةِ واختلافاتِ السَّائقين الفرديّة.

ثانياً: الشَّبكاتُ العصبيَّةُ الالتفافيَّةُ (CNN) ونماذجُ النَّقلِ التَّعلّميّ (Transfer Learning): مع ظهورِ التَّعلّمِ العميقِ، اتّجهتِ العديدُ من الدّراساتِ إلى استخدامِ الشَّبكاتِ الالتفافيّةِ لاستخراجِ الميزاتِ البصريّةِ بشكلٍ آليٍّ ودقيقٍ. وقدْ أظهرَتْ هذهِ النَّماذجُ أداءً مُرتفعاً في مهامِّ كشفِ النُّعاسِ، حيثُ قدَّمتْ دراسةٌ نظاماً يعتمدُ على دمجِ ميزاتٍ وجهيّةٍ مُتعدّدةٍ، مثلُ زمنِ إغلاقِ العينِ، وإمالةِ الرَّأسِ، والتَّثاؤبِ، مُحقّقاً دقّةً بلغَتْ 97.06% باستخدامِ مجموعةِ بياناتٍ NTHU-DDD (Liu et al., 2019) .كما تشيرُ دراساتٌ مُراجِعةٌ حديثةٌ إلى أنَّ نماذجَ CNN المُدعّمةِ بأساليبِ النَّقلِ التَّعلّميّ تُحقّقُ أداءً مُتقدّماً في تحليلِ الصُّورِ مقُارنةً بالأساليبِ التَّقليديَّةِ (Singh and Papanikolopoulos, 2023). وعلى الرَّغمِ من هذا التّفوّقِ، فإنَّ هذهِ النَّماذجَ تُعالجُ كلَّ إطارٍ بشكلٍ مُنفصل، دونَ الأخذِ بعينِ الاعتبارِ العلاقاتِ الزمنيّةِ بينَ الإطاراتِ، وهوَ ما يُعدّ قُصوراً مُهمّاً نظراً للطَّبيعةِ التَّدريجيَّةِ لحالةِ النُّعاسِ.

ثالثاً: الشَّبكاتُ العصبيَّةُ المُتكرّرةُ (RNN) والنَّماذجُ الهجينةُ (CNN+RNN): لمُعالجةِ البُعدِ الزَّمنيّ، وظَّفتْ دراساتٌ أُخرى نماذجَ LSTM و GRU و BiLSTM. وفي عامِ 2025، اقترحتْ دراسةٌ نظاماً يعتمدُ على شبكاتِ الالتواءِ الزَّمنيّةِ (TCN) لتحليلِ التّبعيّاتِ الزَّمنيَّةِ في سُلوكيَّاتِ النُّعاسِ (مدّةُ إغلاقِ العينِ، تكرارُ الرَّمشِ، وضعيَّةُ الرَّأسِ، التَّثاؤبُ)، مُحقّقاً دقّةَ 96.7% على مجموعةِ بياناتِ NTHU-DDD، مُتفوّقاً على نماذجِ LSTM التَّقليديَّة بنسبةِ 2.8% (Sharath Kumar et al., 2025). ومع ذلكَ، يبقى أداءُ هذهِ النَّماذجُ مُعتمداً بشكلٍ كبيرٍ على جودةِ الميزاتِ المكانيّةِ المُدخلةِ.

رابعاً: مُحوّلاتُ الرُّؤيةِ (Vision Transformer – ViT) والنَّماذجُ القائمةُ على آليّةِ الانتباهِ: مع ظهورِ نماذجِ المُحوّلاتِ، قدَّمَتْ دراسةٌ نموذجَ Vis-Net الهجين الَّذي يجمعُ ViT مع شبكاتِ CNN مُتعدّدةٍ (MobileNet-V2, Inception-V3, ResNet152-V2, NASNetLarge, DenseNet)، مُحقّقاً دقّةَ 98.94% في تصنيفِ مُستوياتِ النُّعاسِ المُتعدّدةِ وفقَ مقياسِ Katajima (Phan and Nguyen, 2025) . كما قامتْ دراسةٌ بتطويرِ نموذجٍ هجينٍ CNN-ViT يجمعُ بينَ DenseNet121, VGG16, VGG19, ResNet50 مع ViT، مُحقّقاً دقّةَ 99.76%(Venkateswarlu and Ch, 2025).

خامساً: المنهجيَّاتُ القائمةُ على الإشاراتِ الفيزيولوجيّةِ والكهربائيّةِ (EEG, ECG, EOG): إلى جانبِ المنهجيَّات البصريّة، ظهرَتْ دراساتٌ اعتمدَتْ على إشاراتٍ فيزيولوجيّةٍ. طوَّرتْ دراسةٌ إطاراً يجمعُ CNN و LSTM مع آليَّةِ انتباهٍ ذاتيٍّ لتحليلِ إشاراتِ EEG و EOG، حيثُ تمَّ تقسيمُ حالةِ النُّعاسِ إلى ثلاثِ مراحلٍ فرعيّةٍ (البدايةُ، المدّةُ، النّهايةُ) (Jiao et al., 2025). وقدَّمتْ دراسةٌ شبكةَ ID3RSNet تعملُ على قناةِ EEG واحدةٍ مع تقنيّةِ ECAM لتفسيرِ

القراراتِ النَّموذجيّةِ (Feng et al., 2025). كما أظهرَتْ دراسةٌ حديثةٌ أنَّ الاقترانَ الأمثلَ لخصائصِ PSD alpha و PSD theta يمكنُ أن يحقّقَ تغطيةً 96.1% مع دقّة 95.4%، ممَّا يسمحُ بتقليلِ القنواتِ EEG من ست إلى قناةٍ واحدة (Minhas et al., 2025).

يلاحظُ من خلالِ الاستعراضِ السَّابقِ أنَّ مُعظمَ الدّراساتِ تركّزُ إمّا على التَّمثيلِ المكانيّ (مع إهمالِ البُعدِ الزَّمنيّ) أو على المُعالجةِ الزَّمنيَّةِ (مع اعتمادِ ميزاتٍ مكانيّةٍ محدودةٍ). وعلى الرَّغمِ من التَّفوّقِ الواضحِ لنماذجِ Vision Transformer في التَّمثيلِ المكانيّ، فإنَّ دمجها بشكلٍ منهجيٍّ مع نماذجٍ زمنيّةٍ مُتعدّدةٍ ضمنَ إطارٍ مُقارنٍ يُحقّقُ توازناً بينَ الدّقّةِ والكفاءةِ لا يزالُ محدوداً في الأدبيَّاتِ العلميّةِ، وهوَ ما يشكّلُ الدَّافعَ الرَّئيسيَّ لهذا البحثِ.

