aerospace

Article

An Objective Handling Qualities Assessment Framework of Electric Vertical Takeoff and Landing

Yuhan Li [1,2], Shuguang Zhang [3,*], Yibing Wu [3], Sharina Kimura [2], Michael Zintl [2] and Florian Holzapfel [2]

1 School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China; ffflora@buaa.edu.cn
2 Institute of Flight System Dynamics, Technical University of Munich, 80333 Munich, Germany; sharina.kimura@tum.de (S.K.); michael.zintl@tum.de (M.Z.); florian.holzapfel@tum.de (F.H.)
3 School of Transportation Science and Engineering, Beihang University, Beijing 100191, China; wuyibing@buaa.edu.cn
* Correspondence: gnahz@buaa.edu.cn

Abstract: Assessing handling qualities is crucial for ensuring the safety and operational efficiency of aircraft control characteristics. The growing interest in Urban Air Mobility (UAM) has increased the focus on electric Vertical Takeoff and Landing (eVTOL) aircraft; however, a comprehensive assessment of eVTOL handling qualities remains a challenge. This paper proposed a handling qualities framework to assess eVTOL handling qualities, integrating pilot compensation, task performance, and qualitative comments. An experiment was conducted, where eye-tracking data and subjective ratings from 16 participants as they performed various Mission Task Elements (MTEs) in an eVTOL simulator were analyzed. The relationship between pilot compensation and task workload was investigated based on eye metrics. Data mining results revealed that pilots' eye movement patterns and workload perception change when performing Mission Task Elements (MTEs) that involve aircraft deficiencies. Additionally, pupil size, pupil diameter, iris diameter, interpupillary distance, iris-to-pupil ratio, and gaze entropy are found to be correlated with both handling qualities and task workload. Furthermore, a handling qualities and pilot workload recognition model is developed based on Long-Short Term Memory (LSTM), which is subsequently trained and evaluated with experimental data, achieving an accuracy of 97%. A case study was conducted to validate the effectiveness of the proposed framework. Overall, the proposed framework addresses the limitations of the existing Handling Qualities Rating Method (HQRM), offering a more comprehensive approach to handling qualities assessment.

Keywords: eye metrics; handling qualities; eVTOL; urban air mobility; aircraft design; LSTM

Citation: Li, Y.; Zhang, S.; Wu, Y.; Kimura, S.; Zintl, M.; Holzapfel, F. An Objective Handling Qualities Assessment Framework of Electric Vertical Takeoff and Landing. *Aerospace* **2024**, *11*, 1020. https://doi.org/10.3390/aerospace11121020

Academic Editor: Konstantinos Kontis

Received: 25 September 2024
Revised: 4 December 2024
Accepted: 10 December 2024
Published: 11 December 2024

1. Introduction

Urban Air Mobility (UAM) has garnered increasing interest due to urbanization trends, where the electric Vertical Takeoff and Landing (eVTOL) plays a crucial role. Numerous innovative eVTOL design concepts have emerged, such as wingless configurations, lift-and-cruise models, and vectored thrust systems [1]. These diverse flight mechanics, coupled with advanced flight control surfaces and state-of-the-art human–machine interfaces (HMI), present significant challenges for the certification of eVTOL aircraft. In response, the European Union Aviation Safety Authority (EASA) has modified the Handling Qualities Rating Method (HQRM) to better evaluate eVTOL [2]. Pilot satisfaction, workload, and safety are regarded as important items during civil aircraft certification. The HQRM suggests the Cooper–Harper Rating scale (CHR) to ensure the aircraft can be operated without exceptional piloting skills, following a human-centered approach [3]. However, despite its widespread adoption, the CHR has notable limitations. It is inherently subjective, influenced by individual differences and self-assessment abilities, and continuous evaluation is challenging as participants must pause to complete the rating scale [4]. Therefore, there is a need to enhance the HQRM by developing methods that more objectively integrate human factors.

This study aims to investigate the potential of achieving results equivalent to subjective assessments of handling qualities by exploring correlations between personal eye data and subjective ratings, thus enabling a data-driven approach to validate and simplify HQ assessment. Specifically, the proposed framework leverages eye metrics to assess operators' task workload and establish the relationship between precepted workload and pilot compensation. By combining pilot compensation with flight performance data, the framework offers an objective assessment of aircraft handling qualities. The contributions of this work include the following:

- A database of eye movements at varying levels of handling qualities when operating an eVTOL simulator;
- An analysis of the impact of perceived task workload on subjective handling qualities ratings, supplemented by statistical data mining to reveal key indicators for handling qualities assessment;
- A framework for assessing handling qualities objectively, supported by rigorous analyses and offers novel insights into the use of eye tracking as a supplementary tool for understanding handling qualities and pilot compensation.

2. Background

2.1. Handling Qualities Evaluation

Handling qualities refers to the controllability and maneuverability of an aircraft, allowing pilots to perform specific tasks [3]. It is closely linked to the pilot's perception of how the aircraft handles during various maneuvers. Many studies have investigated methods for evaluating handling qualities, as listed in Table 1.

Table 1. Summary of handling qualities assessment methods.

Method	Examples	Material	Advantages	Disadvantages
Subjective methods	[5,6]	CHR/Bedford scale/NASA TLX	Human-centered evaluation; Easy to understand and apply.	Relies on the pilot's subjective feelings; difficult to quantify.
Flight model	[7,8]	Flight test/simulation data	Adopting quantitative indicators makes the assessment objective.	Largely influenced by model accuracy. Ignore human factor.
Flight test	[9,10]	Flight test reports	Assessment results are realistic and comprehensive.	Costly and risky; affected by pilots' ability to fly and assess.
Pilot model	[11,12]	Motion data and subjective report	Qualitative and quantitative evaluation.	Influenced by pilot subjectivity and environmental factors.

According to EASA [2], no specific generally recognized method for handling qualities evaluation exists, though CHR is commonly employed. CHR guides pilots in assessing HQ through a series of questions (Figure 1). However, as a subjective method, CHR has inherent limitations, particularly in the subjective interpretation of its two key discriminants: "performance" and "pilot compensation" [13]. While "performance" can be quantified using flight parameters, quantifying "pilot compensation" remains challenging [14,15]. To overcome the limitations of CHR and enhance the comprehensiveness of handling qualities assessment, supplementary approaches are necessary.

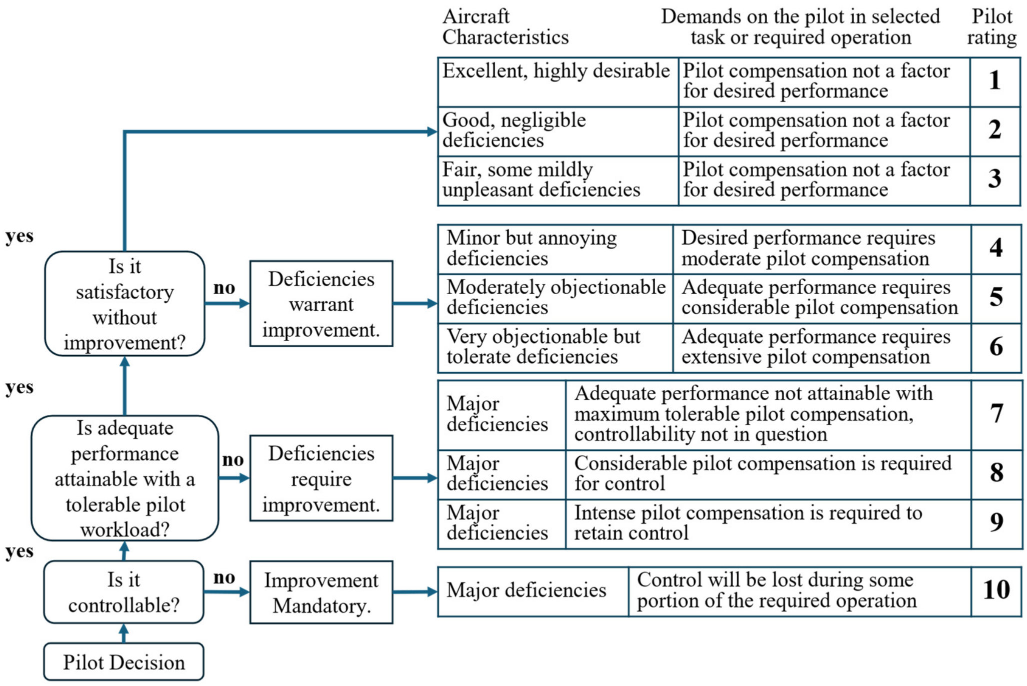

Figure 1. The Cooper–Harper Rating scale [3].

Pilot compensation, as defined by [3], refers to the additional workload a pilot must exert to enhance aircraft performance, whereas workload includes both the compensation for aircraft deficiencies and the effort required to complete a given task [16]. Several studies have explored the relationship between handling qualities and physiological measurements. For example, Klyde et al. [4] attempted to involve electroencephalogram (EEG) and electrocardiogram (ECG) data to assess physical workload and predict handling qualities; Bachelder et al. utilized pilot neuromuscular feedback to estimate handling qualities [17]; Li et al. used EMG and eye data to investigate aircraft control interface deficiencies [18]. However, significant advancements in this field are challenging due to the complex interactions between physiological feedbacks and handling qualities. Therefore, our study aims to establish a method that can deal with the uncertainty and unstable interactions between these data and ratings in a feasible way, thus enhancing the reliability of handling qualities assessments.

