EXAMINING THE ROLE OF EYE MOVEMENTS IN FACIAL EXPRESSION AND EMOTIONAL INTENSITY RECOGNITION VIA MACHINE LEARNING

DİLARA DENİZ TÜRK

JUNE, 2022

EXAMINING THE ROLE OF EYE MOVEMENTS IN FACIAL EXPRESSION

AND EMOTIONAL INTENSITY

RECOGNITION VIA MACHINE LEARNING

BY

DİLARA DENİZ TÜRK

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF MASTERS OF ARTS

IN

COGNITIVE SCIENCE

YEDITEPE UNIVERSITY

JUNE, 2022

PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results are not original to this work.

Date: 06.06.2022

Name/Surname: Dilara Deniz Türk

ABSTRACT

Recognition and classification of emotional face expressions is one of the most advanced abilities of humans. Many studies have investigated recognition processes and scanning strategies for different facial expressions but there is still no consensus on a particular scanning pattern for emotion types. Most of the previous studies also did not include the level of emotional intensity of facial expressions. The aim of this study is to investigate which eye movement parameters are important for emotional expression recognition and emotional intensity recognition processes. Eye movements provide important clues about face scanning strategies. Therefore, we examined the eye movements of 60 participants and fed machine learning models with eye tracking data. Neutral, surprised, sad, happy, and fearful facial expressions were presented to the participants with 0%, 25%, %50, %75, and %100 emotional intensity. Half of the participants were asked to classify the emotion type and half of them were asked to rate the intensity level of the emotion. Total number of fixations, pupil size and number of fixations on areas of interest (AOIs: eyes, nose and mouth) were collected for eye tracking analysis. Results showed that the fixation number on facial expressions decreased as the intensity of the expression increased. It was also found that participants fixated most to the eyes and least to the mouth regardless of the task. There was no difference between fixation numbers to eyes and nose during intensity recognition tasks for all emotion types. However, the fixation number towards the eye, nose and mouth differed between emotion types during the emotion recognition task. Higher emotional intensity resulted in less fixations in the mouth area. Analysis of pupillary response showed that fearful faces resulted in larger pupil size compared to all other expressions during the emotion recognition task but this difference was not evident during the intensity recognition task. Data collected from the eye tracking device has also been implemented in Random Forest (RF), Decision Tree and k-Nearest Neighbors (k-NN) models. All models achieved an accuracy rate at above chance level for both emotion classification and intensity classification. The highest accuracy rate was achieved with the RF method. The model was able to classify emotion types with an average accuracy of 85.8% for emotion recognition task and with an average accuracy of 89.2% for intensity recognition task when implemented with fixation point of gaze, fixation duration and pupil size. Emotion and intensity classification remained similar for the intensity recognition task with the same parameters. Effect of task, emotion type and intensity level were observed for pupil size and fixation numbers in the eye tracking analysis. However, the task did not affect the classification performance of the model. Also, intensity level of expressions or emotion type of an intensity level did not affect machine's performance as well. We interpreted these results as scanning strategies implemented by viewers while reading emotional expressions with different intensities. Results of the current study further discussed possible application areas.

Keywords: emotional intensity, emotion recognition, eye tracking, facial expression, machine learning

