WILEY

IET The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Evaluating Driver Readiness in Conditionally Automated Vehicles From Eye-Tracking Data and Head Pose

Mostafa Kazemi | Mahdi Rezaei | Mohsen Azarmi

Institute for Transport Studies, University of Leeds, Leeds, UK

Correspondence: Mahdi Rezaei (m.rezaei@leeds.ac.uk)

ABSTRACT

As automated driving technology advances, the role of the driver to resume control of the vehicle in conditionally automated vehicles becomes increasingly critical. In the SAE level 3 or partly automated vehicles, the driver needs to be available and ready to intervene when necessary. This makes it essential to evaluate their readiness accurately. This article presents a comprehensive analysis of driver readiness assessment by combining head pose features and eye-tracking data. The study explores the effectiveness of predictive models in evaluating driver readiness, addressing the challenges of dataset limitations and limited ground truth labels. Machine learning techniques, including LSTM architectures, are utilised to model driver readiness based on the spatio-temporal status of the driver's head pose and eye gaze. The experiments in this article revealed that a bidirectional LSTM architecture, combining both feature sets, achieves a mean absolute error of 0.363 on the DMD dataset, demonstrating superior performance in assessing driver readiness. The modular architecture of the proposed model also allows the integration of additional driver-specific features, such as steering wheel activity, enhancing its adaptability and real-world applicability.

1 | Introduction

According to data published by the World Health Organisation (WHO) [1], over 1.35 million people lose their lives annually due to road-related accidents, and millions more suffer severe injuries. Within these unfortunate events, it has been observed that 90% of all accidents are attributed to human errors [2]. The development of technology has given rise to the hope that the concept of automated driving can address these challenges [3]. Removing the driver from the driving process and transferring control of the vehicle to an automated driving function can be a step towards eliminating human error in accidents [4].

In L3 conditionally automated vehicles, the system temporarily takes responsibility for steering, acceleration, and braking, along with the vehicle-related duties, while the driver has the option to engage in non-driving related tasks (NDRTs) like reading the news, watching a movie, or relaxing in the driver's seat [5]. However, the human drivers must be prepared to regain control in cases of emergencies, like when the system malfunctions and prompts a take-over request (TOR) or when the situation goes beyond the system's capabilities [6]. During such scenarios, the ability to perceive and understand road and traffic conditions, essentially situational awareness, becomes crucial in determining immediate actions. At this stage, the question arises whether the driver is ready to take back control of the vehicle. If the enabler determines that the driver is ready, the vehicle control will be transferred to the driver; otherwise, a safe, low-risk manoeuvre such as safe stopping, pulling over to the side of the road, or activating hazard lights will be activated to prevent any potential danger [7, 8].

Prior to initiating a TOR, it is crucial to gather both internal and external data regarding the driver's condition and the traffic context [9] to ensure a secure outcome [10]. This is especially important when dealing with take-over procedures necessitated by sudden emergencies. Thus, the development of driver monitoring systems (DMS) becomes imperative [11]. These systems should possess the capability to assess the driver's physical and mental condition and subsequently correlate this data with the specific road scenario [10]. To achieve this objective, the DMS accumulates baseline metrics for various driver state indicators, including factors like the driver's gaze, posture, and physiological measurements [12]. These baseline metrics are then employed to compute a measure of deviation, enabling the system to evaluate the extent of the driver's readiness.

In this paper, we investigated the strengths and weaknesses of the existing models and datasets for assessing driver readiness in conditionally automated vehicles. Building on this understanding, we develop a non-intrusive computer vision-based framework for evaluating driver readiness using facial landmark features. The proposed model tracks facial features and analyses their geometric configuration as features to estimate the driver's head pose and gaze directions.

Our main research contribution can be summarised as follows:

- We introduce an extended dataset with new ground truth labels with numerical values for driver readiness, addressing the challenge of limited ground truth availability in the context of driver readiness assessment studies.
- We leverage state-of-the-art convolutional neural network (CNN) models in computer vision to enhance the accuracy of feature extraction from eye-tracking data and head pose, establishing them as reliable indicators for driver readiness evaluation.
- We employ a recurrent neural network (RNN) architecture that utilises long short-term memory (LSTM) modules to analyse spatio-temporal dependencies within the video frames recorded from drivers.

The rest of this paper is organised as follows: Section 2 presents the literature review, which examines the existing knowledge on automated driving systems, the importance of NDRTs, the challenges of transitioning between automated and manual driving, and the landscape of vision-based driver monitoring techniques. It also delves into previous research on driver readiness and identifies research gaps. Section 3 outlines the dataset used for the study and explains the creation of a ground truth framework to measure driver readiness. In Section 4, the methodology is detailed, covering the extraction of head pose and gaze direction features, the implementation of machine learning models, and the training and validation strategies. Section 5 evaluates the performance of the proposed models using various metrics, analyses the impact of different configurations, and presents the obtained results. Finally, in Section 6, the conclusion synthesises the findings, reiterates the research objectives, discusses the implications of the study's outcomes, and suggests potential directions for further research in automated driving and driver readiness assessment.

2 | Related Work

Decision-making to pass the driving control to a human driver would need to rely on empirical data regarding drivers' abilities and behaviours in specific situations. This is normally done by utilisation of camera-based systems and computer vision-based methodologies to analyse the driver's direction of attention. For instance, if the analysis of the driver's visual attention indicates that the driver is completely disconnected from the driving task, or if the analysis reveals a challenging driving scenario, the vehicle might execute a low-risk safety manoeuvre to position itself on the roadside [7]. Thus, it is crucial for a DMS to consistently evaluate the driver's readiness to take-over, which is essential to ensure a safe and timely transfer of control [11]. The most accurate depiction of the driver's state can be achieved using physiological-based DMS; however, these systems are excessively intrusive for application in commercial vehicles [13]. To address this, the development of DMS has largely shifted towards leveraging computer vision and deep learning technologies [14].

2.1 | Vision-Based Driver Monitoring

Driver behaviour analysis based on vision sensors has been a popular topic of research, with a large body of literature focusing on the driver's gaze estimation as a useful cue for estimating the driver's attention [15]. By analysing visual cues from the driver's face, these systems can detect signs of drowsiness, distraction, or other cognitive impairments [16]. Initial research relied on assessing the driver's gaze through head pose estimation. In this context, Lee et al. [17] introduce a vision-based approach to instantly determine the driver's head orientation and gaze direction. This technique remains effective despite changes in facial appearance due to wearing glasses and functions proficiently during both daylight and nighttime scenarios. The method employs facial features and a geometric facial model to gauge the driver's yaw and pitch angles. Furthermore, the approach includes a gaze estimation technique that employs support vector machines (SVM) to accurately calculate both the gaze directions and gaze regions. In the study by Tawari et al. [18], they introduced a distributed camera system that analyses head movements. The main strength of the model is its capability to work reliably and consistently, even when dealing with significant head motions. The system's core function involves tracing facial features and examining their spatial arrangement. This process helps gauge head orientation through a three-dimensional model.

Recent methods mainly utilise a combination of both head and eye features. Vasli et al. [19] presented a system for estimating a driver's gaze by considering cues from both the head and eye poses. They employed a multi-plane geometrical setting and integrated this with a data-driven learning approach. The researchers evaluated their methods using real-world driving data, encompassing diverse drivers and vehicles, to assess how well the techniques could be applied broadly. Similarly, Fridman et al. [20] investigated the improvement in classifying driver gaze by utilising both head and eye poses instead of relying solely on the head pose. They also explored whether individual-specific gaze patterns were linked to the extent of gaze classification enhancement when including eye pose data. The key insight

from their work is metaphorically explained using the concepts of an 'owl' and a 'lizard.' This metaphor illustrates that when the head exhibits substantial movement ('owl'), the addition of eye pose information does not significantly enhance gaze classification. Conversely, when only the eyes move while the head remains still ('lizard'), including eye pose data substantially boosts classification accuracy. The authors also investigated how this accuracy variation relates to different individuals, gaze strategies, and gaze regions.

Recent studies involving CNNs have enabled the accurate prediction of driver gaze areas, which can be applied universally to various drivers and slight deviations in camera positioning. Vora et al. [21] addressed the challenge of creating a generalised gaze zone estimation system that remains consistent across various subjects, perspectives, and scales. To assess the effectiveness of their system, they gathered extensive real-world driving data and trained their CNN model, achieving high accuracy. Similarly, Lollett et al. [22] tackled the challenge of accurately gauging a driver's gaze in real-world scenarios, where factors like facial obstructions and varying distances from the camera can complicate the process.

While existing studies have shown promise in accurately predicting driver gaze areas using CNNs and computer vision techniques, only a limited number of studies have focused on the application of gaze zone estimation for evaluating driver readiness in driving scenarios.

2.2 | The Driver Readiness Studies

In order for the human driver to take back control of the vehicle safely, the driver must be in an appropriate state of readiness prior to the TOR alarm. According to ISO/TR 20195-1 [23], driver readiness refers to the condition of the driver while engaged in automated driving, which impacts their ability to effectively take back control of the vehicle from the system in order to resume manual driving.

Braunagel et al. [5] present an advanced driver assistance system that predicts the driver's take-over readiness in conditionally automated vehicles. The system proactively warns the driver if low take-over readiness is anticipated. It determines readiness based on traffic complexity, the driver's secondary task, and gaze direction. An evaluation using a driving simulator with 81 participants demonstrated a 79% accuracy in predicting take-over readiness.