2.2. Physiological Measurements in Aviation

Physiological metrics have been shown to provide valuable insights into an operator's engagement, distraction, and workload across various tasks [19]. In our previous research [20], electroencephalography (EEG) has been utilized to investigate pilots' trust in automated Urban Air Mobility (UAM), where excessive signal noise in EEG data posed challenges. Additionally, ECG signals are highly susceptible to fluctuations in operators' emotional and psychological states promoting us to explore other physiological indicators that may be more responsive to compensation rather than overall task load.

Among these indicators, eye metrics are promising because they reflect compensatory behaviors as they provide visual patterns that are directly related to operations and decision making [21]. The application of eye metrics in the aviation industry is well established, including measurements of fixation, visual search, pupil, saccades, and blink [22]. Numerous studies have utilized eye-tracking data to assess pilot performance [23]. For instance, ref. [24] employed areas of interest (AOIs) to measure pilots' attention distribution during flight, while ref. [25] added gaze, fixations, and the corresponding fixation hit maps. These studies indicate that decreased situational awareness often correlates with deteriorating visual search strategies. The pilot's workload was investigated in [26]; selected eye metrics include pupil dilation, saccadic, fixation, and saccades. Increased workload is generally associated with larger pupil dilation, shorter gaze duration, and fewer saccades. Furthermore, pilots' experience levels can be reflected in their eye gaze patterns and fixation behavior [27]. Beyond performance assessments, eye metrics have also been applied to aircraft design, particularly in enhancing aviation displays [28] and improving safety [29]. These studies have demonstrated strong correlations between eye metrics and factors such as situational awareness, pressure, attention, and workload—all of which contribute to pilot compensation. Therefore, this study aims to incorporate eye metrics into the assessment of pilot compensation and handling qualities.

However, most existing studies utilizing eye-tracking data primarily rely on traditional statistical analyses [30]. While such methods yield useful information like gaze position, they often fall short in providing deeper insights into cognitive and behavioral processes. Additionally, eye-tracking data are abundant and complex, requiring considerable effort to analyze [31]. Its relationship with handling qualities is intricate and variable, potentially leading to contradictory results. For instance, while pupil diameter typically decreases as fatigue increases, individual differences and environmental factors can introduce conflicting effects [32]. Consequently, it is crucial to develop an approach capable of capturing the deep features of eye-tracking data in relation to handling qualities and revealing the underlying interaction mechanisms. Rapid advancements in artificial intelligence (AI) present an opportunity to address this challenge. Deep learning (DL) techniques, capable of handling large datasets, modeling nonlinear relationships, and performing end-to-end learning, have been increasingly applied in related studies [29]. For example, ref. [33] adopted various methods including 1D convolutional neural networks (CNNs), Support Vector Machines (SVMs), Random Forests (RFs), and AdaBoost (AB), to recognize eye-movement trajectories in patients with neglect syndrome. Ref. [34] combined autoencoder neural networks with SVMs to extract features from eye data and assess user preferences for humanoid robots. Inspired by these promising results, this study utilizes DL techniques to assess handling qualities levels and ratings.

3. Material and Methods

3.1. Handling Qualities Assessment Framework

The proposed handling qualities assessment framework consists of three key components, as depicted in Figure 2. *Pilot Compensation* is the primary focus of this study. It establishes a correlation between pilot compensation and task workload, facilitating an objective assessment of pilot compensation using eye metrics through statistical data mining and deep learning (DL) network modeling. *Flight Performance* is evaluated using flight data, consistent with methodologies described in previous studies [4]. Lastly, Qualitative Interviews offer descriptions of aircraft deficiencies and help explain discrepancies between the results derived from eye metrics and subjective ratings.

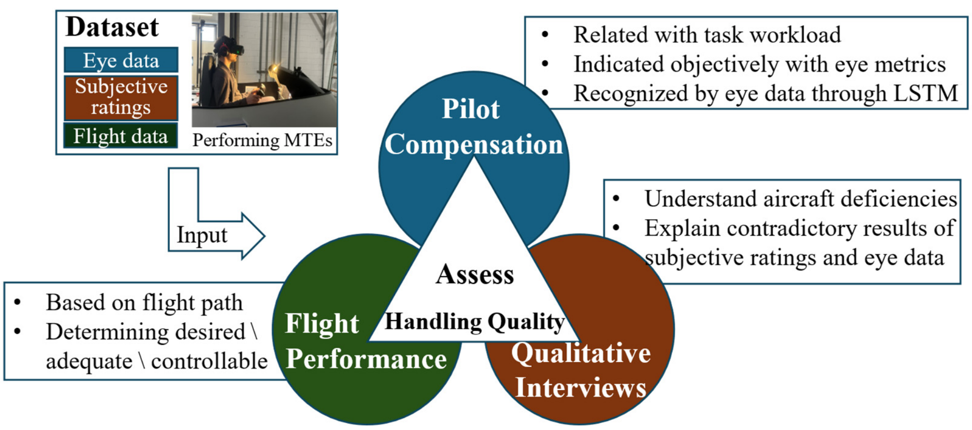

Figure 2. The handling quality assessment framework.

3.2. Dataset

The dataset comprises flight trajectory data, eye-tracking data from operators, and their subjective ratings of handling qualities and task workload across nine Mission Task Elements (MTEs).

3.2.1. Participants

The experiment involved 23 licensed pilots (commercial, helicopter, or private), all experienced in performing MTEs using the eVTOL simulator. All participants were informed of the experimental protocols and signed the informed consent form. Data from two participants who experienced simulator sickness and two others who voluntarily withdrew from the experiment were excluded. Eye-tracking data from three participants were discarded due to poor recording quality. Ultimately, usable data were obtained from 16 participants (15 males, 1 female). The study protocol was approved by the Ethics Committee of the Technical University of Munich.

3.2.2. Simulator Setup

The *Institute of Flight System Dynamics at the Technical University of Munich (TUM-FSD)* has developed an eVTOL simulator based on Mixed Reality (MR) technology and a motion platform [35], as shown in Figure 3. The eVTOL was designed based on the concept of Simplified Vehicle Operation (SVO), relying on two joysticks to achieve unified control, without throttle levers and rudder pedals in cockpit. More details about aircraft configuration and control inceptor can be found in [36].

Figure 3. The motion-based Mixed Reality simulator [35].

The cockpit replica and motion system complied with the EASA standards, providing a professional and realistic environment for research [37]. The development process of the simulator is available in [35–38].

3.2.3. Mission Task Elements

Participants were required to perform a series of MTEs and assess the handling qualities and task workload for each mission. MTEs are a set of standardized flight tasks designed to test an aircraft's response to pilot inputs, typically tailored to the specific aircraft types, ensuring that the aircraft's performance meets established standards [39]. Currently, there are no generally recognized MTEs for eVTOL. With reference to [40], *TUM-FSD* proposed MTEs for eVTOL by analyzing tasks that cover nominal missions and be served as certification candidates, as follows:

- Vertical step: From a stable hover at 10 feet, the eVTOL ascends to a reference altitude (40–50 feet), stabilizes for at least 2 s, then descends back to hover at 10 feet;
- Acceleration/deceleration (Acc/Dec): Starting from a stable hover, participants rapidly increase speed to 50 knots, then decelerate back to a hover, adjusting pitch to maintain altitude;
- Sidestep: From a stable hover, the aircraft moves laterally to a set point, maintaining constant altitude throughout the maneuver;
- Diagonal hover to stop (diagonal): The aircraft moves diagonally while maintaining altitude, beginning from a stable hover with the longitudinal axis set at a 45° angle to a reference line;

- Slalom: Performs a series of smooth turns along the centerline of the test route at 500-ft intervals;
- Takeoff and transition (takeoff): The eVTOL takes off from a stable hover, climbs vertically to 100 feet, accelerates, and transitions into wingborne mode while staying above marked boundaries;
- Re-transition and landing (landing): From wingborne mode at 80 knots, the aircraft decelerate to transition mode, following marked boundaries until coming to a hover and landing;
- Hover turn: Execute a 180° turn from a stable hover at an altitude under 20 feet;
- Pirouette: The aircraft moves laterally around a 100-feet radius circle while keeping the nose pointing toward the center and maintaining it at 10 feet.

Criteria for these MTEs include hover time, lateral and longitudinal position, altitude, and heading. Detailed requirements for each MTE are provided in [41]. The virtual scene from some participants when performing MTEs is shown in Figure 4, where participants achieve the mission objectives by aligning green dots and lines within yellow regions. Flight performance was evaluated by the flight trajectory recorded by X-plane. Flight performance was classified as "desired" when the trajectory remained within the yellow markers, "adequate" when within the red markers, and "uncontrollable" if it exceeded the red markers at any point during the flight.

