applied sciences

MDPI

Article

Assessing Feature Importance in Eye-Tracking Data within Virtual Reality Using Explainable Artificial Intelligence Techniques

Meryem Bekler [1,†], Murat Yilmaz [1,†] and Hüseyin Emre Ilgın [2,*,†]

[1] Department of Computer Engineering, Faculty of Engineering, Gazi University, 06570 Ankara, Turkey; meryem.bklr@gmail.com (M.B.); my@gazi.edu.tr (M.Y.)
[2] School of Architecture, Faculty of Built Environment, Tampere University, P.O. Box 600, FI-33014 Tampere, Finland
* Correspondence: emre.ilgin@tuni.fi
† These authors contributed equally to this work.

Abstract: Our research systematically investigates the cognitive and emotional processes revealed through eye movements within the context of virtual reality (VR) environments. We assess the utility of eye-tracking data for predicting emotional states in VR, employing explainable artificial intelligence (XAI) to advance the interpretability and transparency of our findings. Utilizing the VR Eyes: Emotions dataset (VREED) alongside an extra trees classifier enhanced by SHapley Additive ExPlanations (SHAP) and local interpretable model agnostic explanations (LIME), we rigorously evaluate the importance of various eye-tracking metrics. Our results identify significant correlations between metrics such as saccades, micro-saccades, blinks, and fixations and specific emotional states. The application of SHAP and LIME elucidates these relationships, providing deeper insights into the emotional responses triggered by VR. These findings suggest that variations in eye feature patterns serve as indicators of heightened emotional arousal. Not only do these insights advance our understanding of affective computing within VR, but they also highlight the potential for developing more responsive VR systems capable of adapting to user emotions in real-time. This research contributes significantly to the fields of human-computer interaction and psychological research, showcasing how XAI can bridge the gap between complex machine-learning models and practical applications, thereby facilitating the creation of reliable, user-sensitive VR experiences. Future research may explore the integration of multiple physiological signals to enhance emotion detection and interactive dynamics in VR.

Keywords: emotion recognition; emotion models; eye-tracking; virtual reality; feature importance; explainable artificial intelligence; SHAP values; LIME

check for updates

Citation: Bekler, M.; Yilmaz, M.; Ilgın, H.E. Assessing Feature Importance in Eye-Tracking Data within Virtual Reality Using Explainable Artificial Intelligence Techniques. *Appl. Sci.* **2024**, *14*, 6042. https://doi.org/10.3390/app14146042

Academic Editor: João M. F. Rodrigues

Received: 3 June 2024
Revised: 7 July 2024
Accepted: 8 July 2024
Published: 11 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Emotions significantly shape human behavior, influencing cognitive processes, decision-making, social interactions, and overall well-being. Although the concept of emotion is commonly recognized, there is no consensus on its definition, resulting in diverse interpretations. According to Scherer [1], emotions can be described from a componential perspective, which views them as a series of interconnected and synchronized changes occurring across most or all of the five subsystems of an organism. These changes are triggered in response to the organism's evaluation of an internal or external stimulus as significant to its core interests. This perspective underscores emotion as a multifaceted construct involving complex interactions among mental and physiological components, thereby enriching our understanding of its influence on human behavior.

Emotions are linked to transient physiological changes in the body, driven by underlying cognitive processes and emotional responses. The capacity to accurately detect and interpret these emotional states is critically vital across various domains, including psychology [2], physiology [3], healthcare [4], safe driving [5], education [6], and marketing [7].

Recent advancements in machine learning (ML) and deep learning have significantly enhanced emotion recognition capabilities. These technologies facilitate the development of complex algorithms capable of analyzing diverse data sources to infer and categorize human emotions. Such classification techniques are crucial, as they align with established models of emotion to systematically categorize emotional states, thereby broadening our understanding and application of affective computing.

Despite the efficacy of ML techniques, they often need to catch up in terms of interpretability, which is crucial for understanding the outcomes they generate. Interpretability is defined as the extent to which a human can comprehend the reasons behind decisions made by a classifier. Models known as 'glass-box' models, such as linear regression, logistic regression, and decision tree (DT), provide transparency into their operational mechanisms, thereby facilitating human understanding. Conversely, 'black-box' models, which include deep neural networks, ensemble methods like random forests (RF) and gradient boosting machines (GBM), as well as support vector machines (SVM), possess complex internal workings that are less accessible and harder to interpret by humans [8].

In critical domains such as medicine and healthcare research, understanding the rationale behind a model's specific predictions is essential. This necessity extends to business environments, where stakeholders demand clarity on the underlying reasons for each prediction made by machine learning models [9]. To address these challenges, the field of explainable artificial intelligence (XAI) has gained prominence, focusing on making the decision-making processes of opaque 'black-box' models more transparent and comprehensible to human users [10]. Research in this area includes the development of methods like SHapley Additive ExPlanations (SHAP) [11] and local interpretable model agnostic explanations (LIME) [12], which elucidate the contribution of input variables to model predictions. These techniques enhance the interpretability of machine learning models, thereby increasing their transparency and reliability and providing insights into the feature significance that underpins model outputs [13].

Despite significant advances in emotion recognition and the application of XAI to quantify feature importance, there remains a notable gap in integrating these methodologies with established emotion models. This study aims to bridge this gap by leveraging the publicly available VR Eyes: Emotions Dataset (VREED) [14] to detect gaze behaviors associated with various emotional states. By incorporating XAI techniques such as SHAP and LIME in conjunction with the emotion model of the Circumplex Model of Affect (CMA) [15], we systematically assess the contributions of eye-tracking features—saccades, micro-saccades, blinks, and fixations—and their interactions in understanding human affect. This approach allows us to systematically identify and analyze the specific eye movements associated with differing emotions, thereby enhancing our understanding of affective responses in virtual reality settings.

The structure of this paper is organized as follows: Section 2 explains fundamental concepts and provides a detailed literature review. Section 3 outlines the methodology. In Section 4, the results of the applied ML algorithm and XAI techniques SHAP and LIME are presented. Finally, Section 5 presents the conclusions and suggests directions for future research.

2. Literature Review

2.1. Affective Computing and Virtual Reality

The advent of affective computing marks a significant intersection of emotional study and technological advancements. This interdisciplinary field is dedicated to developing systems and devices capable of recognizing, interpreting, processing, and responding to human emotional states [16]. Central to enhancing the effectiveness of affective computing are the processes of emotion recognition and emotion induction [17]. Emotion recognition entails the accurate detection of emotional states using various data sources. These include facial expressions [18], gestures [19], speech [20], eye movements [21], heart rate [22], skin temperature (SKT), blood volume pulse (BVP) [23], wrist pulse signal (WPS) [24], electroen-

cephalogram (EEG) [25], electrocardiogram (ECG) [26], electromyogram (EMG) [27], and galvanic skin response (GSR) [22]. These modalities provide a comprehensive array of physiological and behavioral cues that are instrumental in identifying and interpreting the complex landscape of human emotions.

Emotion induction is a deliberate process designed to elicit specific emotional responses through a variety of techniques, including video clips [28], pictures [21], gameplay [29], music [30], virtual reality (VR) [27], and autobiographical recall [23]. These methods are commonly categorized as passive induction, where participants engage passively within a controlled setting. The passive induction approach offers several benefits, notably the capability to control the stimuli presented, standardize the measurement conditions, and reduce the influence of external variables on the experimental results [31].

Recent studies have increasingly explored virtual reality (VR) across various domains. VR distinguishes itself from other emotion induction techniques through its ability to immerse individuals in highly realistic and meticulously controlled environments [32]. Leveraging advanced VR technologies, researchers can simulate a broad spectrum of scenarios that evoke authentic emotional responses and behaviors. Such profound immersion provides invaluable insights into human cognition and affective states, elucidating the genuine reactions of individuals across diverse emotional contexts. Building on this foundation, Somarathna et al. [32], gathered data on facial expressions, heart activity, EDA, SKT, and ECG during 28 pre-labeled VR games, each tagged with a dominant emotion term. This data was analyzed using SVM, RF, XGBoost, and DT classifiers. The results demonstrated a correlation between the experienced emotions and the pre-labeled emotions based on their frequencies, indicating that VR environments effectively elicit the intended emotions. These findings underscore the vast potential of VR as a powerful tool for studying and understanding human emotions in a controlled yet dynamic setting.

2.2. Emotion Models

To understand human emotions, researchers have developed various models, broadly categorized into discrete and dimensional types. Discrete models identify distinct, universally recognized basic emotions, each characterized by unique features [16]. In contrast, dimensional models describe emotions using continuous scales that reflect changes in psychological and physiological parameters, such as valence and arousal [33]. Arousal refers to the intensity of an emotional state, ranging from high to low, while valence describes the emotion's positivity or negativity.

Dimensional models can capture the continuous nature of emotional states, which are extensively utilized in research. One prominent dimensional model, the Circumplex Model of Affect (CMA), organizes emotions on a two-dimensional circular plane. The vertical axis represents arousal, and the horizontal axis corresponds to valence. This configuration divides the plane into four quadrants: high arousal/positive valence, low arousal/positive valence, low arousal/negative valence, and high arousal/negative valence [15]. Each of these quadrants represents different emotional states based on their arousal and valence levels. For instance, emotions such as alert, excited, elated, and happy are related to high arousal/positive valence; contented, serene, relaxed, and calm are related to low arousal/positive valence; bored, depressed, and sad emotions are referred to as low arousal/negative valence; and finally, emotions such as upset, stressed, nervous, and tense are related to high arousal/negative valence [34]. The CMA, applied in this study, provides a framework for systematically categorizing emotions and is illustrated in Figure 1.

This two-dimensional framework provides a clear and intuitive means of mapping a broad spectrum of emotional states, making it particularly useful for studies involving complex and dynamic environments like VR. Its ability to capture a wide range of emotional experiences with relatively few parameters makes it an efficient and effective tool for large-scale data analysis, as required in our study involving eye-tracking.

While more recent models have emerged, often incorporating additional dimensions or focusing on specific aspects of emotional experience, the CMA's simplicity and gen-

eralizability remain advantageous. The model's enduring use in contemporary research attests to its ongoing relevance. For instance, it continues to be cited and applied in studies exploring affective computing, human-computer interaction, and psychological assessments, providing a common framework for comparing results across different studies and contexts.

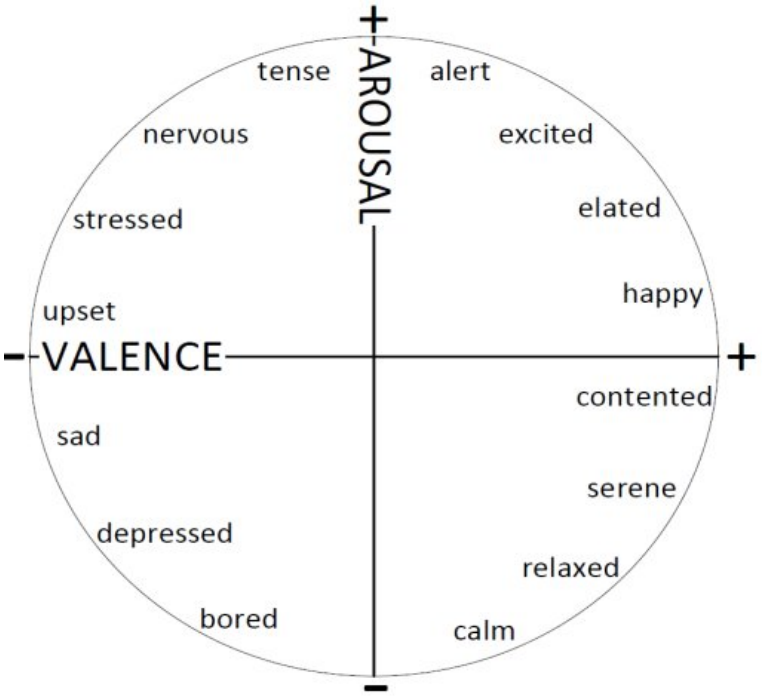

Figure 1. The Circumplex Model of Affect [34].

Utilizing these arousal and valence dimensions to measure emotions, Saffaryazdi et al. [18] integrated facial micro-expressions with EEG and physiological signals collected during a video task designed to elicit six basic emotions and a neutral state. For the analysis, they employed a 3D convolutional neural network (3D-CNN) to classify micro-expression sequences by arousal and valence. Additionally, they used SVM, RF, K-nearest neighbors (KNN), and long short-term memory (LSTM) networks to process other signal data. This multifaceted approach enhances the robustness and accuracy of emotion classification, showcasing the potential of combining multiple data sources and analytical techniques in affective computing.

In their innovative work, Goshvarpour et al. [25] developed an EEG-based emotion recognition system that utilizes a two-dimensional valence-arousal emotional space to categorize emotions. To enhance accuracy, they utilized the publicly available datasets SEED [35] and DEAP [36], classifying emotions into four distinct classes via SVM, KNN, and naïve Bayes (NB). In a related study, Garg et al. [24] explored the efficacy of WPS data in identifying emotions such as anxiety and boredom, employing SVM for their analysis. Furthermore, Miyamoto et al. [37] leveraged CNN to design a system that recommends music based on an individual's emotional state detected through EEG, applying meta-learning techniques to optimize performance. These studies demonstrate the wide range of applications and the effectiveness of emotion recognition systems that incorporate emotion models. By facilitating a deeper understanding of emotional states, emotion models not only improve theoretical insights but also enhance practical applications, ranging from personalized music recommendations to mental health assessments.