Marberger et al. [24] research focused on human performance during take-over situations and how the driver's state impacts these scenarios. The study established a comprehensive model for the transition process from automated to manual driving. This model defined specific time points and time intervals critical to this transition. The researchers introduced the concept of 'driver availability,' quantitatively measuring the time needed for a safe take-over against the available time frame.

Deo and Trivedi [13] conducted a study to assess driver readiness in conditionally automated vehicles by analysing information from in-vehicle cameras. Human evaluators examined the cam-

era data and gave their subjective rates for the driver's ability to take-over control. The research introduces the term 'observable readiness index (ORI)', which quantifies the driver's readiness based on these evaluations. Furthermore, the study presents an LSTM model that continuously estimates the driver's ORI by considering various aspects of the driver's state, including gaze, hand position, posture, and foot activity.

In the research by Arslanyilmaz et al. [25], they explored factors influencing the duration and quality of take-overs during system failures before intersections. These factors encompass secondary task characteristics, traffic density, time buffer before TOR, and the type of TOR warning. The study's primary aim was to investigate how different time buffers (3 and 7 s) impact take-over duration and quality in case of system failures near intersections. They also aimed to analyse drivers' eye-tracking behaviour in these time buffers, comparing how drivers focus on information on the dashboard and the road.

In the study by Kim, J. et al. [26], the focus was on investigating how NDRTs impact a driver's readiness and take-over performance. Through a driving simulator experiment, the researchers assessed driver take-over readiness during the transition from automated to manual driving while participants engaged in various NDRTs. The study established that driver readiness had a negative correlation with take-over time but a positive correlation with the quality of vehicle control.

In the study conducted by Greer et al. [27], the focus was on achieving safe transitions from automated to manual control in vehicles by understanding the driver's situational awareness. The researchers introduced two metrics, the ORI and take-over time, which quantify this awareness state. Their model used feature vectors that included hand location, foot activity, and gaze location as input. Additionally, the study introduced two new metrics to assess the quality of control transitions after the take-over event, namely the maximal lateral deviation and velocity deviation.

The current state of research on driver readiness in the context of transitioning from automated to manual driving has taken significant steps forward in understanding the factors influencing this critical process. The literature review highlights the importance of driver engagement in NDRTs, the challenges posed by the transition to manual driving, and the essential role of vision-based DMS in assessing driver readiness. However, certain challenges and research gaps necessitate attention. First, the lack of comprehensive naturalistic driving datasets capturing a wide range of driver behaviours in conditionally automated vehicles hinders the development of data-driven approaches to map driver activity to take-over readiness effectively. Such datasets would significantly enhance the applicability of research findings. Second, defining a clear and objective ground truth for take-over readiness remains a complex task. As data-driven approaches heavily rely on accurate ground-truth information, the absence of a universally accepted measure for readiness poses a challenge in designing robust models and evaluations. Addressing these gaps would lead to more accurate and reliable methods for assessing driver readiness during the transition from automated to manual driving, ultimately contributing to safer and more efficient automated driving systems.
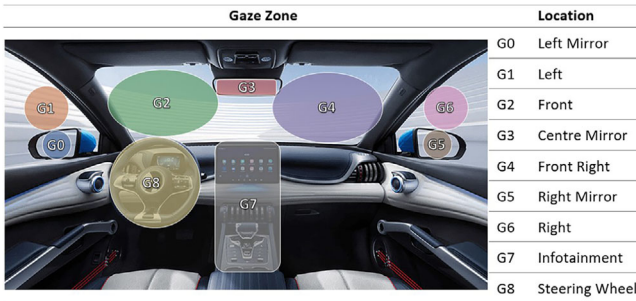
FIGURE 1 | Classification and positioning of gaze zones within the vehicle cabin.

3 | Dataset and Ground Truth

In this study, we randomly selected 15 video clips, each spanning approximately 12 s, from the S6 category of the Driver Monitoring Dataset (DMD) [28]. This selection meets the following prerequisites of this study:

- High-resolution videos: The dataset includes 1280×720 pixels RGB videos that are zoomed on the driver's face only. This ensures capturing detailed facial features and expressions which are crucial for assessing driver readiness.

- Gaze estimation relevance: The scenarios depicted in the dataset finely correspond to the gaze estimation requirements necessary for our study, particularly for both owl-style and lizard-style glances. This alignment between the dataset's content and our research focus enhances the dataset's suitability for our investigation.

The selected video clips, captured at a rate of 30 FPS, result in a total of 5640 frames available for evaluation. These frames have been split into 90 sections, each lasting 2 s. Furthermore, the chosen videos were recorded in a controlled environment inside a real car cabin. Participants were directed to focus their attention on specific regions of interest in the car (regions G0 to G8), as shown in Figure 1. This deliberate focus on distinct regions facilitates the subsequent estimation of gaze positions, which is crucial for the analysis conducted in our study.

Acknowledging the limitations of the available dataset and to the best of our knowledge, this is the first attempt to investigate driver readiness by establishing relevant ground truth labels using the DMD dataset. Given the dataset size limit, we tried to maximise the utilisation of the available dataset for training, validation, and test phases by considering a cross-validation approach.

3.1 | Human Ratings for Driver Readiness

One of the fundamental challenges in driver readiness assessment for take-over control is the absence of publicly available data that includes relevant ground truth information regarding driver readiness. To address the issue of ground truth, we employed a rating method by two human evaluators, similar to the approach taken by studies [29] and [13]. Human evaluators watch frame feeds and assign a value to the driver readiness based on their head movements and eye gaze. This methodology relies on the ability of experienced drivers to identify and assess driver

TABLE 1 | Driver readiness assessment levels based on head pose and eye direction.

Rating	Driver readiness	Description
1	Non-attentive	The driver's head pose and eye direction indicate a complete lack of readiness to take control.
2	Low attentiveness	The driver's head pose and eye direction suggest minimal awareness and readiness to take control.
3	Partially ready	The driver's head pose and eye direction show some awareness, but their readiness is limited.
4	Moderately ready	The driver's head pose and eye direction demonstrate moderate readiness to take immediate control.
5	Fully ready	The driver's head pose and eye direction indicate full attentiveness and readiness to take control.

readiness from video recordings of the driver's face. By analysing the driver's head movements and eye gaze, the evaluators can provide consistent ratings of the driver's readiness levels. A substantial level of agreement among evaluators would indicate a metric for assessing the driver's readiness for take-over that could be determined solely through visual cues [13]. The average value offered by the evaluators can serve as the ground truth for training and testing a machine learning algorithm for take-over readiness estimation.

### 3.2 | Protocol for Collecting Ratings

Two individuals holding UK driving licenses, and familiar with the functioning of level 3 (L3) automated driving systems were selected as evaluators. A time span of 2 s was utilised to evaluate driver readiness. To achieve this, a sequence of 12 frames, equivalent to 2 s at a frame rate of five FPS, was presented to each evaluator. Independently, they assigned a readiness score ranging from 1 (lowest driver readiness) to five (highest driver readiness) to the driver's state at the culmination of that 2-s interval. The assessment criteria for these readiness levels are outlined in Table 1.

During the rating process, the evaluators could review each segment and update their previous ratings. Although selecting 2-s intervals minimises evaluators' confusion and accelerates the rating process, it leads to the creation of a discrete rating scale.

3.3 | Readiness Index as Ground Truth

We take the average of the ratings provided by the evaluators to create a single rating score for each 2-s interval. We employ
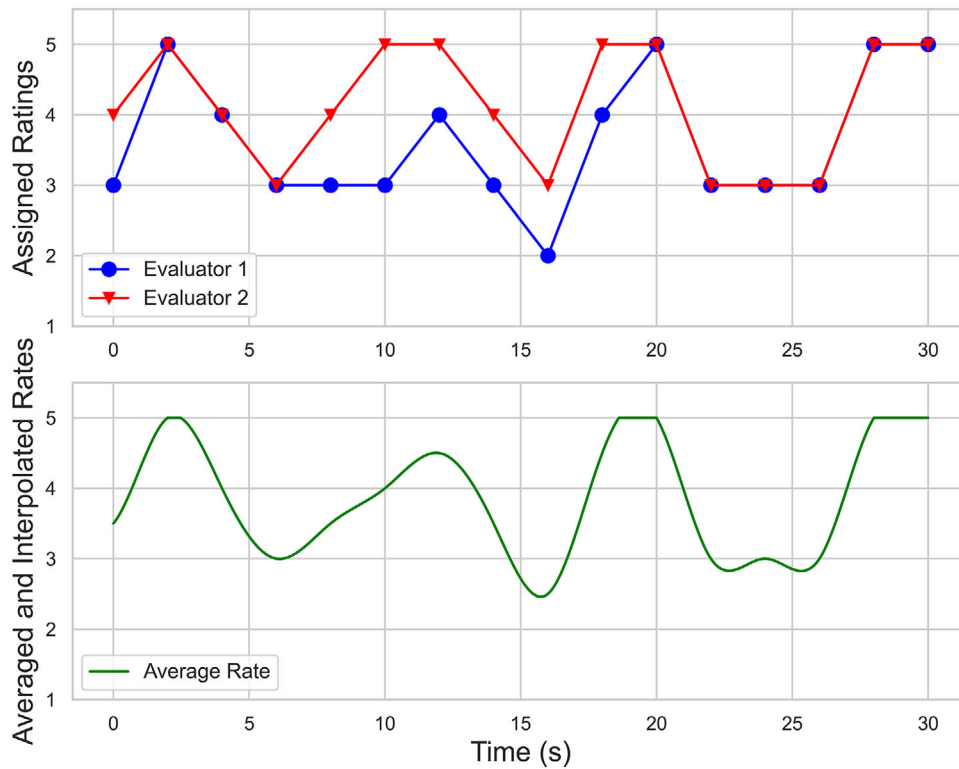
FIGURE 2 | Assigned ratings and derived readiness index for a 30-s time interval.

the cubic interpolation method, as described in [30], to extend this discrete rating set to ensure a continuous representation of driver readiness for take-over. This interpolation method is chosen based on its ability to create smooth and accurate curves between data points, capturing the gradual changes in driver readiness over time. While other interpolation methods, such as linear, polynomial, spline, and nearest neighbour, were considered as alternatives, cubic interpolation was selected due to its capacity to handle complex trends, reduce oscillations, and align with the expected gradual changes in human readiness state [31].

