Article

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

Yuhan Li , Shuguang Zhang * , Yibing Wu , Sharina Kimura , Michael Zintl , Florian Holzapfel

Posted Date: 27 September 2024

doi: 10.20944/preprints202409.2177.v1

Keywords: Eye metrics; handling qualities; eVTOL; Urban air mobility; aircraft design; LSTM

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; michael.zintl@tum.de; florian.holzapfel@tum.de
[3] School of Transportation Science and Engineering, Beihang University, Beijing 100191, China; gnahz@buaa.edu.cn; 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 Take-Off 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 23 participants as they performed various Mission Task Elements (MTEs) in an eVTOL simulator were recorded. 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

1. Introduction

Urban Air Mobility (UAM) has garnered increasing interest due to urbanization trends, where the electric vertical take-off 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 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 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 HQRM by developing methods that more objectively integrate human factors.

This study investigates the potential of achieving results equivalent or superior to subjective assessments of handling qualities through physiological indicators. 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:

- 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, supplementing the existing HQRM.

2. Background

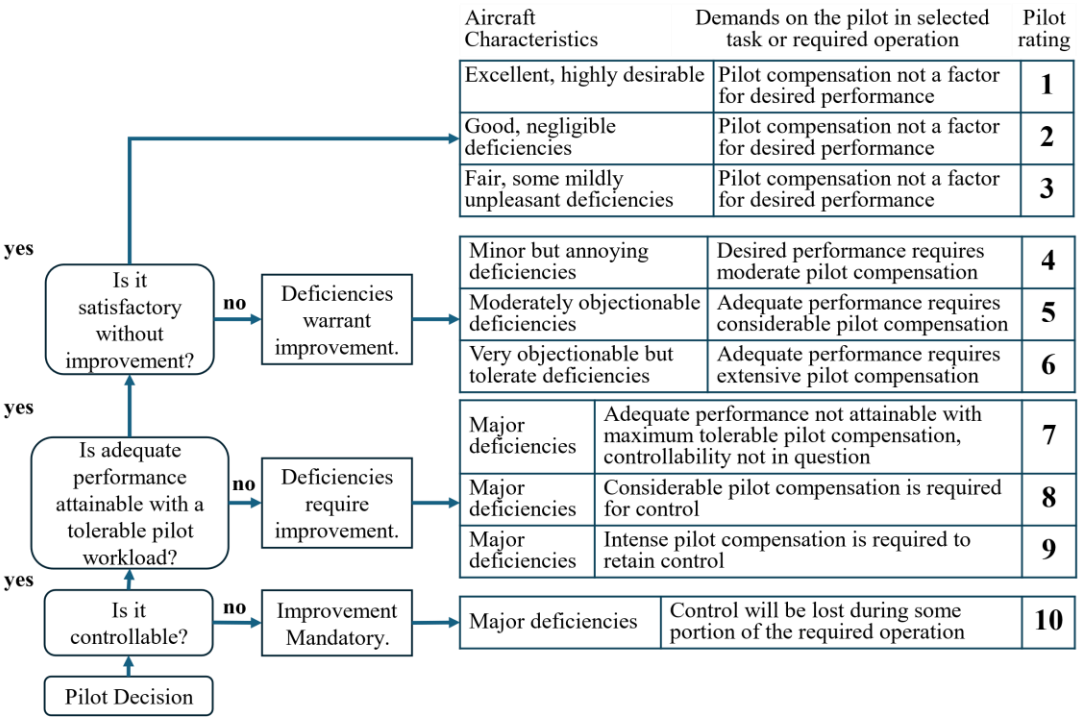
Handling qualities refers to the controllability and maneuverability of an aircraft, allowing pilots to perform specific tasks [3]. It is closely linked to the pilot's perception of how the aircraft handles during various maneuvers. Many studies have investigated methods for evaluating handling qualities,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

Table 1. The 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 & 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.

Pilot compensation, as defined in [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]. One novel study linked physiological measurements to pilot compensation by using electroencephalogram (EEG) and electrocardiogram (ECG) data to assess physical workload and predict handling qualities [4]. However, it involved only five subjects, making it insufficient to establish a robust quantitative correlation between workload and HQ ratings. In addition, EEG data collection is difficult during real flights. Therefore, our study aims to establish a correlation between workload and pilot compensation in a more objective and feasible way, thus enhancing the reliability of handling qualities assessments.


yes
Is it satisfactory without improvement?
no
Deficiencies warrant improvement.
yes
Is adequate performance attainable with a tolerable pilot workload?
no
Deficiencies require improvement.
yes
Is it controllable?
no
Improvement Mandatory.
Pilot Decision

Aircraft Characteristics	Demands on the pilot in selected task or required operation	Pilot rating
Excellent, highly desirable	Pilot compensation not a factor for desired performance	1
Good, negligible deficiencies	Pilot compensation not a factor for desired performance	2
Fair, some mildly unpleasant deficiencies	Pilot compensation not a factor for desired performance	3
Minor but annoying deficiencies	Desired performance requires moderate pilot compensation	4
Moderately objectionable deficiencies	Adequate performance requires considerable pilot compensation	5
Very objectionable but tolerate deficiencies	Adequate performance requires extensive pilot compensation	6
Major deficiencies	Adequate performance not attainable with maximum tolerable pilot compensation, controllability not in question	7
Major deficiencies	Considerable pilot compensation is required for control	8
Major deficiencies	Intense pilot compensation is required to retain control	9
Major deficiencies	Control will be lost during some portion of the required operation	10

Figure 1. The Cooper-Harper Rating Scales [3].