2.3. Explainable Artificial Intelligence (XAI)

XAI has emerged as a crucial research field with the objective of demystifying the complex operations and decision-making processes of AI models, making them transpar-

ent and understandable for human users. This is especially important in the healthcare sector, where the reliability of ML methods for diagnostic and treatment strategies is paramount [38].

XAI methodologies can be categorized based on their agnosticity to models and the scope of their explanations [39]. These methods are divided into model-agnostic and model-specific categories. Model-agnostic approaches, such as SHapley Additive ExPlanations (SHAP) and local interpretable model agnostic explanations (LIME), are versatile and applicable across various types of machine learning models, regardless of their underlying architecture. In contrast, model-specific methods like layer-wise relevance propagation (LRP) [40] and integrated gradients (IntGrad) [41] are designed to work with specific models, providing insights tailored to the unique characteristics of those models.

Additionally, the scope of explanation in XAI methods can be classified as either local or global. Local explanations focus on individual instances or records, clarifying why a model made a particular decision in a specific case. Conversely, global explanations aim to illuminate the overall behavior and decision-making process of a model across all instances. This distinction is crucial for users who need to understand whether the explanations reflect isolated cases or generalize across the entire model.

Zhang et al. [42] have developed a deep-learning-based EEG emotion recognition framework that utilizes XAI to assess the impact of individual features on prediction outcomes, focusing mainly on SHAP values. Similarly, Khalane et al. [43] employ SHAP to elucidate significant features across various modalities—audio, video, and text—in multimodal emotion recognition. Torres et al. [44] delve into the interpretability of deep neural networks for EEG-based emotion recognition in autism, analyzing the reliability of various relevance mapping techniques such as LRP, PatternNet, pattern attribution, and smooth-grad squared. In a related study, Kim et al. [45] compare multiple explainable models, including Grad CAM and LIME, for speech emotion recognition.

Liew et al. [46] applied SHAP values to map physiological features onto a binary classification system of arousal and valence using ECG data, highlighting the role of individual features and their interactions in describing emotions. Zhao et al. [47] propose an interpretable classification framework with XGBoost for emotion recognition, utilizing food images and EEG data within an arousal-valence model to provide both local and global interpretations through SHAP values.

Interpretability in emotion recognition models is crucial, as it allows researchers to understand which features or signals contribute most significantly to the classification of emotions, thereby validating the model's decisions and ensuring its reliability across different datasets. In this study, we examined the impact of applying different techniques, namely SHAP and LIME, sequentially on the model's behavior. By employing SHAP, we assessed the global feature importance to understand the overall behavior of the model. Subsequently, we employed LIME to analyze the local explanations for individual predictions. By aggregating these local explanations, we gained valuable insights into the model's general behavior. We then compared these findings with the results obtained from SHAP to evaluate the consistency and robustness of the explanations provided by both methods.

3. Materials and Methods

In this section, we describe the VREED dataset, which is categorized into four unique classes. The preprocessing stage involved adapting the data for binary classification and addressing missing values and imbalances to maintain the dataset's integrity. For the machine learning analysis, we chose the extra trees classifier due to its high predictive performance. Initially, SHAP and LIME were applied to the entire dataset to interpret the impact of individual features on the model's predictions. Subsequently, principal component analysis (PCA) was employed to reduce the number of features by grouping them as blinks, saccades, fixations, and micro-saccades. SHAP and LIME were then reapplied to the reduced dataset. Each of these steps is detailed in this section.

3.1. Data Description

The present study utilized the publicly available VREED (VR Eyes: Emotions dataset), a multimodal affective dataset where emotions were elicited using immersive 360-degree video-based virtual environments (360-VEs) collected through VR headsets [14]. VREED is particularly useful for addressing our research question due to several key factors. Firstly, it is one of the pioneering multimodal VR datasets specifically designed for emotion recognition, incorporating both behavioral and physiological signals, including eye-tracking, ECG, and GSR data. This multimodal approach provides a comprehensive view of emotional responses, making it well-suited for our goal of analyzing visual attention and gaze patterns in VR environments. Secondly, the dataset was meticulously curated, with environments selected based on feedback from focus groups and pilot trials, ensuring that the stimuli used effectively elicit the intended emotional responses. This rigorous selection process enhances the validity of the data, making it a reliable source for examining the correlation between eye-tracking metrics and emotional states. Additionally, VREED includes data from a diverse group of participants across various age ranges, which improves the generalizability of our findings. The balanced representation of different emotional quadrants within the CMA further strengthens the dataset's relevance to our research.

While ECG and GSR data provide valuable physiological insights, in the present study, we chose to focus exclusively on eye-tracking metrics since our primary objective was to analyze visual attention and gaze patterns in VR environments. Eye-tracking data offers a unique and direct window into cognitive and emotional processes, providing rich insights into how individuals interact with virtual environments. Given that eye movements are closely linked to visual attention and can be precisely quantified, they serve as a reliable indicator of emotional states. Moreover, eye-tracking metrics such as saccades, fixations, and blinks offer specific and actionable data that can be directly correlated with emotional responses. This focus aligns with our goal of advancing the understanding of affective computing within VR by leveraging XAI techniques to interpret these specific metrics.

The eye-tracking, ECG, and GSR data were collected from 34 healthy participants, comprising 17 males and 17 females, aged between 18 and 61 years. All participants were required to sign a consent form and complete a pre-exposure questionnaire. They interacted with 12 distinct VEs selected based on a focus group and a pilot trial.

During the selection phase of VEs, the focus group—comprising six experts in human-computer interaction and psychology—identified the potential 126 VEs, which they manually found on the Youtube platform [14]. The focus group selected 21 VEs out of all 126, adhering to specific exclusion criteria. Subsequently, a pilot trial involving 12 volunteers (6 males and 6 females) aged between 19 and 33 years further refined these 21 environments using two established psychological measurement tools: the Self-Assessment Manikin (SAM) and the Visual Analog Scale (VAS) [48,49]. SAM, utilizing cartoon-shaped manikins, helps visualize the arousal-valence dimensions of emotions. At the same time, VAS is a 0 to 100 linear scale for quantifying various emotional states such as joy, happiness, calmness, relaxation, anger, disgust, fear, anxiousness, and sadness. Since the objective of the pilot trial was to select three 360-VEs corresponding to each quadrant of the CMA, ensuring a comprehensive representation of emotional states, a total of 12 VEs were chosen for the final experiment according to SAM (arousal, valence) and VAS (joy, anger, calmness, sadness, disgust, relaxation, happiness, fear, anxiousness, and dizziness) ratings.

In the experiment phase, the eye-tracking, ECG, and GSR data were collected using the final 12 VEs, each trio representing one quadrant of the CMA and designed to evoke a range of emotional responses corresponding to the different quadrants. Each VE included diverse elements of sights, sounds, and activities to create immersive and engaging experiences. In certain environments, participants were exposed to stimuli such as monsters, zombies, or a possessed woman to induce emotions associated with high arousal and negative valence, such as fear, stress, or anger. To evoke emotions of high arousal and positive valence, like excitement, participants were immersed in scenarios where they walked on a tightrope or danced with exotic performers. For inducing low arousal and

positive valence emotions, such as calmness and relaxation, participants encountered settings like a farm with bunnies, a forest with bird sounds, or various spas and tranquil locations around the world. Lastly, to elicit low arousal and negative valence emotions, such as depression and sadness, participants were immersed in environments depicting mourning scenes, refugee camps, or war zones. A total of 34 participants (17 males and 17 females) aged between 18 and 61 years volunteered for this phase after conducting a survey. 19 of them reported having used VR before. None have reported feeling motion sick amid or post-exposure to VR. Before the experiment, all participants signed a consent form and filled out a pre-exposure questionnaire, including the SAM and VAS. Then, they engaged with these environments in a randomized order, from which eye tracking, ECG, and GSR data were collected. They also repeated the SAM and VAS questionnaires after the exposure to VR. Initially, the interactions resulted in 408 trials; however, due to concerns about data quality and technical issues, the final dataset was narrowed down to 312 trials involving 26 participants.

Subsequent to data collection, the raw eye-tracking data were processed using the GazeParser library in Python [50]. This processing extracted vital features such as fixation, micro-saccade, saccade, and blink. Fixation is defined as a temporary halt in eye movement; micro-saccades as minor; involuntary movements within a fixation; saccades as rapid movements between fixations; and blinks as periods when the eyes are closed. The researchers then computed sub-features for each main feature, utilizing statistical calculations like normalized count (NormCount), mean, standard deviation (SD), skewness (Skew), and maximum (Max). These eye-tracking features are detailed in Table 1.

Table 1. All eye tracking data features included in the present study.

Main Features	Statistical Metrics
Fixation	Number of Fixations (NormCount) per second First Fixation Duration Duration (M, Max, SD, Skew)
Micro-Saccade	Number of Micro-Saccade (NormCount) per second Peak Velocity (M, Max, SD, Skew) Direction (Mean, Max, SD, Skew) Horizontal Amplitude (Mean, Max, SD, Skew) Vertical Amplitude (Mean, Max, SD, Skew)
Saccade	Number of Saccade (NormCount) per second Duration (Mean, Max, SD, Skew) Direction (Mean, Max, SD, Skew)
Blink	Number of Blink (NormCount) per second Duration (Mean, Max, SD, Skew)

Additionally, the eye-tracking dataset includes a target column labeled 'Quadrant Category', which corresponds to a quadrant of the CMA for each trial. Table 2 delineates the nominal categories associated with each CMA quadrant. In total, the dataset comprises 312 rows and 50 columns, capturing a comprehensive set of variables crucial for the subsequent analysis.

Table 2. Nominal categorical variables for CMA quadrants.

Quadrant Category	CMA
0	High Arousal/Positive Valence
1	Low Arousal/Positive Valence
2	Low Arousal/Negative Valence
3	High Arousal/Negative Valence

3.2. Data Preprocessing

In this section, we detailed the essential preprocessing steps undertaken to prepare the eye tracking data for a classification problem, aiming to isolate significant variables effective in predicting elicited emotions and their respective CMA categories.

Throughout the preprocessing phase, we utilized Python and Jupyter Notebook for data manipulation and analysis. Initial checks for missing values revealed 96 instances labeled as 'not a number (NaN)'. These were imputed with the average values from their respective columns to maintain data integrity.

Given our goal to assess each quadrant category individually, we reformatted the 'Quadrant Category' target column, which originally contained four distinct values, each of which constructs a class, into a binary classification format for each category. This restructuring led to an imbalance in the dataset; for instance, quadrant category 0 contained only 78 trials, while the other categories comprised the remaining 234 out of 312 trials. Training models on such imbalanced data can introduce bias, as models tend to favor the majority class, thus misleadingly inflating accuracy metrics. To address this, we implemented the Synthetic Minority Over-sampling Technique (SMOTE) [51], which effectively balanced the data by enhancing the representation of minority classes. After the application of SMOTE, a balanced dataset suitable for unbiased binary classification was achieved.

This preprocessing approach was consistently applied across the remaining three categories to ensure uniform data quality and reliability in subsequent analyses.

3.3. Machine Learning Model Selection and Fitting

In this study, the extra trees classifier (ET) was selected as our predictive model due to its superior performance across various metrics, specifically the F1-score, as it provides a more robust evaluation with imbalanced data. Utilizing the PyCaret library [52], we automated the evaluation process for multiple machine learning models, thereby streamlining our workflow and objectively assessing model performance based on key metrics such as accuracy, AUC, recall, precision, and F1-score. Among the 14 models evaluated—which included the extra trees classifier (ET), light gradient boosting (LightGBM), random forest classifier (RF), quadratic discriminant analysis (QDA), extreme gradient boosting (XGBoost), gradient boosting classifier (GBC), AdaBoost classifier (Ada), Ridge classifier (Ridge), linear discriminant analysis (LDA), K-neighbors classifier (KNN), decision tree classifier (DT), logistic regression (LR), naive Bayes (NB), and support vector machines (SVM) with a linear kernel-the extra trees classifier consistently demonstrated superior performance compared to other models in terms of average metric values across each quadrant. The results for the mean values of the calculated metrics in all four quadrants are shown in Table 3.

Table 3. Results for the model selection process.

Model	Accuracy	AUC	Recall	Precision	F1-Score
ET	0.8794	0.9471	0.8797	0.8800	0.8744
RF	0.8659	0.9263	0.866	0.8734	0.8645
LightGBM	0.8458	0.9257	0.8745	0.8319	0.8480
GBC	0.8429	0.9235	0.8723	0.828	0.8459
XGBoost	0.8448	0.9243	0.8589	0.842	0.8453
QDA	0.8689	0.9412	0.7387	0.9947	0.8438
Ridge	0.8155	0.8783	0.8682	0.7896	0.8235
LDA	0.8146	0.8737	0.8719	0.7805	0.8207
Ada	0.7949	0.8726	0.8196	0.7898	0.7971
LR	0.7585	0.8223	0.7831	0.7521	0.7615
KNN	0.6868	0.7516	0.8111	0.6502	0.7193
DT	0.7184	0.7178	0.73	0.7137	0.7166
NB	0.6733	0.7797	0.7654	0.6649	0.6844
SVM	0.5736	0.6693	0.5023	0.5726	0.4514

It is critical to select a model that exhibits high predictive performance to ensure the reliability of XAI techniques SHAP and LIME. These values aim to accurately reflect the significance and contribution of each feature to the model's predictions. Choosing a poorly performing model could result in misleading interpretations, as the SHAP values would be describing an inaccurate representation of the model's behavior.