Consequently, for each frame, which corresponds to a time of 1/30 s, an interpolated value is generated for the readiness index using cubic interpolation. This value serves as the foundation for training and evaluating the driver readiness evaluation model, acting as a representative ground truth.

3.4 | Qualitative Analysis of Readiness Ratings

In Figure 2, an illustration of a representative evaluation during a 30-s timeframe is displayed. The upper row showcases the ratings designated by individual evaluators. It is evident from the illustration that the ratings display a favourable level of convergence. While there seems to be a lower level of strictness in the rating assignment by evaluator 2, the overall rating trend remains consistent. The lower row, on the other hand, portrays the readiness index in the form of a continuous variable. This index is derived through the process of averaging and cubic interpolation of the provided rates.

4 | Methodology

We aim to evaluate the readiness index continuously based on processing recorded driver videos. This issue is addressed as a regression problem. Drawing inspiration from study [13], at each moment, the video recorded from the past 2 s is taken as input to the model, and the output is an estimated measure of driver readiness for take-over. CNN-based models are utilised to extract head pose and eye-tracking features from frames related to each two-second interval. Subsequently, these frame-level features are concatenated to represent a frame-by-frame depiction of the driver's state.

While CNNs extract general features from a single frame, the decision-making basis for assessing driver readiness should consider spatio-temporal data via multiple frames. Therefore, we need a model capable of reasoning about temporal dependencies to estimate the readiness index. Given the effectiveness of LSTM networks in modelling long-term temporal dependencies in various sequential modelling tasks within the domain of automated vehicles [13, 32, 33], it is utilised to estimate the readiness index. As depicted in Figure 3, the CNN component extracts complex spatial features from individual frames. Meanwhile, the LSTM component effectively captures the temporal relationships across a sequence of frames.

By employing consecutive LSTM layers, we aim to enhance the model's ability to capture and understand temporal dependencies in the input data, as relevant studies show that deeper neural network architectures can better handle complex patterns in sequential data [34, 35]. The final step includes a dense
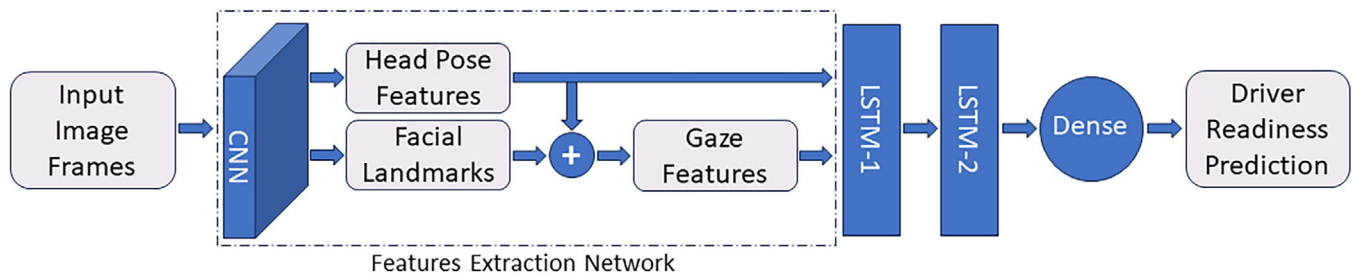
FIGURE 3 │ Deep learning architecture for driver readiness evaluation.

layer, which integrates the combined insights from the entire architecture to generate a precise numerical value that serves as the predicted driver readiness index.

4.1 │ Frame-Level Feature Extraction

We utilise the SPIGA model presented by Prados-Torreblanca et al. [36], for extracting head pose and eye-tracking features due to its demonstrated higher accuracy compared to other state-of-the-art (SoTA) models. SPIGA stands out as a leading model for head pose estimation on the WFLW dataset [37], having achieved the lowest MAE mean (°) among its counterparts. Moreover, in landmark extraction, the SPIGA model is also considered SoTA for face alignment on datasets like 300 W [38] and MERL-RAV [39]. This CNN-based model takes recorded facial frames as input and calculates three head pose-related features: Yaw, pitch, and roll, and 98 Facial Landmarks for each frame. Its CNN architecture excels at extracting relevant features, making it ideal for processing the facial frames in our study. Additionally, SPIGA has been successfully tested in challenging lighting conditions, such as those encountered in in-cabin monitoring scenarios, further boosting its suitability for our research context.

4.2 │ Head Pose Estimation

In computer vision, head pose estimation primarily refers to measuring the orientation of the driver's head with respect to the camera or the global coordinate system. However, achieving this requires understanding the camera's built-in characteristics to correct the perceptual bias caused by perspective distortion [40]. A common approach in head pose estimation is predicting the relative orientation in a 3D environment using Euler yaw, pitch, and roll angles [41].

Although head pose estimation has been a longstanding and extensively explored issue, attaining satisfactory results has only become feasible due to recent advancements in deep learning. Conditions that pose challenges, such as extreme angles, poor lighting, occlusions, and the presence of additional faces within the frame, complicate the task of detecting and estimating head poses for data scientists [41]. This study utilised the SPIGA model, pre-trained on the WFLW dataset [37], to extract yaw, pitch, and roll features. Figure 4 illustrates the model's output for a frame within the DMD dataset.

**FIGURE 4** │ Head pose estimation for a sample frame of DMD dataset using SPIGA model.

4.3 │ Eye-Tracking Data

Eye-tracking metrics serve as signals of where the visual focus is directed, and this aspect is extremely important for understanding, capturing, and predicting the evolving take-over procedure [42]. Automated driving systems should be able to measure and monitor drivers' situational awareness during the take-over transition period using eye-tracking data with machine learning models [43]. The eye openness ratio provides an understanding of eyelid closure, potentially indicating driver drowsiness or fatigue [44]. Meanwhile, the horizontal and vertical ratios provide insights into gaze direction and fixation patterns [45], offering valuable information about the driver's interaction with the environment.

To initiate the analysis of eye-tracking data, the process begins with extracting facial landmark points through the utilisation of the SPIGA model [36], as visualised in Figure 5. Among the 98 extracted facial landmarks, special significance is placed on landmarks 60 to 75, pinpointing the localisation of the eyes. Noteworthy are landmarks 96 and 97, crucial for determining the precise positions of the irises. Figure 6 illustrates the spatial arrangement of these key points on the driver's face.

4.3.1 │ Eye Aspect Ratio Metric

Eye aspect ratio (EAR) provides an approximate estimation of the eye openness state as the ratio of the vertical distance to the horizontal distance of the eye [46]. As we employ 98 facial landmarks, the EAR can be defined by the equations below [47]:
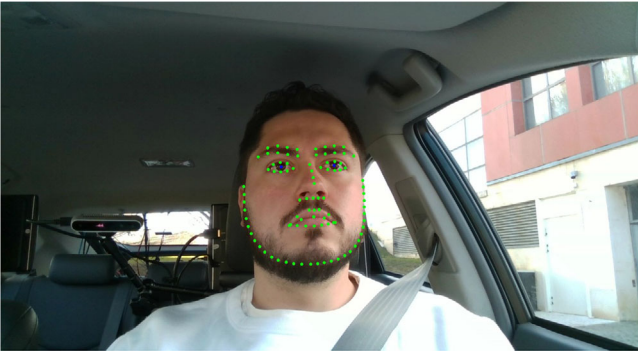
FIGURE 5 | Extracted facial landmarks via SPIGA model for a sample frame.
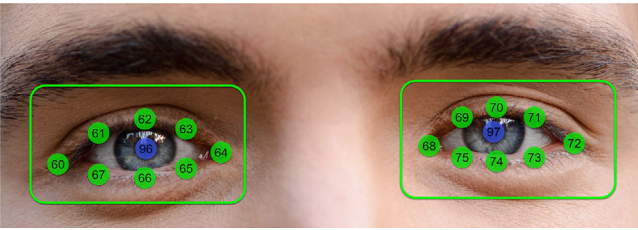
FIGURE 6 | Visual representation of eye and iris landmarks.

$$EAR_R = \frac{|P_{66} - P_{62}|}{|P_{64} - P_{60}|} \quad (1)$$

$$EAR_L = \frac{|P_{74} - P_{70}|}{|P_{72} - P_{68}|} \quad (2)$$

where EAR_R and EAR_L represent right and left eye aspect ratios, respectively and P_{dd} shows the keypoint number dd. EAR values below the threshold of 0.2 indicate that the eyes are closed [48]. Furthermore, when there is an extreme head rotation leading one of the eyes to fall outside the camera's field of view, a noticeable difference in the calculated EAR values between the right and left eyes becomes apparent.