2. موادُّ البحثِ وَ طرائقهِ:

2-1 إعدادُ البياناتِ والمُعالجةُ المُسبقَةُ:

تمَّ اعتمادُ قاعدةِ البياناتِ العامَّةِ NTHU-DDD لتقييمِ أداءِ النِّظامِ المُقترَحِ في بيئاتِ قيادةٍ واقعيَّةٍ، نظراً لاحتوائها على تسجيلاتِ فيديو لسائقينَ ضمنَ ظروفِ تصويرٍ متنوّعةٍ تشملُ اختلافَ الإضاءةِ، القيادةَ اللَّيليّةَ والنَّهاريّة، ارتداءَ النّظاراتِ، وتغيّرَ زوايا الوجهِ ووضعيَّاتِ الرَّأسِ، ممَّا يجعلها من أكثرِ قواعدِ البياناتِ استخداماً في أبحاثِ كشفِ نُعاسِ السَّائقِ (Weng et al.,2016). تتضمَّنُ القاعدةُ حالاتٍ سلوكيّةٍ مُتعدّدةٍ مثلُ اليقظةِ الطَّبيعيّةِ، النُّعاسِ التَّدريجيّ، التَّثاؤبِ، وإغلاقِ العينِ لفتراتٍ طويلةٍ، مع توثيقٍ زمنيٍّ دقيقٍ يسمحُ بدراسةِ الأنماطِ الدّيناميكيّةِ للتَّعبِ أثناءَ القيادةِ. وفي هذا البحثِ، تمَّ استخراجُ تتابعاتٍ زمنيَّةٍ ثابتةِ الطُّولِ مُكوّنةٍ من 16 إطاراً متتالياً لكلّ عيّنةٍ بهدفِ تمثيلِ التَّطوّرِ التَّدريجيّ لحالةِ السَّائقِ زمنيَّاً. قُسّمتِ البياناتُ إلى مجموعاتِ تدريبٍ وتحقّقٍ واختبارٍ بنسبةِ 70% و15% و15% على التَّوالي، وفقَ بروتوكولٍ تجريبيّ يضمنُ عدالةَ التَّقييمِ ويمنعُ تداخلَ البياناتِ بينَ مراحلِ التَّعلّمِ. كما طُبّقتْ خطواتُ مُعالجةٍ مُسبقةٍ شملَتْ استخراجَ الإطاراتِ بمُعدّلٍ ثابتٍ، كشفَ الوجهِ وقصّهِ، توحيدَ أبعادِ الصُّورِ بما يتوافقُ مع مُتطلَّباتِ نموذجِ ViT، وتطبيعَ القيمِ اللَّونيّةِ ضمنَ المجالِ [0,1]، قبلَ تجميعِ الإطاراتِ ضمنَ تسلسلاتٍ زمنيّةٍ تمثّلُ المُدخلاتِ النّهائيّةِ للنَّموذجِ.

تسهمُ هذهِ الإجراءاتُ في تحسينِ جودةِ البياناتِ وتعزيزِ قدرةِ النَّموذجِ على تعلّمِ التَّمثيلاتِ المكانيّةِ والزَّمنيَّةِ المُرتبطةِ بمؤشّراتِ النُّعاسِ بشكلٍ أكثرَ دقّةً وموثوقيّةً. يوضّحُ الشَّكلُ (1) عيّناتٍ مُختارةٍ من قاعدةِ البياناتِ المُستخدمةِ .



(الشَّكلُ (1): عيّناتٌ من قاعدةِ بياناتِ NTHU–DDD المُستخدمةِ في تدريبِ النّظامِ المُقترحِ، وتوضّح حالاتِ اليقظةِ والنُّعاسِ في ظروفِ إضاءةٍ وزوايا تصويرٍ مختلفةٍ،(Weng et al.,2016)).

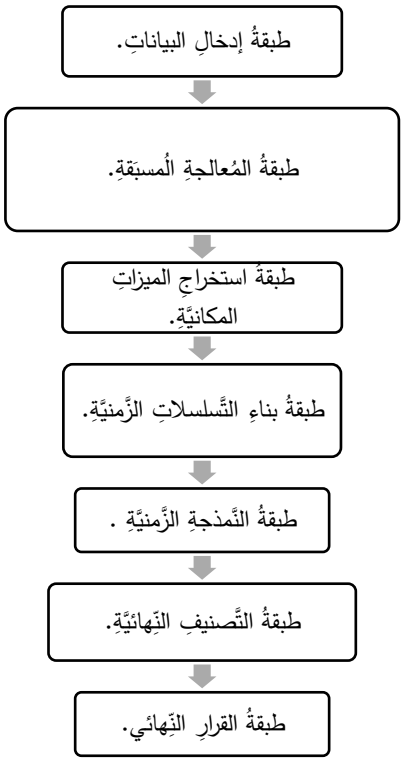
2-2 منهجيَّةُ البحثِ و بُنيةُ النّظامِ المُقترحِ:

يقترحُ هذا البحثُ نظاماً ذكيّاً لكشفِ نُعاسِ السَّائقِ يعتمدُ على تحليلِ تسلسلِ الإطاراتِ الوجهيَّةِ المُلتقطةِ من كاميرا داخلَ المركبةِ باستخدامِ تقنيَّاتِ التَّعلُّمِ العميقِ. وترتكزُ المنهجيَّةُ على بُنيةٍ هجينةٍ تجمعُ بينَ التَّمثيلِ المكانيّ العميقِ والتَّحليلِ الزَّمنيّ، بما يتيحُ نمذجةَ التَّغيُّرِ التَّدريجيّ في حالةِ السَّائقِ والكشفِ المُبكّرِ عن مُؤشّراتِ النُّعاسِ.

تعتمدُ البُنيةُ المُقترحةُ على دمجِ نوعين رئيسيّين من النَّماذجِ العصبيّةِ، حيثُ يُستخدمُ نموذجُ Vision Transformer (ViT) لاستخراجِ الميزاتِ المكانيَّةِ من الصُّورِ اعتماداً على آليَّةِ الانتباهِ الذَّاتيّ، الَّتي تمكّنهُ من تمثيلِ العلاقاتِ البصريّةِ بعيدةِ المدى داخلَ الإطارِ الواحدِ بكفاءةٍ عاليةٍ. في المُقابلِ، تُستخدمُ الشَّبكاتُ العصبيَّةُ المُتكرّرةُ مثل LSTM وGRU وBiLSTM لمُعالجةِ البُعدِ الزَّمنيّ، إذ تقومُ بنمذجةِ تسلسلِ الميزاتِ عبرَ الزَّمنِ من خلالِ آلياتٍ تعتمدُ على الحالاتِ المخفيَّةِ وبوَّاباتِ التَّحكّمِ، ممَّا يسمحُ بتتبّعِ التَّغيُّراتِ التَّدريجيَّةِ في سلوكِ السَّائقِ والتَّغلُّبِ على مُشكلةِ تلاشي التَّدرُّجِ في السَّلاسلِ الطَّويلةِ .يُحقّق هذا الدَّمجُ تكاملاً بينَ التَّحليلِ المكانيّ الدَّقيقِ والتَّحليلِ الزَّمنيّ الدّيناميكيّ، وهو أمرٌ أساسيٌّ في مهامّ كشفِ النُّعاسِ الَّتي تعتمدُ على مُؤشّراتٍ بصريّةٍ مُتغيّرةٍ مع الزَّمنِ.

يوضّحُ الشَّكلُ (2) البُنيةَ العامَّةَ للنّظامِ المُقترحِ، والَّتي تتكوَّنُ من مراحلَ مُتتابعةٍ تبدأُ باستخلاصِ الإطاراتِ الوجهيّةِ من مقاطعِ الفيديو، يليها استخراجُ الميزاتِ المكانيَّةِ باستخدامِ نموذجِ Vision Transformer، ثمَّ تمريرُ تسلسلاتِ الميزاتِ إلى نماذجَ زمنيَّةٍ مُتخصّصةٍ لتحليلِ الاعتماداتِ الزَّمنيَّةِ والتَّعرّفِ على أنماطِ النُّعاسِ المُتدرّجةِ لدى السَّائقِ. يمثّل هذا التَّكاملُ أحدَ الفروقاتِ الجوهريَّةِ عن الدّراساتِ السَّابقةِ الَّتي غالباً ما عالجّتِ البُعدين المكانيّ والزَّمنيّ بشكلٍ مُنفصلٍ.