After the initial preprocessing, the dataset was partitioned into training and testing sets in an 80:20 ratio, yielding 374 rows for training and 94 rows for testing. Subsequently, we applied the ET with its default parameters to each quadrant individually within our dataset. This step was crucial to evaluating the model's performance across different segments of the data and ensuring robustness in its predictive capabilities.

3.4. Application of SHAP for Model Interpretation

Upon establishing a robust machine learning model as the foundation for our analysis, we focused on integrating XAI techniques to enhance the interpretability of the model. Specifically, we utilized the Shapley Additive Explanations (SHAP) algorithm as a tool for assessing the model's global explainability. SHAP provides a comprehensive framework for interpreting the predictions made by machine learning models [11]. This method assigns a SHAP value to each feature, quantifying its relative impact on the model's decision-making process. These values facilitate both local and global explanations, offering insights into the influence of individual features on specific predictions and the model as a whole [53]. The mathematical formulation of SHAP values ensures that the contribution of each feature is accurately represented, thereby enabling a deeper understanding of the model's internal mechanics.

$$\phi_i(f, x) = \sum_{f' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)] \quad (1)$$

where x represents a certain sample requiring explanation, f denotes the model under consideration, i identifies the feature being assessed, and M signifies the total count of features. Additionally, x' encompasses every conceivable variation or disturbance of x.

In this study, we leveraged the SHAP library [11] in Python to compute SHAP values for the features in our dataset. Central to the SHAP framework is the assignment of a SHAP value to each feature, quantifying its incremental impact on the model's prediction for a given instance. Originally adapted from cooperative game theory, where SHAP values ensure a fair allocation of benefits among players, these values are reinterpreted within the SHAP framework to distribute the output of the model among features based on their relative contribution [53]. This approach is particularly advantageous for complex models, where discerning the impact of individual features can be challenging.

Our analysis focused on evaluating the relative importance of features within the dataset across the individual CMA quadrants. Given the unique emotional profiles of each quadrant, specific eye-tracking metrics likely influence distinct emotional states. By applying the SHAP algorithm to our model, we aimed to uncover which eye-tracking features are most influential in evoking specific emotions, thereby providing insights into how human emotions can be effectively identified from eye-tracking data. Our computation of SHAP values for each feature allowed us to identify key features that significantly affect the model's decision-making process in each quadrant. Notably, some of the features assessed either negatively impacted the model's decisions or had no discernible effect. This observation suggests the potential for developing a similarly effective model with a reduced set of features.

3.5. Application of LIME for Model Interpretation

In this subsection, we extended our model interpretability analysis by applying the LIME (local interpretable model-agnostic explanations) algorithm [12] to our dataset. While SHAP provided insights into feature importance at a global level, LIME offers a complementary approach by explaining individual predictions locally. LIME achieves this by

approximating complex machine learning models with interpretable models (such as linear models) on small, locally perturbed subsets of the data. By focusing on local explanations, LIME allows us to understand how specific instances are influenced by features in our model, providing a nuanced understanding of predictions that may not be captured by global feature importance measures alone. Similarly to our use of SHAP, we employed LIME to assess the four quadrants of the CMA separately, utilizing all 49 features. Although LIME is designed to provide local explanations for individual instances within the dataset, we aggregated these local explanations to derive insights into the model's overall behavior. For this purpose, we identified the most significant feature for each instance within our testing data using LIME library [12] in Python and subsequently calculated the two most frequently occurring features for each quadrant. This methodology enabled us to delineate the relationship between specific eye-tracking features and their corresponding CMA quadrants.

Additionally, we conducted a comparative analysis of the two XAI models, SHAP and LIME, to evaluate their consistency and robustness in explaining the model's predictions.

3.6. Principal Component Analysis (PCA)

After applying SHAP and LIME to each of the 49 features of the dataset and analyzing their impact on the model, we employed principal component analysis (PCA) as a dimensionality reduction technique to validate our findings. PCA not only mitigates the curse of dimensionality but also unveils the underlying structure of the data by transforming the original features into a set of orthogonal components. With PCA, we created new features by grouping fixations, blinks, saccades, and micro-saccades. As a result of this, we had 15 features named: Number of Micro-Saccade, Number of Blink, Number of Saccade, Number of Fixations, Blink Duration, Fixation Duration, Saccade Duration, Saccade Direction, Saccade Amplitude, Saccade Length, Micro-Saccade Amplitude, Micro Saccade Direction, Micro Saccade Peak Velocity, Micro Saccade Vertical Amplitude, and Micro Saccade Horizontal Amplitude.

We applied both SHAP and LIME, respectively, to these newly derived features for interpretability following ET and thoroughly analyzed the results. The findings corroborated our previous results obtained using the original 49 features, reinforcing the validity and consistency of our initial interpretations.

4. Results and Discussion

In the results section, we presented a comparison of the performance metrics for the extra trees classifier (ET) using both the full set of 49 features and a reduced feature set obtained through principal component analysis (PCA). Subsequently, we provided the SHAP results for all 49 features, followed by a comprehensive analysis of the LIME results. Similarly, the results for the reduced dataset are shared in the same sequence. Finally, we compared the outcomes to assess the impact of dimensionality reduction and different XAI techniques on model performance and interpretability.

The following quality metrics were used to evaluate the ML results.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

AUC (area under the curve) represents the area under the receiver operating characteristic (ROC) curve, which is a plot of the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings, where (TPR) and (FPR) are defined as following:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

4.1. Results for Extra Trees Classifier

The accuracy, AUC, recall, precision, and F1-scores in each quadrant are detailed in Tables 4 and 5, highlighting the classifier's robust performance. Notably, Quadrant 2 achieves the highest accuracy and F1-score, demonstrating superior prediction capability using the full set of 49 features. However, Quadrant 0 has the highest accuracy and F1-score using a reduced feature set after the deployment of PCA.

Table 4. Results for the extra trees classifier for each quadrant individually with all 49 features.

CMA Quadrant	Accuracy	AUC	Recall	Precision	F1-Score
0	0.9361	0.9369	0.9555	0.9148	0.9361
1	0.9148	0.9090	1.0000	0.8620	0.9129
2	0.9574	0.9571	0.9607	0.9607	0.9571
3	0.9042	0.9045	0.9000	0.9183	0.9039

Table 5. Results for the extra trees classifier for each quadrant individually for reduced features after PCA.

CMA Quadrant	Accuracy	AUC	Recall	Precision	F1-Score
0	0.9148	0.9156	0.9333	0.8936	0.9148
1	0.8191	0.8190	0.8200	0.8367	0.8186
2	0.8829	0.8775	0.9411	0.8571	0.8806
3	0.8510	0.8504	0.8600	0.8600	0.8504

4.2. Shap Analysis across 49 Features

4.2.1. Results for Quadrant 0

After the application of the ET Classifier, we employed the SHAP algorithm to generate visual representations of the influence exerted by all 49 features in Quadrant 0 of the CMA. These influences are depicted in the summary plots shown in Figure 2, which arranges the features according to their importance to the model. The results highlight that the 'SD Micro-Saccade Peak Velocity' plays a significant role in influencing the model's predictions. Other features are presented in descending order of importance.

In Figure 2, the y-axis lists the feature names, arranged from the most to the least important from top to bottom. The x-axis displays the SHAP values, which indicate the degree of change in the model's log odds prediction. The color of each dot on the plot represents the value of the corresponding feature, with red indicating high values and blue indicating low values. Each point on the graph represents an observation from the dataset.

The analysis shows that high values of 'SD Micro-Saccade Peak Velocity' correspond to positive SHAP values, suggesting a positive impact on the model's output. Conversely, lower values of 'SD Saccade Amplitude' are associated with positive SHAP values, indicating that these lower values positively impact the model's predictions.

A detailed examination of features associated with blinking reveals that lower values in 'Skew Blink Duration', 'Number of Blink', 'Max Blink Duration', 'SD Blink Duration',

and 'Mean Blink Duration' are correlated with higher SHAP values. This suggests a positive influence on the model's predictions for emotions typical of Quadrant 0, which includes high-arousal/positive-valence emotional states like happiness, excitement, and alertness. Indeed, high arousal states are associated with decreased blinking. Generally, decreased blinking may indicate that an individual is engrossed in activities that require visual attention, such as reading intensely, watching an engaging movie, or participating in stimulating conversations.

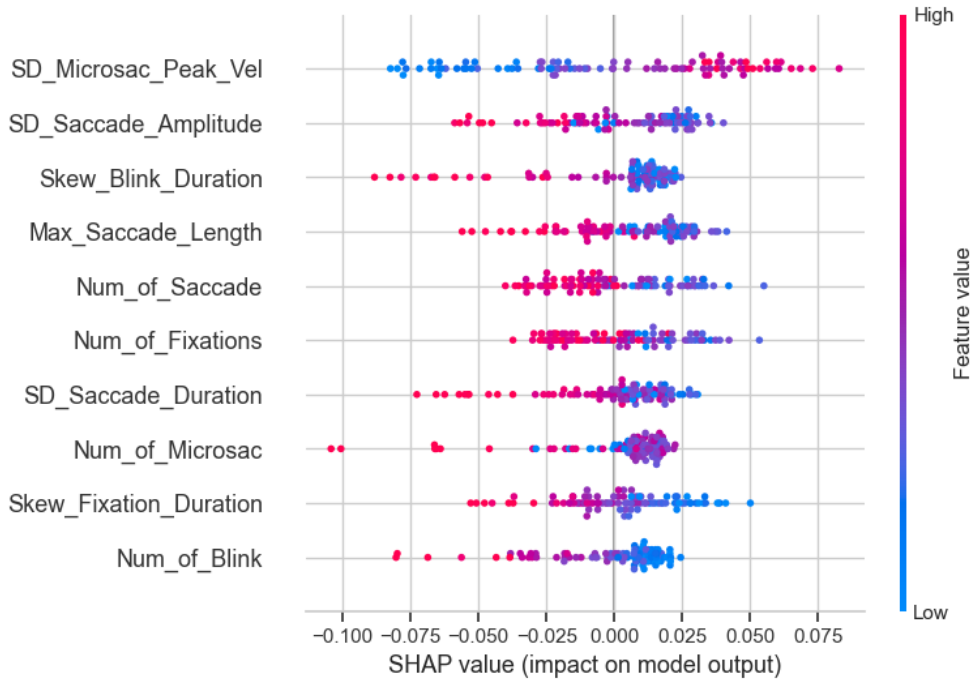

Figure 2. Summary Plot of the First 10 Features for Quadrant 0. (For a complete view of all 49 features, visit: https://doi.org/10.6084/m9.figshare.26130589 (accessed on 10 July 2024)).

Similarly, features related to fixation durations—specifically 'First Fixation Duration', 'Mean Fixation Duration', and 'SD Fixation Duration'—show that higher values positively impact the model, while lower values of 'Skew Fixation Duration', 'Max Fixation Duration', and 'Number of Fixations' have a positive impact. Although the results show some uncertainty, this aligns with observations that individuals experiencing high arousal and positive emotional states, such as during enjoyable activities, tend to show longer fixation durations due to both increased attention and positive emotional engagement. In addition to that, individuals experiencing positive emotions often show an exploratory attentional style with frequent gaze shifts, resulting in shorter fixation durations.

Regarding saccadic movements, lower values in 'Number of Saccade', 'SD Saccade Duration', 'Maximum of Saccade Duration', and 'Skew Saccade Duration' are beneficial for the model, reflecting that heightened emotional arousal often leads to quicker, more frequent eye movements as individuals scan their environment. However, 'Mean Saccade Duration' exhibits both positive and negative effects, depending on its values, indicating a nuanced role in emotional processing. Similarly, lower counts of 'Number of Micro-Saccade' generally positively affect the model, consistent with the reduced micro-saccadic movements seen when individuals are focused and emotionally engaged.

Utilizing all 49 features in the model results in an F1-score of 0.9361 for Quadrant 0. However, refining the model to prioritize features with higher importance could further enhance its performance. Variations in F1-scores with different numbers of selected features for Quadrant 0 are depicted in Figure 3.

Figure 3 clearly shows that using the top 34 important features instead of all 49 can enhance model performance, achieving an F1-score of 0.95. Additionally, the model's

predictions for Quadrant 0 improve significantly when the top 43 features are used, reaching an F1-score of 0.97.