Figure 7 illustrates the results of EAR computations for four different scenarios: (a) closed eyes, (b) partially open eyes, (c) fully open eyes, and (d) a situation where the left eye falls outside the camera's field of view, due to excessive rotation of the driver's head.

4.3.2 | Horizontal Gaze Ratio Metric

The horizontal gaze ratio (HR), ranging from 0.0 to 1.0, indicates horizontal gaze direction. This metric is derived from ref. [49] open-source gaze tracking Python code definitions, utilising the specific landmark x-coordinates (x_{dd}).

The right eye's horizontal gaze ratio (HR_R) is determined by the x offset of the right pupil's landmark 96 from landmark 60, divided by the distance between landmarks 64 and 60:

$$HR_R = \frac{|x_{96} - x_{60}|}{|x_{64} - x_{60}|} \quad (3)$$

Similarly, HR_L represents the left eye's horizontal gaze ratio, calculated by the x offset of the left pupil's landmark 97 from landmark 68, divided by the distance between landmarks 72 and 68:

$$HR_L = \frac{|x_{97} - x_{68}|}{|x_{72} - x_{68}|} \quad (4)$$

The obtained values reflect the horizontal gaze direction. A value within the range [0, 0.33) suggests a rightward gaze, [0.33, 0.67) indicates a straight-ahead gaze, and [0.67, 1.0] implies a leftward gaze tendency.

We evaluated the computed values for the horizontal gaze ratio in each of the five driver readiness classes based on the ground truth values generated in Section 3.3. Figure 8 illustrates the calculated mean and standard deviation for the horizontal gaze ratio in each of these classes. It can be observed that the driver's readiness level increases when the calculated standard deviation decreases. For frames representing the minimum readiness level with a value of 1, we observe the highest standard deviation of 0.158, whereas conversely, for the highest assigned readiness level of 5, the standard deviation reduces more than threefold to 0.049.

This interpretation is straightforward: When a driver possesses a higher level of readiness, their gaze concentration is directed towards a specific forward region, resulting in a slower rate of eye positional changes. This observation aligns with the findings of ref. [50], which demonstrated that the standard deviation of horizontal gaze position is the most sensitive metric to cognitive focus variations, even though it remains one of the simplest metrics to calculate.

4.3.3 | Vertical Gaze Ratio Metric

The vertical gaze ratio (VR) is a metric that returns a value between 0.0 and 1.0, indicating the vertical orientation of the gaze. The calculation involves dividing the vertical difference between the pupil's position and a specified reference point (62 for the right eye, 70 for the left eye) by the height of the corresponding eye area [49]:

$$VR_R = \frac{|y_{96} - y_{62}|}{|y_{66} - y_{62}|} \quad (5)$$

$$VR_L = \frac{|y_{97} - y_{70}|}{|y_{74} - y_{70}|} \quad (6)$$

where VR_R represents the vertical gaze ratio for the right eye, VR_L for the left eye, and y_{dd} denotes the vertical position of a specific eye landmark dd.

In this context, a value in the interval [0, 0.33) corresponds to looking upwards, [0.33, 0.67) indicates a direct gaze, and [0.67, 1] signifies looking downwards. While the primary use of the vertical gaze ratio is in conjunction with the horizontal gaze ratio for gaze zone classification, values tending towards 1—where the driver's gaze is directed towards the lower portion of their field of view—can suggest a decrease in readiness. This situation may arise when the driver is engaged in an NDRT, such as tuning the radio or using a mobile phone.

FIGURE 7 | Variation in eye aspect ratio across different eye states and extreme head rotation.
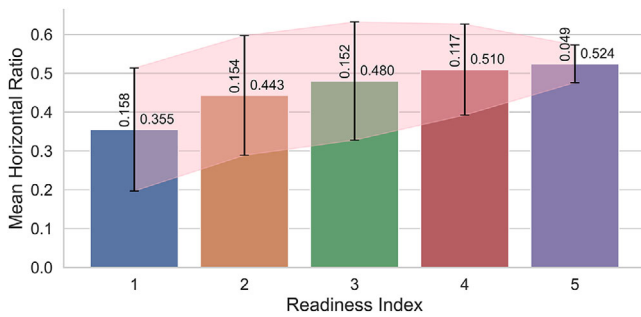
FIGURE 8 | Relationship between driver readiness levels and variability in horizontal gaze ratio.

4.4 | Gaze Zone Classification

Building upon the extracted features related to head pose and eye-tracking data, the driver's gaze area is categorised into nine distinct regions of interest (ROI). The selection of zone boundaries follows the definition of the DMD dataset.

While the decision to adopt a nine-region gaze classification introduces inherent challenges, particularly with gaze regions like $G0$ and $G1$ due to their close proximity, the availability of gaze region annotations for the 2331 frames can provide a robust foundation for training the gaze zone estimation classifier. While we acknowledge that the proposed approach might not compete with the precision of specialised, highly calibrated eye gaze tracker cameras, the introduction of human oversight in the manual review stage serves as a practical strategy to fill the accuracy gap. On this basis, we ensure that a suitable foundation will be established for the study of driver readiness assessment in the following steps.

The main focus is selecting an appropriate classifier based on the extracted features for head pose and eye-tracking data, which can assign the driver's gaze zone to one of these nine ROIs in each frame. Two separate scenarios were considered to estimate the driver's gaze region using a Random Forest-based classifier. In Case 1, the estimation relied solely on extracted head pose features, while in Case 2, the evaluation was achieved by combining head pose features with additional eye-tracking data. The capability of random forest classifiers in diverse machine learning applications is widely acknowledged, attributed to their skill in directly interpreting acquired parameters and their minimal necessity for hyper-parameter tuning [51]. Following the approach outlined in [51], we divide the dataset into training and test sets as 80%–20%. Specifically, 1864 frames were allocated to training purposes, while 467 were reserved for testing.

To assess the performance of the trained model, accuracy metric and confusion matrix are employed. Accuracy is a fundamental metric used to evaluate the performance of classification models. It measures the proportion of correctly predicted instances among the total instances in a dataset. It is defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (7)$$
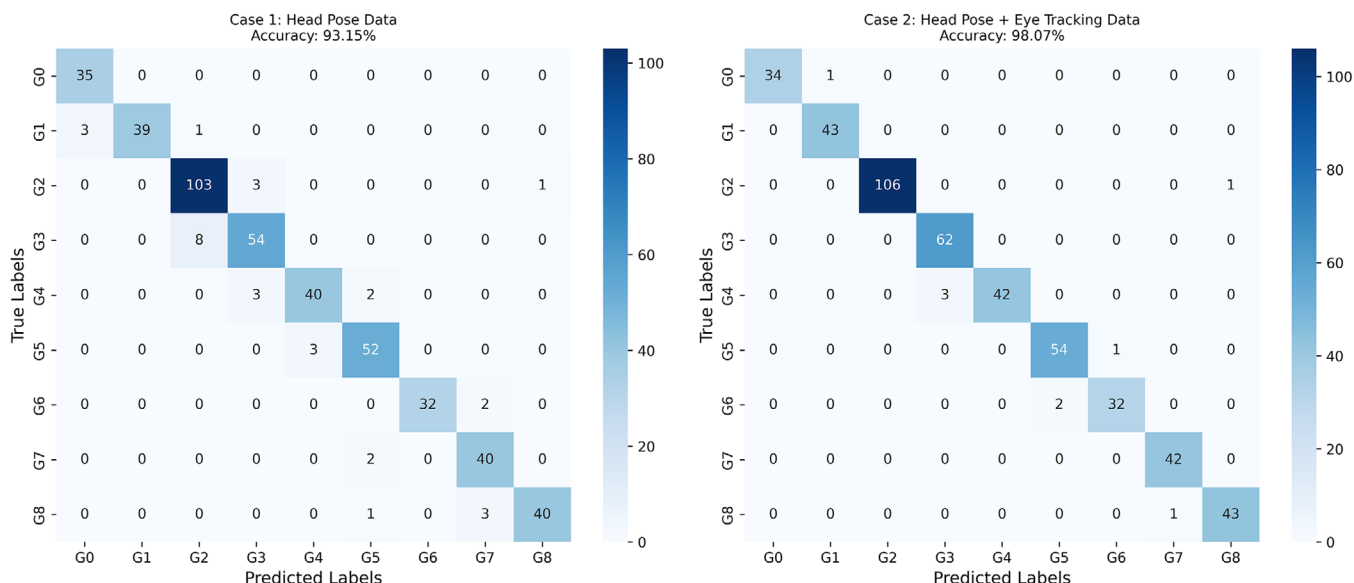
FIGURE 9 | Comparison of confusion matrices and accuracies for gaze zone classification scenarios.

While accuracy provides a general overview of a model's effectiveness, it may not always be sufficient to fully understand its performance, especially in cases where class distributions are imbalanced. This is where a confusion matrix comes into play. The confusion matrix presents a table format that breaks down the model's predictions into four distinct categories: True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These elements provide insights into how well the model distinguishes between classes and where errors occur. By analysing the confusion matrix, it becomes possible to gain a more nuanced assessment of a model's ability to correctly classify instances across different classes [52].