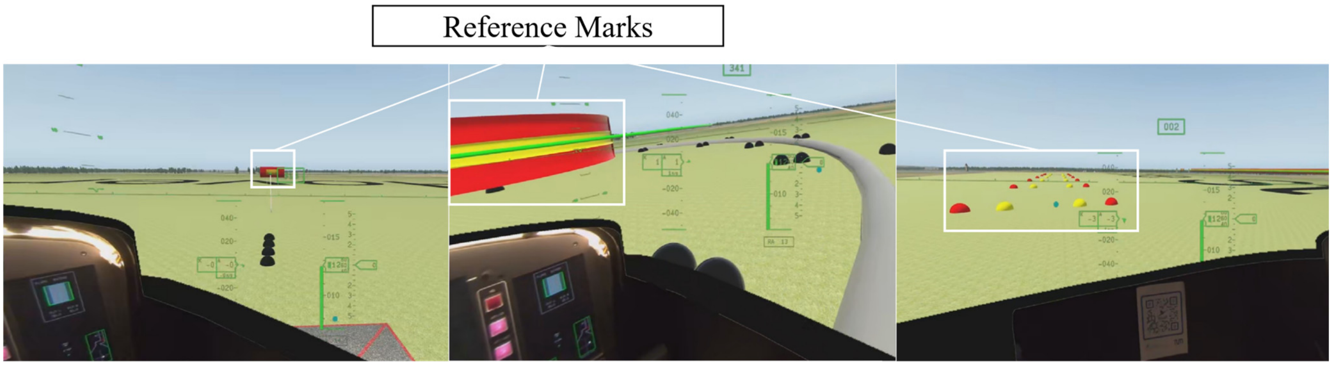

Figure 4. The virtual scene from participants. From left to right, hover turn, pirouette, and sidestep are performing.

3.2.4. Eye Data Collection

Eye-tracking data were collected using the Varjo XR3 headset, equipped with two eye-tracking cameras that capture eye movements with a latency of approximately 20–30 ms. The headset supports an interpupillary distance (IPD) range of 58–72 mm and features a gaze camera resolution of 640 × 400 pixels, with a gaze tracking frequency of 200 Hz [42]. The recorded data were processed using Varjo Base and output in .csv format. The recorded indicators are listed in Appendix A.

In addition, other features associated with workload were computed, as shown in Appendix B. Gaze entropy was calculated using the forward gaze coordinates $(x_f,\ y_f)$ and origin coordinates $(x_o,\ y_o)$ output by the Varjo XR-3 system. Here, $(x_f,\ y_f)$ refers to the eye position coordinates, while $(x_o,\ y_o)$ represents the direction of the gaze vector. Data gaps caused by blinks (defined as missing data for periods exceeding 100 ms) were excluded from the analysis. The visual field was discretized into 2560 partitions based on the resolution of the gaze camera, and Shannon's entropy formula was applied to calculate gaze entropy (G), followed by

$$G = -\sum_{i=1}^{n} p(x_i,\ y_i) \times \log[p(x_i,\ y_i)]$$

where n is the total number of gaze points, $p(x_i,\ y_i)$ is the proportion of the ith gaze point in the (x,y) partition relative to the total gaze points.

Autocorrelation function (ACF) values (ρ_k) were computed to measure the correlation between gaze points at different time intervals, represented by the ratio of self-covariance at a given lag time to the variance of entire time series. Gaze in the same visual field for more than 100 ms was defined as fixations. In this study, fixations were represented by 0 and 1, where 1 indicated that the operator was gazing and the fixations duration was the length of gaze, while 0 the reverse. Saccadic distance (S) was calculated by the distance between two consecutive gaze points:

$$S = \sqrt[2]{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

3.2.5. Subjective Measurements

Subjective ratings for task workload and handling qualities were based on the NASA Raw Task Load Index (NASA-RTLX) and CHR respectively. NASA-RTLX is a simplified version of the NASA Task Load Index (NASA-TLX), which directly averages the scores across six dimensions rather than using comparisons to derive weighted scores [43]. This simplified scale was used to reduce experimental complexity, and results were standardized on a scale of 0 to 10 to represent participants' overall workload perception, including the effort to complete the task itself and compensate for aircraft deficiencies. For handling qualities assessment, CHR was adopted, focusing on controllability, performance, and pilot compensation [14]. The CHR scale used in this study is depicted in Figure 1. Participants were also encouraged to comment on the aircraft qualitatively, including the unsatisfactory designs and suggestions for improvements.

3.2.6. Experimental Procedures

The experimental procedures are illustrated in Figure 5. All participants first reviewed and signed the informed consent form. After confirming the absence of simulator-induced sickness or ophthalmic disorders, they were guided into the laboratory. Participants familiarized themselves with the simulator and practiced the MTEs for approximately one hour. After calibration for eye tracking, participants proceeded with the main experiment, where MTEs were presented in random order. Immediately after each MTE, participants completed the NASA-RTLX and CHR. Considering that participants had already performed the MTE several times in other experiments [35–38], each participant performed all MTEs once, with a two-minute break between tasks. If a task attempt resulted in errors that exceeded the preset safe operational threshold, the experiment was paused, and the participant was required to retry the MTE until it was performed safely. After completing all MTEs, participants were ranked the MTEs based on perceived workload, allowing for the assessment of overall subjective workload. MTEs ranked lowest and highest in workload were assigned scores of 1 and 9, respectively.

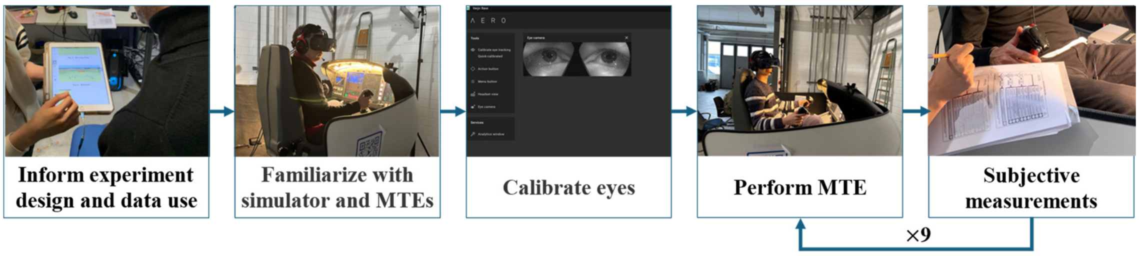

Figure 5. The overall experimental procedures.

3.3. Data Mining

This study applied both statistical analysis methods and deep learning networks to analyze and classify HQ and TW based on eye-tracking data.

3.3.1. Statistical Analysis

To investigate the interaction and mutual influence of handling qualities and task workload, correlation coefficients between subjective task workload and handling qualities scores were calculated. Eye-tracking indicators were tested for normality using the Anderson–Darling test. The Kruskal–Wallis one-way Analysis of Variance (ANOVA) and Dunn's post hoc test was applied to eye-tracking indicators to identify statistically significant differences across various MTEs. In addition, a chi-square test was employed to examine whether eye indicators correlate with handling qualities and task workload levels. Spearman's rank correlation analysis was conducted to explore the monotonic relationship between eye metrics and subjective measurements of handling qualities and task workload. Furthermore, gaze points were visualized to intuitively illustrate variations in eye metrics across different MTEs [44], with heatmaps generated by overlaying data from all participants to account for individual differences.

In addition to these traditional analyses, machine learning methods were employed to identify key indictors related to handling qualities. Given that eye data are temporal and high-dimensional, Random Forest (RF) was adopted to identify key indicators based on their contribution to decision tree splits during the feature selection process.

3.3.2. Deep Learning Networks

Long Short-Term Memory (LSTM) is a neural network architecture designed to process sequential data by capturing long-term dependencies in sequences [45]. It addresses the problem of vanishing or exploding gradients through the use of gating mechanisms. The Forget Gate determines which information from the previous cell state should be discarded, while the Input Gate decides which new information to incorporate from the current input and the hidden state of the previous time step. This is combined with the output of the Forget Gate to update the cell state for the current time step. Finally, the Output Gate selects what information to output from the current cell state as the hidden state for the current time step. The structure of the LSTM cell is depicted in Figure 6.

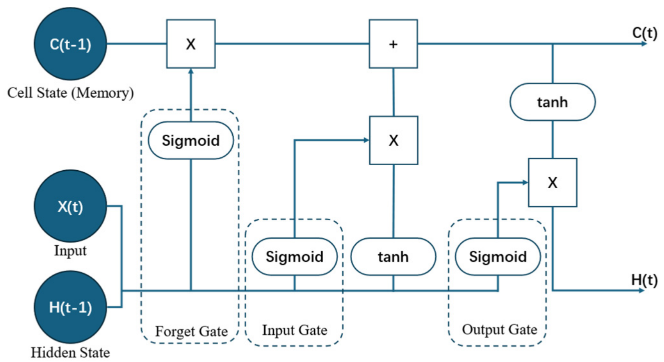

Figure 6. The structure of LSTM cell.

As outlined in Section 3.2.4, a total of 36 eye-tracking features were initially considered. To reduce model complexity and improve training efficiency, feature selection was conducted based on data mining. Features deemed unimportant through both chi-square test and RF importance analysis were removed, resulting in a final set of 30 input features, as shown in Figure 7. The dataset was segmented into 1-second intervals to capture temporal information. With a recording frequency of 200 Hz, the segmentation resulted in 31,820 samples with an input shape of (t, f), where t represents timestamps (200) and f indicates the number of feature (30).