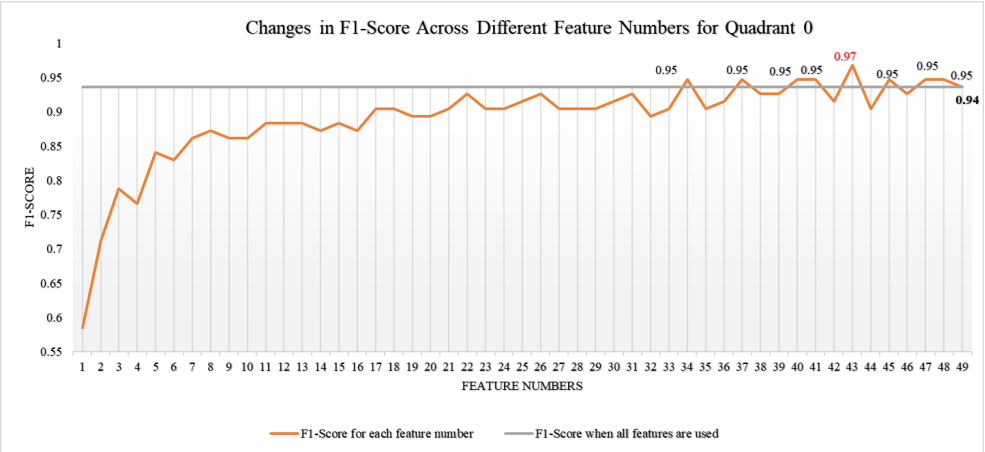

Figure 3. Variations in F1-Score across different feature numbers for Quadrant 0.

4.2.2. Results for Quadrant 1

In Figure 4, we observe that 'Max Saccade Length', which represents the amplitude of saccades, is the most significant feature for predicting emotions in Quadrant 1 of the CMA, associated with low arousal and positive valence, indicative of emotions such as relaxation and calmness. A higher value of this feature positively influences the model's predictions, suggesting that individuals experiencing these emotions tend to exhibit longer saccades. This indicates that during states of calmness and relaxation, the eye travels greater distances between fixation points. Similarly, higher values of 'SD Saccade Length' and 'Mean Saccade Length' also affect the model's output in a positive way.

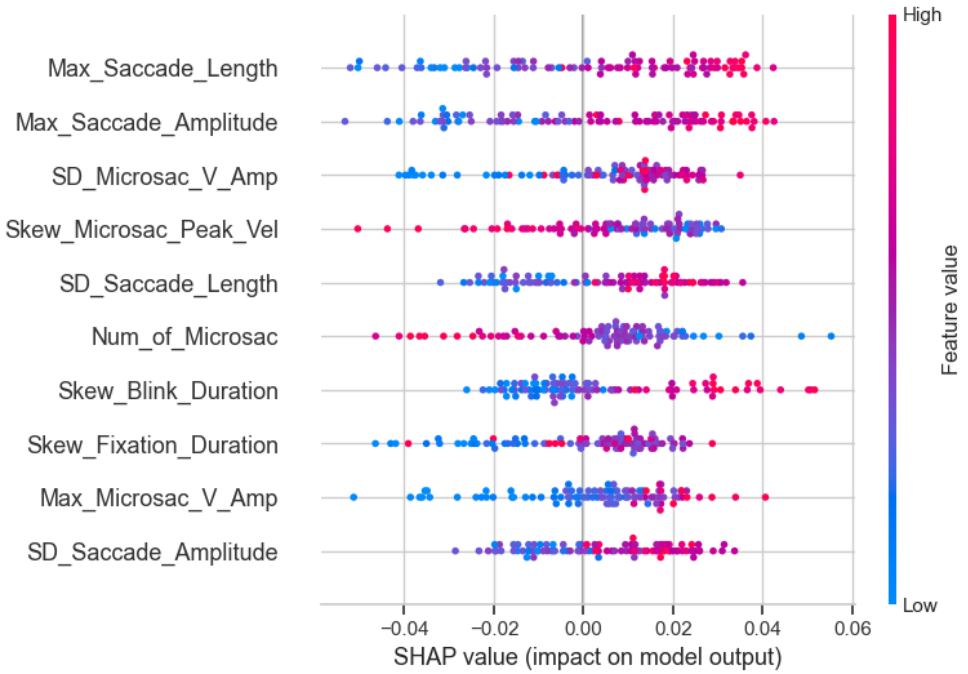

Figure 4. Summary Plot of the First 10 Features for Quadrant 1. (For a complete view of all 49 features, visit: https://doi.org/10.6084/m9.figshare.26130589 (accessed on 10 July 2024)).

The figure also shows that an increased 'Number of Saccade' impacts the model generally in a positive way, but in some cases, it has a negative impact. This variability suggests that the number of saccades alone does not consistently indicate specific emotions

within Quadrant 1. Additionally, higher values for 'Max Saccade Amplitude', 'SD Saccade Amplitude', and 'Mean Saccade Amplitude' positively affect the model, reinforcing the observation that larger saccadic movements are characteristic of individuals feeling positive and relaxed.

Further analysis reveals that lower values of 'Skew Fixation Duration', 'Max of Fixation Duration', 'SD Fixation Duration', and 'Mean Fixation Duration' negatively impact the model's predictions. Conversely, higher values in these fixation duration metrics are associated with states of calmness or relaxation. This pattern supports the hypothesis that during relaxed states, individuals tend to have longer fixation durations, indicating fewer eye movements and a more stable gaze.

Figure provides insight into the role of blinking-related features in the model's predictions for Quadrant 1 of the CMA, which is associated with emotions such as relaxation and calmness. The data show that lower values for 'Skew Blink Duration' and 'Max Blink Duration' negatively influence the model, suggesting that shorter blink durations do not correspond with the emotional states typical of this quadrant. Conversely, longer blink durations are more indicative of relaxation and calmness. Additionally, 'Number of Blink' and 'Standard Deviation of Blink Duration' demonstrate minimal influence on the model, leading to negligible changes in SHAP values, indicating their limited diagnostic value for emotions in this quadrant. Figure provides insight into the role of blinking-related features in the model's predictions for Quadrant 1 of the CMA, which is associated with emotions such as relaxation and calmness. The data show that lower values for 'Skew Blink Duration' and 'Max Blink Duration' negatively influence the model, suggesting that shorter blink durations do not correspond with the emotional states typical of this quadrant. Conversely, longer blink durations are more indicative of relaxation and calmness. Additionally, 'Number of Blink' and 'SD Blink Duration' demonstrate minimal influence on the model, leading to negligible changes in SHAP values, indicating their limited diagnostic value for emotions in this quadrant.

Further analysis in Figure 5 demonstrates the benefits of feature selection. Employing only the top 12 most important features instead of the full set of 49 results in an enhanced model performance with an F1-score of 0.95. This optimized feature set not only improves the model's accuracy but also demonstrates that a reduced number of features can significantly enhance prediction capabilities for Quadrant 1, with F1-scores of 0.92, 0.94, and 0.95 achieved with varying subsets of these features. Notably, the model achieves peak F1-scores with 21, 37 and 46 features, suggesting optimal feature balance.

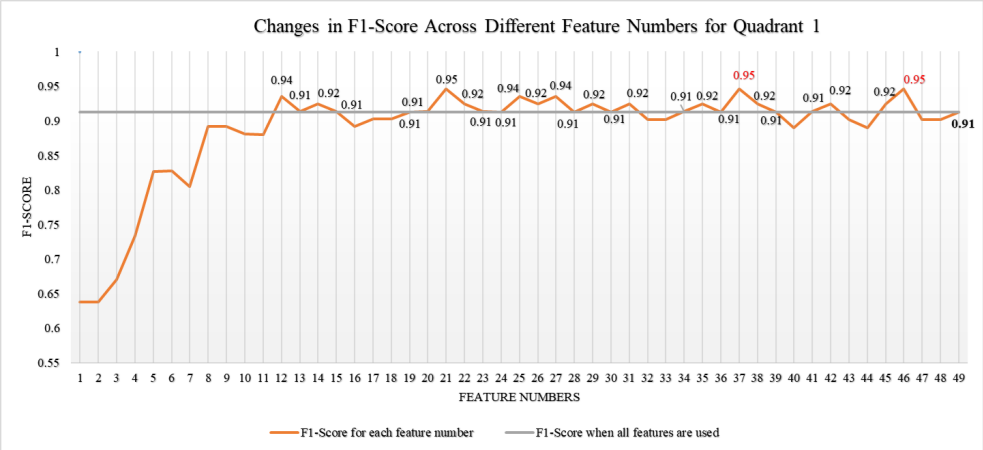

Figure 5. Variations in F1-Score across Different Feature Numbers for Quadrant 1.

Moreover, using just 13 features yields the same F1-score as using all 49 features, indicating that fewer features can train the model just as effectively. This reduction in feature count decreases the computational demand and memory usage, consequently speeding up the model's training process.

4.2.3. Results for Quadrant 2

In Figure 6, it is evident that the 'Number of Micro-Saccade' holds significant importance for predicting emotions in Quadrant 2 of the CMA, which is associated with low arousal and negative valence, indicative of emotions such as sadness, depression, and boredom.

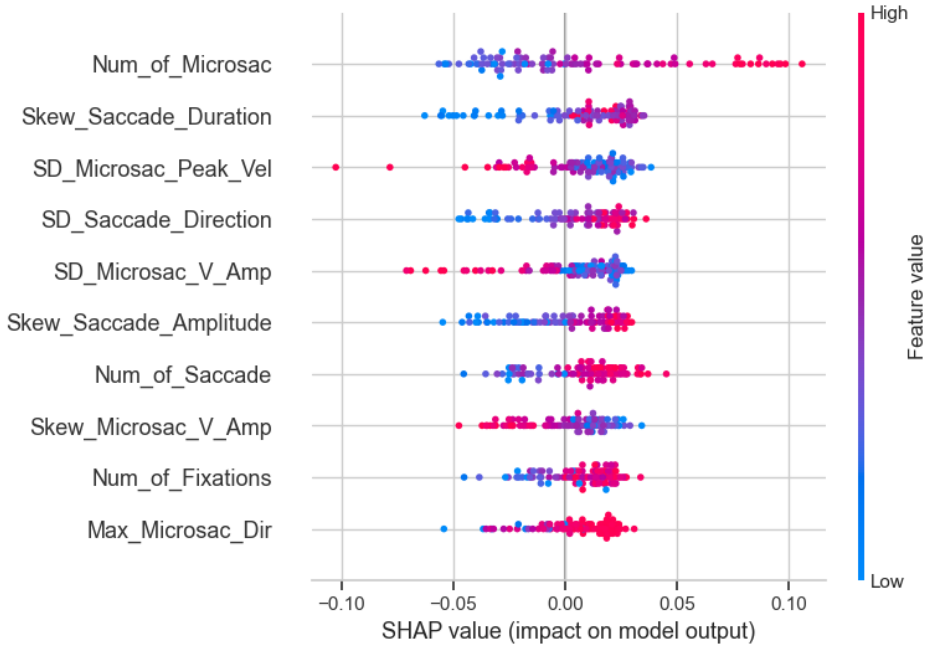

Figure 6. Summary Plot of the First 10 Features for Quadrant 2. (For a complete view of all 49 features, visit: https://doi.org/10.6084/m9.figshare.26130589 (accessed on 10 July 2024)).

The Figure 6 shows that higher values of 'Number of Micro-Saccade' positively influence predictions, suggesting that individuals experiencing emotional states such as sadness or depression exhibit more frequent tiny, involuntary eye movements during visual fixation.

Additionally, an increase in 'Max Micro-Saccade Direction'—which represents the most common orientation or path of a micro-saccade—also positively impacts the model. This implies that not only the frequency but also the specific orientation of micro-saccades is relevant in contexts of sadness or depression.

Observations from Figure 6 further indicate that higher counts of 'Number of Saccade', 'Max Saccade Duration', and 'Mean Saccade Duration' positively affect the model when predicting emotions characteristic of Quadrant 2. This supports the hypothesis that longer saccade durations are linked to decreased engagement with the environment, which is typical in states of depression, sadness, or boredom.

Conversely, lower values for 'Mean Blink Duration', 'SD Blink Duration', and 'Max Blink Duration' positively influence the model, suggesting that blinking frequency and duration decrease in these emotional states. However, a higher 'Number of Blink' impacts the model in a positive way, indicating that an increased blinking rate is associated with low arousal and negative emotional states.

Furthermore, an increase in the 'Number of Fixations' positively impacts model predictions, suggesting that individuals feeling sad or depressed may fixate more frequently, though not necessarily for longer durations. The features 'Mean Fixation Duration' and 'Max Fixation Duration' show average values and do not significantly impact the model, indicating that the duration of fixations may be less indicative of emotional state compared to their frequency.

These insights demonstrate that eye movement metrics like saccades, micro-saccades, and fixations provide critical indicators of emotional states, particularly in contexts characterized by negative valence and low arousal.

Figure 7 illustrates the impact of feature reduction on the model's performance for Quadrant 2. By utilizing only the top 33 important features instead of the full set of 49, the model achieves an enhanced F1-score of 0.97. Moreover, the model maintains the same F1-score of 0.97 with fewer features—specifically 36, 39, 40, and 42 features—demonstrating that a streamlined feature set can achieve comparable or superior performance compared to using all available features. This finding highlights the potential for optimizing computational efficiency without sacrificing accuracy.