Figure 9 provides a visual representation of the confusion matrices and the corresponding calculated accuracies for both analysed scenarios. In the case of the classifier that relies solely on head pose features, the accuracy stands at 93.15%. However, when eye-tracking features are integrated into the existing information, the accuracy experiences a notable enhancement, increasing by approximately 5% to attain a commendable accuracy of 98.07%. This improvement underscores the significant contribution of incorporating eye-tracking data into the model's predictive capabilities, resulting in a more refined and accurate gaze region classification. The augmentation of accuracy by this margin indicates that the integration of eye-tracking data complements and refines the model's decision-making process, enabling it to better distinguish between different ROIs.

Given the accuracy demonstrated by the random forest classifier in gaze zone classification, we proceeded to utilise the model to predict the gaze regions for 3309 frames, which had not been previously annotated.

Figure 10 presents the Pearson correlation coefficient between driver gaze-related features and driver readiness values. The correlation coefficient, a statistical measure, aids in comprehending the strength and direction of a linear relationship between two data sets. Its function is to reveal how closely changes in one
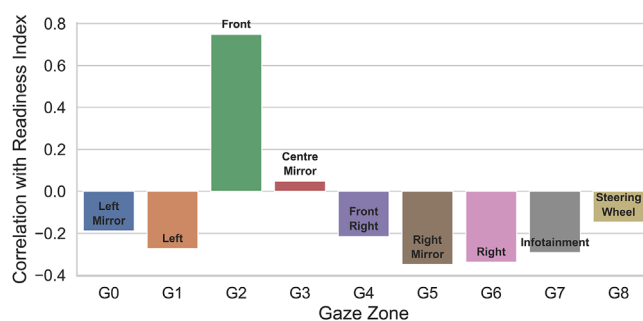
FIGURE 10 | Correlation of driver gaze patterns with readiness levels.

variable match up with changes in another variable [53]. The graph highlights the distinct relationships between different gaze regions and the readiness index. Notably, regions associated with forward gaze demonstrate the highest positive correlation with the readiness index. Conversely, gaze regions linked to driver distractions, such as glancing to the left or right, focusing on infotainment systems, or fixating on the steering wheel, exhibit negative correlations with the readiness index. These observations align with the understanding that these actions may divert the driver's attention away from the road, potentially lowering their readiness to respond effectively to unexpected situations.

However, an interesting exception occurs regarding the negative correlations observed with the left and right mirrors. This difference arises from the specific context in which the DMD dataset was generated. In this scenario, drivers have maintained long-lasting and consistent fixation on the mirrors. Evaluators have interpreted this behaviour as indicative of a driver's readiness level reduction. While unexpected, this observation underscores the complexity of interpreting gaze-related features within their unique context.

The random forest classifier categorises the driver's gaze direction by labelling each frame of data based on the driver's visual focus.

However, for this data to be effectively utilised by the subsequent machine learning model, which evaluates driver readiness, it needs to be pre-processed. This is where the importance of one-hot encoding comes into play.

In the context of the driver's gaze zone classification, employing one-hot encoding translates to generating a set of binary features, precisely nine in number, each representing a distinct gaze region. The core principle behind one-hot encoding is the creation of new columns, each corresponding to a unique category present in the categorical variable. Subsequently, for every data point, a value of 1 is assigned to the appropriate category-specific column, while all other columns are assigned a value of 0. This transformation enables machine learning models to comprehend categorical distinctions without making any assumptions about the inherent order of categories [54]. Every frame's data is encoded such that one of these features holds a value of 1, indicating the driver's focus on a particular gaze region, while the remaining features hold values of 0. By adopting this format, the dataset becomes amenable to harnessing the capabilities of neural networks within the domain of machine learning.

4.5 | Time Series Cross-Validation

With time series data, the concept of cross-validation takes on a new form that adheres to temporal order. Time series carry hidden temporal sequences, predicting future events based on past occurrences. Here, training and validation sets are defined while preserving the temporal order. This approach mimics real-world scenarios where past events guide predictions of future occurrences. Similar to our model, which evaluates the driver's readiness at each moment based on their head movement and gaze position in previous moments. By considering temporal dependencies and sequentially changing training and validation split points, this method ensures that the data from the validation set does not leak into the training set [55]. This encompasses the nature of time series data and provides a more genuine assessment of model performance.

4.6 | Proposed LSTM Models

In this research, the dynamic nature of driver readiness assessment demands a model that can capture temporal dependencies within the sequential data. LSTMs excel in this aspect due to their ability to maintain a memory state that can capture information from earlier time steps and propagate it through time. This is crucial for scenarios where the driver's readiness is influenced not only by recent events but also by past interactions with the driver's state. The integration of both head pose and eye-tracking data as input features aligns well with LSTMs' capacity to handle multiple input streams simultaneously [56].

In this study, the effectiveness of two distinct LSTM architectures was explored in the prediction of driver readiness: the vanilla LSTM [57] and the bidirectional LSTM [58]. The length of the input sequence for the model is determined to be 60 frames (2 s) after evaluating various time window sizes of 30, 60, 90, and 120 frames. Although a slightly lower mean absolute error (MAE)

was observed for 120 frames, the standard deviation was higher compared to 60 frames.

The vanilla LSTM model includes 64 neural units (32 units in each direction for the bidirectional LSTM) to effectively handle the temporal dependencies within this 60-frame input sequence, inspired by [35]. Additionally, we utilised two consecutive LSTM layers for each architecture of vanilla and Bidirectional models as suggested in studies [34, 35]. This makes the network deeper and allows for the extraction of more intricate patterns from the input data, potentially capturing more complex spatio-temporal dependencies and improving prediction capability. Finally, a dense layer with a single output unit was included, utilising linear activation to facilitate regression tasks.

4.7 | Training and Optimisation Strategy

In the training phase of our models, our primary objective was to minimise MAE loss between the predicted values and the actual ground truth values within the training dataset. To optimise the model's performance, we employed the Adam optimiser [59], with a learning rate of 0.001. The learning rate determines the step size taken during each iteration of the optimisation process. A higher learning rate might cause the optimisation process to overshoot the optimal solution, while a lower learning rate might result in slower convergence [60]. The chosen learning rate of 0.001 balances quick convergence and stable optimisation [61], allowing the model to refine its predictions while gradually avoiding drastic oscillations or overshooting.

The model was trained using the Tensorflow deep-learning package on a Linux OS with an NVIDIA T4 GPU utilising CUDA parallel processing APIs. The training batch size was four for 100 epochs, and the training process was halted once the model converged with no significant decrease in MAE loss function value.

4.8 | Validation and Test Strategy

We employed a model selection process to determine the most suitable model for final evaluation. This process involved using a validation set that was distinct from the training set. For each training epoch, we calculated the MAE between its predictions and the true values on the validation set. The model that exhibited the lowest MAE loss on the validation set was chosen as the most optimal one [62]. This selection criterion ensured that we identified the model configuration that provided the best predictive performance on unseen data.

Subsequently, the chosen model was evaluated on the test set, which was separate from both the training and validation sets. This set served as an independent measure of the model's performance and generalisation ability. By reporting the results on the test set using the model with the minimum validation set MAE, we aimed to provide an unbiased and reliable assessment of the model's capability to predict the readiness index accurately in real-world scenarios.

5 | Evaluation

This section highlights the significance of evaluation metrics, particularly the MAE, in quantifying the alignment between predicted and actual values. We analyse the impact of various parameters on the model's predictions, delving into the influence of cross-validation fold configurations and batch sizes on the model's accuracy. Next, we explore the importance of integrating head pose and gaze features within the model architecture for assessing driver readiness. Finally, we present the results obtained from employing different LSTM architectures with distinct feature sets, revealing the model's ability to assess driver readiness in conditionally automated vehicles. Each experiment provides valuable insights into specific aspects contributing to the overall performance and effectiveness of the model.

5.1 | Evaluation of Model Performance

Evaluation metrics are essential tools for quantifying the performance of machine learning models. One common metric used for evaluating regression tasks, such as driver readiness evaluation, is the mean absolute error (MAE). This metric measures the average absolute difference between the predicted values $\hat{y}_i$ and the actual ground truth values y_i, as outlined in Equation (8) [63].

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^{m} |\hat{y}_i - y_i| \tag{8}$$

MAE provides a straightforward and interpretable measure of how well the model's predictions align with the actual driver readiness status. A lower MAE indicates that the model's predictions are, on average, closer to the actual values, implying better prediction performance.

5.2 | N-Fold Configuration

Since ensuring consistent test set sizes across all splits was a key consideration in our analysis, the maximum number of folds was constrained by the ref. [64] formula, and the split of the data was ultimately constrained to a maximum of five folds:

$$\text{Validation size} + \text{Test size} = \frac{\text{Number of samples}}{\text{Number of splits} + 1}$$

$$(400 + 60) + (400 + 60) = \frac{5640}{\text{Number of splits} + 1} \tag{9}$$

$$\text{Number of splits} = 5.13$$

To investigate the impact of selecting an optimal number of folds on the performance of the model, with both head pose and gaze features, a thorough analysis was conducted. The results of this analysis are illustrated in Figure 11 and provide insight into the relationship between the number of folds and the average MAE of the model.