)



الشَّكلُ (2) : البُنيةُ المعماريَّةُ الطَّبقيّةُ للنِّظامِ المُقترحِ لكشفِ نُعاسِ السَّائقِ ، من إعداد الباحثة)

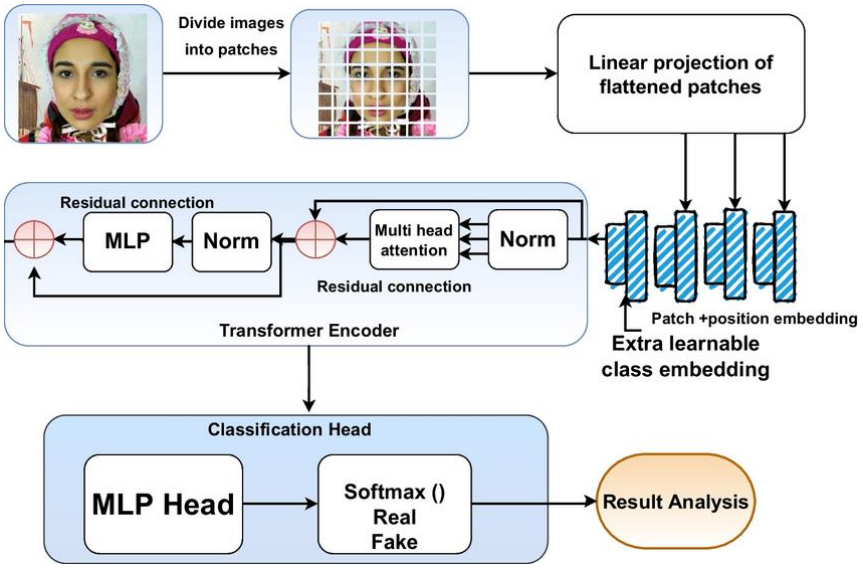
2-3 طبقةُ إدخالِ البياناتِ و المُعالجةِ المُسبقةِ :

تبدأُ المنهجيّةُ بإدخالِ تسلسلاتِ الإطاراتِ الوجهيّةِ المُستخرجةِ من تسجيلاتِ القيادةِ بعدَ تطبيقِ مراحلِ إعدادِ البياناتِ والمُعالجةِ المُسبقةِ الموصوفةِ سابقاً. تمثّلُ هذهِ الإطاراتُ المُدخلاتِ الأساسيّةَ للنّظامِ، حيثُ تُحوَّلُ إلى مصفوفاتٍ رقميّةٍ مُوحّدةِ الأبعادِ ومُتوافقةٍ مع مُتطلّباتِ نماذجِ الرُّؤيةِ العميقةِ، ممَّا يضمنُ استقرارَ عمليَّةِ التَّعلُّمِ وتحسينِ جودةِ التَّمثيلاتِ البصريَّةِ المُستخرجةِ .

2-4 استخراجُ الميزاتِ المكانيَّةِ باستخدامِ Vision Transformer :

بعدَ المُعالجةِ المُسبقةِ، يتمُّ تمريرُ كلَّ إطارٍ وجهيٍّ إلى نموذجِ ViT لاستخراجِ تمثيلاتٍ مكانيّةٍ عاليةِ المُستوى. يعتمدُ هذا النّموذجُ على آليَّةِ الانتباهِ الذَّاتيّ لتمثيلِ العلاقاتِ البعيدةِ داخلَ الصُّورةِ بدلاً من العمليَّاتِ الالتفافيَّةِ التَّقليديّةِ، ممَّا يتيحُ التقاطُ الأنماطِ البصريّةِ الدَّقيقةِ المُرتبطةِ بمؤشّراتِ النُّعاسِ مثلُ إغلاقِ العينِ وحركاتِ الفمِ ووضعيَّةِ الرَّأسِ بكفاءةٍ عاليةٍ . يوضّحُ الشَّكلُ (3) البُنيةَ العامّةَ وآليّةَ عملِ نموذجِ ViT ، حيثُ تُقسَّمُ الصُّورةُ المُدخلةُ إلى رُقعٍ صغيرةٍ (Patches) يتمُّ تحويلها إلى تمثيلاتٍ مُتّجهيّةٍ، ثمَّ تمرُّ عبرَ طبقاتِ الانتباهِ الذَّاتيّ مُتعدّدِ

الرُّؤوسِ لاستخراجِ العلاقاتِ السّياقيّةِ بينَ أجزاءِ الصُّورةِ، ممَّا يسمحُ ببناءِ تمثيلٍ بصريّ شاملٍ يعكسُ السّماتَ المكانيّةَ الدَّقيقةَ المُرتبطةَ بحالةِ السَّائقِ(Dosovitskiy et al.,2021).



(الشَّكلُ (3) البُنيةَ العامَّةَ وآليَّةَ عملِ نموذجِ Vision Transformer ، ((Dosovitskiy et al.,2021

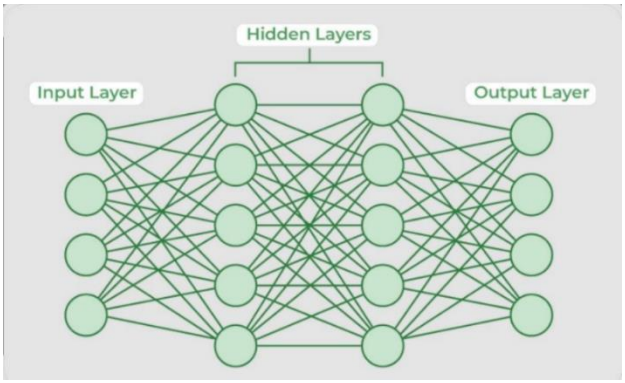
2-5 بناءُ التَّتابعاتِ الزَّمنيّةِ للميزاتِ :

تُجمعُ الميزاتُ المكانيَّةُ المُستخرجةُ من الإطاراتِ ضمنَ تتابعاتٍ زمنيَّةٍ ثابتةِ الطُّولِ تمثّلُ التَّطوّرَ الدّيناميكيّ لحالةِ السَّائقِ. يتيحُ هذا التَّمثيلُ تحويلَ المُشكلةِ من تحليلِ صورٍ مُنفردةٍ إلى تحليلِ سلوكٍ زمنيّ مُتكاملٍ، وهو أمرٌ أساسيٌّ في مهامّ كشفِ النُّعاسِ نظراً لطبيعتهِ التَّدريجيَّةِ (Singh and Papanikolopoulos,2023).

2-6 النَّمذجةُ الزَّمنيّةُ باستخدامِ الشَّبكاتِ العصبيَّةِ المُتكرّرةِ :