2.1. Physiological Measurements in Aviation

Physiological metrics have been shown to provide valuable insights into an operator's engagement, distraction, and workload across various tasks [17]. In our previous research [18], electroencephalography (EEG) has been utilized to investigate pilots' trust in automated Urban Air Mobility (UAM). However, excessive noise in EEG signals occurred, hence we are exploring other physiological indicators. Among these indicators, eye metrics are promising in assessing handling qualities and task workload as the eyes are primary sensory organs in flight tasks [19]. The application of eye metrics in the aviation industry is well-established, including measurements of fixation, visual search, pupil, saccades, and blink [20]. Numerous studies have utilized eye-tracking data to assess pilot performance and mental states [21]. For instance, [22] employed areas of interest (AOIs) to measure pilots' attention distribution during flight, while [23] added gaze, fixations, and hit maps. These studies indicate that decreased situational awareness often correlates with deteriorating visual search strategies. Pilot’s workload has been investigated in [24] 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 [25]. Beyond performance assessments, eye metrics have also been applied to aircraft design, particularly in enhancing aviation displays [26] and improving safety [27]. 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 [28]. 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 is abundant and complex, requiring considerable effort to analyze [29]. Consequently, it is crucial to develop an approach capable of capturing the deep features of eye-tracking data in relation to HQ and TW ratings and revealing the mechanisms of their interactions. 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 [27]. For example, [30] 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. [31] 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.

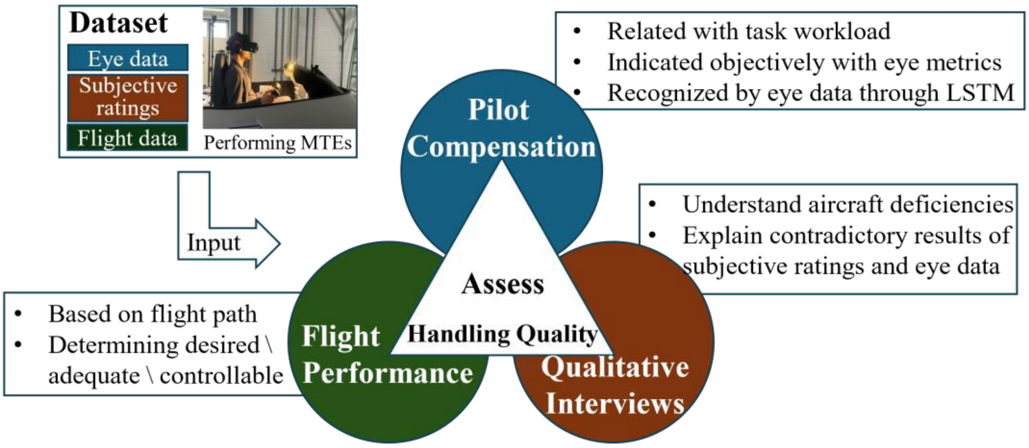

Dataset
Eye data
Subjective ratings
Flight data
Performing MTEs
Input
Related with task workload
Indicated objectively with eye metrics
Recognized by eye data through LSTM
Pilot Compensation
Understand aircraft deficiencies
Explain contradictory results of subjective ratings and eye data
Assess
Handling Quality
Based on flight path
Determining desired \ adequate \ controllable
Flight Performance
Qualitative Interviews

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 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 [32], as shown in Figure 3. The eVTOL was designed based on 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 [33].

The cockpit replica and motion system complied with the EASA standards, providing a professional and realistic environment for research [34]. The development process of the simulator is available in [32–35].

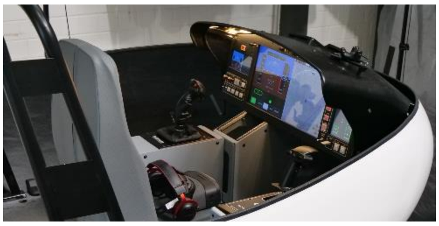


Figure 3. The motion-based mixed reality simulator [32].

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 [36]. Currently, there are no generally recognized MTEs for eVTOL. *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 seconds, 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 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 [37]. The virtual scene from some participants when performing MTEs are 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.