Interestingly, the results show a trend of decreasing average MAE as the number of folds increases from 2 to 4, reaching its lowest value of 0.363 at four-folds. This suggests that, in this particular context, a higher number of folds enhances the model's
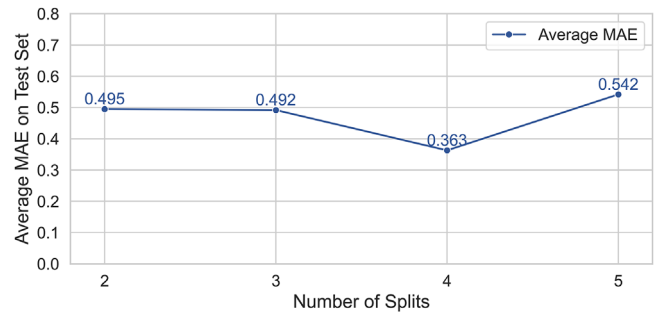

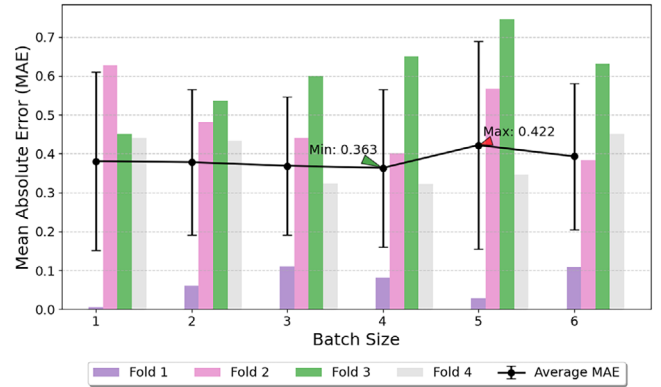**FIGURE 11** | Impact of fold configuration on model performance.

FIGURE 12 | Impact of batch size on model performance.

generalisation capabilities and predictive accuracy. However, it is worth mentioning that when the number of folds is increased to 5, there is a noticeable rise in the average MAE to 0.542.

5.3 | Optimal Batch Size Selection

Batch size is a key hyper-parameter which refers to the number of training examples utilised in a single iteration of the model's learning process. In this study, we evaluated different batch sizes (from 1 to 6) to understand their impact on model performance and training efficiency.

Upon analysing the results, as shown in Figure 12, it is evident that different batch sizes have led to varying levels of model performance. Batch sizes 1 and 2 demonstrated low MAE values for certain folds, suggesting potential over-fitting to the training set. Moreover, the high standard deviation of MAE values across folds for these batch sizes indicated inconsistency in performance. Batch sizes 5 and 6, on the other hand, resulted in higher average MAE values compared to batch sizes 2–4, indicating reduced model accuracy. Despite batch size 6 having a low standard deviation, its relatively high average MAE suggests that the model might not capture the underlying patterns effectively.

Considering the trade-off between training efficiency and model performance, batch sizes 3 and 4 emerge as promising options. Batch size 3 produced an average MAE of 0.369 and a relatively low standard deviation of 0.178. Likewise, batch size 4 achieved an average MAE of 0.362 and a standard deviation of 0.209. Considering larger batch sizes can improve training efficiency

TABLE 2 | MAE calculated for predicted readiness index values in comparison to the ground truth.

Features used			
Head	**Gaze**	**Vanilla LSTM**	**Bidirectional LSTM**
✓		0.935	0.934
	✓	0.441	0.385
✓	✓	**0.375**	**0.363**

due to hardware optimisation and parallelism [65], batch size 4 emerges as the most suitable choice.

5.4 | Results

The obtained results, as presented in Table 2, reveal several key insights. Both LSTM architectures exhibited similar predictive capabilities when using only the head features, like the 'owl' glances, resulting in MAE values of approximately 0.935 for vanilla LSTM and 0.934 for bidirectional LSTM. Moreover, when focusing solely on the gaze features that resemble the 'lizard' glances, the bidirectional LSTM model showed superior performance with an MAE of 0.385, outperforming the vanilla LSTM's MAE of 0.441.

Significantly, the integration of both head and gaze features enhanced prediction accuracy. The bidirectional LSTM model, when considering these combined features set, demonstrated the most robust performance, achieving an MAE of 0.363 compared to the vanilla LSTM's MAE of 0.375. Consequently, these findings highlight the importance of incorporating both head pose and eye-tracking data, indicating that a bidirectional architecture further improves the model's predictive capacity to assess driver readiness in conditionally automated vehicles.

To provide a qualitative assessment of the model's performance when combining features, an example of data analysis utilising the bidirectional LSTM model on the test set of the second fold is presented in Figure 13. This analysis aims to show how the model's predictions evolve when utilising different sets of features. The top row shows predicted ratings based exclusively on head pose features, whereas the middle row presents ratings solely derived from gaze zone features. The bottom row of the graph displays the model's performance when both sets of features are combined. Upon observation, it becomes evident that the model's performance is notably enhanced when these feature sets are combined.

Specifically focusing on the range of frames spanning from 3530 to 3740, corresponding to 7 s of the recorded video, successive changes in the driver's head movement and gaze direction can be observed during this period. Despite these variations, due to the driver's attention to the driving environment remains consistent. The situational awareness has led evaluators to assign a high level of readiness to this interval. However, it is worth noting that while the model does recognise these variations in head pose and gaze direction, its estimation of readiness slightly differs from
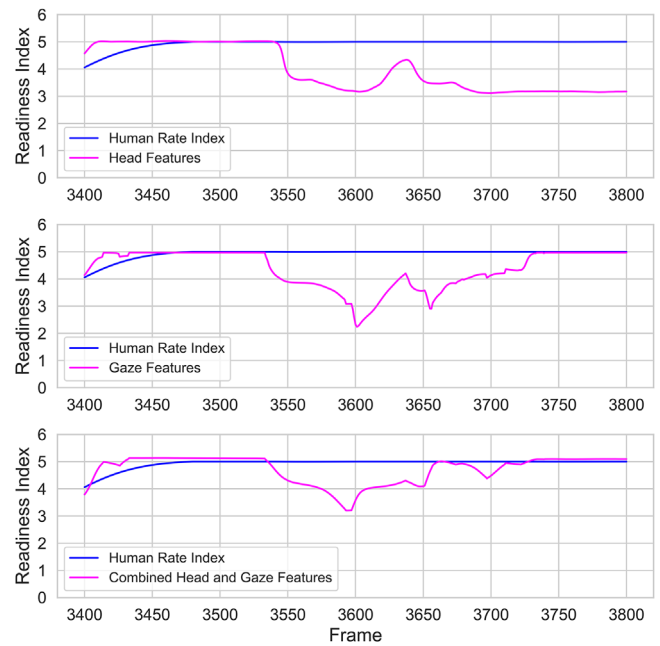

FIGURE 13 | Evolution of model performance through feature integration.

the evaluators' high rating. This difference might be due to the model's sensitivity to small changes that might not necessarily reflect a reduction in readiness but rather a reaction to the feature change.

Nevertheless, the key point is that the model's performance improves when both head pose and eye-tracking data features are combined. The synergy between these two sets of features allows the model to better capture and understand the complex interplay between the driver's head movements, gaze direction, and overall readiness level.

Figure 14 provides a visual representation of the output from the best model in each cross-validation iteration. The graphs are colour-coded to aid in understanding the model's learning process. The goal is to find optimal hyper-parameters as the model progresses through different epochs in the training set, represented by the green colour in the graphs. The model's output is assessed using the validation set, denoted by the yellow sections in the figures throughout each epoch. The model's architecture is designed to select the model with the lowest MAE on the validation set as the best model. Once the best model is identified, it is subjected to testing on the red section, which represents a previously unseen test set.

The MAE results attained from the test set are averaged, and this outcome contributes to the overall assessment of prediction performance. In the case of the best model, which employs a bidirectional LSTM architecture and utilises both head pose and gaze zone features, the achieved MAE is recorded as 0.363. A lower MAE suggests better performance, as it indicates less loss function value in the predictions compared with the ground truth values. The MAE performance improvement of −0.012 the bidirectional LSTM over the vanilla LSTM by is equivalent to a 0.3% error reduction in predicting driver readiness status (1 to 5).
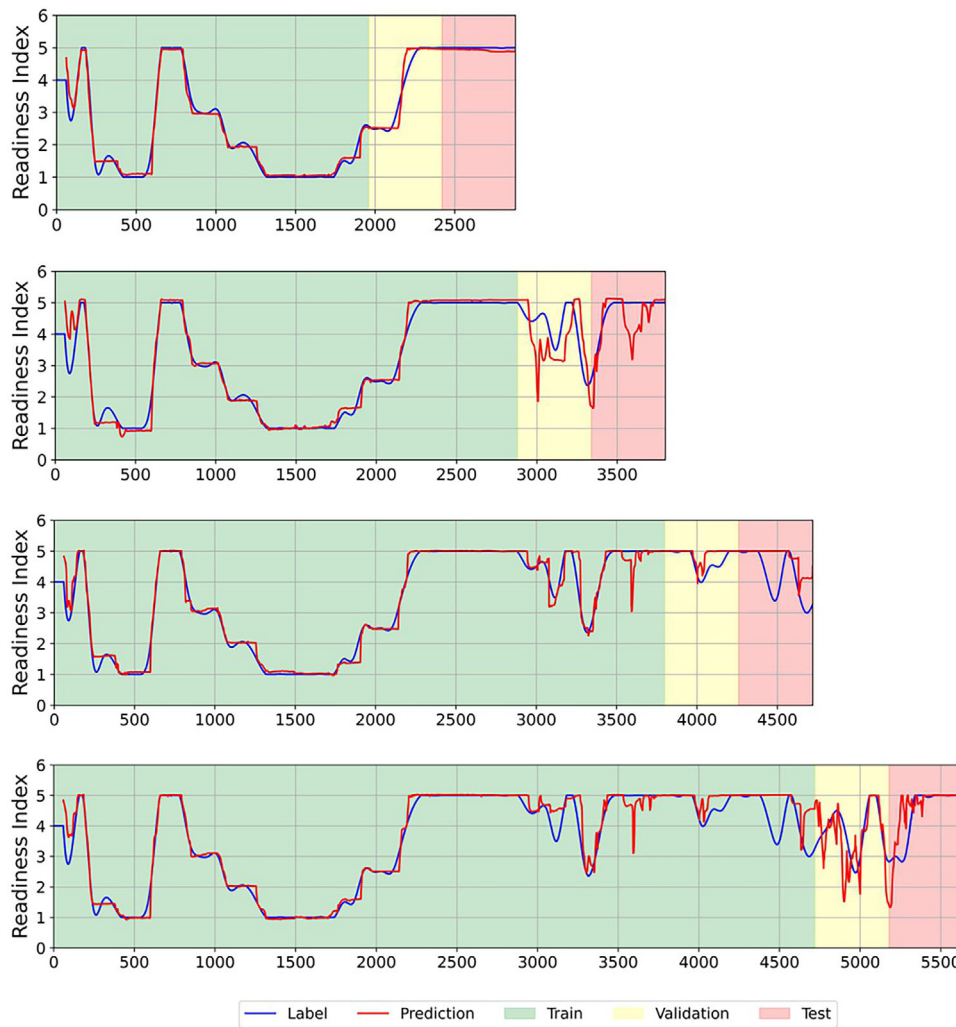
FIGURE 14 | Visualisation of model outputs in different cross-validation folds.