ÖZET

Duygusal yüz ifadelerini tanıma ve sınıflandırma, insanların en gelişmiş yeteneklerinden biridir. Birçok çalışma, farklı yüz ifadeleri için tanıma süreçlerini ve inceleme stratejilerini araştırmıştır, ancak duygu türleri için belirli bir inceleme stratejisi üzerinde hala bir fikir birliğine varılamamıştır. Önceki çalışmaların çoğu, yüz ifadelerinin duygusal yoğunluğunun seviyesini de içermemektedir. Bu çalışmanın amacı, duygusal ifadeleri tanıma ve duygu yoğunluğu tanıma süreçlerinde hangi göz hareketi parametrelerinin önemli olduğunu araştırmaktır. Göz hareketleri yüz tarama stratejileri hakkında önemli ipuçları sağlar. Bu nedenle, 60 katılımcının göz hareketlerini ve göz takip verileriyle beslenmiş makine öğrenimi modellerini inceledik. Katılımcılara %0, %25, %50, %75 ve %100 duygu yoğunluğu ile nötr, şaşırmış, üzgün, mutlu ve korkulu yüz ifadeleri gösterildi. Katılımcıların yarısından duygu türünü sınıflandırmaları, diğer yarısından da duygunun yoğunluk derecesini derecelendirmeleri istenmiştir. Göz takip analizi için toplam odaklanma sayısı, göz bebeği boyutu ve ilgi alanlarındaki odaklanma sayısı (ilgi alanları: gözler, burun ve ağız) verileri toplandı. Sonuçlar, ifade yoğunluğu arttıkça yüz ifadelerindeki odaklanma sayısının azaldığını gösterdi. Katılımcıların görevden bağımsız olarak en çok göze ve en az ağıza odaklandıkları da bulundu. Tüm duygu türleri için yoğunluk tanıma görevleri sırasında gözlere ve buruna odaklanma sayıları arasında fark yoktu. Ancak duygu tanıma işlemi sırasında duygu türleri arasında göze, buruna ve ağza odaklanma sayısı farklılık gösterdi. Daha yüksek duygusal yoğunluk, ağız bölgesinde daha az odaklanma ile sonuçlandı. Göz bebeği boyutunun analizi, korkulu yüzlerin, duygu tanıma görevi sırasında diğer tüm ifadelere kıyasla daha büyük göz bebeği boyutuyla sonuçlandığını gösterdi. Ancak bu fark, yoğunluk tanıma görevinde belirgin değildi. Göz takip cihazından toplanan veriler, Rastgele Orman, Karar Ağacı ve k-En Yakın Komşular (k-NN) algoritmalarına beslenmiştir. Tüm modeller, hem duygu sınıflandırması hem de yoğunluk sınıflandırması için şans seviyesinin üzerinde bir doğruluk oranı elde etti. En yüksek doğruluk oranı Rastgele Orman yöntemi ile elde edilmiştir. Model, odaklanma noktası, odaklanma süresi ve göz bebeği boyutu ile beslendiğinde, duygu tanıma görevi için ortalama %85,8 ve yoğunluk tanıma görevi için ortalama %89,2 doğrulukla duygu türlerini sınıflandırabildi. Aynı parametrelerle yoğunluk tanıma görevi için duygu ve yoğunluk sınıflandırması benzer kaldı. Göz takip analizinde göz bebeği boyutu ve odaklanma sayıları için görevin, duygu türünün ve yoğunluk seviyesinin etkisi gözlemlendi. Ancak görev, modelin sınıflandırma performansını etkilemedi. Ayrıca, ifadelerin yoğunluk seviyesi veya bir yoğunluk seviyesinin duygu türü de makinenin performansını etkilemedi. Bu sonuçları, insanların değişen yoğunluk seviyelerinde duygusal ifadeleri okurken belirli inceleme stratejileri uyguladıkları şeklinde yorumladık. Mevcut çalışmanın sonuçları, olası uygulama alanları ile daha fazla tartışılmıştır.

Anahtar kelimeler: duygusal yoğunluk, duygu tanıma, göz takibi, makine öğrenimi, yüz ifadesi

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor Dr. Funda Yıldırım who guided and encouraged me not only during my thesis but throughout my entire master's degree.

I would like to thank all the CNVP Lab members who have been more helpful than I could have imagined and created an inspiring and loving lab environment.

I would like to thank my father Olgun Türk, my mother Sevim Türk, and my sister Asude Selin Türk. I am very lucky to have parents who stand with me in every decision I make, who always encourage me to go one step further and who never stop supporting me in every way. I also feel very lucky to have a big sister who is also my best friend and who is always there when I face indecisions or when I lose my courage.