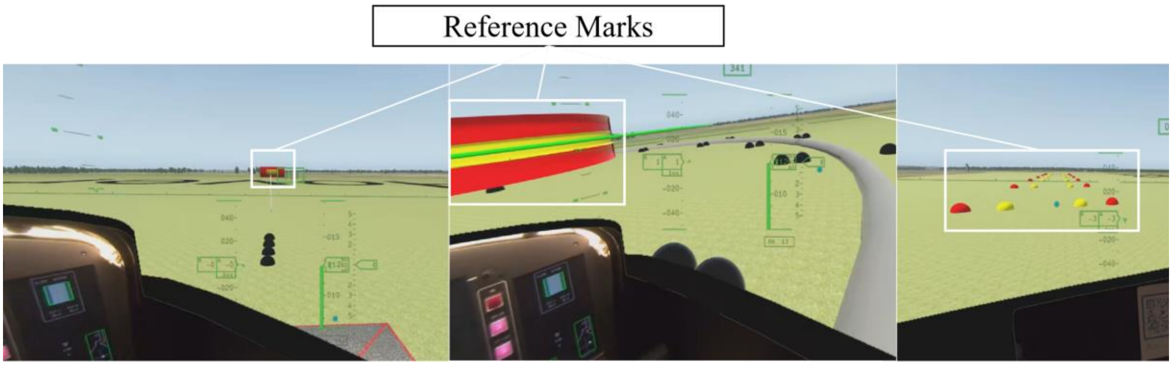


Figure 4. The virtual scene from participants. From left to right, hover turn, pirouette and side step are performing.

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 58-72mm and features the gaze camera resolution of 640 x 400 pixels, with a gaze tracking frequency of 200 Hz [38]. The recorded data were processed using Varjo Base and output in .csv format. The recorded indicators are listed in Appendix 1. In addition, other features associated with workload were computed, as shown in Appendix 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 [39]. 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 the aircraft qualitatively, including the unsatisfactory designs, suggestions for improvements, etc.

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. Each participant performed all MTEs once, with a two-minute break between tasks. If any errors exceeded the preset threshold, the experiment was paused, and the participant was required to retry the MTE until it met the criteria. 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.

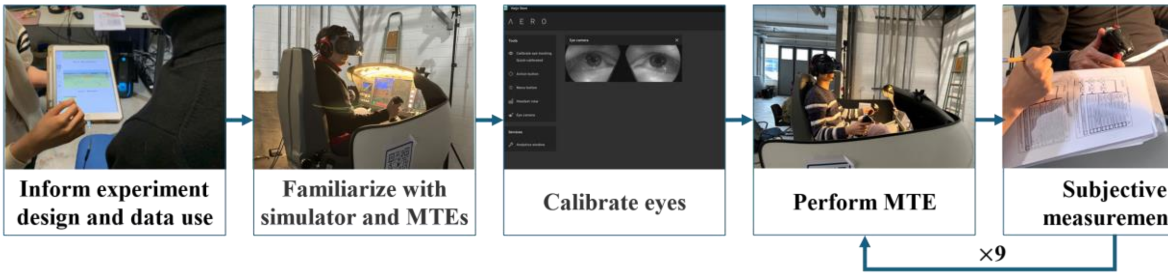

Inform experiment design and data use
Familiarize with simulator and MTEs
Calibrate eyes
Perform MTE
Subjective measuremen
×9

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. Experimental Procedures

To investigate the interaction and mutual influence of handling qualities and task workload, correlation coefficients between task workload and handling qualities scores were calculated. Subsequently, Analysis of Variance (ANOVA) and Honestly Significant Difference (HSD) test [40] was applied on eye-tracking indicators to determine statistically significant differences across various MTEs. Besides, a chi-square test was employed to examine whether eye indicators correlate with and 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 [41]. Gaze point heatmaps were 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 selected to identify key indicators based on their contribution to decision trees splits.

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 [42]. 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.

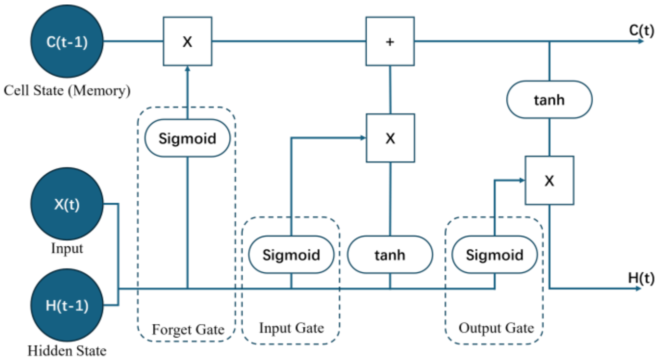
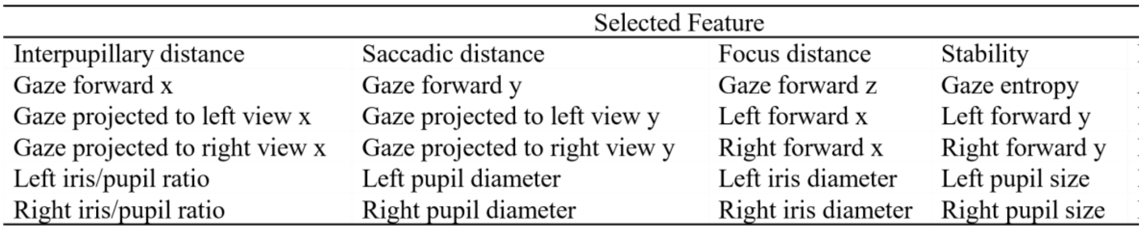


Figure 6. The structure of LSTM cell.