Selected Feature				
Interpupillary distance	Saccadic distance	Focus distance	Stability	Fixations
Gaze forward x	Gaze forward y	Gaze forward z	Gaze entropy	ACF values
Gaze projected to left view x	Gaze projected to left view y	Left forward x	Left forward y	Left forward z
Gaze projected to right view x	Gaze projected to right view y	Right forward x	Right forward y	Right forward z
Left iris/pupil ratio	Left pupil diameter	Left iris diameter	Left pupil size	Left eye openness
Right iris/pupil ratio	Right pupil diameter	Right iris diameter	Right pupil size	Right eye openness

Figure 7. The selected input features.

To identify the most responsive target variables, the proposed model adopted a multi-label classification approach. Each label was predicted by its corresponding output layers, with a specific number of neurons. The multi-label classification targets include the following:

- Subjective Handling Qualities Score (HQS): the rated scores of CHR (9-classes);
- Subjective Handling Qualities Level (HQL): set CHR scores 1–3/4–6/7–9 as level 1/2/3 (3-classes);
- Subjective Task Workload Score (TWS): the standardized NASA-RTLX scores (9-classes);
- Subjective Task Workload Level (TWL): set NASA-RTLX scores 1–3/4–6/7–9 as level 1/2/3 (3-classes);
- Subjective Overall Workload Level (OWL): the ranking results at the end of the experiment (3-classes);
- Pre-defined Handling Qualities Level (PHQL): the pre-defined HQ level according to the task difficulty (3-classes).

The proposed model was built with three LSTM modules, each consisting of 64 units, followed by a fully-connected layer of 128 neurons and 6 distinct output layers, as illustrated in Figure 8. Model parameters were optimized using grid search, with a batch size of 64 and a learning rate of 0.001. Adam optimizer and binary cross-entropy loss function were used. In addition, early stopping was implemented, whereby training was terminated if there was no improvement in accuracy after 100 epochs. The model's performance was evaluated using 5-fold cross-validation, with metrics including accuracy (acc), recall (rec), and precision (pre).

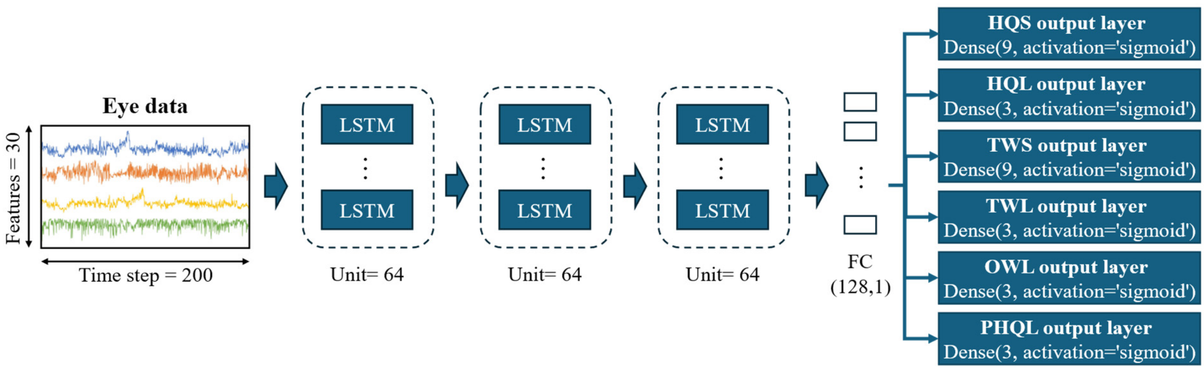

Figure 8. The proposed LSTM model.

4. Results

4.1. Subjective Measurements Results

Detailed subjective measurements are presented in Table 2. The results reveal that participants generally rated *landing*, *pirouette*, and *hover turn* as inducing a higher workload and worse handling qualities. Meanwhile, the ratings of *pirouette* and *slalom* varied among individuals, emphasizing the importance of human factors in assessing handling qualities.

Table 2. The subjective measurements from Cooper–Harper Rating scale and NASA-TLX.

	ID	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	Ave	Std
CHR	Vertical	2	4	1	1	2	1	3	2	1	2	2	2	1	1	1	2	1.75	0.86
	Acc/dec	2	1	1	3	2	3	3	2	1	4	1	7	2	5	3	4	2.75	1.65
	Sidestep	4	5	2	3	2	1	4	2	2	4	1	2	3	4	1	1	2.56	1.31
	Diagonal	2	1	1	3	2	4	1	3	3	3	1	2	4	3	1	7	2.56	1.59
	Slalom	4	6	2	2	1	1	1	4	1	2	1	1	6	8	2	3	2.81	2.20
	Takeoff	3	1	4	3	2	2	1	2	1	3	1	1	7	2	4	4	2.56	1.63
	Landing	5	5	4	6	2	3	1	5	2	8	5	2	8	8	2	2	4.25	2.38
	Hover turn	4	7	4	5	6	3	4	5	4	8	4	1	8	7	5	5	5.00	1.86
	Pirouette	2	4	1	2	5	1	4	5	4	3	4	4	5	6	1	7	3.63	1.82
NASA-TLX	Vertical	1	5	1	1	2	1	4	2	1	2	2	3	2	2	2	2	2.06	1.12
	Acc/dec	2	4	2	1	3	3	1	3	1	3	4	5	1	1	2	2	2.38	1.26
	Sidestep	3	4	2	3	1	1	2	4	2	3	3	1	3	3	4	3	2.63	1.02
	Diagonal	4	3	3	2	2	2	1	5	2	6	3	3	3	6	3	5	3.31	1.49
	Slalom	4	8	3	4	2	2	2	5	1	1	4	2	5	8	3	6	3.75	2.21
	Takeoff	3	5	3	3	1	1	1	4	1	3	3	1	4	4	3	4	2.75	1.34
	Landing	4	8	4	5	2	2	4	6	2	7	6	6	8	7	4	5	5.00	2.00
	Hover turn	3	6	5	5	7	2	6	7	3	2	6	2	7	6	5	7	4.94	1.91
	Pirouette	8	6	5	6	6	2	6	9	1	3	7	9	5	5	6	8	5.75	2.29

To explore the interaction and mutual influence of handling qualities and task workload, the correlation between subjective ratings of handling qualities and task workload was studied by calculating the correlation coefficient and regression analysis, with results shown in Table 3. Data marked in red indicate no correlation, and data marked in bold represent a strong correlation.

Table 3. The correlation and regression analysis results for CHR scores and NASA-TLX scores.

ID	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	Ave
Correlation	−0.03	0.66	0.45	0.44	**0.94**	0.52	0.55	**0.88**	0.60	0.41	**0.82**	0.58	**0.90**	0.68	0.06	0.59	0.60
Regression	R-squared = 0.361, F-statistic = 80.22, p-value = 1.70^{-15}																

4.2. Eye Measurements

4.2.1. Statistical Differences

The Kruskal–Wallis ANOVA and Dunn's post hoc test were performed to assess the statistical differences in eye-tracking features among various MTEs. The ANOVA results showed p-values below 0.05 for all features, indicating significant differences in eye metrics across MTEs. Further intergroup variability was examined through Dunn's test. It was found that fixations and saccadic distance did not differ significantly between hover turn and pirouette; left and right projected x, gaze projected to left and right view x showed no statistical difference in diagonal and sidestep; and left eye openness did not vary statistically between acc/dec and takeoff.

4.2.2. Associations with Handling Qualities

Table 4 shows the results of chi-square test, which determines if eye features have significant associations with handling qualities and task workload. Features with p-values less than 0.05 rejected the null hypothesis and their associations with handling qualities and task workload were confirmed.

Table 4. The results of the chi-square test between eye metrics and the target variables.