لمُعالجةِ الاعتماداتِ الزَّمنيَّةِ في سُلوكِ السَّائقِ، يتمُّ تمريرُ التَّتابعاتِ الإطاريَّةِ المُستخرجةِ إلى نماذجِ الشَّبكاتِ العصبيَّةِ المُتكرّرةِ، والَّتي تشملُ LSTM وGRU وBiLSTM، لما تتمتَّعُ بهِ هذهِ النَّماذجُ من قُدرةٍ عاليةٍ على تمثيلِ العلاقاتِ الزَّمنيَّةِ طويلةِ المدى والتَّغلُّبِ على مشكلةِ تلاشي التَّدرُّجِ في التَّسلسلاتِ الطَّويلةِ، وهوَ ما يجعلها مناسبةً لتحليلِ التَّغيُّرِ التَّدريجيّ في حالةِ انتباهِ السَّائقِ مع مرورِ الزَّمنِ. تمَّ اختيارُ شبكةِ LSTM لقُدرتها على الاحتفاظِ بالمعلوماتِ المُهمّةِ عبرَ الزَّمنِ من خلالِ بُنيةِ الذَّاكرةِ المُعتمدةِ على بوّاباتِ التَّحكّمِ، الأمرُ الَّذي يتيحُ تتبّعَ المُؤشّراتِ المُتراكمةِ للنُّعاسِ مثلُ ازديادِ مدّةِ إغلاقِ العينِ أو تباطؤِ حركاتِ الرَّأسِ بصورةٍ دقيقةٍ ومستقرّةٍ. أمّا شبكةُ GRU فقد تمَّ اعتمادُها نظراً لبنيتها الأبسطِ مُقارنةً بـ LSTM، حيثُ تعتمدُ على عددٍ أقلّ من البوّاباتِ، ممَّا يؤدّي إلى تقليلِ الكلفةِ الحسابيّةِ وتسريعِ عمليّةِ التَّدريبِ مع الحفاظِ على أداءٍ تنبؤيّ مُرتفعٍ، وهوَ عاملٌ مهمٌّ في تطبيقاتِ المُراقبةِ الزَّمنيّةِ الفعليّةِ داخلَ المركباتِ . في حينٍ تمَّ توظيفُ شبكةِ BiLSTM لتمكينِ النّموذجِ من تحليلِ التَّتابعاتِ الزَّمنيّةِ في الاتّجاهين الأماميّ والخلفيّ، الأمرُ الَّذي يسمحُ بفهمٍ أعمقٍ لسياقِ التَّغيُّراتِ السُّلوكيَّةِ المُتتابعةِ وتحسينِ دقّةِ اكتشافِ الأنماطِ المُرتبطةِ بحالةِ النُّعاسِ.

يوضّحُ الشَّكلُ (4) البُنيةَ العامَّةَ للشَّبكاتِ العصبيَّةِ المُتكرّرةِ المُستخدمةِ في هذهِ الدّراسةِ، حيثُ تُظهرُ آليَّةُ انتقالِ المعلوماتِ الزَّمنيّةِ بينَ الحالاتِ المخفيّةِ عبرَ تسلسلِ الإطاراتِ، إضافةً إلى دورِ بوّاباتِ التَّحكّمِ في تنظيمِ تدفّقِ المعلوماتِ وحفظِ الخصائصِ المُهمّةِ عبرَ الزَّمنِ. كما يبيّنُ الشّكلُ الفرقَ المفاهيميَّ بينَ النَّماذجِ المُتكرّرةِ التَّقليديّةِ والبنى المُحسّنةِ مثلُ LSTM وGRU، الَّتي تمَّ تصميمها خصّيصاً لمُعالجةِ مشكلةِ فقدانِ المعلوماتِ في التَّسلسلاتِ الطَّويلةِ، ممَّا يفسّرُ فعاليتها في نمذجةِ التّغيّرِ التَّدريجيّ لحالةِ السَّائقِ أثناءَ القيادةِ (Bengio et al.,2015).



(الشَّكلُ (4) البُنيةَ العامَّةَ للشَّبكاتِ العصبيَّةِ المُتكرّرةِ RNN ، (Bengio et al.,2015))

2-7 طبقةُ التَّصنيفِ النّهائيّ :

في المرحلةِ الأخيرةِ، تُمررُ المُخرجاتُ الزَّمنيَّةُ إلى طبقاتٍ كاملةِ الاتّصالِ متبوعةٍ بدالّةِ Softmax لإنتاجِ احتمالاتِ الفئاتِ، حيثُ يتمُّ تصنيفُ حالةِ السَّائقِ إلى يقظةٍ أو نُعاسٍ وفقَ أعلى احتمالٍ مُتوقّعٍ .

2-8 آليّةُ التَّدريبِ والتَّحسينِ :

تمَّ تدريبُ النّموذجُ باستخدامِ خوارزميَّةِ الانتشارِ العكسيّ مع مُحسّن Adam لتقليلِ دالّة الخسارةِ، كما أُستخدمَتْ تقنيَّاتُ تنظيمٍ مثلُ Dropout والتَّوقّفِ المُبكّرِ لمنعِ فرطِ التَّكيّفِ وتحسينِ قُدرةِ النّموذجِ على التَّعميمِ .

3. النَّتائجُ والتَّقييمُ التَّجريبيّ:

3-1 بروتوكولُ التَّقييمِ التَّجريبيّ:

تمَّ تقييمُ أداءِ النِّظامِ المُقترَحِ باستخدامِ قاعدةِ بياناتِ NTHU-DDD وذلكَ وفقَ بروتوكولٍ تجريبيّ صارمٍ يضمنُ عدالةَ التَّقييمِ ومنعَ أيّ تَّداخلٍ بينَ بياناتِ التَّدريبِ وَالاختبار. تمَّ تقسيمُ البياناتِ إلى ثلاثِ مجموعاتٍ مُستقلّةٍ كما يلي: مجموعةُ التَّدريبِ (Training Set)، مجموعةُ التَّحقُّقِ (Validation Set)، مجموعةُ الاختبارِ (Test Set). تمَّ اعتمادُ تقسيمٍ بنسبةِ (70% تدريب – 15% تحقُّق – 15% اختبار)، مع ضمانِ عدمِ تداخلِ عيّناتِ السَّائقِ نفسهِ بينَ المجموعاتِ المختلفة (Subject-independent split) ، وذلكَ للحفاظِ على حياديَّةِ التَّقييمِ وقياسِ قُدرةِ النَّموذجِ على التَّعميمِ على سائقين جددٍ لم يتمُّ تدريبُ النَّموذجِ عليهم مُسبقاً (Weng et al., 2016).

كما تمَّ استخدامُ مجموعةِ التَّحقُّقِ فقط لضبطِ مُعاملاتِ النَّموذجِ ومنعِ فرطِ التَّكيّفِ، في حين تمَّ حجزُ مجموعةِ الاختبارِ بالكاملِ وعدمُ استخدامها خلال أيّ مرحلةٍ من مراحلِ التَّدريبِ أو الضَّبطِ. وقد أُجريَ التَّقييمُ النهائيُّ للنَّماذجِ حصراً على مجموعةِ الاختبارِ غيرِ المرئيّةِ (Unseen Test Set)، والَّتي لم تُستخدمْ إطلاقاً خلالَ مرحلتي التَّدريبِ أو الضَّبطِ، وذلكَ لضمانِ قياسٍ واقعيٍّ لقُدرةِ النَّموذجِ على التَّعميمِ.

تمَّ تدريبُ النِّظامِ المُقترحِ باستخدامِ ثلاثةِ نماذجٍ زمنيَّةٍ مختلفةٍ مُدمجةٍ مع Vision Transformer وهيَ:

- ViT–LSTM.
- ViT–GRU.
- ViT–BiLSTM.

وتمَّت مُقارنةُ أدائها باستخدامِ مجموعةٍ من المقاييسِ الإحصائيَّةِ القياسيَّةِ المُعتمدةِ في مهامِّ التَّصنيفِ الثُّنائيّ.

3–2 مقاييسُ التَّقييمِ المُعتمدةِ:

لتقييمِ كفاءةِ النَّماذجِ المُقترَحةِ، تمَّ اعتمادُ مجموعةٍ من المقاييسِ الإحصائيّةِ الشَّائعةِ في مجالِ التَّعلُّمِ العميقِ والرُّؤيةِ الحاسوبيّة، وهي: الدقّةُ (Accuracy) ،دقّة التّنبؤ (Precision) ،استرجاعُ المعلومات (Recall) ،مقياسُ F1–Score ،زمنُ التَّدريبِ (Training Time) ،زمنُ الاستدلال (Inference Time). تُعدُّ هذهِ المقاييسُ معياريّةً في تقييمِ نماذجِ التَّعلّمِ العميقِ لمهامِّ التَّصنيفِ الثُّنائيّ، حيثُ توفّرُ تقييماً شاملاً لدقّةِ النّموذجِ وقدرتهِ على التَّعميمِ.