As outlined in 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 200Hz, 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
Gaze forward x	Gaze forward y	Gaze forward z	Gaze entropy
Gaze projected to left view x	Gaze projected to left view y	Left forward x	Left forward y
Gaze projected to right view x	Gaze projected to right view y	Right forward x	Right forward y
Left iris/pupil ratio	Left pupil diameter	Left iris diameter	Left pupil size
Right iris/pupil ratio	Right pupil diameter	Right iris diameter	Right pupil size

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:

- 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 terminating training 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).

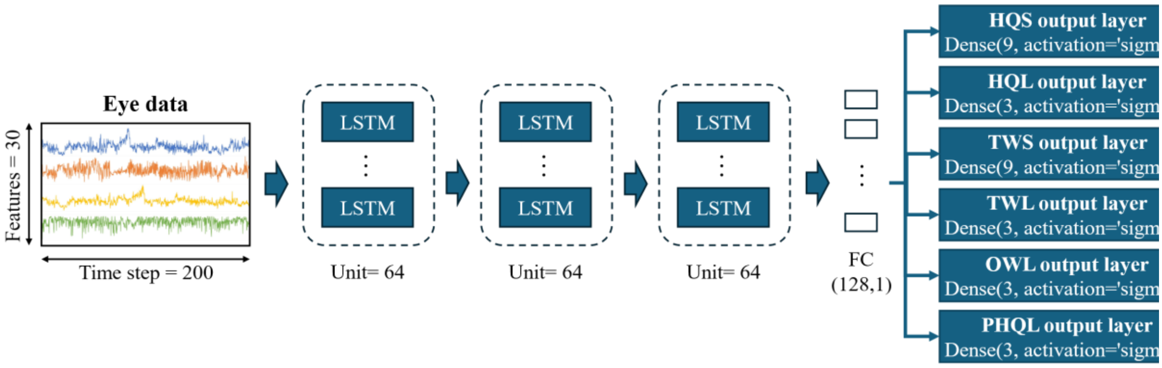


Figure 8. The proposed LSTM Model.

4. Results

4.1. Subjective Measurements

Detailed subjective measurements are presented in Figure 9. The results reveal that participants generally rated *landing* and *hover turn* as inducing higher workload and worse handling qualities. In contrast, the ratings of *pirouette* and *slalom* varied among individuals, emphasizing the importance of human factors in assessing handling qualities.

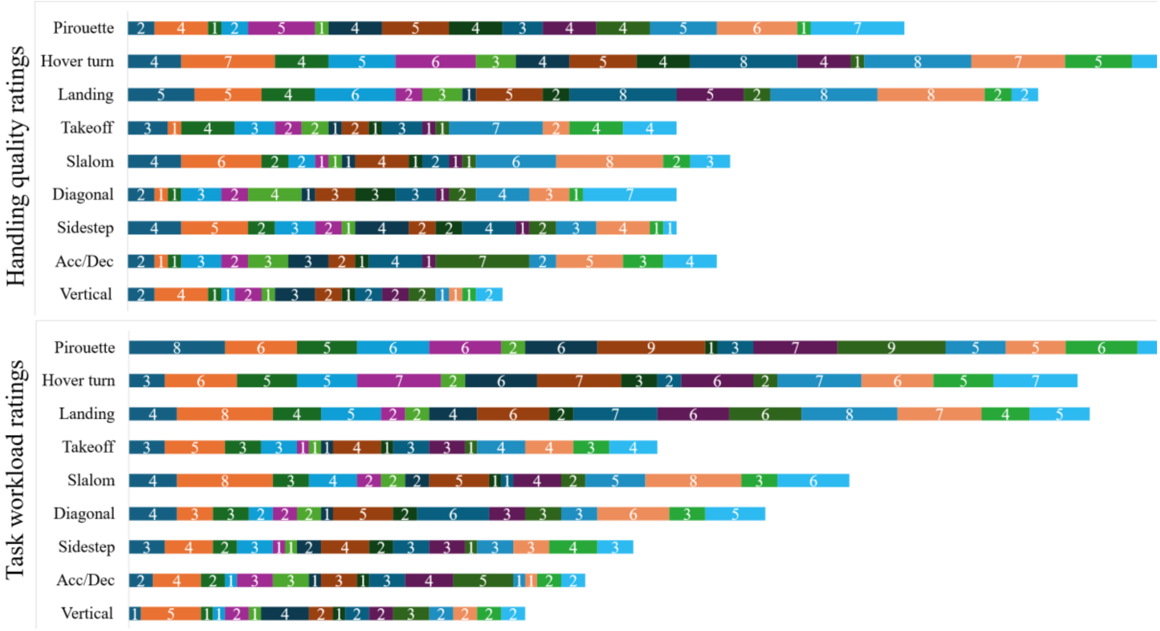


Figure 9. The subjective rating results. The scores assigned by different participants are indicated by the colors in the legend on the right. The stack length represents the overall score for the MTE.