In addition to the MAE metric, we have also considered the accuracy and F1 score metrics to demonstrate the performance of the best model in the driver readiness classification. We categorised driver readiness status into binary classes: 'not ready' and 'ready'. The not ready class consists of 'non-attentive', 'low attentiveness', and 'partially ready'; while the ready class includes 'moderately ready' and 'fully ready' as defined in Table 1. The $F1 = \frac{2TP}{2TP + FP + FN}$, considers both false positives and false negatives, thus assessing the model's effectiveness in binary classifications. The model achieves an accuracy of 0.89 and an F1 score of 0.94, indicating that the model is able to correctly predict the driver's readiness status in almost nine out of ten cases, demonstrating strong classification performance.

It is essential to acknowledge that the presence of noise in the output from the test section is evident. This noise can be attributed to several factors. First, the dataset has limited frames, which could affect the model's ability to generalise effectively. Additionally, an imbalance in driver behaviour across different dataset segments might contribute to the observed noise. For example, more straightforward scenarios at the beginning of the video and more complex situations with rapid variations in head and eye positions at the end of the videos can challenge the model's performance. Despite these challenges, the model demonstrates a reasonable level of performance in its predictions when compared to the assigned readiness index values.

6 | Conclusion

Efficient feature extraction and integration for driver readiness evaluation was one of the objectives of this study. By combining head pose and eye-tracking features, this model demonstrated significant capabilities in interpreting complex relations between the driver's physical head orientation and direction of attention and focus. LSTMs were selected due to their effective temporal pattern recognition. Bidirectional LSTM architecture, more specifically when initialised with this feature combination, performed optimally, achieving a substantial mean absolute error (MAE) of 0.363. This emphasises that a holistic perspective on driver behaviour provides a more accurate representation of their readiness level.

A significant challenge in this research was the absence of a comprehensive driving dataset that encompasses a wide range of driver behaviours in automated vehicles. Additionally, the lack of a clear ground truth for readiness assessment posed significant obstacles. To address these challenges, we created a ground truth

set in order to accurately evaluate our model and also for the benefit of the research community in future studies.

This was done based on the qualitative observations and ratings of multiple human evaluators. The model's demonstrated performance at an acceptable level suggests its potential for real-world level 3 automated driving applications. As technology progresses and data accessibility improves, these obstacles can be reduced, leading to even greater precision and resilience in predictive models.

The modular model architecture makes it easy to integrate new driver-specific features, such as hand status, body posture, and steering wheel activity, into the model for future research. By combining these different features, the accuracy and predictive power of the model can be improved. This will allow for a more comprehensive assessment of driver readiness in conditionally automated vehicles. The model's adaptability and real-world applicability will also improve, making it more useful for ensuring a safe and reliable automated driving experience.

Author Contributions

Mostafa Kazemi: conceptualisation, formal analysis, investigation, methodology, software, visualisation, writing – original draft, writing – review and editing. **Mahdi Rezaei:** conceptualisation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualisation, writing – original draft, writing – review and editing. **Mohsen Azarmi:** software, visualisation, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in the DMD repository at https://dmd.vicomtech.org/.

References

1. "Road Traffic Injuries," World Health Organisation, accessed June 21, 2021, https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

2. M. Amrun, T. Kamaruddin, N. A. Husain, et al., "A Review on the Autonomous Technology in ASEAN Ncap," *AIP Conference Proceedings* 2643, no. 1 (2023): 1–10.

3. A. M. Boggs, B. Wali, and A. J. Khattak, "Exploratory Analysis of Automated Vehicle Crashes in California: A Text Analytics and Hierarchical Bayesian Heterogeneity-Based Approach," *Accident Analysis & Prevention* 135 (2020): 105354.

4. A. Boelhouwer, A. P. van den Beukel, M. C. van der Voort, and M. H. Martens, "Should I Take Over? Does System Knowledge Help Drivers in Making Take-Over Decisions While Driving a Partially Automated Car?," *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (2019): 669–684.

5. C. Braunagel, W. Rosenstiel, and E. Kasneci, "Ready for Take-Over? A New Driver Assistance System for an Automated Classification of Driver Take-Over Readiness," *IEEE Intelligent Transportation Systems Magazine* 9, no. 4 (2017): 10–22.

6. W. Morales-Alvarez, O. Sipele, R. Léberon, H. H. Tadjine, and C. Olaverri-Monreal, "Automated Driving: A Literature Review of the Take Over Request in Conditional Automation," *Electronics* 9, no. 12, (2020): 2087.

7. T. L. Louw, "The Human Factors of Transitions in Highly Automated Driving," (PhD diss., University of Leeds, 2017).

8. D. Yu, C. Park, H. Choi, D. Kim, and S.-H. Hwang, "Takeover Safety Analysis with Driver Monitoring Systems and Driver-Vehicle Interfaces in Highly Automated Vehicles," *Applied Sciences* 11, no. 15 (2021): 6685.

9. M. Rezaei, M. Azarmi, and F. M. P. Mir, "3D-Net: Monocular 3D Object Recognition for Traffic Monitoring," *Expert Systems with Applications* 227 (2023): 120253, https://www.sciencedirect.com/science/article/pii/S0957417423007558

10. N. Du, F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Predicting Driver Takeover Performance in Conditionally Automated Driving," *Accident Analysis & Prevention* 148 (2020): 105748.

11. D. Dogan, S. Bogosyan, and T. Acarman, "Evaluation of Takeover Time Performance of Drivers in Partially Autonomous Vehicles Using a Wearable Sensor," *Journal of Sensors* 2022 (2022): 7924444.

12. H.-B. Kang, "Various Approaches for Driver and Driving Behavior Monitoring: A Review," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (IEEE, 2013), 616–623.

13. N. Deo and M. M. Trivedi, "Looking at the Driver/Rider in Autonomous Vehicles to Predict Take-Over Readiness," *IEEE Transactions on Intelligent Vehicles* 5, no. 1, (2019): 41–52.

14. J. D. Ortega, M. Nieto, P. Cañas, O. Otaegui, and L. Salgado, "A Real-Time Software Framework for Driver Monitoring Systems: Software Architecture and Use Cases," in *Real-Time Image Processing and Deep Learning*, Vol. 11736 (SPIE, 2021), 76–88.

15. X. Wang, J. Zhang, H. Zhang, S. Zhao, and H. Liu, "Vision-Based Gaze Estimation: A Review," *IEEE Transactions on Cognitive and Developmental Systems* 14, no. 2 (2021): 316–332.

16. A. C. Hayley, B. Shiferaw, B. Aitken, F. Vinckenbosch, T. L. Brown, and L. A. Downey, "Driver Monitoring Systems (DMS): The Future of Impaired Driving Management?," *Traffic Injury Prevention* 22, no. 4 (2021): 313–317.

17. S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim, "Real-Time Gaze Estimator Based on Driver's Head Orientation for Forward Collision Warning System," *IEEE Transactions on Intelligent Transportation Systems* 12, no. 1 (2011): 254–267.

18. A. Tawari, S. Martin, and M. M. Trivedi, "Continuous Head Movement Estimator for Driver Assistance: Issues, Algorithms, and On-Road Evaluations," *IEEE Transactions on Intelligent Transportation Systems* 15, no. 2 (2014): 818–830.

19. B. Vasli, S. Martin, and M. M. Trivedi, "On Driver Gaze Estimation: Explorations and Fusion of Geometric and Data Driven Approaches," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2016), 655–660.

20. L. Fridman, J. Lee, B. Reimer, and T. Victor, "Owl and 'Lizard': Patterns of Head Pose and Eye Pose in Driver Gaze Classification," *IET Computer Vision* 10, no. 4 (2016): 308–314.