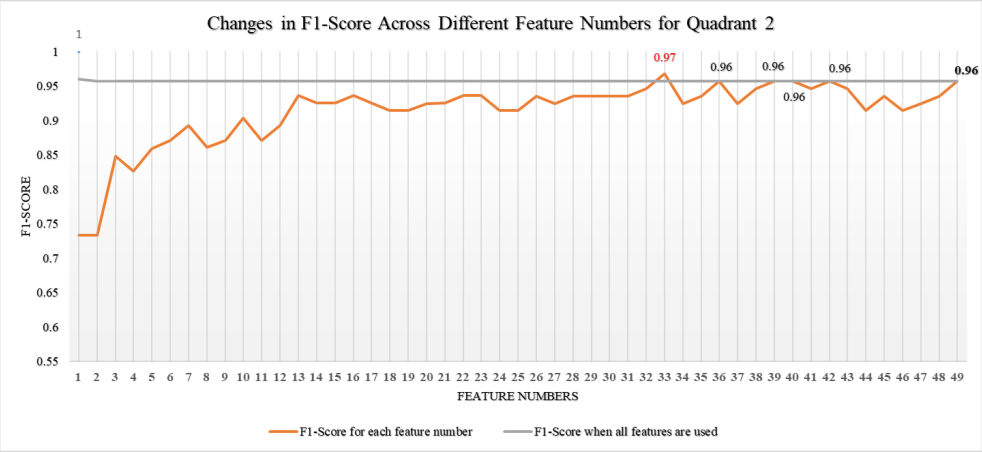

Figure 7. Variations in F1-Score Across Different Feature Numbers for Quadrant 2.

4.2.4. Results for Quadrant 3

Figure 8 identifies 'Max Micro-Saccade Direction' as the most impactful feature on the model's predictions for Quadrant 3, which is characterized by high arousal and negative valence, associated with emotions such as anger, fear, stress, and disgust. The figure reveals that higher values of this feature negatively impact the model's predictions, suggesting that larger angular movements in micro-saccades are indicative of these intense emotional states. This pattern may reflect a behavioral response where individuals experiencing strong negative emotions such as disgust, fear, or stress exhibit rapid, large-angle micro-saccades as a part of their visual scanning or threat detection processes.

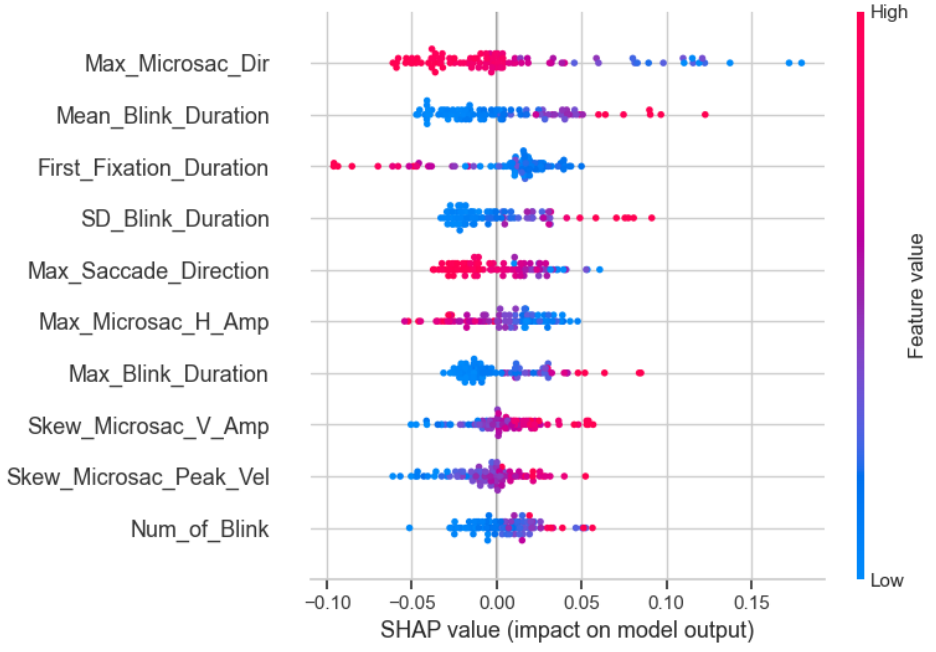

Figure 8. Summary Plot of the First 10 Features for Quadrant 3. (For a complete view of all 49 features, visit: https://doi.org/10.6084/m9.figshare.26130589 (accessed on 10 July 2024)).

In Quadrant 3, characterized by high arousal and negative valence associated with emotions like anger, fear, and disgust, specific blinking and fixation features significantly influence the model's predictions. It is observed that higher values of 'Mean Blink Duration', 'SD Blink Duration', 'Max Blink Duration', and 'Number of Blink' positively impact the model's predictions. This pattern indicates that individuals tend to blink more frequently and for longer durations when experiencing intense negative emotions, likely as a physiological response to stress or discomfort.

Additionally, shorter 'First Fixation Duration' has a positive effect on the model's accuracy in Quadrant 3. This suggests that when confronted with disturbing or frightening stimuli, individuals are likely to make brief initial fixations, possibly as a way to avoid distressing visual content.

However, the influence of the remaining features on the model's predictions shows variability, affecting outcomes both positively and negatively. This complexity underscores the challenge of making generalizations about the relationship between certain eye tracking metrics and specific emotional states in this quadrant. The data indicate that while some eye movement patterns are consistently linked with high-arousal negative emotions, others do not show a clear trend and may depend on additional contextual or individual factors.

Figure 9 depicts how the F1-scores fluctuate as the number of features used in the model for Quadrant 3 is varied. Notably, the results demonstrate that the model's predictive accuracy remains stable even when the number of features is substantially reduced. Specifically, reducing the feature set from all 49 features to just the top 9 most significant ones does not compromise performance. Furthermore, a model utilizing only 19 features performs on par with one that incorporates the full feature set. Additionally, the model reaches its peak performance by leveraging 42 features with a f1-score of 0.97. These findings suggest that a more streamlined model, which prioritizes the most impactful features, can achieve comparable accuracy to a more complex model that uses a larger array of features.

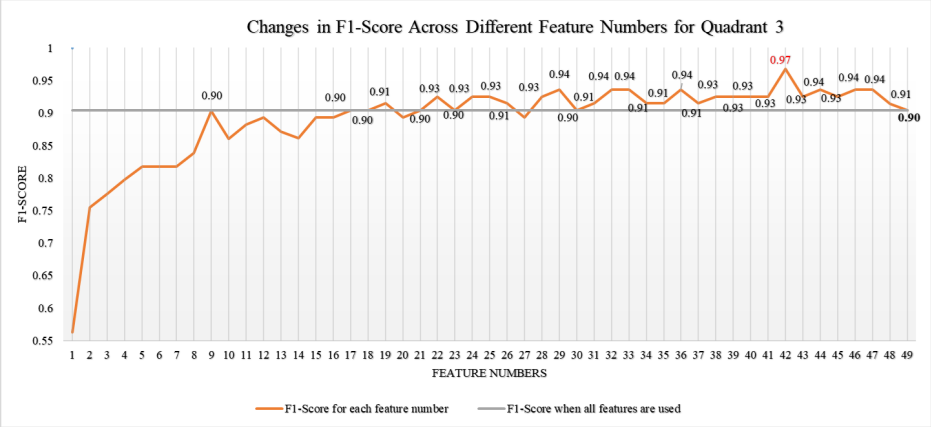

Figure 9. Variations in F1-Score Across Different Feature Numbers for Quadrant 3.

4.2.5. Key Findings of SHAP Analysis across 49 Features

Key findings for four quadrants after SHAP analysis, including all 49 features, can be seen in Table 6. Key features refer to those attributes that hold the highest significance as determined by summary plots. These summary plots provide a visual representation of the importance of each feature in the model. The analysis reveals that among the various attributes examined, saccadic movements stand out as the most crucial. This indicates that saccadic movements have the greatest influence on the model's predictions within the context of the four quadrants.

Table 6. Key findings from the SHAP analysis for all features.

Quadrant	Key Features	Influence on Model Predictions
Quadrant 0	SD Micro Saccade Peak Velocity	Higher values of 'SD Micro Saccade Peak Velocity' positively impact predictions, correlating with high arousal/positive valence emotions (e.g., happiness, excitement).
	SD Saccade Amplitude	Lower values of 'SD Saccade Amplitude' also positively impact the model predictions, indicating smaller saccadic movements.
Quadrant 1	Max Saccade Length	Higher 'Max Saccade Length' and 'SD Saccade Length' are cruical for predicting low arousal/positive valence emotions (e.g., relaxation).
	SD Saccade Amplitude	Longer saccades may indicate a state of calmness or relaxation.
Quadrant 2	Number of Micro-Saccade	'Number of Micro-Saccade' is a significant predictor for low arousal/negative valence emotions (e.g., sadness).
	Skew Saccade Duration	Higher values for 'Skew Saccade Duration' positively affect the model, indicating longer saccade durations linked to low arousal/negative valence.
Quadrant 3	Max Micro-Saccade Direction	Lower values of 'Max Micro-Saccade Direction' impact the model positively, demonstrating high arousal/negative valence emotions (e.g., stress).
	Mean Blink Duration	Higher values of 'Mean Blink Duration' have a positive impact on model predictions, indicating that individuals tend to blink longer durations when they are stressed.

4.3. Lime Analysis across 49 Features

In this section, we present the results of LIME applied to our dataset, comprising 49 features. LIME offers valuable insights into the local interpretability of machine learning models by generating explanations for individual predictions. By applying LIME to each instance within our dataset, we aimed to uncover specific insights into how these features influence model predictions across diverse instances. This local focus allows us to see how the influence of a feature can vary from one instance to another, providing a nuanced understanding of the model's behavior in specific contexts. We observed that LIME's local explanations can significantly differ from one observation to the next, highlighting the variability in feature importance depending on the instance being examined.

Then, we aggregated these local explanations to derive insights into the model's overall behavior to be able to compare with the results of SHAP. For this purpose, we identified the most significant feature for each instance within our testing data and subsequently calculated the two most frequently occurring features for each quadrant. Table 7 shows the most influential features for each quadrant when LIME is applied.

Comparing these findings from LIME with the SHAP results in Table 6, we note that 'SD Micro-Saccade Peak Velocity' in Quadrant 1, 'Number of Micro-Saccade' in Quadrant 2, and 'Max Micro-Saccade Direction' in Quadrant 3 emerge as the most influential features across both algorithms. This comparison highlights the consistent significance of certain features across both LIME and SHAP in different quadrants of the dataset. In contrast, the most influential features in Quadrant 2 are distinct and unrelated to each other. These results demonstrate that while both SHAP and LIME are effective tools in XAI and sometimes produce similar results, their different methodologies, scopes, and applications can lead to varied outcomes.

Table 7. Key findings from the LIME analysis across all features.

Quadrant	Key Features	Influence on Model Predictions
Quadrant 0	SD Micro-Saccade Peak Velocity	These two features are the most frequently occurring features identified as the most significant on model's predictions by local explanations of LIME. Depending on the individual instances, they are indicative for high arousal/positive valence emotions (e.g., happiness, excitement).
	Number of Blink	
Quadrant 1	Max Saccade Direction	These features are the most frequently occurring features identified as the most significant on model's predictions for emotions characterized by low arousal/positive valence (e.g., relaxation) according to LIME.
	Number of Micro-Saccade	This impact can be both positive or negative depending on the specific instance of the dataset.
Quadrant 2	Number of Micro-Saccade	Different values of these features, whether lower or higher depending on the instance, signify emotions associated with low arousal/negative valence emotions (e.g., sadness).
	SD Micro-Saccade Peak Velocity	These two features are the most influential features by local explanations of LIME.
Quadrant 3	Max Micro-Saccade Direction	According to local explanations of LIME, these two features are the most frequently occurring explanations for emotions characterized by high arousal/negative valence (e.g., stress).
	SD Blink Duration	Their impact can be both positive or negative depending on the specific instance of the dataset.

4.4. Shap Analysis Across Reduced Features

4.4.1. Results for Quadrant 0

After the application of the ET classifier on the reduced features, we employed the SHAP algorithm to generate visual representations of the influence exerted by these 15 features for Quadrant 0 of the CMA. These influences are depicted in the summary plot shown in Figure 10.

The results highlight that the 'Number of Micro-Saccade' plays a significant role in influencing the model's predictions. The analysis shows that lower values of 'Number of Micro-Saccade' generally correspond to positive SHAP values, suggesting a positive impact on the model's output, aligned with the decrease in micro-saccadic movements observed when individuals are concentrated and emotionally involved. Similarly, low values of 'Micro-Saccade Peak Velocity', referring to the highest speed reached by micro-saccades during their rapid movements, impact the model's decisions in a positive way. This means that individuals tend to make slower eye movements when they experience the emotional states typical of this quadrant. On the other hand, other micro-saccade-related features have an inconsistent impact on the model output.

The figure also shows that a decreased 'Number of Blink' impacts the model in a positive way by causing higher SHAP values. Similarly, lower values for 'Blink Duration' also result in higher SHAP values. This suggests a positive influence on the model's predictions for emotions typical of Quadrant 0, which include happiness, excitement, and alertness. These results correlate with the situation when all 49 features were included in the predictions.

Similarly, a lower 'Number of Fixations' influences the model positively, indicating that less frequent eye movements are linked to positive emotional states represented in quadrant 0. The 'Fixation Duration' feature has a mixed impact, indicating an impact in both ways that lower values of 'Fixation Duration' result in both positive and negative SHAP values. Although these results show that individuals who experience high arousal

and positive emotional states tend to show shorter fixation durations, this aligns with the exploratory and dynamic attentional styles of individuals.