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 2. Data marked in red indicate no correlation, and data marked in bold represent a strong correlation.

Table 2. The correlation and regression analysis results for handling qualities and task workload.

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																

4.2. Eye Measurements

4.2.1. Statistical Differences

ANOVA and HSD tests 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. Detailed intergroup variability was examined through HSD test. It was found that *fixations*, *saccadic distance*, and *focus distance* did not differ significantly between *hover turn* and *pirouette*; while *left* and *right projected x* showed no statistical difference in *diagonal* and *side step*.

4.2.2. Associations with Handling Qualities

Table 3 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 were confirmed their associations with handling qualities and task workload.

Table 3. 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	2.94e4	0.00	1.63e4	0.00	Stability	1.85e4	0.00	1.99e4	0.00
Gaze forward x	1.53e2	1.04e-29	4.95e2	7.52e-33	Gaze forward y	1.14e3	4.25e-242	1.70e3	8.34e-285
Gaze projected to left view y	5.19e2	7.53e-108	7.65e2	5.79e-160	Gaze projected to left view x	5.45e1	1.90e-11	2.14e2	7.86e-42
Gaze projected to right view y	5.19e2	7.53e-108	7.65e2	5.79e-160	Gaze projected to right view x	6.79e1	3.96e-12	2.10e2	5.88e-41
Gaze forward z	5.92e2	1.20e-123	4.16e2	8.37e-85	Left forward x	2.34e2	5.69e-47	4.34e2	9.65e-89
Left forward y	1.48e3	0.00	1.48e3	0.00	Left forward z	1.54e2	7.38e-30	6.63e1	2.69e-11
Left origin x	5.20e-1	9.99e-1	3.99e-2	1.00e0	Left pupil size	1.10e3	8.18e-234	2.65e3	4.77e-280
Left projected x	8.85e-1	9.96e-1	1.98e0	9.82e-1	Left projected y	1.86e1	9.56e-3	1.52e1	5.58e-2
Right forward x	2.16e2	4.33e-43	6.10e2	1.40e-126	Right forward y	9.21e2	1.35e-194	8.37e2	2.12e-175
Right forward z	1.55e2	4.37e-30	6.53e1	4.13e-11	Right origin x	3.26e-2	1e0	5.62e-1	1.00e0
Right pupil size	1.25e3	1.15e-266	2.95e3	0.00	Right projected x	3.92e0	8.57e-1	4.27e0	8.32e-1
Right projected y	6.89e-1	9.98e-1	4.13e-1	9.89e-1	Interpupillary distance	2.84e4	0.00	7.06e4	0.00
Left iris diameter	3.11e4	0.00	1.52e4	0.00	Left pupil diameter	9.17e2	1.08e-193	2.49e3	0.00
Left iris/pupil ratio	4.04e2	4.09e-83	1.86e3	0.00	Left eye openness	1.75e3	4.20e-266	2.47e3	0.00
Right iris diameter	6.33e4	0.00	6.87e4	0.00	Right pupil diameter	9.26e2	1.18e-195	2.67e3	0.00
Right iris/pupil ratio	4.64e2	4.93e-96	1.10e3	5.06e-232	Right eye openness	1.67e3	7.30e-263	5.81e3	3.52e-120
Gaze entropy	3.95e3	0.00	5.63e3	0.00	ACF values	3.08e4	0.00	9.41e4	0.00

Fixations	6.94e3	0.00	2.21e4	0.00	Saccadic distance	3.78e2	1.17e-77	5.46e2	1.08e-112

To further explore eye metrics’ associations with handling qualities, Spearman's rank correlation coefficients were calculated, as shown in Figure 10, to reveal their monotonic relationship. Additionally, Figure 11 presents the results of RF feature importance analysis, identifying key eye metrics related to handling qualities and task workload.

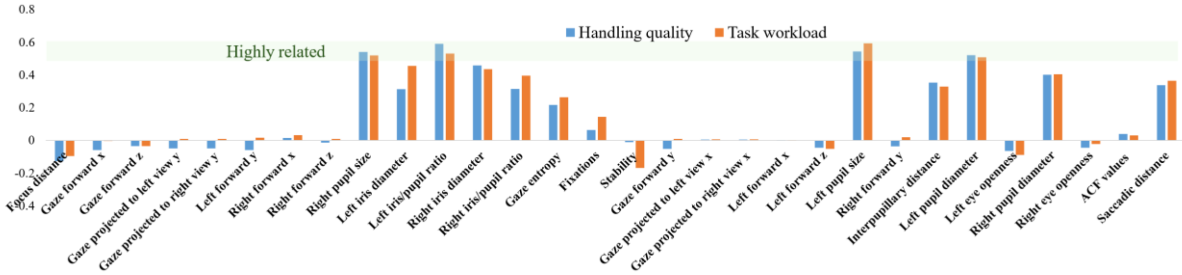


Figure 10. The Spearman’s rank correlation coefficients. Coefficients reaching the green region represent highly related features.