Feature	HQ		TW		Feature	HQ		TW	
	Score	p Value	Score	p Value		Score	p Value	Score	p Value
Focus distance	2.94×10^{4}	0.00	1.63×10^{4}	0.00	Stability	1.85×10^{4}	0.00	1.99×10^{4}	0.00
Gaze forward x	1.53×10^{2}	1.04×10^{-29}	4.95×10^{2}	7.52×10^{-33}	Gaze forward y	1.14×10^{3}	4.25×10^{-242}	1.70×10^{3}	8.34×10^{-285}
Gaze projected to left view y	5.19×10^{2}	7.53×10^{-108}	7.65×10^{2}	5.79×10^{-160}	Gaze projected to left view x	5.45×10^{1}	1.90×10^{-11}	2.14×10^{2}	7.86×10^{-42}
Gaze projected to right view y	5.19×10^{2}	7.53×10^{-108}	7.65×10^{2}	5.79×10^{-160}	Gaze projected to right view x	6.79×10^{1}	3.96×10^{-12}	2.10×10^{2}	5.88×10^{-41}
Gaze forward z	5.92×10^{2}	1.20×10^{-123}	4.16×10^{2}	8.37×10^{-85}	Left forward x	2.34×10^{2}	5.69×10^{-47}	4.34×10^{2}	9.65×10^{-89}
Left forward y	1.48×10^{3}	0.00	1.48×10^{3}	0.00	Left forward z	1.54×10^{2}	7.38×10^{-30}	6.63×10^{1}	2.69×10^{-11}
Left origin x	5.20×10^{-1}	9.99×10^{-1}	3.99×10^{-2}	1.00	Left pupil size	1.10×10^{3}	8.18×10^{-234}	2.65×10^{3}	4.77×10^{-280}
Left projected x	8.85×10^{-1}	9.96×10^{-1}	1.98	9.82×10^{-1}	Left projected y	1.86×10^{1}	9.56×10^{-3}	1.52×10^{1}	5.58×10^{-2}
Right forward x	2.16×10^{2}	4.33×10^{-43}	6.10×10^{2}	1.40×10^{-126}	Right forward y	9.21×10^{2}	1.35×10^{-194}	8.37×10^{2}	2.12×10^{-175}
Right forward z	1.55×10^{2}	4.37×10^{-30}	6.53×10^{1}	4.13×10^{-11}	Right origin x	3.26×10^{-2}	1.00	5.62×10^{-1}	1.00
Right pupil size	1.25×10^{3}	1.15×10^{-266}	2.95×10^{3}	0.00	Right projected x	3.92	8.57×10^{-1}	4.27	8.32×10^{-1}
Right projected y	6.89×10^{-1}	9.98×10^{-1}	4.13×10^{-1}	9.89×10^{-1}	Interpupillary distance	2.84×10^{4}	0.00	7.06×10^{4}	0.00
Left iris diameter	3.11×10^{4}	0.00	1.52×10^{4}	0.00	Left pupil diameter	9.17×10^{2}	1.08×10^{-193}	2.49×10^{3}	0.00
Left iris/pupil ratio	4.04×10^{2}	4.09×10^{-83}	1.86×10^{3}	0.00	Left eye openness	1.75×10^{3}	4.20×10^{-266}	2.47×10^{3}	0.00
Right iris diameter	6.33×10^{4}	0.00	6.87×10^{4}	0.00	Right pupil diameter	9.26×10^{2}	1.18×10^{-195}	2.67×10^{3}	0.00
Right iris/pupil ratio	4.64×10^{2}	4.93×10^{-96}	1.10×10^{3}	5.06×10^{-232}	Right eye openness	1.67×10^{3}	7.30×10^{-263}	5.81×10^{3}	3.52×10^{-120}
Gaze entropy	3.95×10^{3}	0.00	5.63×10^{3}	0.00	ACF values	3.08×10^{4}	0.00	9.41×10^{4}	0.00
Fixations	6.94×10^{3}	0.00	2.21×10^{4}	0.00	Saccadic distance	3.78×10^{2}	1.17×10^{-77}	5.46×10^{2}	1.08×10^{-112}

To further explore eye metrics' associations with handling qualities, Spearman's rank correlation coefficients were calculated, as shown in Figure 9, to reveal their monotonic relationship. Additionally, Figure 10 presents the results of RF feature importance analysis, identifying key eye metrics related to handling qualities and task workload.

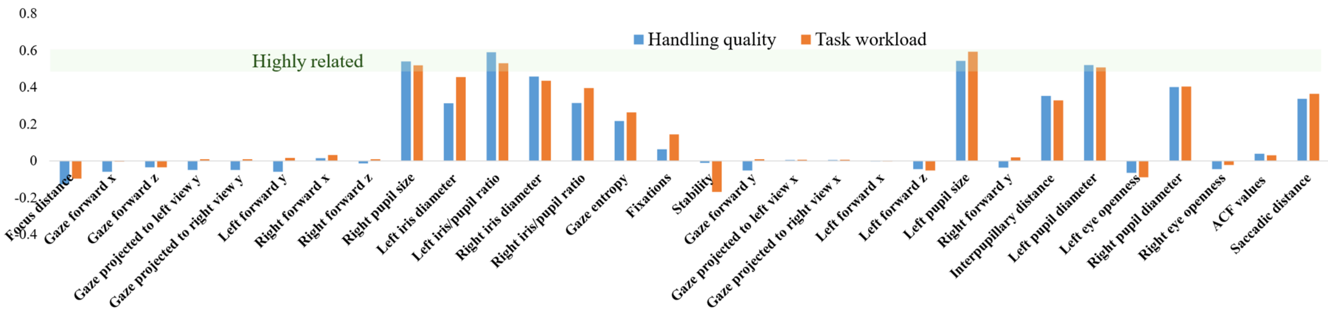

Figure 9. The Spearman's rank correlation coefficients. Coefficients reaching the green region represent highly related features.

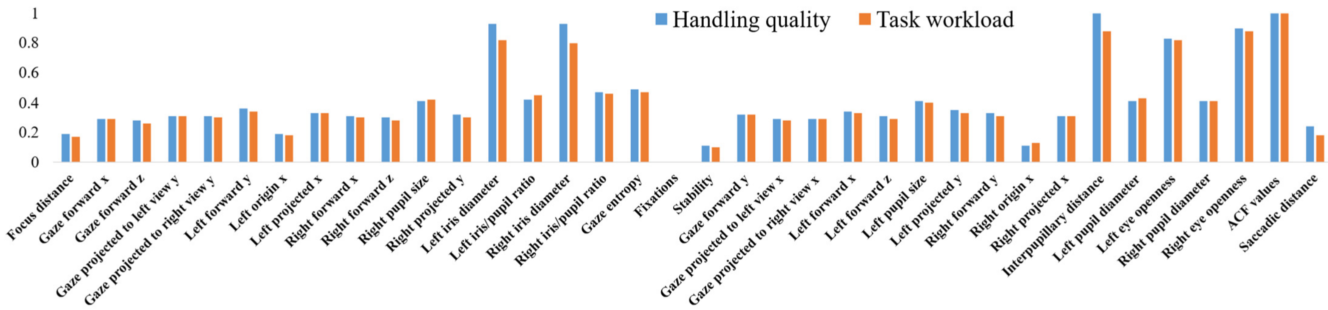

Figure 10. The results of Random Forest importance analysis.

4.2.3. Gaze Heatmaps Visualization

Areas of interest (AOIs) are defined as selected regions containing the key information. However, due to the variability in MTE criteria, such as different reference marks, it is hard to pre-define AOIs. Thus, gaze heatmaps were generated by overlapping participants' gaze direction data, as depicted in Figure 11.

4.3. Deep Networks

The training results for the proposed model are shown in Table 5. To evaluate the effectiveness of the proposed LSTM model, an ablation study was conducted, comparing its performance with other state-of-the-art models. The structures and parameters of these comparison models remained consistent with their original configuration, but the input layer and output layer were adjusted to fit the specific characteristics of this dataset.

Table 5. Comparison of the proposed LSTM with other state-of-the-art models.

Model	Metrics	HQS	HQL	TWS	TWL	OTW	PHQL
Proposed LSTM	Accuracy	0.94	0.97	0.93	0.98	0.94	0.90
	Recall	0.94	0.97	0.93	0.97	0.93	0.90
	Precision	0.94	0.97	0.93	0.97	0.94	0.89
Convolutional neural network (CNN) [46]	Accuracy	0.89	0.93	0.89	0.94	0.91	0.90
	Recall	0.89	0.92	0.89	0.93	0.91	0.90
	Precision	0.90	0.92	0.89	0.94	0.91	0.89
Multi-Layer Perceptron (MLP) [47]	Accuracy	0.83	0.89	0.81	0.91	0.86	0.87
	Recall	0.84	0.88	0.81	0.91	0.86	0.86
	Precision	0.83	0.87	0.83	0.90	0.85	0.85