21. S. Vora, A. Rangesh, and M. M. Trivedi, "On Generalizing Driver Gaze Zone Estimation Using Convolutional Neural Networks," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2017), 849–854.

22. C. Lollett, M. Kamezaki, and S. Sugano, "Single Camera Face Position-Invariant Driver's Gaze Zone Classifier Based on Frame-Sequence Recognition Using 3D Convolutional Neural Networks," *Sensors* 22, no. 15 (2022): 5857.

23. BSI, "Road Vehicles Human Performance and State in the Context of Automated Driving: Common Underlying Concepts," ISO/TR 21959-1:2020 (International Organization for Standardization, 2020).

24. C. Marberger, H. Mielenz, F. Naujoks, J. Radlmayr, K. Bengler, and B. Wandtner, "Understanding and Applying the Concept of 'Driver Availability' in Automated Driving," in *Proceedings of the Advances in Human Aspects of Transportation (AHFE) 2017 International Conference on Human Factors in Transportation* (Springer, 2018), 595–605.

25. A. Arslanyilmaz, S. Al Matouq, and D. V. Doner, "Driver Readiness in Autonomous Vehicle Take-Overs," *International Journal of Transport and Vehicle Engineering* 14, no. 8 (2020): 558–563.

26. J. Kim, W. Kim, H.-S. Kim, S.-J. Lee, O.-C. Kwon, and D. Yoon, "A Novel Study on Subjective Driver Readiness in Terms of Non-Driving Related Tasks and Take-Over Performance," *ICT Express* 8, no. 1 (2022): 91–96.

27. R. Greer, N. Deo, A. Rangesh, P. Gunaratne, and M. Trivedi, "Safe Control Transitions: Machine Vision Based Observable Readiness Index and Data-Driven Takeover Time Prediction," preprint, arXiv:2301.05805, January 14, 2023, https://doi.org/10.48550/arXiv.2301.05805.

28. J. D. Ortega, N. Kose, P. Cañas, et al., "DMD: A Large-Scale Multi-Modal Driver Monitoring Dataset for Attention and Alertness Analysis," in *Proceedings of the ECCV 2020 Workshops on Computer Vision* (Springer, 2020), 387–405.

29. N. Li and C. Busso, "Predicting Perceived Visual and Cognitive Distractions of Drivers with Multimodal Features," *IEEE Transactions on Intelligent Transportation Systems* 16, no. 1 (2014): 51–65.

30. S. Waykole, N. Shiwakoti, and P. Stasinopoulos, "Interpolation-Based Framework for Generation of Ground Truth Data for Testing Lane Detection Algorithm for Automated Vehicle," *World Electric Vehicle Journal* 14, no. 2 (2023): 48.

31. Z. Deng and T. Tian, "A Continuous Optimization Approach for Inferring Parameters in Mathematical Models of Regulatory Networks," *BMC Bioinformatics* 15, no. 1 (2014): 1–12.

32. M. Bonyani, M. Rahmanian, S. Jahangard, and M. Rezaei, "DipNet: Driver Intention Prediction for a Safe Takeover Transition in Autonomous Vehicles," *IET Intelligent Transport Systems* 15, no. 1 (2023): 1769–1783.

33. R. Singh, S. Mozaffari, M. Rezaei, and S. Alirezaee, "Lstm-Based Preceding Vehicle Behaviour Prediction During Aggressive Lane Change for ACC Application," in *Proceedings of the 2023 International Symposium on Signals, Circuits and Systems (ISSCS)* (IEEE, 2023), 1–4.

34. Y. Wang, Q. Chen, M. Ding, and J. Li, "High Precision Dimensional Measurement with Convolutional Neural Network and Bi-Directional Long Short-Term Memory (LSTM)," *Sensors* 19, no. 23 (2019): 5302.

35. Z. Gu, Z. Li, X. Di, and R. Shi, "An LSTM-Based Autonomous Driving Model Using a Waymo Open Dataset," *Applied Sciences* 10, no. 6 (2020): 2046.

36. A. Prados-Torreblanca, J. M. Buenaposada, and L. Baumela, "Shape Preserving Facial Landmarks With Graph Attention Networks," preprint, arXiv:2210.07233, October 13, 2022, https://doi.org/10.48550/arXiv.2210.07233.

37. W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou, "Look at Boundary: A Boundary-Aware Face Alignment Algorithm," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), 2129–2138.

38. C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces In-The-wild Challenge: Database and Results," *Image and Vision Computing* 47 (2016): 3–18.

39. A. Kumar, T. K. Marks, W. Mou, et al., "Luvli Face Alignment: Estimating Landmarks' Location, Uncertainty, and Visibility Likelihood," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2020), 8236–8246.

40. E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, no. 4 (2008): 607–626.

41. A. Asperti and D. Filippini, "Deep Learning for Head Pose Estimation: A Survey," *SN Computer Science* 4, no. 4 (2023): 349.

42. F. Zhou, X. J. Yang, and J. C. De Winter, "Using Eye-Tracking Data to Predict Situation Awareness in Real Time During Takeover Transitions in Conditionally Automated Driving," *IEEE Transactions on Intelligent Transportation Systems* 23, no. 3 (2021): 2284–2295.

43. C. Braunagel, D. Geisler, W. Rosenstiel, and E. Kasneci, "Online Recognition of Driver-Activity Based on Visual Scanpath Classification," *IEEE Intelligent Transportation Systems Magazine* 9, no. 4 (2017): 23–36.

44. P. Singh, S. Chauhan, and E. Rajesh, "Real-Time Driver Drowsiness Detection Using Dlib and OpenCV," in *Proceedings of the 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (IEEE, 2022), 956–960.

45. R. Zemblys, D. C. Niehorster, O. Komogortsev, and K. Holmqvist, "Using Machine Learning to Detect Events in Eye-Tracking Data," *Behavior Research Methods* 50 (2018): 160–181.

46. C. B. S. Maior, M. J. das Chagas Moura, J. M. M. Santana, and I. D. Lins, "Real-Time Classification for Autonomous Drowsiness Detection Using Eye Aspect Ratio," *Expert Systems With Applications* 158 (2020): 113505.

47. Q. Zhuang, Z. Kehua, J. Wang, and Q. Chen, "Driver Fatigue Detection Method Based on Eye States with Pupil and Iris Segmentation," *IEEE Access* 8 (2020): 173440–173449.

48. C. Dewi, R.-C. Chen, X. Jiang, and H. Yu, "Adjusting Eye Aspect Ratio for Strong Eye Blink Detection Based on Facial Landmarks," *PeerJ Computer Science* 8 (2022): e943.

49. A. Lamé, "Gazetracking," Github, accessed 2019, accessed February 19, 2025, https://github.com/antoinelame/GazeTracking.

50. Y. Wang, B. Reimer, J. Dobres, and B. Mehler, "The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand," *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (2014): 227–237.

51. A. Tawari, K. H. Chen, and M. M. Trivedi, "Where is the Driver Looking: Analysis of Head, Eye and Iris for Robust Gaze Zone Estimation," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2014), 988–994.

52. H. Kim, W. Kim, J. Kim, S.-J. Lee, D. Yoon, O.-C. Kwon, and C. H. Park, "Study on the Take-Over Performance of Level 3 Autonomous Vehicles Based on Subjective Driving Tendency Questionnaires and Machine Learning Methods," *ETRI Journal* 45, no. 1 (2023): 75–92.

53. J. Lee Rodgers and W. A. Nicewander, "Thirteen Ways to Look at the Correlation Coefficient," *American Statistician* 42, no. 1 (1988): 59–66.

54. C. Seger, "An Investigation of Categorical Variable Encoding Techniques in Machine Learning: Binary Versus One-Hot and Feature Hashing," (Bachelor Thesis, KTH, School of Electrical Engineering and Computer Science, 2018).

55. I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Millán, *Artificial Intelligence in Education* (Springer, 2020).

56. P. Patil, "A Comparative Study of Different Time Series Forecasting Methods for Predicting Traffic Flow and Congestion Levels in Urban Networks," *International Journal of Information and Cybersecurity* 6, no. 1 (2022): 1–20.

57. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation* 9, no. 8 (1997): 1735–1780.

58. A. Graves and J. Schmidhuber, "Framewise Phoneme Classification With Bidirectional LSTM and Other Neural Network Architectures," *Neural Networks* 18, no. 5 (2005): 602–610, https://www.sciencedirect.com/science/article/pii/S0893608005001206.

59. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," preprint, arXiv:1412.6980, December 22, 2014, https://doi.org/10.48550/arXiv.1412.6980.

60. R. Sasse, "Investigating the Influence of Learning Rates on the Learning Speed of Neural Networks" (BS Thesis, University of Stuttgart, 2021).

61. J. Zhou, J. Wan, and F. Zhu, "Transfer learning based long short-term memory car-following model for adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems* 23, no. 11 (2022): 21345–21359.

62. M. K. Tekin, "Vehicle Path Prediction Using Recurrent Neural Network," in *Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)* (IEEE, 2020), 1–5.

63. C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning* (Springer Science & Business Media, 2011).

64. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

65. S. K. Venkataramanaiah, H.-S. Suh, S. Yin, et al., "FPGA-Based Low-Batch Training Accelerator for Modern CNNS Featuring High Bandwidth Memory," in *Proceedings of the 39th International Conference on Computer-Aided Design* (Association for Computing Machinery, 2020), 1–8.