I would like to thank my friends, my petits-beurre, who tirelessly participated in all my experiments and supported me at all times. I would also like to express my special gratitude to Fuat Şahin who relentlessly helped me through all the troubles I faced during my thesis.

Finally, I would like to thank my cats Hera and Nemo for staying by my side during the long nights I spent in front of my computer.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Faces are invaluable sources of information that help us operate our social lives. They provide us with clues about the mental and internal state of others, as faces usually carry emotional information (Ekman, 1993; Torre & Cohn, 2011). Ekman (1971) suggested that emotional information on faces is universal and persistent across different cultures and we communicate through these facial expressions which are reflected with simple muscle movements in a very plain but powerful manner. As a result of this, a slight upward pull of lip corners on a person's lips indicates that person is having happy feelings or a minor frown on the eyebrows tells us that the person feels angry about something (Calder et al., 2001). In fact, we have become experts in encoding the information carried by human faces. We identify, categorize and react to emotional faces very quickly and accordingly. Whether the facial expression is positive or negative, whether it is a strong or a weak expression are rapidly recognized. This expertise might have developed due to increasing the chances of survival with an avoid/approach behavior from the evolutionary perspective and it enables us to behave properly during social encounters and helps us 'survive' in the social world.

1.1.Emotional Face Expressions

It has been shown by several studies that stimuli with emotional content are processed faster and capture attention more compared to the stimuli without emotional value (Vuilleumier et al., 2011; Pessoa et al., 2002). Similar processing approach appears for emotional facial expressions as well. These expressions enter through our visual system quickly and we can recognize them as fast as 50 milliseconds (Calvo & Lundqvist, 2008).

This enhanced response to the emotional stimuli and emotional expressions was investigated by various studies further for understanding differences in recognition mechanisms across basic emotions. Recognition performance for happy faces is shown to be much higher compared to other emotional expressions (Calvo & Lundqvist, 2008; Calvo & Beltran, 2013; Calvo & Nummenmaa, 2008; Leppänen & Hietanen; 2004; Palermo & Coltheart, 2004). Addition to the higher performance for happy faces, a faster recognition time has been found for happy expressions compared to other expressions as well (Calvo & Beltran, 2013; Calvo & Nummenmaa, 2008; Palermo & Coltheart, 2004). This superior recognition performance towards happy faces reveals itself even when the faces are not located at our central vision (Calvo et al., 2014). In contrast to the superiority of happy faces for recognition, the poorest and slowest recognition performance was mostly found for fearful faces during emotion recognition tasks (Calvo & Lundqvist, 2008; Calvo & Nummenmaa, 2008; Palermo & Coltheart, 2004; Russel, 1994). While the literature often agrees on the best and the worst recognized facial expressions, there is not always a consensus. For example, neutral faces can be recognized as quickly and successfully as happy faces (Calvo & Beltran, 2013; Calvo & Lundqvist, 2008), but some studies continue to show the superior performance towards happy expressions (Sun et al., 2017). Also, fearful faces can share the high recognition performance for happy faces (Sun et al., 2017). Recio et al. (2013) showed not only happy expressions but also angry and surprised expressions were recognized with a higher performance compared to sad, fearful and disgusted expressions. However, indifference of recognition among angry, sad, disgusted and surprised faces was also reported (Calvo & Lundqvist, 2008). In other words, there is still no definitive conclusion on the matter of processing of emotional expressions.