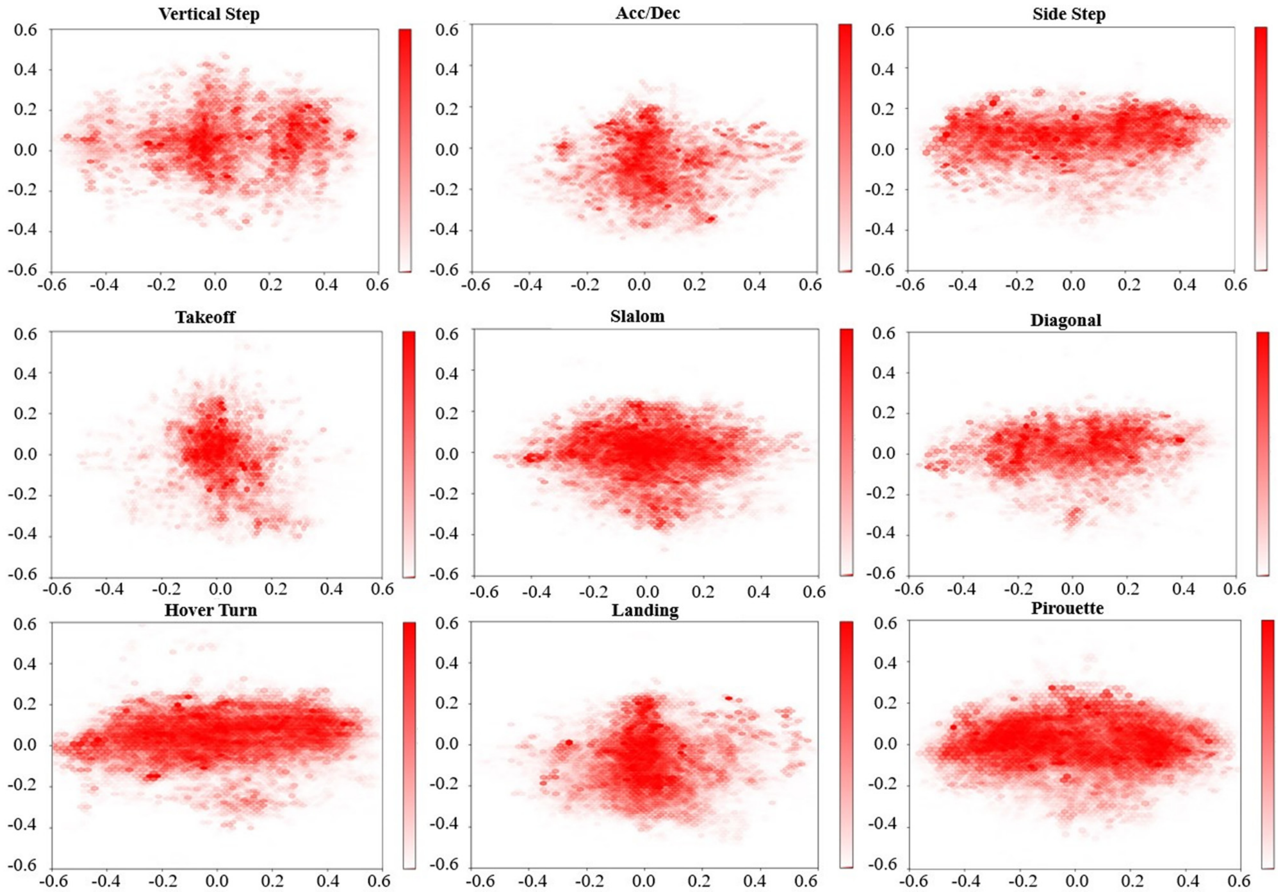

Figure 11. The heatmaps of participants' gaze directions.

5. Discussions

5.1. The Associations Between Handling Qualities and Task Workload

The variability in pilot experience, such as the difference between fixed-wing and helicopter pilots, might have an impact on their subjective ratings. Overall, a similar trend was observed between subjective ratings of handling qualities and task workload, where MTEs rated higher in handling qualities also tended to have higher task workload. Although correlation coefficients varied across participants, the overall correlation coefficient (0.60) suggests a moderate positive correlation between these ratings across the samples. Regression analysis yielded an F-statistic of 80.22 with a significant p-value (1.70^{-15}), indicating a model fit. However, a more complex model is required to see how pilot compensation for aircraft deficiencies contributes to workload.

5.2. Data Mining of Eye Measurements

The Kruskal–Wallis ANOVA and Dunn's tests on eye metrics revealed significant differences ($p < 0.05$) in most features across MTEs, while some features, such as fixations and saccadic distances, showed no significant differences between certain MTEs. The low variability in these features between certain MTEs may stem from similarities in reference marks, or from lesser relevance of these features to the gaze points and eye movement paths [48].

The results of Spearman correlation and RF analysis (Figures 9 and 10) suggest that pupil size, pupil diameter, iris diameter, interpupillary distance, iris/pupil ratio, and gaze entropy exhibit an upward trend with increasing HQ and TW ratings, while other eye features did not exhibit a clear monotonic relationship.

5.3. Gaze Heatmaps

Prior research has associated wider and darker heatmaps with higher workload, stress, and pressure [49]. In this study, wider and darker heatmaps are observed in pirouette and hover turn, which were assigned high scores. High-task-workload MTEs involve stricter criteria, requiring participants to process more information and control the aircraft across more dimensions. This leads to a broader visual scan pattern. Darker areas in the heatmaps indicate longer gaze durations, reflecting the additional time required to process information and navigate complex scenarios during high-task-workload MTEs.

However, flight experiences and operational preferences could lead to varied patterns. For instance, helicopter pilots favored stick maneuvers, while fixed-wing pilot typically used stirrup steering. It was also found that gaze heatmaps under takeoff and landing were narrower, although they had high scores. Prolonged fixation may also reflect participants intentionally focusing on critical elements. During MTEs with detailed visual references, such as the pirouette and vertical step, participants tended to adopt a specific visual pattern. Conversely, in MTEs with fewer visual cues, such as during takeoff and landing, where participants knew the path but lacked clear cues for acceleration or deceleration, their gaze might focus on areas lacking cues. This phenomenon aligns with the "attention tunneling", where an individual's focus narrows in response to high-task demands [50]. This phenomenon suggests that the HMI involved in takeoff and landing needs improvement, as operators faced challenges in assessing information. Furthermore, investigating eye metrics that contradict subjective ratings can provide detailed insights into aircraft design, offering comprehensive evaluations.

5.4. The Proposed LSTM Model

The accuracy of the proposed LSTM model reached an accuracy of 97% in classifying HQL and TWL, indicating that DL techniques are promising in learning eye features. It outperformed other state-of-the-art models due to its explicit memory mechanisms, which capture historical information in time-series data.

Notably, the accuracy of the three-class classification was consistently higher than that of the nine-class classification, demonstrating that the LSTM model performs better with fewer classes. Moreover, in three-class classifications, the accuracies for subjective handling qualities and task workload levels were consistently higher than those for overall task workload levels and pre-defined handling qualities levels, suggesting that eye metrics are more closely associated with individuals' momentary perceptions of task workload and handling qualities, rather than objective task difficulty or overall assessments made afterward.

5.5. Comparison with Existing Handling Qualties Assessment Methods

Integrating eye metrics into the handling qualities assessment process offers a more objective approach. Although our proposed approach still involves subjective ratings, the physiological feedback provides more stability than pilot-provided handling qualities ratings. Additionally, priori expert knowledge on eye metrics helps to reduce the discrepancies caused by participants' varying assessment abilities. For example, the subjective ratings for *slalom* exhibited a significant standard deviation, yet the gaze heatmap indicated a consistent focus. This observation may suggest that subjective ratings capture different facets of task engagement compared to objective eye metrics. Rather than indicating improper ratings, this variability may reflect diverse personal perceptions of task difficulty and engagement, underlining the complementary insights provided by combining subjective and objective measures.

Furthermore, a detailed investigation of eye metrics reveals aircraft design deficiencies, as discussed in Section 5.3. Eye movement data vary among individuals, compensating for the limitations of traditional flight modeling analysis, which often neglects human factors. Since eVTOL aircraft designs and control modes are not yet standardized, the proposed method provides insights into how different designs impact pilot maneuverability, offering valuable data for future flight testing. However, it should be noted that while the proposed method serves as a quantification tool for pilot compensation in CHR, it cannot replace HQRM independently.

6. Case Study

Participant 07 was randomly selected to demonstrate the proposed handling qualities assessment framework. The assessment process is illustrated in Figure 12.

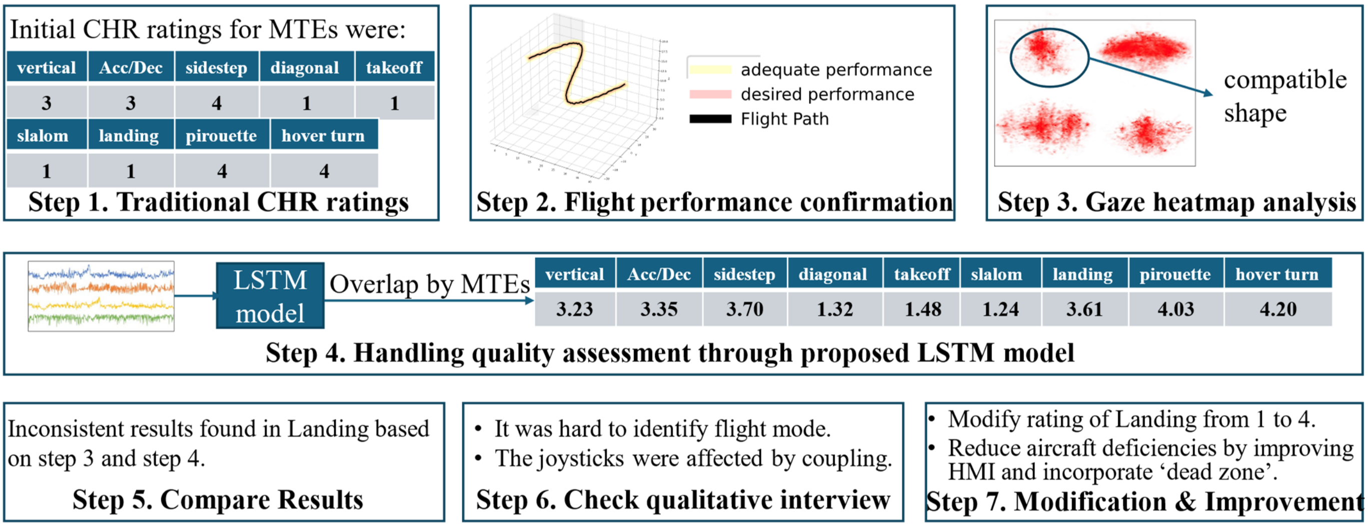

Figure 12. The handling quality assessment process based on the proposed framework.