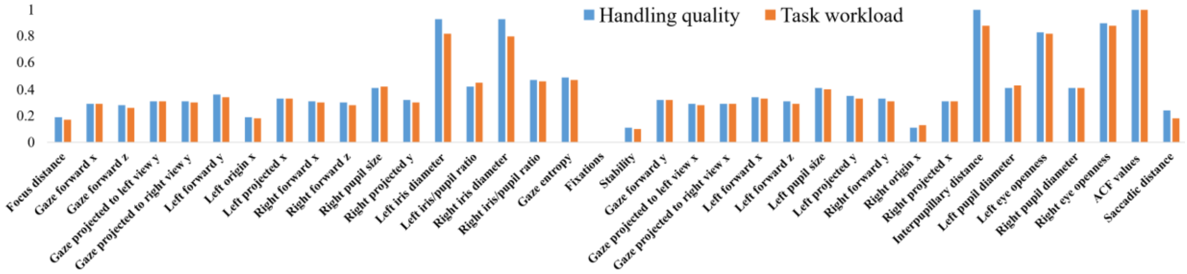


Figure 11. The results of random forest importance analysis.

4.2.3. Gaze Heat Maps

Due to the variability in MTE criteria, it is hard to predefine AOIs. Thus, gaze heatmaps were generated based on participants' gaze direction data. These heatmaps are depicted in Figure 12.

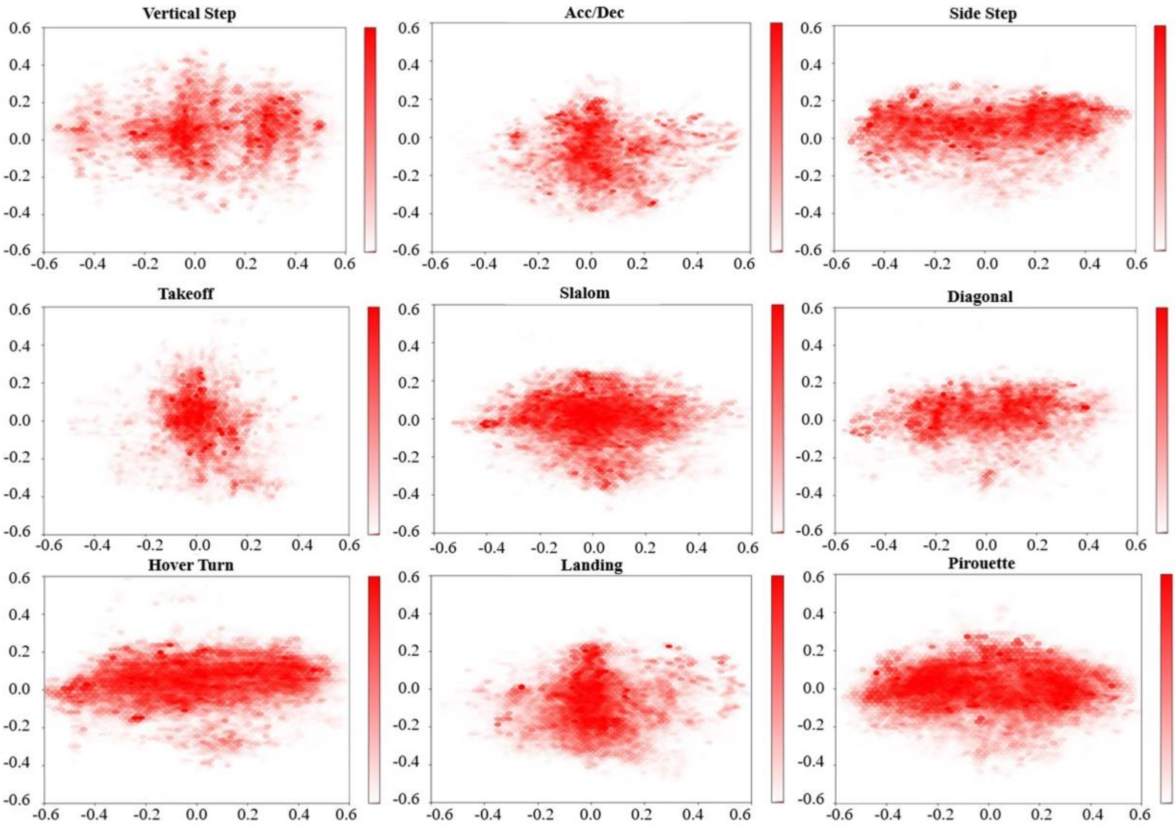


Figure 12. The heat maps of participants' gaze directions.