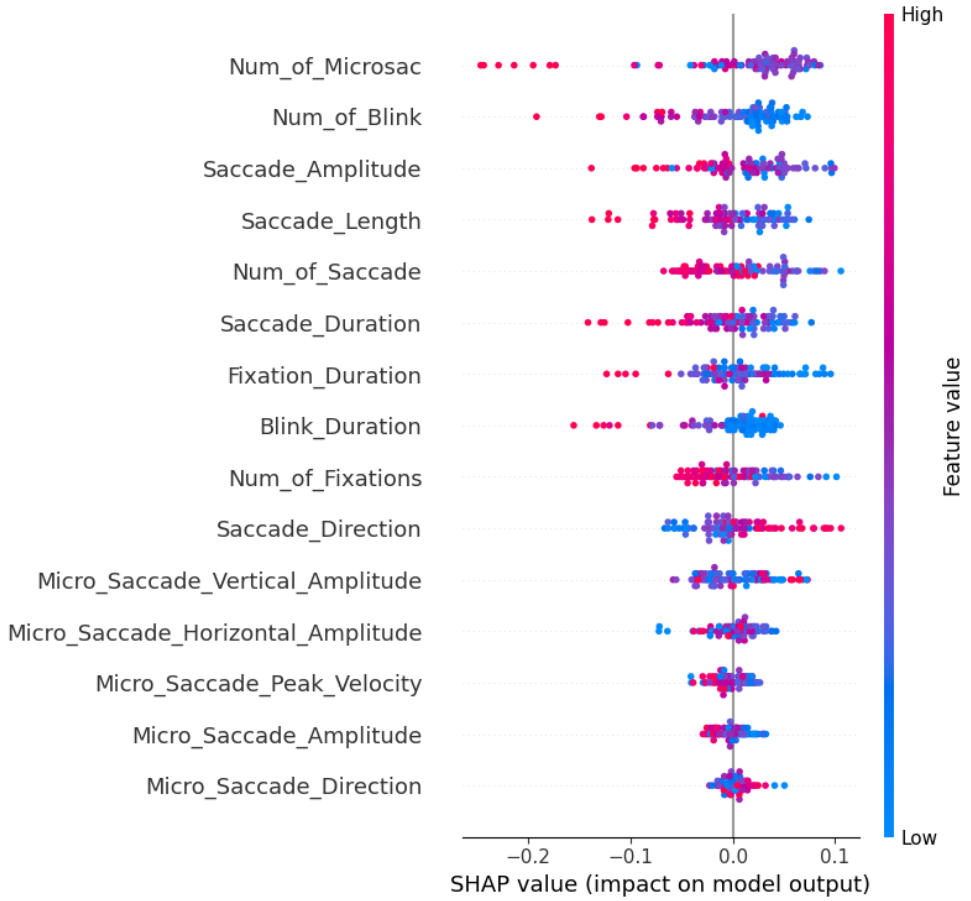

Figure 10. Summary Plot for Quadrant 0 for 15 features.

Regarding saccadic movements, lower values in 'Number of Saccade', 'Saccade Amplitude', and 'Saccade Duration' are advantageous for the model, demonstrating that increased emotional arousal typically results in faster and more frequent eye movements as individuals survey their surroundings. However, higher values for 'Saccade Direction' impact the model positively, indicating that individuals' eyes move within a wide visual field when they experience high arousal and positive emotions. These findings align with the scenario where the model utilized all features.

4.4.2. Results for Quadrant 1

In Figure 11, we observe that 'Number of Micro-Saccade' is the most significant feature for predicting emotions in Quadrant 1 of the CMA, associated with low arousal and positive valence, indicative of emotions such as relaxation and calmness. Lower values for this feature have a positive impact on the model's decisions, indicating that individuals tend to perform less rapid involuntary eye movements while fixating on a particular point when they are calm. Conversely, higher values of 'Micro-Saccade Horizontal Amplitude', 'Micro-Saccade Vertical Amplitude', 'Micro-Saccade Direction', and 'Micro-Saccade Amplitude' cause positive SHAP values, suggesting that individuals experiencing these emotions tend to exhibit longer micro-saccades. This indicates that during states of calmness and relaxation, the eye travels greater distances between fixation points from all angles.

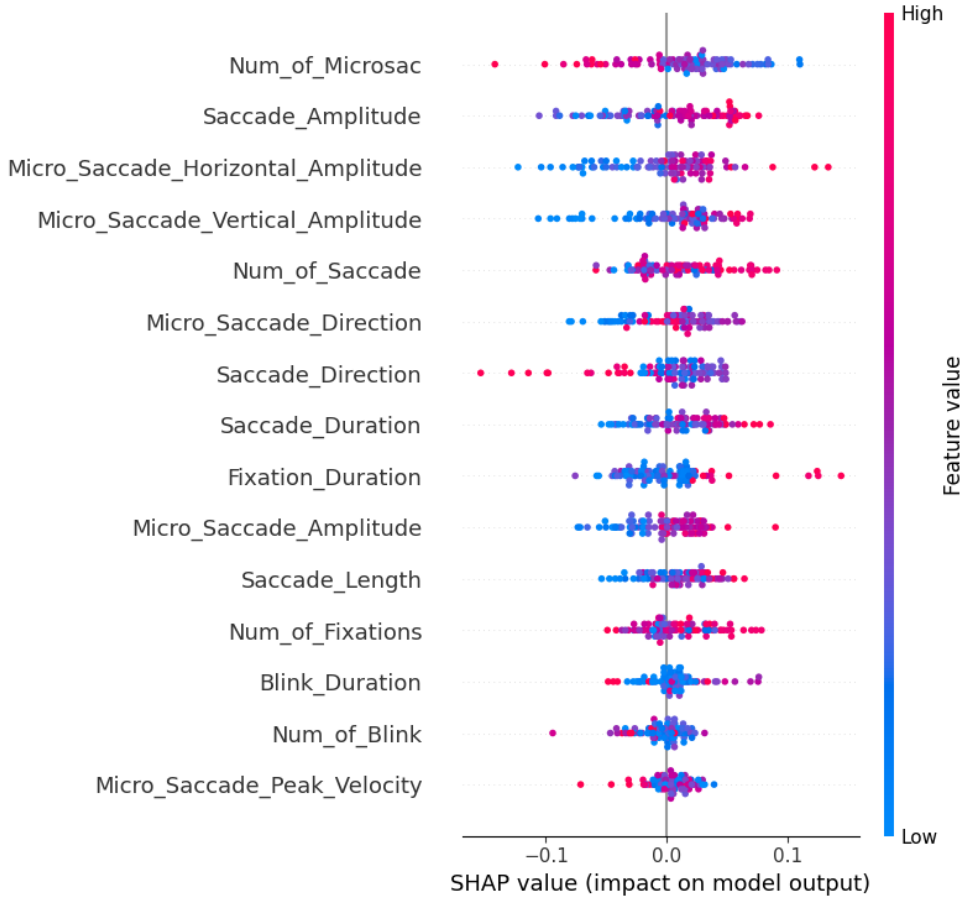

Figure 11. Summary Plot for Quadrant 1 for 15 features.

Similarly, higher values of 'Number of Saccade', 'Saccade Amplitude', 'Saccade Length', and 'Saccade Duration' generally cause positive SHAP values, indicating that people feeling these emotions, such as relaxation and calmness, are likely to display longer saccades. However, lower values for 'Saccade Direction' impact the model positively, suggesting that smaller angular movements in saccades are indicative of intense emotional states in Quadrant 1.

Additionally, lower values of 'Fixation Duration' have a negative impact on the model, while higher values impact positively. In contrast, higher values in these fixation duration metrics are linked to states of calmness or relaxation. This pattern corroborates the hypothesis that individuals in relaxed states tend to have longer fixation durations, signifying fewer eye movements and a steadier gaze. Yet, increased 'Number of Fixations' impacts the model ambiguously, indicating both positive and negative effects.

Figure 11 demonstrates that lower values of 'Number of Blink' and 'Blink Duration' impact the model ambiguously, indicating both positive and negative effects. This variability results in insignificant alterations in SHAP values, suggesting their minimal diagnostic significance for emotions in this quadrant. All these findings for Quandrant 1 generally correlate with the scenario with all 49 features.

4.4.3. Results for Quadrant 2

In Figure 12, it is clear that the 'Number of Micro-Saccade' plays a crucial role in predicting emotions in Quadrant 2 of the CMA. This quadrant is linked to low arousal and negative valence, reflecting emotions like sadness, depression, and boredom. Higher values of 'Number of Micro-Saccade' positively influence predictions, indicating that individuals feeling sadness or depression show a higher frequency of small, involuntary eye movements during visual fixation. Conversely, lower values of 'Micro-Saccade Vertical Amplitude', and

'Micro-Saccade Amplitude' positively impact the model, suggesting that smaller angular movements in vertical direction in micro-saccades are indicative of these intense emotional states, while higher values of 'Micro-Saccade Horizontal Amplitude' create higher SHAP values, indicating that the larger angular movements in horizontal direction are indicative of these emotions. According to the figure, 'Micro Saccade Peak Velocity' is not indicative of these emotions. Additionally, a decrease in 'Micro-Saccade Direction'—which represents the most common orientation or path of a micro-saccade—also positively impacts the model, indicating the specific orientation of micro-saccades is relevant in contexts of sadness or depression.

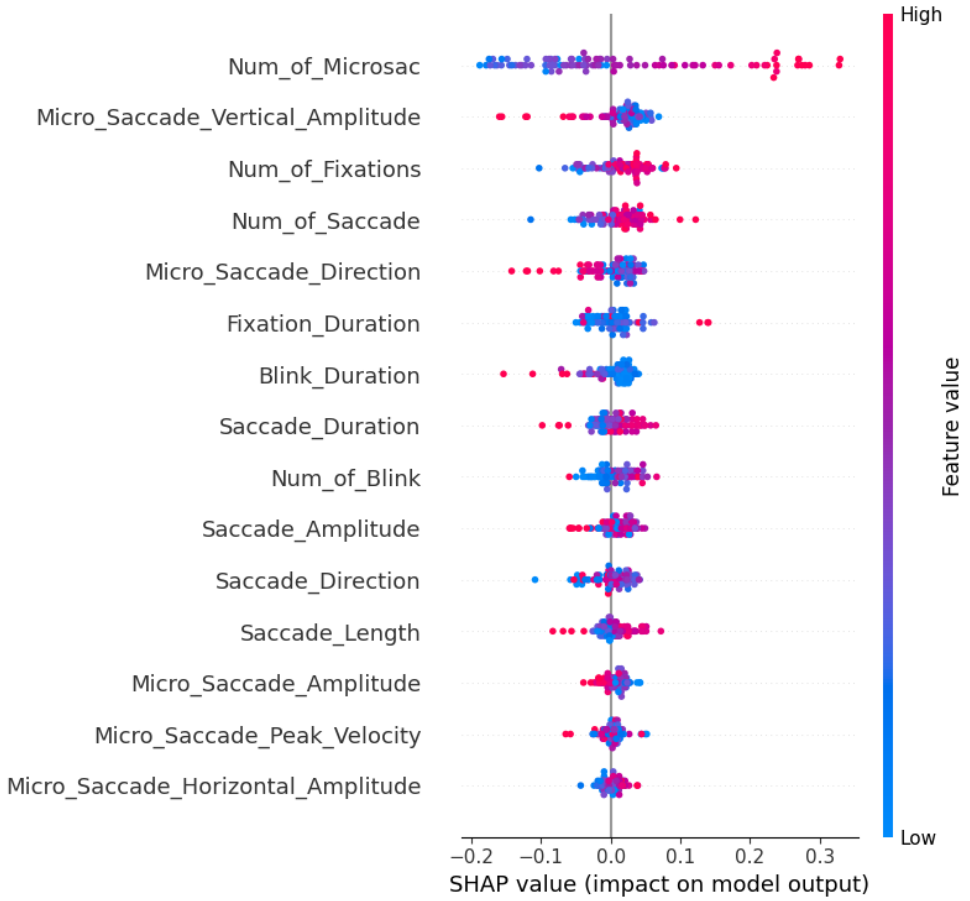

Figure 12. Summary Plot for Quadrant 2 for 15 features.

Moreover, an increase in the 'Number of Fixations' has a positive effect on model predictions, while lower values of 'Fixation Duration' do not specifically affect the model, indicating that individuals experiencing sadness or depression may engage in more frequent fixations, even if not necessarily for longer periods.

Observations from Figure 12 further indicate that higher counts of 'Number of Saccade, 'Saccade Duration', and 'Saccade Length' affect the decisions of the model positively, supporting the hypothesis that longer saccade durations are linked to decreased engagement with the environment, which is typical in states of depression, sadness, or boredom.

Conversely, lower values for 'Blink Duration' positively influence the model, suggesting that blinking frequency and duration decrease in these emotional states. However, a decrease in the 'Number of Blink' adversely affects the model, implying that a higher blinking rate is linked to negative emotional states.

4.4.4. Results for Quadrant 3

In Quadrant 3, marked by high arousal and negative valence related to emotions such as anger, fear, and disgust, 'Blink Duration' and 'Number of Blink' are the most

significant features influencing the model's predictions. In Figure 13, it is observed that higher values of these features negatively affect the model's predictions, suggesting that individuals tend to blink more frequently and for longer periods when experiencing strong negative emotions.