3–3 النَّتائجُ وَ المُقارنةُ بينَ النَّماذجِ:

تمَّ تلخيصُ النَّتائجِ للنَّماذجِ الثَّلاثةِ في الجدولِ (2)، حيثُ تمَّ احتسابُ جميعِ المقاييسِ باستخدامِ مجموعةِ الاختبار فقط (Unseen Data) لضمانِ تقييمٍ موضوعيٍّ وغير متحيّز لقدرةِ النَّموذجِ على التَّعميم. وتمَّ عرضُ قيمِ الدّقّةِ، و تمَّ عرضُ قيمِ الدّقّةِ، والاسترجاعِ، والدّقّةِ الإيجابيّةِ، ومقياسِ F1، إضافةً إلى زمنِ التَّدريبِ وزمنِ الاستدلالِ لكلِّ نموذجٍ.

الجدول (1): المُقارنةُ بينَ نماذجِ ViT–LSTM وViT–GRU وViT–BiLSTM

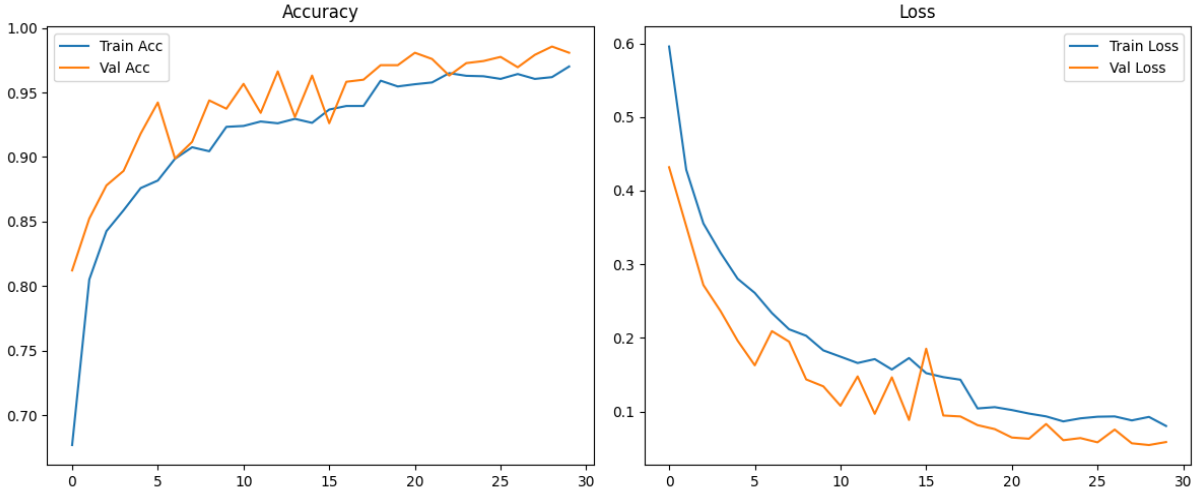
Model	Accuracy	Precision	Recall	F1–Score	Training time (s)	Average Inference time (ms)
LSTM	98.07%	0.98	0.98	0.98	61.41	3.53
GRU	97.11%	0.97	0.97	0.97	38.75	1.33
BiLSTM	97.75%	0.98	0.98	0.98	97.28	3.01

تشيرُ النَّتائجُ إلى أنَّ دمجَ Vision Transformer مع النَّماذجِ الزَّمنيَّةِ يحقّقُ أداءً مُرتفعاً في كشفِ حالاتِ نُعاسِ السَّائقِ ضمنَ سيناريوهاتِ قيادةٍ واقعيَّةٍ. وقد حقَّقَ نموذجُ ViT–LSTM أعلى دقّةِ تصنيفٍ ومقياسِ F1، ممَّا يدلُّ على قدرتهِ العاليةِ في نمذجةِ الاعتماداتِ الزَّمنيَّةِ طويلةِ الأمدِ المُرتبطةِ بتطوّرِ حالةِ النُّعاسِ التَّدريجيّ. في المُقابلِ، أظهرَ نموذجُ ViT–GRU زمناً أقلَّ للاستجابةِ مُقارنةً بباقي النَّماذجِ، ممَّا يجعلهُ خياراً عمليّاً للتَّطبيقاتِ ذاتِ القيودِ الزَّمنيَّةِ الصَّارمةِ. أمَّا نموذجُ ViT–BiLSTM فقد أظهرَ أداءً جيّداً في التَّعرّفِ على الأنماطِ الزَّمنيَّةِ ثُنائيَّةِ الاتّجاهِ، إلَّا أنَّ زمنَ التَّدريبِ كانَ أعلى نسبيّاً نتيجةَ تعقيدِ بنيتهِ الحسابيَّةِ.

ولا تعبّرُ هذهِ النَّتائجُ عن أداءِ النَّموذجِ على بياناتِ التَّدريبِ أو التَّحقُّقِ، وإنّما تعكسُ أداءهُ الحقيقيّ على بياناتٍ جديدةٍ غيرِ مُستخدمةٍ سابقاً.

4-3 نتائجُ التَّدريبِ وَ مُنحنياتُ التَّعلّم:

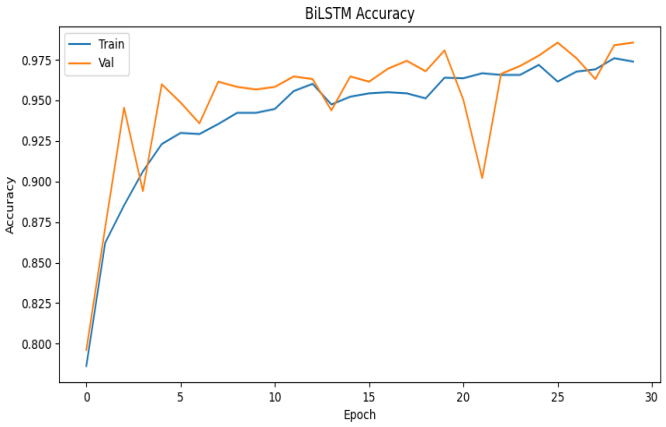
تُظهرُ الأشكالُ الآتيةُ مُنحنياتِ الدقّةِ ودالّةِ الخسارةِ للنماذجِ المُقترَحةِ خلالَ مرحلتي التَّدريبِ والتَّحقُّقِ و تُستخدمُ هذه المُنحنياتُ لتحليلِ سلوكِ التَّعلّمِ أثناءَ التَّدريبِ فقط، ولا تُستخدمُ لتقييمِ الأداءِ النهائيّ للنَّموذجِ.