1. The operators perform several pre-designed MTEs and r provides handling qualities ratings based on CHR, as described in the current HQRM.
2. The recorded flight trajectory is analyzed to assess flight performance, corresponding to the adequate performance and desired performance criteria described in the CHR scale.
3. Gaze heatmaps are examined in terms of their depth and width to verify if they align with the operator's CHR ratings. Wider and darker heatmaps typically indicate a higher workload, which should correspond to lower CHR ratings.
4. Eye-tracking data are processed through the proposed LSTM model. The model outputs for each interval are averaged to provide an overall output for each MTE.
5. Any inconsistencies between the gaze heatmap analysis, model outputs, and the initial CHR ratings are identified and compared. This step highlights any discrepancies that may require further investigation.
6. Qualitative interview feedback is reviewed to analyze inconsistencies between the gaze data, model outputs, and the initial subjective CHR ratings. These interviews provide insight into the reasoning behind the operator's subjective assessments.
7. Based on qualitative feedback and data mining results, the CHR rating for the landing task was adjusted from 1 to 4. This adjustment is accompanied by suggestions for improving the HMI and incorporating a "dead zone" design to mitigate aircraft deficiencies and improve pilot control.

7. Limitations

While one potential limitation for this study is the use of subjective ratings as training labels, the purpose is to predict these ratings from objective physiological signals, thereby reducing the bias and inconsistency of human ratings. Eye metrics are not subject to an individual's subjective will, and therefore, using these signals as inputs improves the objectivity of the assessment.

Additionally, the dataset comprised only 16 participants, limiting the results robustness. Future research will need to validate these findings with a larger sample size. Moreover, the dataset was imbalanced. Despite efforts to address this issue using Random Sampling and the Synthetic Minority Over-sampling Technique (SMOTE), the results were not satisfactory, indicating that more effective data-balancing methods are necessary. Lastly, despite the simulator meeting most of EASA's requirements, there are differences between a simulator and a real aircraft. HQ assessments must ultimately be validated on actual eVTOL to determine if the proposed method is applicable in real-flight conditions.

During the experiment, data from three subjects were excluded due to poor quality, highlighting a key challenge in the reliability of eye-tracking data. Moreover, the consistency observed in gaze heatmaps may reflect both higher workload and task-specific demands, as wider and darker heatmaps suggest prolonged focus but can also indicate intentional attention allocation. Therefore, rather than entirely replacing subjective ratings, eye metrics serve as complementary indicators, enriching the overall understanding of handling qualities.

8. Conclusions

This study investigated the potential for incorporating eye metrics to better identify pilot compensation and handling qualities based on statistical data mining and the DL network. The results demonstrated that handling qualities are positively associated with task workload and can be reflected by key eye metrics, including pupil size, pupil diameter, iris diameter, interpupillary distance, iris/pupil ratio, and gaze entropy. The proposed LSTM model achieves an accuracy of 97%, indicating the utility of eye-tracking data to supplement or support subjective CHR ratings with a degree of objectivity.

A handling qualities assessment framework is proposed based on these findings, integrating pilot compensation, flight performance, and qualitative interviews. The framework incorporates eye metrics to supplement existing HQRM, providing deeper insights into pilot compensation and identifying aircraft design deficiencies. Compared to purely pilot-provided HQ ratings, integrating eye metrics provides more stability, advancing simplified and data-driven approaches for CHR applications. By combining physiological measures and subjective input, we aim to achieve a more comprehensive understanding of task engagement and compensation dynamics. A case study demonstrated the framework's effectiveness in offering a more comprehensive assessment of handling qualities.

The authors also recognize the current limitations in the dataset and eye-tracking technology and intend to further investigate the intricate relationships between eye metrics, task workload, pilot compensation, and handling qualities in future research.

Author Contributions: Conceptualization, Y.L., S.Z. and Y.W.; methodology, Y.L.; software, Y.L., S.K. and M.Z.; validation, Y.L.; formal analysis, Y.L.; investigation, Y.L.; resources, S.Z. and F.H.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, S.Z. and M.Z.; visualization, Y.L.; supervision, S.Z. and F.H.; funding acquisition, S.Z. and F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (Grant No. 52472353); and China Scholarship Council (Nr: 202206020008).

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Acknowledgments: The authors would like to thank our beloved Sharina Kimura for her contributions to this publication and the research team, who was always willing to help others and enrich our lives with her positive energy. Sharina passed away in a tragic accident. As this is the last publication she contributed to, we want to dedicate this article to her—we will keep her alive in our minds and memories.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Indicator	Description
Focus Distance	the distance between the eye and the focus point
Stability	the stability of the user's focus for both eyes
Pupil Size	the size of the pupil for both eyes
Inter-Pupillary Distance In MM	an estimate of the user's inter-pupillary distance measured in millimeters
Pupil Iris Diameter Ratio	ratio of the user's pupil diameter estimate to an estimated iris diameter
Pupil Diameter In MM	an estimated diameter of the tracked pupils in millimeters for both eyes
Iris Diameter In MM	an estimated diameter of the tracked irises in millimeters for both eyes
Eye Openness	estimated openness ratios of the eyes for both eyes
Forward (x, y, z) Origin (x, y, z)	eye position coordinates origin (x, y, z) and the direction of the vector forward (x, y, z) for both eyes
left/right projected (x, y)	Left/right eye gaze vector projected on the video showing the left/right eye image
Gaze projected to left/right view (x, y)	Combined (left+right) eye gaze vector projected on the video showing the left/right eye image

Figure A1. The indicators output by Varjo XR 3 [42].

Appendix B

Indicator	Description
Gaze entropy	the measure of the randomness of the distribution of gaze points across the visual field, indicating the diversity of visual attention
ACF values	the values that represent the correlation of an eye movement pattern with itself at different time lags, reflecting the pattern's self-similarity over time.
Fixations	the periods during which the eyes are relatively stationary
Fixations duration	the amount of time that the eyes remain stationary in a single fixation
Saccadic distance	the spatial extent covered by the eyes during a saccade, which is the rapid shift from one fixation to another.

Figure A2. The calculated eye indicators [23].

References

1. Zhang, J.; Liu, Y.; Zheng, Y. Overall eVTOL aircraft design for urban air mobility. *Green Energy Intell. Transp.* **2024**, *3*, 100–150. [CrossRef]
2. *Means of Compliance with the Special Condition VTOL*; European Union Aviation Safety Authority: Cologne, Germany, 2021.
3. Harper, R.P.; Cooper, G.E. Handling qualities and pilot evaluation. *J. Guid. Control Dyn.* **1986**, *9*, 500–505. [CrossRef]
4. Klyde, D.H.; Lampton, A.K.; Mitchell, D.G.; Berka, C.; Rhinehart, M. A new approach to aircraft handling qualities prediction. In Proceedings of the AIAA SciTech 2021 Forum, Virtual Event, 11–15 & 19–21 January 2021.
5. Bailey, R.E.; Jackson, E.B.; Bilimoria, K.D.; Mueller, E.R.; Frost, C.R.; Alderete, T.S. *Cooper-Harper Experience Report for Spacecraft Handling Qualities Applications*; NASA Center for AeroSpace Information: Hanover, MD, USA, 2009.
6. Harris, D. Measurement of pilot opinion when assessing aircraft handling qualities. *Meas. Control* **2000**, *33*, 239–243. [CrossRef]
7. Ji, H.L.; Chen, R.L.; Li, P. Distributed Atmospheric Turbulence Model for Helicopter Flight Simulation and Handling-Quality Analysis. *J. Aircr.* **2017**, *54*, 190–198. [CrossRef]
8. Dussart, G.; Lone, M.; Bailey, R. Development of a Multi-Directional Manoeuvre for Unified Handling Qualities Investigation. *Aerospace* **2019**, *6*, 70. [CrossRef]
9. Brieger, O.; Kerr, M.; Leissling, D.; Postlethwaite, I.; Sofrony, J.; Turner, M.C. Flight testing of a rate saturation compensation scheme on the ATTAS aircraft. *Aerosp. Sci. Technol.* **2009**, *13*, 92–104. [CrossRef]
10. Kong, X.W. Evaluation of Flight Test Data Quality Based on Rough Set Theory. In Proceedings of the 13th International CISP-BMEI, Chengdu, China, 17–19 October 2020.
11. Zhang, X.; Yao, S.; Zhu, H. Handling Quality Evaluation Method in Aircraft Cockpit Based on Pilot Performance. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Orlando, FL, USA, 21–25 July 2018.
12. Tan, W.Q.; Wu, Y.; Qu, X.J.; Efremov, A.V. A Method for Predicting Aircraft Flying Qualities Using Neural Networks Pilot Model. In Proceedings of the 2ND ICSAI, Shanghai, China, 15–17 November 2014.
13. Herrington, S.M.; Zahed, M.J.H.; Fields, T.D. Handling Qualities Assessment and Performance Evaluation for Unmanned Aerial Systems and Pilots. *Unmanned Syst.* **2024**, *12*, 99–131. [CrossRef]
14. Cooper, G.E.; Harper, R.P. *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*; National Aeronautics and Space Administration: Washington, DC, USA, 1969.
15. Wilson, D.; Riley, D. Cooper-harper pilot rating variability. In Proceedings of the 16th Atmospheric Flight Mechanics Conference, Boston, MA, USA, 14–16 August 1989.