1.1.1. Expression intensity

Literature on the recognition of facial expressions has mostly used the peak of the emotion being expressed which is the most intense state of an expression (Figure 1). However, facial expressions do not always convey the most extreme form of emotions in daily life. They differ in their level of intensity as much as they differ in the emotion type (Carrol & Russel, 1997). Intensity level of facial expressions affects the processing of expressions. When the emotional intensity of an expression increases, the recognition performance towards that expression increases as well (Herba et al., 2006; Montagne et al., 2007; Recio et al., 2014). Studies using six basic emotions (happy, sad, fearful, angry, surprised, disgusted) (Ekman & Friesen, 1975) varying in their intensity level found that categorization performance of emotional expressions increases for every emotion type parallel with the intensity level (Guo, 2012; Montagne et al., 2007) and categorization time decreases when level of emotional intensity increases (Guo, 2012). This intensification of emotion recognition applies on both static and dynamic faces expressing happiness, sadness, fear and anger (Torro-Alves et al., 2013). Muscle movements on faces affects the performance for emotion recognition but varies with the intensity level, specific emotion and even sex of the participant. In a previous study, we asked participants to rate how congruent the facial and vocal expressions were while presenting in/-congruent face-voice pairs in terms of emotional intensity (Türk et al., 2022). We observed that pairs at low intensity level were not rated as congruent even though they are, and congruence ratings were more realistically evaluated with high emotional intensity. We interpreted these results as expressions with low intensity were difficult for people to recognize so that they could not judge the congruency of face-voice pairs accurately.

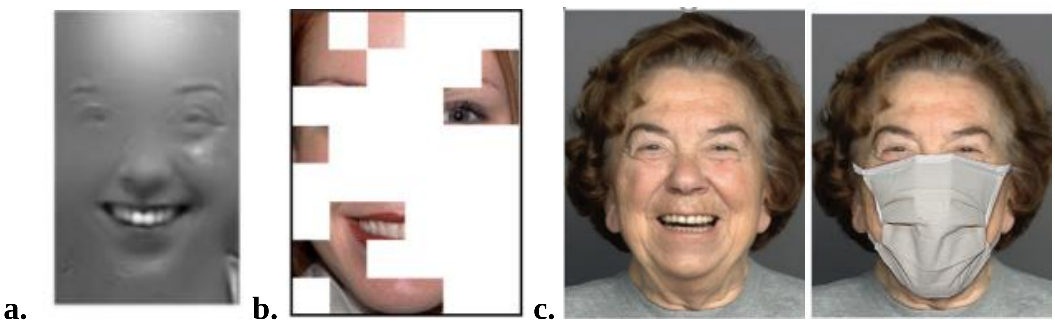
1.1.2. Reading facial expressions

Processing of emotion recognition on faces and viewing strategies vary within emotion types as well. Facial region that provides the most valuable information differs with the expression conveyed by a face. The question of which facial regions transfer the crucial information about an emotion has been investigated by several behavioral studies by using different methods (Figure 1). Calvo and Nummenmaa (2008) conducted a detailed study to answer this question. First, participants were exclusively presented with eyes or mouths with six basic emotions. They observed that only the eyes did not vary the performance of emotion recognition, and that the high performance for happy faces disappeared when only the eyes were given. On the other hand, participants recognized happy and surprised faces even when only the mouth of the expression was presented. Taking these findings, a step further, they examined emotion recognition performance by blackening the mouth and eye region on the face images. Absence of the mouth region increased the recognition time and decreased the recognition performance for all expressions except sad ones. However, the absence of the eye region did not affect the emotion recognition performance. Another study used bubble technique (Smith et al., 2005) to mask the mouth and eye region with both static and dynamic face expressions with six basic emotions and neutral and painful expressions (Blais et al., 2012) and showed the important role of the mouth region for discriminating emotions for static and dynamic face images. A more recent study used a more fine-grained method, breaking up face images into 48 tiles, revealing more of the face in a sequential order, asking participants to stop the sequence when they recognized the emotional expression on the face (Wergrzyn et al., 2017). It was observed that the participants mostly depended on the information from the mouth and eye regions, more specifically, they used the information from eyes for fearful and sad faces and the mouth for disgusted and happy faces. Our previous study examining the change detection performance for facial expressions by creating subtle manipulations on the eyes, noses and mouths of neutral, sad and angry expressions revealed a high performance but no distinguishing effect of the mouth region (Türk & Yildirim, 2022). Changes on the eye region were more easily detected on neutral and sad faces, while performance on the nose region of angry faces was higher.