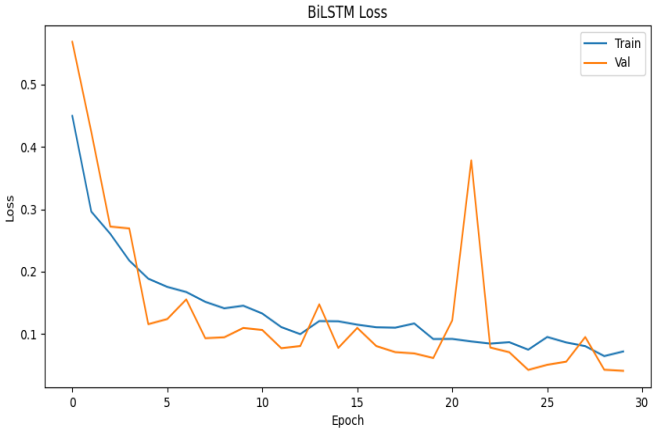
مُنحنى الدِّقَّة لنموذجِ ViT-LSTM **مُنحنى الخسارة لنموذجِ ViT-LSTM**

(الشَّكلُ (5): مُنحنياتُ التَّدريبِ والتَّحقُّقِ لنموذجِ ViT-LSTM، من إعداد الباحثة)

يوضّحُ الشَّكلُ (5) تطوّرَ مُنحنياتِ الدّقّةِ وقيمةَ دالّةِ الخسارةِ لنموذجِ ViT–LSTM أثناءَ مرحلتي التَّدريبِ والتَّحقّقِ، حيثُ يُلاحظُ تحسّنٌ تدريجيٌّ في الدّقّةِ وانخفاضٌ مُستقرٌ في الخسارةِ مع تقدّمِ الحِقَبِ التَّدريبيَّةِ، ممّا يدلُّ على استقرارِ عمليَّةِ التَّعلّمِ خلالَ مرحلتي التَّدريبِ و التَّحقُّقِ دونَ مؤشراتٍ واضحةٍ على فرطِ التَّكيُّفِ.



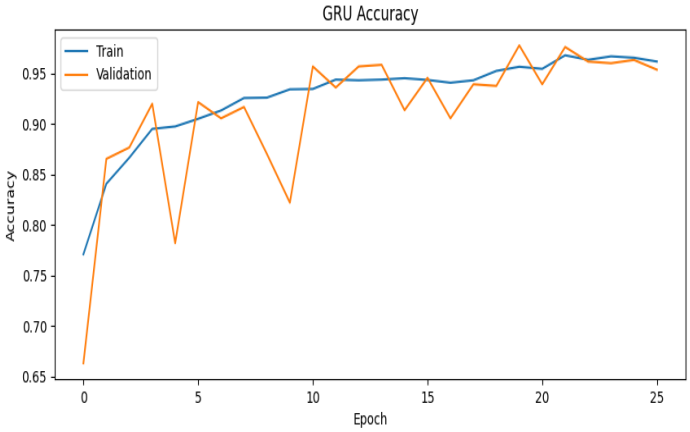
مُنحنى الدِّقَّة لنموذجِ ViT–GRU



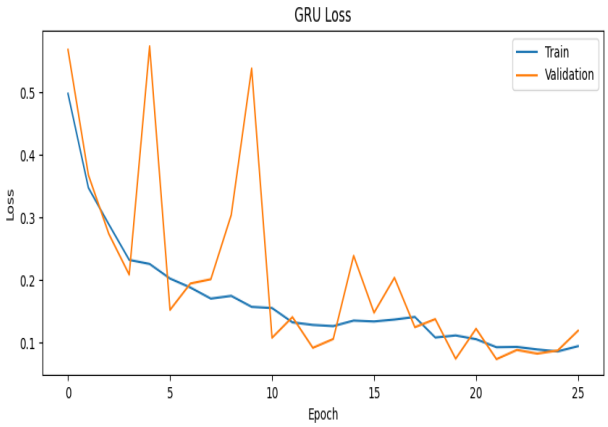
مُنحنى الخسارة لنموذجِ ViT–GRU

(الشَّكلُ (6): مُنحنياتُ التَّدريبِ والتَّحقُّقِ لنموذجِ ViT–GRU ،من إعداد الباحثة)

يوضّحُ الشَّكلُ (6) مُنحنياتِ الدّقّةِ وقيمةَ دالّةِ الخسارةِ لنموذجِ ViT–GRU أثناءَ مرحلتي التَّدريبِ والتَّحقّقِ، حيثُ يظهرُ تحسّنٌ تدريجيٌّ في الدّقّةِ وانخفاضٌ مُنتظمٌ في الخسارةِ مع تقدّمِ الحِقَبِ التَّدريبيَّةِ، ممّا يعكسُ استقرارَ عمليّةِ التَّعلّمِ وكفاءةَ النّموذجِ في تحقيقِ أداءٍ تنبّؤيّ جيّدٍ مع زمنِ استجابةٍ مُنخفضٍ.



مُنحنى الدِّقَّة لنموذجِ ViT–BiLSTM



مُنحنى الخسارة لنموذجِ ViT–BiLSTM

(الشَّكلُ (7): مُنحنياتُ التَّدريبِ والتَّحقُّقِ لنموذجِ ViT–BiLSTM، من إعداد الباحثة)

يوضّحُ الشَّكلُ (7) مُنحنياتِ الدّقّةِ وقيمةَ دالّةِ الخسارةِ لنموذجِ ViT–BiLSTM أثناءَ مرحلتي التَّدريبِ والتَّحقّقِ، حيثُ يُلاحظُ تحسّنٌ تدريجيٌّ في الدّقّةِ مع انخفاضٍ مُستقرٍ في الخسارةِ عبرَ الحِقَبِ التَّدريبيَّةِ، ممّا يدلُّ على قُدرةِ النّموذجِ على تعلّمِ الأنماطِ الزَّمنيّةِ ثُنائيّةِ الاتّجاهِ بكفاءةٍ وتحقيقِ أداءٍ تنبّؤيّ مُستقرٍ.

يُلاحظُ من مُنحنياتِ التَّعلّمِ أنَّ جميعَ النَّماذجِ أظهرتْ تقارباً جيّداً واستقراراً واضحاً خلالَ عمليّةِ التَّدريبِ، دونَ ظهورِ فجوةٍ كبيرةٍ بينَ مُنحنياتِ التَّدريبِ والتَّحقّقِ، ممَّا يشيرُ إلى استقرارِ عمليّةِ التَّعلّمِ وعدمِ وجودِ مؤشّراتٍ واضحةٍ على فرطِ التكيّفِ خلالَ مرحلتي التَّدريبِ والتَّحقّقِ. ومع ذلك، فإنَّ تقييمَ قدرةِ النَّموذجِ على التَّعميمِ تمَّ بالاعتمادِ على نتائجِ مجموعةِ الاختبار كما تمَّ توضيحُهُ في القسمِ السَّابقِ.

3-5 التَّقييمُ على بياناتٍ خارجيَّةٍ :

تمَّ إجراءُ اختبارٍ إضافيٍّ للنِّظامِ المُقترحِ باستخدامِ بياناتٍ بصريّةٍ خارجيّةٍ لم يتمْ تضمينها ضمن بياناتِ التَّدريبِ أو التَّحقّقِ أو الاختبار، وذلك بهدفِ تقييمِ قدرةِ النموذجِ على التَّعميم في بيئاتٍ واقعيّةٍ مختلفة.

تمَّ استخدامُ مقطعِ فيديو حقيقيٍّ يتضمّنُ ظروفَ تصويرٍ متنوّعةٍ من حيثِ الإضاءةِ، وزوايا الوجهِ، وحركةِ السَّائقِ، وتمَّ تطبيقُ نفسِ مراحلِ المُعالجةِ المُسبقةِ المُعتمدةِ في النظام.

أظهرتِ النتائجُ أنَّ النِّظامَ قادرٌ على التَّعرُّفِ على الأنماطِ العامّةِ المرتبطةِ بحالةِ النُّعاس، حيثُ نجحَ في تمييزِ عددٍ من الحالاتِ بشكلٍ صحيح كما هو موضّحٌ في الشَّكلِ (8)، إلا أنّ الأداءَ تأثّرَ نسبياً باختلافِ خصائصِ البياناتِ الخارجيّةِ مُقارنةً ببياناتِ التَّدريب، وهو ما يُعرفُ في الأدبيّاتِ العلميّةِ بمشكلةِ (Domain Shift).