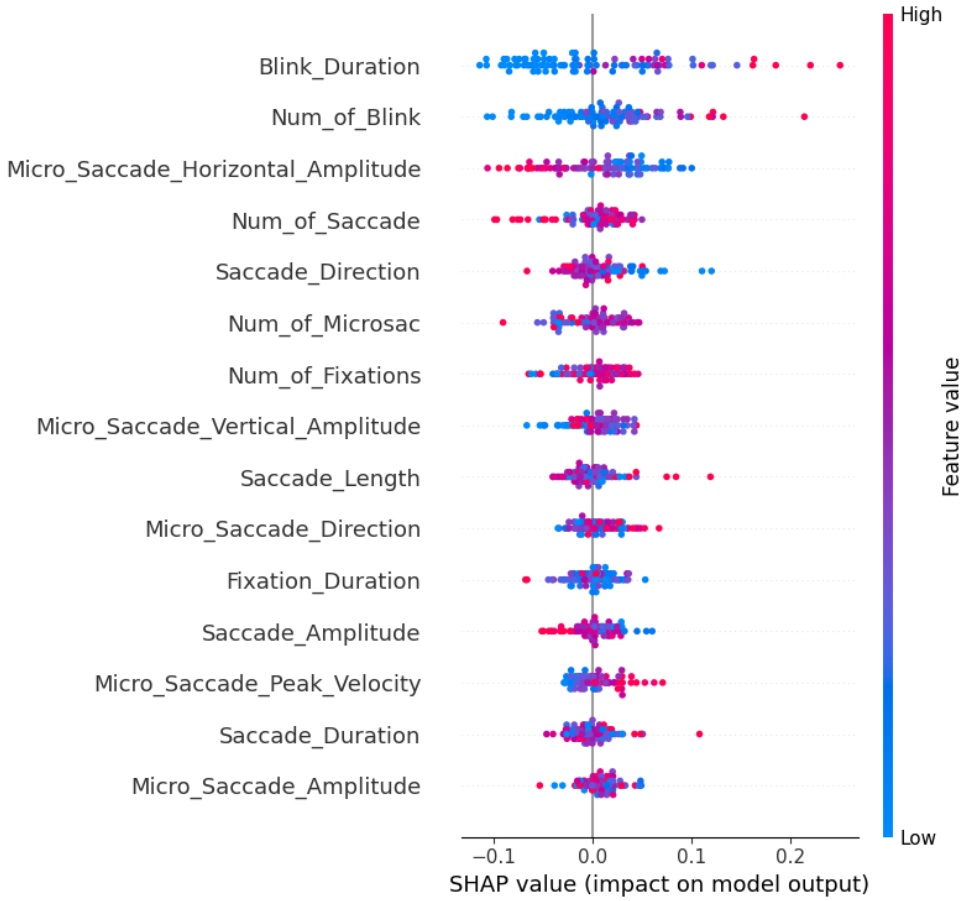

Figure 13. Summary Plot for Quadrant 3 for 15 features.

Additionally, an increased number of 'Number of Fixations' generally has a positive impact on the model. However, the values for 'Fixation Duration' can influence the model in both positive and negative ways, making it difficult to draw definitive conclusions about the emotions being represented.

Lower values of 'Micro-Saccade Horizontal Amplitude' are indicative of higher SHAP values, meaning individuals' eyes travel a shorter distance in the horizontal direction when they experience an emotional state related to Quadrant 3. Higher values of 'Micro-Saccade Peak Velocity' and 'Micro-Saccade Direction' have a positive impact on the model's decisions, indicative of higher speed values and movements in large angels during micro-saccades. This aligns with emotions such as anger, fear, and disgust, as these emotions often involve heightened arousal, which can lead to more pronounced and faster eye movements.

Concerning saccadic movements, lower values of 'Saccade Amplitude' and 'Saccade Direction' impact the model positively, while 'Number of Saccade', 'Saccade Length', and 'Saccade Duration' have a mixed impact, indicating they can be directly indicative of a specific emotion within Quadrant 3.

All these evaluations generally align with the scenario in which all 49 features were utilized for the analysis. This alignment of results between the full and reduced feature sets validates the selected features' importance.

4.4.5. Key Findings of SHAP Analysis for Reduced Features

Table 8 illustrates the variations in F1-scores across all quadrants as the number of features changes. Employing these 15 reduced features-Number of Micro-Saccade, Number of Blink, Number of Saccade, Number of Fixations, Blink Duration, Fixation Duration, Saccade Duration, Saccade Direction, Saccade Amplitude, Saccade Length, Micro-Saccade Amplitude, Micro Saccade Direction, Micro Saccade Peak Velocity, Micro Saccade Vertical Amplitude, and Micro Saccade Horizontal Amplitude- yields F1-scores of 0.9148, 0.8186, 0.8806, and 0.8504 for Quadrants 0, 1, 2, and 3, respectively. However, optimizing the model to prioritize features with greater importance values may lead to further improvements in performance. Table 8 indicates that Quadrant 0 achieves 0.9254 with 13 features, Quadrant 1 reaches 0.8399 with 10 features, Quadrant 2 matches its F1-score with just 5 features and surpasses it with 6, and Quadrant 3 achieves 0.8721 with 10 features. This examination reveals that a reduced feature set can maintain high performance while enhancing computational efficiency.

Table 8. F1-scores across different numbers of features using reduced dataset.

Feature Number	F1-Score for Q0	F1-Score for Q1	F1-Score for Q2	F1-Score for Q3
2	0.7765	0.7127	0.7009	0.6906
3	0.7763	0.7976	0.7552	0.6479
4	0.9041	0.7967	0.7445	0.7436
5	0.8935	0.7763	0.8186	0.7847
6	0.8828	0.7745	0.8818	0.7530
7	0.8936	0.7436	0.8714	0.7967
8	0.8936	0.8181	0.9139	0.8402
9	0.9042	0.8181	0.8918	0.8186
10	0.9042	0.8399	0.8818	0.8721
11	0.9777	0.8399	0.9033	0.8404
12	0.9148	0.8399	0.9023	0.8826
13	0.9254	0.8290	0.9023	0.9041
14	0.9148	0.8186	0.8813	0.8613
15	0.9148	0.8186	0.88066	0.8504

Key findings for four quadrants after SHAP analysis, including reduced 15 features, can be seen in Table 9, referring to the attributes that exhibit the greatest importance according to summary plots. The table 9 shows that 'Number of Micro-Saccade' is the most significantly impacting feature for the first three quadrants, whereas features related to blinking are crucial for Quadrant 3. Both Tables 6 and 9 infer that saccadic eye movements and blinking are the most significant main features according to SHAP.

Understanding these features can significantly enhance applications in affective computing and human-computer interaction. By identifying key eye-tracking features that correlate with specific emotional states, we can develop more responsive and adaptive user interfaces that adjust in real-time to a user's emotional state. For instance, in VR environments, knowing that certain eye movement patterns indicate stress or relaxation can allow the system to modify the virtual experience accordingly, improving user comfort and engagement. Additionally, in educational technologies, detecting frustration or confusion through eye-tracking can trigger adaptive learning systems to provide additional support or alter instructional strategies, thereby enhancing learning outcomes.

Table 9. Key findings from the SHAP analysis for reduced features.

Quadrant	Key Features	Influence on Model Predictions
Quadrant 0	Number of Micro-Saccade	Lower counts of 'Number of Micro-Saccade' generally impact predictions in a positive way. This aligns with descreased micro-saccadic movements, correlating with high arousal/positive valence emotions (e.g., happiness, excitement).
	Number of Blink	Lower values of 'Number of Blink' also positively impact the model predictions, indicating decreased blinking rate.
Quadrant 1	Number of Micro-Saccade	Lower 'Number of Micro-Saccade' values impact the predictions positively for low arousal/positive valence emotions (e.g., relaxation), indicating less rapid eye movements.
	Saccade Amplitude	Longer saccades may indicate a state of calmness or relaxation.
Quadrant 2	Number of Micro-Saccade	Higher values of 'Number of Micro-Saccade' is a significant predictor for low arousal/negative valence emotions (e.g., sadness and depression).
	Micro-Saccade Vertical Amplitude	Higher counts of micro-saccades and smaller vertical movements are linked to low arousal/negative valence emotions.
Quadrant 3	Blink Duration	Lower values of 'Max Micro-Saccade Direction' impact the model positively, demonstrating high arousal/negative valence emotions (e.g., stress).
	Number of Blink	'Blink Duration' and 'Number of Blink' are key indicators of high arousal/negative valence emotions (e.g., anger, fear, stress). Frequent and longer blinks correlate with strong negative emotions.

4.5. Lime Analysis across Reduced Features

In this section, we presented the outcomes of applying the LIME algorithm to our reduced dataset, which consists of 15 features. These features are categorized into groups related to blinks, saccades, micro-saccades, and fixations. Like our application of LIME to the full dataset, we noted that LIME's local explanations can vary considerably between observations, highlighting the variability in feature importance depending on the instance being examined.

Using an aggregation method similar to the one applied to the full dataset, Table 10 demonstrates the principal features identified as the most significant that influence the model's decisions for each quadrant by the local explanations of LIME. A comparative analysis with the SHAP results, as depicted in Table 9, reveals a consistent pattern in the influential features for Quadrants 0 and 2. Similarly, 'Number of Saccade' is identified as the most significant feature for Quadrant 1, while 'Blink Duration' is the predominant factor in Quadrant 3, according to both LIME and SHAP analyses. These findings indicate a substantial agreement between the SHAP and LIME algorithms regarding the key features impacting the model's decisions.

Furthermore, Table 10 underscores the significance of saccadic eye movements, particularly the 'Number of Micro-Saccade', in influencing the model's decision-making process. This observation is corroborated by the SHAP analysis conducted on the reduced feature set, highlighting the consistent and pivotal role these eye movements play in influencing predictive outcomes.

It is important to note that LIME, as a local explanation algorithm, provides insights into the model's decision-making process for specific instances rather than offering a comprehensive global perspective. Consequently, while the results discussed above indicate a substantial agreement between LIME and SHAP regarding the influential features, the

localized nature of LIME's explanations may not fully capture the overall behavior of the model.

Table 10. Key findings from the LIME Analysis for reduced features.

Quadrant	Key Features	Influence on Model Predictions
Quadrant 0	Number of Micro-Saccade Number of Blink	These two features are the most frequetly occurring features identified as the most significant on model's predictions by local explanations of LIME. Depending on the specific observation, they are indicative for high arousal/positive valence emotions (e.g., happiness, excitement).
Quadrant 1	Number of Micro-Saccade Micro-Saccade Horizontal Amplitude	Accoding to local explanations of LIME, they are the most frequently occurring features for low arousal/positive valence emotions (e.g., relaxation). Their impact can be positive or negative on model's decisions depending on the specific observation.
Quadrant 2	Number of Micro-Saccade Micro-Saccade Vertical Amplitude	These two feature are significant predictors for low arousal/negative valence emotions (e.g., sadness and depression) by LIME. Their impact can be positive or negative on model's predictions depending on the specific observation.
Quadrant 3	Blink Duration Micro-Saccade Horizontal Amplitude	'Blink Duration' and 'Micro-Saccade Horizontal Amplitude' are the most frequently occurring features identified the as most significant on model's decisions by LIME. Depending on the specific instances of the dataset, they are indicative for high arousal/negative valence emotions (e.g., stress).

5. Conclusions

The primary objective of this study was to identify the critical features within eye tracking data that significantly influence predictive modeling in the context of emotional state detection, utilizing XAI methodologies such as SHAP and LIME. The empirical findings from our analysis have highlighted the robustness of the SHAP algorithm in pinpointing pivotal features that considerably affect the model's predictions on a global scale. This capability allows for the development of more streamlined models that not only maintain high accuracy—as evidenced by heightened F1-scores—but also benefit from a reduction in complexity. Concurrently, aggregating LIME's local explanations to derive global insights yields results that substantially align with SHAP.

This paper significantly contributes to the ongoing discourse in XAI research, particularly demonstrating how models with a reduced set of features can mitigate overfitting risks and enhance model generalizability. By focusing solely on the most impactful features, these models not only become easier to interpret and understand but also promote greater transparency. This enhanced clarity facilitates deeper trust and accountability from users towards the model, as stakeholders can more readily grasp how decisions are being made. Additionally, the reduced computational demand results in lower operational costs, contributing to more efficient use of resources.

Another yet crucial finding of this study relates to the specific eye tracking features—such as blinks, saccades, micro-saccades, and fixations—that have shown significant predictive power regarding the emotional states of individuals. These features provide profound insights into cognitive processes and emotional reactions, underscoring their importance in the development of affective computing systems. For instance, variations in blink duration and saccadic movements were linked to different emotional states across the Circumplex Model of Affect (CMA) quadrants, offering an understanding of how such physiological responses correlate with emotional experiences.

Our findings indicate that efficient feature selection not only preserves but can enhance the predictive performance of models. For example, reducing the number of features to 9 or 19, from an initial set of 49, retained high levels of accuracy, thereby underscoring the effectiveness of targeted feature selection in XAI implementations. Furthermore, our analysis reveals that saccadic eye movements and blinking are the most crucial main eye features.

In addition to these findings, it is important to acknowledge that human data is inherently subject to various biases and pre-existing emotional states. Such biases can stem from individual differences in emotional baseline levels, past experiences, and personal predispositions, all of which can influence eye tracking metrics. For example, an individual's current mood or stress level might affect their saccadic movements or blink rate, potentially introducing variability in the data. Recognizing and accounting for these biases is essential for developing more accurate and generalizable predictive models.