Since surgical masks have been used in daily life all over the world due to Covid-19 pandemic, it has also been investigated how face masks affect viewing patterns and recognition performance for facial expressions. Use of masks makes it difficult to read emotions from faces in general (Carbon, 2020; Grahlow et al., 2021; Proverbio & Cerri, 2021), but some studies showed that angry face recognition was not affected by face masks (Proverbio & Cerri, 2021). Others found use of masks damaged the recognition performance for angry, sad and disgusted faces (Grahlow et al., 2021). These studies shed light on which emotions can be recognized with the information from the eyes alone.

Figure 1

Different techniques for facial regions and emotional expressions



Note. (a) bubble technique, (b) tile technique, (c) surgical mask.

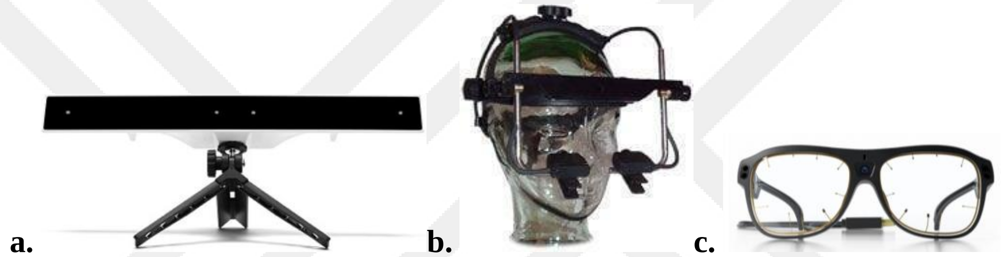
1.2. Eye Movements

1.2.1. Eye Movement Parameters

Eye tracking mainly refers to the measuring where a person's gaze is directed. Eye tracking devices allow us to understand various eye movements involving the user's visual exploration processes. There are different parameters collected from the devices. Saccades are the user's rapid eye movements to relocate their gaze. Fixation is directing the gaze to an object and reflecting that object in the fovea of the retina and eye trackers provide the duration of fixations as well. While saccades and fixations are voluntary eye movements of users for visual scanning, eye tracker devices can also measure autonomous pupillary responses. The responsiveness of pupils gives us a substantial amount of information about a person's current state. Gaze of point gives information about where we direct our attention and where we focus. Fixation duration shows how long we focus on one point. Pupil, on the other hand, is sensitive not only to light but also emotional changes, cognitive loads, defensive mechanisms and pain (Beatty, 1982, Kahneman et al., 1969).

There are various sensor devices to collect such information from the eyes. Electrooculography (EOG) measures the resting electrical potential on cornea and retina by placing electrodes around the eye (Fenn & Hursh, 1934). Videooculography collects eye movements based on video recordings either with respect to the user's head or surroundings which might be sensitive to light changes. There are also infrared oculography using infrared light reflections on pupil and cornea and use this reflection to calculate the eye movements (Duschowski, 2007). This kind of eye tracking device marks the border between pupil and iris and calculates where gaze and pupil are positioned by the reflections. Infrared light is invisible for people so do not create a distraction. Also, changes in the surrounding light affect the data much less compared to video-based tracking technologies. Eye tracking devices can be head-mounted like EyeLink II (SR Research, Ontario, Canada) and thus less sensitive to motion compared to desktop mounted eye trackers like Gazepoint GP3 (Vancouver, Canada) (Figure 2a, b). Researchers usually use chinrest with desktop mounted eye trackers to limit head movements of users. Apart from these, there are also mobile eye tracking devices that can be used like glasses which are less motion sensitive but will be more highly priced (Figure 2c).

Figure 2
Eye tracking devices



Note. (a) Desktop mounted eye tracker, Gazepoint GP3, (b) Head mounted eye tracker, Eyelink II, (c) Mobile eye tracker, Tobii Pro Glasses 3 (Stockholm, Sweden).