وعلى الرَّغمِ من ذلك، فإنَّ النِّظامَ أظهرَ سلوكاً مقبولاً في التَّعاملِ مع البياناتِ غيرِ المرئيّة، ممّا يشيرُ إلى امتلاكهِ درجةً جيّدةً من القابليّةِ للتَّعميم، مع الحاجةِ إلى تحسيناتٍ إضافيّةٍ لزيادةِ الاستقرارِ في البيئاتِ الواقعيّةِ المتنوّعة.



(a) (b)

(الشَّكلُ (8): نتائجُ التَّقييمِ النّوعيّ على بياناتٍ خارجيّةٍ ، من إعدادِ الباحثةِ)

(a) حالة يقظة (Alert): يظهرُ السَّائقُ بعينين مفتوحتين ونظرةٍ مُستقرّة، ممّا يدلُّ على مستوى انتباهٍ طبيعيّ أثناءَ القيادةِ.

(b) حالة نُعاس (Drowsy): يظهرُ انخفاضٌ في فتحِ العينِ وملامحِ تعبٍ واضحةٍ، ممَّا يشيرُ إلى بدايةِ حالةِ النُّعاسِ.

وبناءً على ذلكَ، يمكنُ اعتبارُ هذا التَّقييمِ دليلاً إضافيّاً على قُدرةِ النّموذجِ على العملِ في سيناريوهاتٍ واقعيّةٍ، مع التَّأكيدِ على أهميَّة تحسينِ استقرارِ الأداءِ من خلالِ استخدامِ بياناتِ تدريبٍ أكثر تنوّعاً في الأعمالِ المُستقبليّةِ.

3-6 مُناقشةُ النَّتائجِ:

تؤكّدُ النَّتائجُ التَّجريبيّةُ فعاليّةَ النَّهجِ المُقترحِ القائمِ على دمجِ نموذجِ Vision Transformer مع النَّماذجِ الزَّمنيّةِ المُتقدِّمةِ في كشفِ حالاتِ نُعاسِ السَّائقِ ضمنَ بيئاتِ قيادةٍ واقعيّةٍ. ويعكسُ هذا الأداءُ المُتقدّمُ قدرةَ البُنيةِ الهجينةِ المُقترحةِ على الاستفادةِ من التَّمثيلِ المكانيّ العميقِ والتَّحليلِ الزَّمنيّ في آنٍ واحدٍ.

يُلاحظُ أنَّ استخدامَ Vision Transformer بدلاً من الشَّبكاتِ العصبيّةِ الالتفافيّةِ التَّقليديّةِ قد أتاحَ للنّظامِ استخراجَ تمثيلاتٍ مكانيّةٍ أكثرَ شمولاً وتعبيراً لملامحِ الوجهِ، وذلك بفضلِ آليّةِ الانتباهِ الذَّاتيّ القادرةِ على نمذجةِ العلاقاتِ بعيدةِ المدى داخلَ الصُّورة، وهو ما يسهمُ في تحسينِ تمثيلِ المُؤشّراتِ البصريّةِ المُرتبطةِ بالنُّعاسِ مثلُ إغلاقِ العينِ وحركاتِ الرَّأسِ (Dosovitskiy et al.,2021).

في المُقابلِ، مكّنتِ الشَّبكاتُ العصبيّةُ المُتكرّرةُ من تحليلِ التَّطوّرِ الزَّمنيّ لحالةِ السَّائقِ وتمثيلِ التَّغيّراتِ التَّدريجيّةِ في السُّلوكِ البصريّ، وهو ما يُعدُّ عاملاً حاسماً في مهامّ الكشفِ المُبكّرِ عن النُّعاسِ، نظراً لقُدرتها على نمذجةِ الاعتماداتِ الزّمنيّةِ طويلةِ المدى والتَّغلّبِ على مشكلةِ تلاشي التَّدرّجِ .

كما أظهرَتِ النَّتائجُ أنَّ اعتمادَ تسلسلاتٍ زمنيّةٍ بطولِ 16 إطاراً متتالياً أسهمَ في تحقيقِ توازنٍ فعّالٍ بينَ دقّةِ التَّمثيلِ الزَّمنيّ والتَّعقيدِ الحاسوبيّ، مما مكّن النَّماذجَ من التقاطِ المُؤشّراتِ التَّدريجيّةِ للنُّعاسِ بصورةٍ أدقّ من الأنظمةِ المُعتمدةِ على تحليلِ إطارٍ مُنفردٍ فقط. وتبرزُ هذه النَّتائجُ الأهميّةَ الجوهريّةَ لدمجِ البُعدِ الزَّمنيّ في أنظمةِ مُراقبةِ السَّائقِ الذَّكيّةِ لتحقيقِ أداءٍ أكثرَ موثوقيةً في التَّطبيقاتِ الواقعيّة.

وبشكلٍ عامّ، تؤكّدُ هذه المُعطياتُ أنَّ البُنيةَ الهجينةَ المُقترحةَ تمثّلُ اتّجاهاً واعداً في تطويرِ أنظمةِ الكشفِ المُبكّرِ عن نعاسِ السَّائقِ، لما توفّرهُ من توازنٍ فعّالٍ بين دقّةِ التَّنبّؤِ والكفاءةِ الحاسوبيّةِ وإمكانيّةِ التَّطبيقِ في الزَّمنِ الحقيقيّ.

وبالمُقارنةِ مع الدِّراساتِ السَّابقةِ، يتبيّنُ أنَّ الأنظمةَ المُعتمدةَ على الشَّبكاتِ الالتفافيّةِ (CNN) تُحقّقُ أداءً جيّداً في استخراجِ الميزاتِ المكانيّةِ، إلَّا أنَّها تبقى محدودةً في تمثيلِ الاعتماداتِ الزَّمنيَّةِ، ممّا يؤثّرُ على دقّةِ الكشفِ المُبكّرِ عن النُّعاسِ في تسلسلاتِ الفيديو. في المُقابلِ، أظهرَتِ النَّماذجُ الهجينةُ الَّتي تجمعُ بينَ التَّمثيلِ المكانيّ والتَّحليلِ الزَّمنيّ تحسّناً ملحوظاً في الأداءِ، وهو ما يتوافقُ مع نتائجِ هذا البحثِ.

وتُظهرُ النَّتائجُ الَّتي تمَّ الحصولُ عليها باستخدامِ بياناتِ الاختبارِ غيرِ المرئيَّةِ أنَّ النَّموذجَ المُقترحَ يحقّقُ توازناً فعّالاً بينَ دقّةِ التَّنبؤِ والكفاءةِ الحسابيَّةِ مُقارنةً بالمُقارباتِ التَّقليديَّةِ، مع قُدرةٍ أفضلَ على التَّعميمِ في بيئاتِ قيادةٍ واقعيَّةٍ. ويُعزى ذلكَ إلى تكاملِ تمثيلِ Vision Transformer مع النَّمذجةِ الزَّمنيَّةِ، ممَّا يسمحُ بالتقاطِ الأنماطِ التَّدريجيَّةِ للنُّعاسِ بشكلٍ أكثرَ دقّةٍ.