Finally, this study not only advances the field of affective computing by integrating eye tracking metrics with advanced XAI techniques but also sets a benchmark for developing cost-effective, accurate, and user-trustworthy predictive models. Future research on this topic could expand this study to include multi-class classification across the four quadrants of CMA, offering further improvements in model performance and interpretability. Moreover, extending the dataset with additional physiological signals, namely, ECG and GSR data, could provide a more holistic view of the emotional landscape and improve the performance of emotion detection models.

Author Contributions: Conceptualization, M.B. and M.Y.; formal analysis, M.B.; investigation, M.B.; data curation, M.Y.; methodology, M.B. and M.Y.; project administration, H.E.I.; resources, H.E.I.; software, H.E.I.; supervision, M.Y.; validation, H.E.I.; visualization, M.B.; writing—original draft, M.B.; writing—review and editing, M.Y. and H.E.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset utilized in this study were sourced from the publicly available VREED (VR Eyes: Emotions Dataset) at https://www.kaggle.com/datasets/lumaatabbaa/vr-eyes-emotions-dataset-vreed (accessed on 1 July 2024). The relevant codes for this research are available upon request.

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

Abbreviations

The following abbreviations are used in this manuscript:

VREED	VR Eyes: Emotions Dataset
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model Agnostic Explanations
LRP	Layer-wise Relevance Propagation
IntGrad	Integrated Gradients
SKT	Skin temperature
BVP	Blood volume pulse
WPS	Wrist pulse signal
EEG	Electroencephalogram
ECG	Electrocardiogram
EMG	Electromyogram
GSR	Galvanic skin response
VR	Virtual reality
CMA	Circumplex Model of Affect
CNN	Convolutional Neural Network

SVM	Support Vector Machine
KNN	K-Nearest Neighbors
RF	Random Forest Classifier
DT	Decision Tree Classifier
LSTM	Long Short-Term Memory
GBM	Gradient Boosting Machines
LightGBM	Light Gradient Boosting
ET	Extra Trees Classifier
QDA	Quadratic Discriminant Analysis
LDA	Linear Discriminant Analysis
XGBoost	Extreme Gradient Boosting
GBC	Gradient Boosting Classifier
LR	Logistic Regression
NB	Naive Bayes
SEED	SJTU Emotion EEG Dataset
DEAP	Dataset for Emotion Analysis using Physiological Signals
XAI	eXplainable Artificial Intelligence
ML	Machine Learning
SAM	Self-Assessment Manikin
VAS	Visual Analog Scale
VE	Visual Environment
SMOTE	Synthetic Minority Over-sampling Technique

References

1. Scherer, K.R. Toward a dynamic theory of emotion: The component process model of affective states. *Geneva Stud. Emot. Commun.* **1987**, *1*, 1–98.
2. Rodríguez, L.F.; Ramos, F. Development of computational models of emotions for autonomous agents: A review. *Cogn. Comput.* **2014**, *6*, 351–375. [CrossRef]
3. Xu, H.; Plataniotis, K.N. Affect recognition using EEG signal. In Proceedings of the 2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP), Banff, AB, Canada, 17–19 September 2012 ; IEEE: Piscataway, NJ, USA, 2012; pp. 299–304.
4. Hermanis, A.; Cacurs, R.; Nesenbergs, K.; Greitans, M.; Syundyukov, E.; Selavo, L. Wearable Sensor System for Human Biomechanics Monitoring. In Proceedings of the EWSN, Graz, Austria, 15–17 February 2016; pp. 247–248.
5. Chen, L.l.; Zhao, Y.; Ye, P.f.; Zhang, J.; Zou, J.z. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst. Appl.* **2017**, *85*, 279–291. [CrossRef]
6. Krithika, L.; Venkatesh, K.; Rathore, S.; Kumar, M.H. Facial recognition in education system. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Kunming, China, 27–29 October 2017; IOP Publishing: Bristol, UK, 2017; Volume 263, p. 042021.
7. Yadava, M.; Kumar, P.; Saini, R.; Roy, P.P.; Prosad Dogra, D. Analysis of EEG signals and its application to neuromarketing. *Multimed. Tools Appl.* **2017**, *76*, 19087–19111. [CrossRef]
8. Burkart, N.; Huber, M.F. A survey on the explainability of supervised machine learning. *J. Artif. Intell. Res.* **2021**, *70*, 245–317. [CrossRef]
9. Matrenin, P.V.; Gamaley, V.V.; Khalyasmaa, A.I.; Stepanova, A.I. Solar Irradiance Forecasting with Natural Language Processing of Cloud Observations and Interpretation of Results with Modified Shapley Additive Explanations. *Algorithms* **2024**, *17*, 150. [CrossRef]
10. Ahmed, I.; Jeon, G.; Piccialli, F. From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where. *IEEE Trans. Ind. Inform.* **2022**, *18*, 5031–5042. [CrossRef]
11. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–10.
12. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
13. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *Acm Comput. Surv. (CSUR)* **2018**, *51*, 1–42. [CrossRef]
14. Tabbaa, L.; Searle, R.; Bafti, S.M.; Hossain, M.M.; Intarasisrisawat, J.; Glancy, M.; Ang, C.S. Vreed: Virtual reality emotion recognition dataset using eye tracking & physiological measures. *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–20.
15. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
16. Arya, R.; Singh, J.; Kumar, A. A survey of multidisciplinary domains contributing to affective computing. *Comput. Sci. Rev.* **2021**, *40*, 100399. [CrossRef]

17. Marín-Morales, J.; Higuera-Trujillo, J.L.; Greco, A.; Guixeres, J.; Llinares, C.; Scilingo, E.P.; Alcañiz, M.; Valenza, G. Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. *Sci. Rep.* **2018**, *8*, 13657. [CrossRef] [PubMed]
18. Saffaryazdi, N.; Wasim, S.T.; Dileep, K.; Nia, A.F.; Nanayakkara, S.; Broadbent, E.; Billinghurst, M. Using facial micro-expressions in combination with EEG and physiological signals for emotion recognition. *Front. Psychol.* **2022**, *13*, 864047. [CrossRef] [PubMed]
19. Keshari, T.; Palaniswamy, S. Emotion recognition using feature-level fusion of facial expressions and body gestures. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1184–1189.
20. Li, W.; Xue, J.; Tan, R.; Wang, C.; Deng, Z.; Li, S.; Guo, G.; Cao, D. Global-local-feature-fused driver speech emotion detection for intelligent cockpit in automated driving. *IEEE Trans. Intell. Veh.* **2023**, *8*, 2684–2697. [CrossRef]
21. Wu, Q.; Dey, N.; Shi, F.; Crespo, R.G.; Sherratt, R.S. Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks. *Appl. Soft Comput.* **2021**, *110*, 107752. [CrossRef]
22. Somarathna, R.; Bednarz, T.; Mohammadi, G. An exploratory analysis of interactive VR-based framework for multi-componential analysis of emotion. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops), Pisa, Italy, 21–25 March 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 353–358.
23. Mattern, E.; Jackson, R.R.; Doshmanziari, R.; Dewitte, M.; Varagnolo, D.; Knorn, S. Emotion Recognition from Physiological Signals Collected with a Wrist Device and Emotional Recall. *Bioengineering* **2023**, *10*, 1308. [CrossRef] [PubMed]
24. Garg, N.; Kumar, A.; Ryait, H. Analysis of wrist pulse signal: Emotions and physical pain. *IRBM* **2022**, *43*, 391–404. [CrossRef]
25. Goshvarpour, A.; Goshvarpour, A. Innovative Poincare's plot asymmetry descriptors for EEG emotion recognition. *Cogn. Neurodyn.* **2022**, *16*, 545–559. [CrossRef]
26. Zhang, J.; Yuan, G.; Lu, H.; Liu, G. Recognition of the impulse of love at first sight based on electrocardiograph signal. *Comput. Intell. Neurosci.* **2021**, *2021*, 6631616. [CrossRef]
27. Mateos-García, N.; Gil-González, A.B.; Luis-Reboredo, A.; Pérez-Lancho, B. Driver Stress Detection from Physiological Signals by Virtual Reality Simulator. *Electronics* **2023**, *12*, 2179. [CrossRef]
28. Goshvarpour, A.; Goshvarpour, A. Novel high-dimensional phase space features for EEG emotion recognition. *Signal Image Video Process.* **2023**, *17*, 417–425. [CrossRef]
29. Siqueira, E.S.; Fleury, M.C.; Lamar, M.V.; Drachen, A.; Castanho, C.D.; Jacobi, R.P. An automated approach to estimate player experience in game events from psychophysiological data. *Multimed. Tools Appl.* **2023**, *82*, 19189–19220. [CrossRef]
30. Sheykhivand, S.; Mousavi, Z.; Rezaii, T.Y.; Farzamnia, A. Recognizing emotions evoked by music using CNN-LSTM networks on EEG signals. *IEEE Access* **2020**, *8*, 139332–139345. [CrossRef]
31. Meuleman, B.; Rudrauf, D. Induction and profiling of strong multi-componential emotions in virtual reality. *IEEE Trans. Affect. Comput.* **2018**, *12*, 189–202. [CrossRef]
32. Somarathna, R.; Bednarz, T.; Mohammadi, G. Virtual reality for emotion elicitation—A review. *IEEE Trans. Affect. Comput.* **2022**, *14*, 2626–2645. [CrossRef]
33. Sacharin, V.; Schlegel, K.; Scherer, K.R. *Geneva Emotion Wheel Rating Study*; NCCR Affective Sciences ; Center for Person, Kommunikation, Aalborg University: Aalborg, Denmark, 2012.
34. Alexandros, L.; Michalis, X. The physiological measurements as a critical indicator in users' experience evaluation. In Proceedings of the 17th Panhellenic Conference on Informatics, Thessaloniki, Greece, 9–21 September 2013; pp. 258–263.
35. Zheng, W.L.; Lu, B.L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]
36. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]
37. Miyamoto, K.; Tanaka, H.; Nakamura, S. Applying Meta-Learning and Iso Principle for Development of EEG-Based Emotion Induction System. *Front. Digit. Health* **2022**, *4*, 873822. [CrossRef]
38. Metta, C.; Beretta, A.; Guidotti, R.; Yin, Y.; Gallinari, P.; Rinzivillo, S.; Giannotti, F. Advancing Dermatological Diagnostics: Interpretable AI for Enhanced Skin Lesion Classification. *Diagnostics* **2024**, *14*, 753. [CrossRef]
39. Patil, A.; Patil, M. A Comprehensive Review on Explainable AI Techniques, Challenges, and Future Scope. In Proceedings of the International Conference on Intelligent Computing and Networking, Mumbai, India, 24–25 February 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 517–529.
40. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **2015**, *10*, e0130140. [CrossRef] [PubMed]
41. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In Proceedings of the International Conference on Machine Learning (PMLR), Sydney, Australia, 6–11 August 2017; pp. 3319–3328.
42. Zhang, C.; Su, L.; Li, S.; Fu, Y. Differential Brain Activation for Four Emotions in VR-2D and VR-3D Modes. *Brain Sci.* **2024**, *14*, 326. [CrossRef] [PubMed]
43. Khalane, A.; Makwana, R.; Shaikh, T.; Ullah, A. Evaluating significant features in context-aware multimodal emotion recognition with XAI methods. *Expert Syst.* **2023**, e13403.

44. Torres, J.M.M.; Medina-DeVilliers, S.; Clarkson, T.; Lerner, M.D.; Riccardi, G. Evaluation of interpretability for deep learning algorithms in EEG emotion recognition: A case study in autism. *Artif. Intell. Med.* **2023**, *143*, 102545. [CrossRef] [PubMed]
45. Kim, T.W.; Kwak, K.C. Speech Emotion Recognition Using Deep Learning Transfer Models and Explainable Techniques. *Appl. Sci.* **2024**, *14*, 1553. [CrossRef]
46. Liew, W.S.; Loo, C.K.; Wermter, S. Emotion recognition using explainable genetically optimized fuzzy ART ensembles. *IEEE Access* **2021**, *9*, 61513–61531. [CrossRef]
47. Zhao, K.; Xu, D.; He, K.; Peng, G. Interpretable emotion classification using multi-domain feature of EEG signals. *IEEE Sens. J.* **2023**, *23*, 11879–11891. [CrossRef]
48. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [CrossRef]
49. Hawker, G.A.; Mian, S.; Kendzerska, T.; French, M. Measures of adult pain: Visual analog scale for pain (vas pain), numeric rating scale for pain (nrs pain), mcgill pain questionnaire (mpq), short-form mcgill pain questionnaire (sf-mpq), chronic pain grade scale (cpgs), short form-36 bodily pain scale (sf-36 bps), and measure of intermittent and constant osteoarthritis pain (icoap). *Arthritis Care Res.* **2011**, *63*, S240–S252.
50. Sogo, H. GazeParser: An open-source and multiplatform library for low-cost eye tracking and analysis. *Behav. Res. Methods* **2013**, *45*, 684–695. [CrossRef]
51. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
52. Ali, M. *PyCaret: An Open Source, Low-Code Machine Learning Library in Python*, PyCaret Version 2 ; PyCaret: San Francisco, CA, USA, 2020.
53. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]

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.