4.3. Deep Networks

The training results for the proposed model are shown in Table 4. 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 4. This is a table. Tables should be placed in the main text near to the first time they are cited.

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) [43]	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) [44]	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

5. Discussions

5.1. The Associations Between Handling Qualities and Task Workload

A similar trend was observed between 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

ANOVA and HSD tests on eye metrics revealed significant differences 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 [45].

The results of Spearman correlation and RF analysis 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 Heat Maps

Previous studies have found that wider and darker heatmaps are typically associated with higher workload, stress, and pressure [46]. 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, gaze heatmaps under takeoff and landing were narrower although they had high scores. This is likely due to the MTEs' reference mark. 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 [47]. This phenomenon suggests 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 is 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 3-class classification was consistently higher than that of the 9-class classification, demonstrating that the LSTM model performs better with fewer classes. Moreover, in 3-class classifications, the accuracies for subjective handling qualities and task workload levels were consistently higher than those for overall task workload levels and predefined 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 HQ Assessment Methods

Integrating eye metrics into the handling qualities assessment process offers a more objective approach, reducing the discrepancies caused by participants' varying assessment abilities. For example, while the subjective ratings for *slalom* exhibited a significant standard deviation, the gaze heatmap indicated a consistent focus, suggesting that participants might have made improper ratings. This inconsistency calls for further confirmation or follow-up interviews.

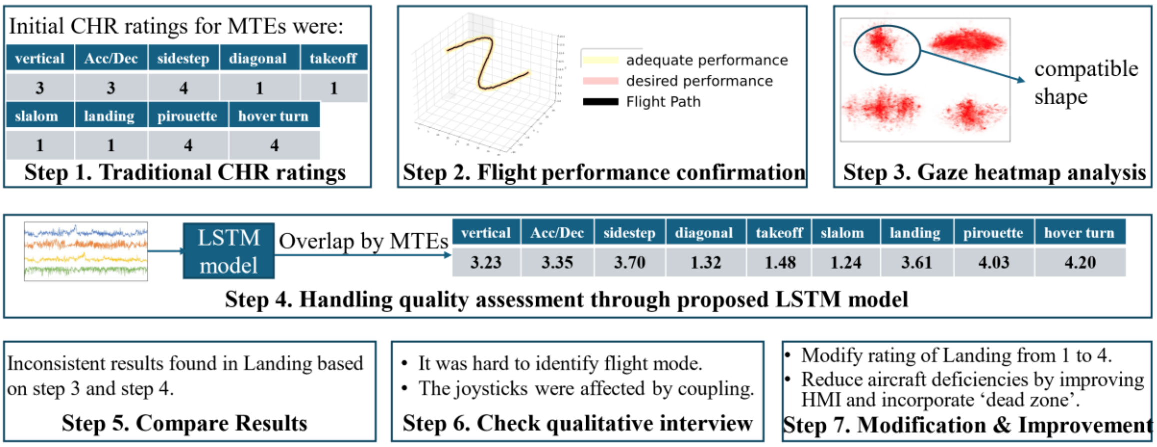
Furthermore, a detailed investigation of eye metrics reveals aircraft design deficiencies, as discussed in 5.3. Eye movement data varies among individuals, compensating for the limitations of traditional flight modeling analysis, which often neglects human factors. Since eVTOL aircraft designs and control 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 13.

1. The operators perform several pre-designed MTEs and r provides handling qualities ratings based on CHR, as described in 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 higher workload, which should correspond to lower CHR ratings;
4. Eye-tracking data is 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.



Initial CHR ratings for MTEs were:

vertical	Acc/Dec	sidestep	diagonal	takeoff
3	3	4	1	1
slalom	landing	pirouette	hover turn	
1	1	4	4	

Step 1. Traditional CHR ratings

Step 2. Flight performance confirmation

Step 3. Gaze heatmap analysis

vertical	Acc/Dec	sidestep	diagonal	takeoff	slalom	landing	pirouette	hover turn
3.23	3.35	3.70	1.32	1.48	1.24	3.61	4.03	4.20

Step 4. Handling quality assessment through proposed LSTM model

Inconsistent results found in Landing based on step 3 and step 4.

Step 5. Compare Results

- It was hard to identify flight mode.
- The joysticks were affected by coupling.

Step 6. Check qualitative interview

- Modify rating of Landing from 1 to 4.
- Reduce aircraft deficiencies by improving HMI and incorporate 'dead zone'.

Step 7. Modification & Improvement

Figure 13. The handling quality assessment process based on the proposed framework.

7. Limitations

The dataset comprised only 16 participants, limiting the results robustness. Future research will need to validate these findings with a larger sample size. Besides, the dataset was imbalanced. Despite efforts to address this issue using Random Sampling and 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. This instability in eye movement measurements could pose significant challenges in putting the method into practice. Additionally, relying solely on eye metrics may not provide a comprehensive assessment of HQ. Future research will integrate additional physiological data, such as ECG signals, to create a more robust assessment framework.