1.2.2. Fixation patterns and emotional expressions

Gaze patterns, number of fixations, location of fixations and fixation durations collected from eye tracking devices have been used quite frequently in order to better understand the emotional face viewing when categorizing human expressions. It has been shown by most studies that people fixate their gaze mostly around the eye and mouth region during emotion identification (Eisenbarth & Alpers, 2011; Jack et al., 2009; Scurgin et al., 2014, Vaidya et al., 2014). Some of these studies showed that these fixation patterns are not influenced by emotion types (Jack et al., 2009; Vaidya et al., 2014) and these findings were supported by the findings for both static and dynamic six basic emotions (Blais et al., 2017). It was also observed that the number of fixations around the mouth and eyes decreased only in dynamic faces. A study that wanted to examine in detail this tendency of relying the information from mouth and eye presented six basic emotional face expressions with no mouth, congruent mouth, incongruent mouth (Calvo & Fernandez-Martin, 2012). Mouth was found to be more salient than other features, as the mouth of happy faces were found to attract more attention even when matched with eyes of other emotions and damage recognition of other emotional face expressions.

1.2.2.1. Fixation patterns and type of the emotion.

Although the importance of the mouth and eye area continues to reveal itself in emotion recognition in general, contrary to studies above, there are also studies that reveal emotion-specific fixation patterns. Eisenbarth & Alpers (2011) found the eye region was looked at more for angry and sad faces, mouth region was looked at more for happy faces, and mouth and eye were equally important for neutral and fearful. Another study using dynamic expressions also observed the same fixation pattern between emotions (Calvo et al., 2018). In addition to these findings, authors showed that the fixation distribution in the mouth and eyes is balanced for both fearful and surprised facial expressions, and that the nose-cheek region is looked at more and earlier for disgusted faces. Another study investigating whether these distinguishing features used in emotion recognition are independent of the task and the location of the stimulus found that the subjects looked more at the eyes for neutral and fearful faces and more at the mouth for happy faces. They observed these findings across different stimulus locations and different experimental tasks such as emotion recognition, gender recognition, or passive viewing (Scheller et al., 2012).

1.2.2.2. Fixation patterns and intensity of the emotion.

These fixation patterns have also been investigated with facial expressions of varying emotional intensity. A study presenting an emotion classification task with six emotional facial expressions with five different intensity levels showed that the fixation number decreased as the intensity level increased, but the percentage of fixation and fixation duration towards facial features were not affected (Guo, 2012). Another study using three different intensity levels found that as the emotional intensity increased, the number of fixations on the eyes of angry faces increased, except at the 60% intensity level (Schurgin et al., 2014). Vaidya et al. (2014) examined the fixation patterns with emotional facial expressions at %20, %30, 40% (subtle) and %100 (extreme) intensity level and used a simple prediction model. While this model predicted subtle fearful, disgusted, and surprised faces from fixations on the eyes, it was not able to predict subtle happy faces or extreme expressions other than fear. These findings provided a perspective based on a predictive model on the investigation of fixation patterns for emotional expressions with varying intensity levels.

1.2.3. Pupil size and emotional expressions

It is known that pupil size shows reactivity that can be affected by difficult tasks (Kahneman da Beatty, 1966) or dilate by up to 20% with the effect of cognitive loads such as created by decision-making processes (Hess, 1972). Along with these, effect of emotional stimuli such as images (Bradley et al., 2008), sounds (Partala & Surakka, 2003) and words (Siegle et al., 2003) has also been suggested to result in heightened autonomous response (Bradley, 2009; Lang & Bradley, 2010). Many studies have shown that negatively or positively charged emotional images result in enlarged pupil size of viewers (Bradley et al., 2008; Partala & Surakka, 2003; Van

Steenbergen et al., 2011). Thus, pupil dilation was observed with emotional stimuli regardless of its valence. Van Steenbergen et al. (2011) also found an attentional selectivity towards negative pictures while increase in pupil size for both negative and positive pictures indicating that pupillary responses are autonomous and independent from motivational attentional processes.