16. Memon, W.A.; White, M.D.; Padfield, G.D.; Cameron, N.; Lu, L. Helicopter Handling Qualities: A study in pilot control compensation. *Aeronaut. J.* **2023**. [CrossRef]
17. Bachelder, E.; Berger, T.; Aponso, B.L. Predicting Pilot Performance, Workload and Handling Qualities using SCOPE. In Proceedings of the Vertical Flight Society 2023 Rotorcraft Handling Qualities Technical Meeting, Huntsville, AL, USA, 7–8 February 2023.
18. Li, Y.H.; Zhang, S.G.; Wu, Y.B. Assessing Handling Qualities of an SVO-based eVTOL through EMG and Eye Data. *Acta Aeronaut. Et. Astronaut. Sin.* **2024**, 46. [CrossRef]
19. Cunningham, K.; Cox, D.; Murri, D.; Riddick, S. A piloted evaluation of damage accommodating flight control using a remotely piloted vehicle. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Portland, Oregon, 8–11 August 2011.
20. Li, Y.H.; Zhang, S.G.; He, R.C.; Holzapfel, F. Objective Detection of Trust in Automated Urban Air Mobility: A Deep Learning-Based ERP Analysis. *Aerospace* **2024**, *11*, 174. [CrossRef]
21. Charles, R.L.; Nixon, J. Measuring mental workload using physiological measures: A systematic review. *Appl. Ergon.* **2019**, *74*, 221–232. [CrossRef]
22. Mahanama, B.; Jayawardana, Y.; Rengarajan, S.; Jayawardena, G.; Chukoskie, L.; Snider, J.; Jayarathna, S. Eye Movement and Pupil Measures: A Review. *Front. Comput. Sci.* **2022**, *3*, e733531. [CrossRef]
23. Peissl, S.; Wickens, C.D.; Baruah, R. Eye-Tracking Measures in Aviation: A Selective Literature Review. *Int. J. Aerosp. Psychol.* **2018**, *28*, 98–112. [CrossRef]
24. Skvarekova, I.; Skultety, F. Objective Measurement of Pilot's Attention Using Eye Track Technology during IFR Flights. In Proceedings of the TRANSCOM, Novy Smokovec, Slovakia, 29–31 May 2019.
25. Greiwe, D.H.; Friedrich, M. Gaze Movements of Helicopter Pilots during Real and Simulated Take-Off and Landing Maneuvers. *Aerospace* **2024**, *11*, 429. [CrossRef]
26. Lu, T.J.; Lou, Z.S.; Shao, F.; Li, Y.; You, X.Q. Attention and Entropy in Simulated Flight with Varying Cognitive Loads. *Aerosp. Med. Hum. Perform.* **2020**, *91*, 489–495. [CrossRef] [PubMed]
27. Harris, D.J.; Arthur, T.; de Burgh, T.; Duxbury, M.; Lockett-Kirk, R.; McBarnett, W.; Vine, S.J. Assessing Expertise Using Eye Tracking in a Virtual Reality Flight Simulation. *Int. J. Aerosp. Psychol.* **2023**, *33*, 153–173. [CrossRef]
28. Niu, Y.F.; Zhou, T.Y.; Bai, L. Research on color coding of fighter jet head-up display key information elements in air-sea flight environment based on eye-tracking technology. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **2022**, *236*, 2010–2030. [CrossRef]
29. Lyu, M.T.; Li, F.; Xu, G.Y.; Su, H. Leveraging eye-tracking technologies to promote aviation safety—A review of key aspects, challenges, and future perspectives. *Saf. Sci.* **2023**, *168*, 106295.
30. Di Stasi, L.L.; McCamy, M.B.; Martinez-Conde, S.; Gayles, E.; Hoare, C.; Foster, M.; Macknik, S.L. Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiol. Behav.* **2016**, *153*, 91–96. [CrossRef]
31. Pathirana, P.; Senarath, S.; Meedeniya, D.; Jayarathna, S. Eye gaze estimation: A survey on deep learning-based approaches. *Expert Syst. Appl.* **2022**, *199*, e116894. [CrossRef]
32. Ferreira, F.; Ferreira, S.; Mateus, C.; Barbosa-Rocha, N.; Coelho, L.; Rodrigues, M.A. Advancing the understanding of pupil size variation in occupational safety and health: A systematic review and evaluation of open-source methodologies. *Saf. Sci.* **2024**, *175*, 106490. [CrossRef]
33. Franceschiello, B.; Di Noto, T.; Bourgeois, A.; Murray, M.M.; Minier, P.; Pouget, J.; Richiardi, P.; Bartolomeo, F.; Anselmi, F. Machine learning algorithms on eye tracking trajectories to classify patients with spatial neglect. *Comput. Methods Programs Biomed.* **2022**, *221*, 106929. [CrossRef] [PubMed]
34. Li, F.; Chen, C.H.; Liu, Y.S.; Chang, D.N.; Cui, J.; Sourina, O. Autoencoder-enabled eye-tracking data analytics for objective assessment of user preference in humanoid robot appearance design. *Expert Syst. Appl.* **2024**, *249*, 123857. [CrossRef]
35. Zintl, M.; Marb, M.M.; Wechner, M.A.; Seiferth, D.; Holzapfel, F. Development of a virtual reality simulator for eVTOL flight testing. In Proceedings of the AIAA Aviation 2022 Forum, Chicago, IL, USA, 27 June–1 July 2022.
36. Dollinger, D.; Reiss, P.F.; Angelov, J.; Loebl, D.; Holzapfel, F. Control Inceptor Design for Onboard Piloted Transition VTOL Aircraft Considering Simplified Vehicle Operation. In Proceedings of the AIAA SciTech Forum, Virtual Event, 11–15, 19–21 January 2021.
37. Zintl, M.; Kimura, S.; Holzapfel, F. A Mixed Reality Research Flight Simulator for Advanced Air Mobility Vehicles. In Proceedings of the AIAA Aviation Forum and Ascend 2024, Las Vegas, NV, USA, 30 July–1 August 2024.
38. Liedtke, S.; Zintl, M.; Holzapfel, F.; Klinker, G. A Mixed Reality Setup for Prototyping Holographic Cockpit Instruments. In *Virtual Reality and Mixed Reality*; EuroXR 2023; Lecture Notes in Computer Science; Springer: Cham, Germany, 2023; Volume 14410.
39. Baskett, B.J.; Daniel, D.L.O. *Aeronautical Design Standard Performance Specification Handling Qualities Requirements for Military Rotorcraft*; United States Army Aviation and Missile Command: Huntsville, AL, USA, 2000.
40. European Organisation for Civil Aviation Equipment (EUROCAE). *VTOL Flight Control Handling Qualities Verification*; 11 July 2021; European Organisation for Civil Aviation Equipment (EUROCAE): Saint-Denis, France, 2021.
41. Wechner, M.A.; Marb, M.M.; Zintl, M.; Seiferth, D.; Holzapfel, F. Design, Conduction and Evaluation of Piloted Simulation Mission Task Element Tests for Desired Behavior Validation of an eVTOL Flight Control System. In Proceedings of the AIAA Aviation 2022 Forum, Chicago, IL, USA, 27 June–1 July 2022.
42. Varjo. Varjo XR3. Available online: https://developer.varjo.com/docs/native/eye-tracking (accessed on 1 January 2024).
43. Friesen, D.; Brost, C.; Pavel, M.D.; Masarati, P.; Mulder, M. Human-automation interaction for helicopter flight: Comparing two decision-support systems for navigation tasks. *Aerosp. Sci. Technol.* **2022**, *129*, 107719. [CrossRef]

44. Jiang, H.Y.; Gao, M.D.; Huang, J.Q.; Tang, C.; Zhang, X.Q.; Liu, J. DCAMIL: Eye-tracking guided dual-cross-attention multi-instance learning for refining fundus disease detection. *Expert Syst. Appl.* **2024**, *243*, 122889. [CrossRef]
45. Lazzara, M.; Chevalier, M.; Colombo, M.; Garcia, J.G.; Lapeyre, C.; Teste, O. Surrogate modelling for an aircraft dynamic landing loads simulation using an LSTM AutoEncoder-based dimensionality reduction approach. *Aerosp. Sci. Technol.* **2022**, *126*, 107629. [CrossRef]
46. Yin, Y.; Juan, C.; Chakraborty, J.; McGuire, M.P. Classification of Eye Tracking Data Using a Convolutional Neural Network. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018.
47. Lee, Y.; Shin, C.; Plopski, A.; Ltoh, Y.; Piumsomboon, T.; Dey, A.; Lee, G.; Kim, S.; Billinghurst, M. Estimating Gaze Depth Using Multi-Layer Perceptron. In Proceedings of the 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR), Nara, Japan, 27–29 June 2017.
48. Strenzke, R.; Uhrmann, J.; Benzler, A.; Maiwald, F.; Rauschert, A.; Schulte, A. Managing cockpit crew excess task load in military manned-unmanned teaming missions by dual-mode cognitive automation approaches. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Portland, Oregon, 8–11 August 2011.
49. Das, S.; Maiti, J. Assessment of cognitive workload based on information theory enabled eye metrics. *Saf. Sci.* **2024**, *176*, 106567. [CrossRef]
50. Wickens, C.D.; Alexander, A. Attentional Tunneling and Task Management in Synthetic Vision Displays. *Int. J. Aviat. Psychol.* **2009**, *19*, 182–199. [CrossRef]