8. Conclusion

This study investigated the potential for incorporating eye metrics in identifying pilot compensation based on statistical data mining and DL network, thus objectively assessing handling qualities. Results demonstrated that handling qualities is associated with task workload, being 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%, providing results equivalent or superior to subjective CHR ratings through eye indicators.

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 Handling Qualities Rating Methods (HQRM), providing deeper insights into pilot compensation and identifying aircraft design deficiencies. 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 the household funding of the Institute of Flight System Dynamics, Technical University of Munich; 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. The indicators output by Varjo XR 3.

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

Appendix B. The calculated eye indicators.

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.

References

1. Zhang, J.; Liu, Y.; Zheng, Y. Overall eVTOL aircraft design for urban air mobility. *Green Energy and Intelligent Transportation* **2024**, *3*, 100-150.
2. European Union Aviation Safety Authority. Means of Compliance with the Special Condition VTOL. 2021.
3. Harper, R.P.; Cooper, G.E. Handling qualities and pilot evaluation. *Journal of Guidance, Control and Dynamics* **1986**, *9*, 500-505.
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, 2009.
6. Harris, D. Measurement of pilot opinion when assessing aircraft handling qualities. *Measurement & Control* **2000**, *33*, 1-5.
7. Ji, H.L.; Chen, R.L.; Li, P. Distributed Atmospheric Turbulence Model for Helicopter Flight Simulation and Handling-Quality Analysis. *Journal of Aircraft* **2017**, *54*, 190-198.
8. Dussart, G.; Lone, M.; Bailey, R. Development of a Multi-Directional Manoeuvre for Unified Handling Qualities Investigation. *Aerospace* **2019**, *6*, 700-710.
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. *Aerospace Science and Technology* **2009**, *13*, 92-104.
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, 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 Systems* **2024**, *12*, 99-131.
14. Cooper, G.E.; Harper, R.P. The use of pilot rating in the evaluation of aircraft handling qualities. National Aeronautics and Space Administration, 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. *The Aeronautical Journal* **2023**.

17. 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, 08-11 August 2011.
18. 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(3)*, 174.
19. Charles, R.L.; Nixon, J. Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics* **2019**, *74*, 221-232.
20. Mahanama, B.; Jayawardana, Y.; Rengarajan, S.; Jayawardena, G.; Chukoskie, L.; Snider, J.; Jayarathna, S. Eye Movement and Pupil Measures: A Review. *Frontiers in Computer Science* **2022**, *3*, e733531.
21. Peissl, S.; Wickens, C.D.; Baruah, R. Eye-Tracking Measures in Aviation: A Selective Literature Review. *International Journal of Aerospace Psychology* **2018**, *28*, 98-112.
22. 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, 2019.
23. Greiwe, D.H.; Friedrich, M. Gaze Movements of Helicopter Pilots during Real and Simulated Take-Off and Landing Maneuvers. *Aerospace* **2024**, *11*, 429-429.
24. Lu, T.J.; Lou, Z.S.; Shao, F.; Li, Y.; You, X.Q. Attention and Entropy in Simulated Flight with Varying Cognitive Loads, *Aerospace Medicine and Human Performance* **2020**, *91*, 489-495.
25. 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. *International Journal of Aerospace Psychology* **2023**, *33*, 153-173.
26. 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. Proceedings of the Institution of Mechanical Engineers part G-Journal of Aerospace Engineering **2022**, 236, 2010-2030.
27. 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. *Safety Science* **2023**, 168.
28. 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. *Physiology & Behavior* **2016**, *153*, 91-96.
29. Pathirana, P.; Senarath, S.; Meedeniya, D.; Jayarathna, S. Eye gaze estimation: A survey on deep learning-based approaches. *Expert Systems with Applications* **2022**, *199*, e116894.
30. 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. *Computer Methods and Programs in Biomedicine* **2022**, *221*.
31. 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 Systems with Applications 2024, 249.
32. 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 & Virtual, June 27-July 1, 2022.
33. 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.
34. 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, Nevada, USA, 30 July- 1 August, 2024.
35. 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, vol 14410; Springer, Cham, 2023.
36. 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, 2000.
37. 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 & Virtual, June 27-July 1, 2022.
38. Varjo. Varjo XR3. Available online: https://developer.varjo.com/docs/native/eye-tracking (accessed on [01 January 2024]).
39. 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. *Aerospace Science and Technology*.
40. Abdi, H.; Williams, L.J. Tukey's honestly significant difference (HSD) test. *Encyclopedia of Research Design* **2010**, *3*, 1-5.

41. 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 Systems with Applications* **2024**, *243*.
42. 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. *Aerospace Science and Technology* **2022**, *126*.
43. 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.
44. 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.
45. 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, 08-11 August 2011.
46. Das, S.; Maiti, J. Assessment of cognitive workload based on information theory enabled eye metrics. *Safety Science* **2024**, *176*.
47. Wickens, C.D.; Alexander, A. Attentional Tunneling and Task Management in Synthetic Vision Displays. *International Journal of Aviation Psychology* **2009**, *19*, 182-199.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.