There have been studies that take emotional stimuli beyond the valence attribute and investigate the pupillary responses towards human faces conveying specific emotional expressions (Kleberg et al., 2019; Kret et al., 2013). These studies showed different results compared to previous studies having a consensus on larger pupil size towards negative and positive emotional stimuli. For example, a study investigating the difference in pupil dilation between people with social anxiety disorder and healthy subjects found that healthy subjects showed larger pupil size for angry facial expressions than happy facial expressions (Kleberg et al., 2019). Another study examined the eye movements and pupillary responses to emotional face-emotional body and emotional body-emotional scene pairs (Kret et al., 2013). Angry, happy and neutral faces, bodies and scenes were used in the study and higher enlargement in pupil size was observed for angry face-angry body and angry body-aggressive scene pairs than other pair combinations. There have been those who suggested that these reactions are due to the activation of the flight part of the body's fight or flight response (Porges, 2003). Considering the sensitivity of pupillary responses to external factors, the effect of viewing neutral and fearful images on the pupil size was investigated more in detail by including different experimental conditions such as presentation time of the stimulus, familiarity level with the stimulus and the difficulty of the task (Snowden et al., 2016). Pupil dilation to fearful images was found to be independent of these experimental conditions. Thus, these findings showed us that findings on pupil size and emotional stimuli are consistent and explanatory even though conditions vary in each study.

Pupillary response to emotional stimuli has also been frequently studied on infants to examine the developmentally learned or acquired response to these emotionally charged stimuli. 6- and 12-months old infants were given happy, stressed and neutral infant sounds and infant pictures (Geangu et al., 2011). Both stressed and happy expressions resulted in larger pupil size, but the response to happy expressions lasted shorter. Further, Jessen et al. (2016) investigated whether this pupil dilation occurs only during conscious processing and presented the stimuli 50 milliseconds or 900 milliseconds to 7-month-old infants. They found happy faces caused a larger pupil size than fearful faces, regardless of the stimulus presentation time indicating indifference of conscious/unconscious processing. More recent study included all 6 basic emotions and observed a significant pupillary response only for happy and fearful facial expressions (Prunty et al., 2021).

1.3. Machine Learning

Machine learning emerged under the name of “intelligent machinery” and has evolved since the 1950’s and existed not just in various research fields. The main purpose of machine learning is to make computers as intelligent as humans. As the name suggests, the aim is to learn the given data. Based on machine learning models, computers find certain patterns in the data and can make classifications, predictions and detections. Thus, implementing machine learning models has been very valuable in many research areas such as computer vision for object-recognition, -detection, -classification and -processing, linguistics for natural language processing and semantic analysis and computer scientists improves the models accordingly. In fact, some computational models have become able to perform these tasks better than humans.

1.3.1. Machine learning with eye tracking data

It is now known that machine learning models are successful in detecting and understanding eye movements as well (Zemblys et al., 2018). Thus, they have also been applied frequently in the investigating, understanding and analysis of eye tracking data (Schmidhuber, 2015). When people examine their surroundings, if they do not have a specific task such as detecting an object, they direct their attention according to the level of saliency or relevance of regions in the scene, so they use certain scanning patterns (Rensink, 2000). Machine learning models have been used to detect patterns for finding the salient regions in an environment or first region of focus (Itti & Koch, 2001, Le Meur et al., 2006, Walter & Koch, 2006). Psychology, neuromarketing, engineering and even art benefit from detecting such salient features to examine people's attention mechanisms (Ducshowski, 2002). Eye tracking studies provide information on fixation location, fixation sequence and fixation duration during a scene scanning tasks and the obtained data can be fed into computational models to find similarities and make predictions about our visual processes. These scanning strategies have been valuable in the field of human-computer interaction (HCI). Especially since emotion recognition is a frequently studied field in HCI, there are many studies using different machine learning models with eye movement data which obtained while people view emotion evoking stimuli and different models were found to be successful at predicting the emotion felt by people (Alsehri & Alghowinem, 2013; Babiker et al., 2015; Lu et al., 2015; Tarnowski et al., 2020; Guo et al., 2019). Moreover, eye movement patterns provide a lot of information about people's health, general condition and cognitive processes. Computational models fed by eye movement data can also predict mild cognitive impairment that can transform into Alzheimer’s disease (Raatikainen et al., 2021), developmental dyslexia (Lagun et al., 2011; Prabha & Bhargav, 2019) or autistic traits (Lin et al., 2021). Eye movement patterns were even used for detecting drivers’ fatigue as a sensor technology in engineering (Horng et al., 2004).