4. الاستنتاجاتُ وَ التَّوصياتُ المُستقبليَّةُ :

قدَّمَ هذا البحثُ إطاراً منهجيّاً لتطويرِ نظامٍ ذكيٍّ لكشفِ نُعاسِ السَّائقِ اعتماداً على بُنيةٍ هجينةٍ تجمعُ بينَ نموذجِ Vision Transformer لاستخراجِ الميزاتِ المكانيّةِ العميقةِ، ونماذجِ الشَّبكاتِ العصبيّةِ المُتكرّرةِ لمُعالجةِ الاعتماداتِ الزَّمنيّةِ ضمنَ تسلسلاتِ الإطاراتِ الوجهيّةِ.

أظهرتِ النَّتائجُ التَّجريبيّةُ فعاليّةَ هذا التَّكاملِ في تمثيلِ التَّطوّرِ التَّدريجيّ لحالةِ النُّعاسِ، حيثُ حقَّقَ النِّظامُ المُقترَحُ أداءً مرتفعاً في ظروفِ قيادةٍ واقعيّةٍ متنوّعة. كما بيَّنت المُقارنةُ أنَّ نموذجَ ViT–LSTM قدَّم أفضلَ أداءٍ من حيثُ دقّةِ التَّصنيفِ ومقياسِ F1 نتيجةَ قدرتهِ على نمذجةِ الاعتماداتِ الزَّمنيّةِ طويلةِ الأمدِ، في حين تميَّز نموذج ViT–GRU

بكفاءةٍ حسابيّةٍ أعلى وزمنِ استجابةٍ أقلّ، بينما وفَّر نموذج ViT–BiLSTM تمثيلاً أكثر شمولاً للسِّياقِ الزَّمنيّ على حسابِ زيادةِ التَّعقيدِ الحسابيّ.

تؤكِّد هذهِ النَّتائجُ أنَّ دمجَ التَّمثيلِ المكانيّ القائمِ على آليّةِ الانتباهِ مع النَّمذجةِ الزَّمنيّةِ يُعدُّ نهجاً واعداً لتطويرِ أنظمةِ مُراقبةِ السَّائقِ الذكيّةِ، ويُسهمُ في تحقيقِ توازنٍ فعّالٍ بينَ دقّةِ الكشفِ ومتطلّباتِ الاستجابةِ الآنيّةِ في تطبيقاتِ أنظمةِ مُساعدةِ السَّائقِ المُتقدِّمة .

ويمكنُ توسيعُ هذا العملِ مستقبلاً من خلالِ:

- توسيعُ البياناتِ والتَّسلسلاتِ المُستخدمةِ ودمجُ مصادرِ استشعارٍ إضافيّةٍ.
- تحسينُ كفاءةِ النِّظامِ وتقييمُهُ عمليّاً على منصّاتٍ واقعيّةٍ.

ختاماً، يُقدِّمُ هذا البحثُ نموذجاً عمليًّاً يمكنُ الاعتمادُ عليه في تطويرِ حلولٍ ذكيّةٍ لكشفِ نُعاسِ السَّائقِ، بما يدعمُ تعزيزَ السَّلامةِ المروريّةِ والحدَّ من الحوادثِ المُرتبطةِ بالإرهاقِ وفقدانِ الانتباهِ أثناءَ القيادةِ.

5. المراجعُ:

[1] ARORA, G.; TIWARI, A.; YADAV, A. S.; MAURYA, D.; KUMAR, M., 2024, *Driver drowsiness detection: CNN based approach using OpenCV and Keras. In International Conference on Smart Systems and Advanced Computing (ICSSAC 2024),* CRC Press/Taylor & Francis, DOI: 10.1201/9781003593034-61

[2] BENGIO, Y.; LECUN, Y.; HINTON, G., 2015, *Deep learning. Nature*, Vol 521, No 7553, pp 436-444, DOI: 10.1038/nature14539

[3] BIOULAC, S.; ARNAUD, M.; MICOULAUD-FRANCHI, J.; Sagaspe, P.; Moore, N.; Salvo, F., 2022, *Risk of motor vehicle accidents related to sleepiness: A systematic review. Sleep Medicine Reviews,* Vol 62, 101596, DOI: 10.1016/j.smrv.2021.101596

[4] DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GHANI, M.; USZKOREIT, J.; HOUBEN, J., 2021, *An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).*

[5] ED-DOUGHMI, Y.; IDRISSI, N.; HBALI, Y., 2025, *3D convolutional recurrent neural network for driver drowsiness behavior analysis. Heliyon* .

[6] FENG, X.; GUO, Z.; KWONG, S., 2025, *ID3RSNet: cross-subject driver drowsiness detection from raw single-channel EEG with an interpretable residual shrinkage network. Frontiers in Neuroscience,* Vol 18, 1508747

[7] JIAO, Y.; ZHANG, Y.; LIU, W.; JIAO, Z., 2025, *Detecting driver sleepiness from physiological indicators using a CNN-LSTM self-attention model. IEEE Journal of Biomedical and Health Informatics,* DOI: 10.1109/JBHI.2025.3629974

[8] LIU, W.; QIAN, J.; YAO, Z.; JIAO, X.; PAN, J., 2019, *Convolutional Two-Stream Network Using Multi-Facial Feature Fusion for Driver Fatigue Detection. Future Internet,* Vol 11, No 5, 115, DOI: 10.3390/fi11050115

[9] MAIOR, C. B.; MOURA, M. J. C.; SANTANA, J. M. M.; LINS, I. D., 2020, *Real-time classification for autonomous drowsiness detection using eye aspect ratio. Expert Systems with Applications,* Vol 158, 113505, DOI: 10.1016/j.eswa.2020.113505

[10] MINHAS, R.; PEKER, N. Y.; HAKKOZ, M. A.; ARBATLI, S.; CELIK, Y.; EROGLU ERDEM, C.; PEKER, Y.; SEMIZ, B., 2025*, Improved drowsiness detection*

in drivers through optimum pairing of EEG features using an optimal EEG channel comparable to a multichannel EEG system. Medical & Biological Engineering & Computing, Vol 63, No 10, pp 3019-3036, DOI: 10.1007/s11517-025-03375-1

[11] NHTSA (National Highway Traffic Safety Administration), 2023, *Drowsy Driving — Road Safety Facts. U.S. Department of Transportation, Washington,* USA.

[12] PHAN, T. C.; NGUYEN, N. H. Q., 2025, *A novel approach of drowsiness levels detection using Vis-Net combined with facial emotion. Systems and Soft Computing,* Vol 7, 200288, DOI: 10.1016/j.sasc.2025.200288

[13] SHARATH KUMAR, G. N.; HANUMANTHAPPA, J.; RAJ, C. C.; NAVEEN KUMAR, P., 2025, *Temporal Convolutional Network Approach for Video-Based Driver Drowsiness Detection. International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, DOI: 10.22399/ijcesen.4609

[14] SINGH, S.; PAPANIKOLOPOULOS, N., 2023, *Driver fatigue detection using deep learning: A comprehensive review. Expert Systems with Applications,* Vol 234, 119930, DOI: 10.1016/j.eswa.2023.119930

[15] VENKATESWARLU, M.; CH, V. R. R., 2025, *CNN-ViT: A multi-feature learning based approach for driver drowsiness detection. Results in Engineering*, Vol 27, 100425, DOI: 10.1016/j.rineng.2025.100425

[16] WENG, C.; LAI, Y.; LAI, S., 2016, *NTHU driver drowsiness detection dataset. National Tsing Hua University, Taiwan*, Available at: https://cv.cs.nthu.edu.tw/php/callforpaper/dataset/DDD/

[17] WHO (World Health Organization), 2023, *Global Status Report on Road Safety.* Geneva, Switzerland.