1.3.2. Machine learning and emotion classification

Ekman’s (Ekman, 1992) definition of emotional expressions is based on certain cues from faces and the activation of certain facial regions to provide necessary information (Tian et al., 2005). Results of both behavioral and eye tracking studies on the recognition of emotional expressions also support this definition (Blais et al., 2012; Calvo & Nummenmaa, 2008; Calvo et al., 2018; Eisenbarth & Alpers (2011) Grahlow et al., 2021). It can be suggested that people use certain strategies when classifying facial expressions based on previous studies and computational models provide further evidence (Abdallah et al., 2020; Tavakoli et al., 2015; Vaidya et al. 2014). For example, it has been found that gaze duration and fixation frequency can predict the valence of the images (Tavakoli et al., 2015). Models fed with eye tracking data were able to classify whether the scene viewed by a person is positive, negative or neutral (Tamuly et al., 2019). Eye movements from different tasks on face images were calculated with Bayesian approach and showed that people focus on specific regions that will increase their performance for whatever task is in their hands and follow a strategy accordingly (Peterson & Eckstein, 2012). Furthermore, a recent study applied classification models on fixation time for area of interests (AOIs) on six basic expressions and classified the valence of the images (Cao & Elliott, 2021) and reached an accuracy rate of over 80%, showing that fixation patterns in AOIs are important factors for rapid recognition of emotional expressions.

1.4. Aim of the Study

Emotional facial expressions are one of the most common stimuli we encounter in daily life, and understanding the perception of emotional expressions has been broadly covered in the literature. Many behavioral and eye tracking studies investigated recognition performance for emotional expressions, effect of emotional intensity on the recognition, diagnostic features carrying crucial information that help us to recognize emotions quickly and rapidly. However, there is still no agreement about the responsible facial features or eye movements on this fast and almost automatic human ability. More than that, most of the studies neglected the fact that facial expressions can convey different emotional intensity and that the level of intensity can affect processing of facial expressions.

The aim of the current study is to explore the informative eye movement parameters while identifying the emotion or the intensity of a facial expression varying in emotional intensity. To our knowledge, there is no study examining the perceptual strategies for recognition of emotional expressions while considering the type of the emotion, the level of emotional intensity and the nature of the task. In this study, different emotional facial expressions will be presented with different intensity levels and two tasks will be created: emotion recognition task and intensity recognition task. As eye movements provide valuable information about these strategies, eye movements of participants will be collected during the tasks. Eye tracking data will also be fed into different machine learning models in order to investigate the eye movement features responsible for recognition processes and how they differentiate between tasks. In this way, we aim to find an answer to the question of which eye movement features have a predictive value on the process of emotion recognition and intensity recognition on faces. As a secondary aim of the study, we will observe how fixation numbers and pupil size of the viewers will change across emotion types and emotional intensities.

2. METHOD

2.1.Participants

Sixty-six participants volunteered in the current study. Thirty-three of the participants attended the emotion recognition task (13 male, mean age= 24.72, *SD*= 4.37) and thirty-three of the participants attended the intensity recognition task (13 male, mean age= 26.33, *SD*= 6.37).

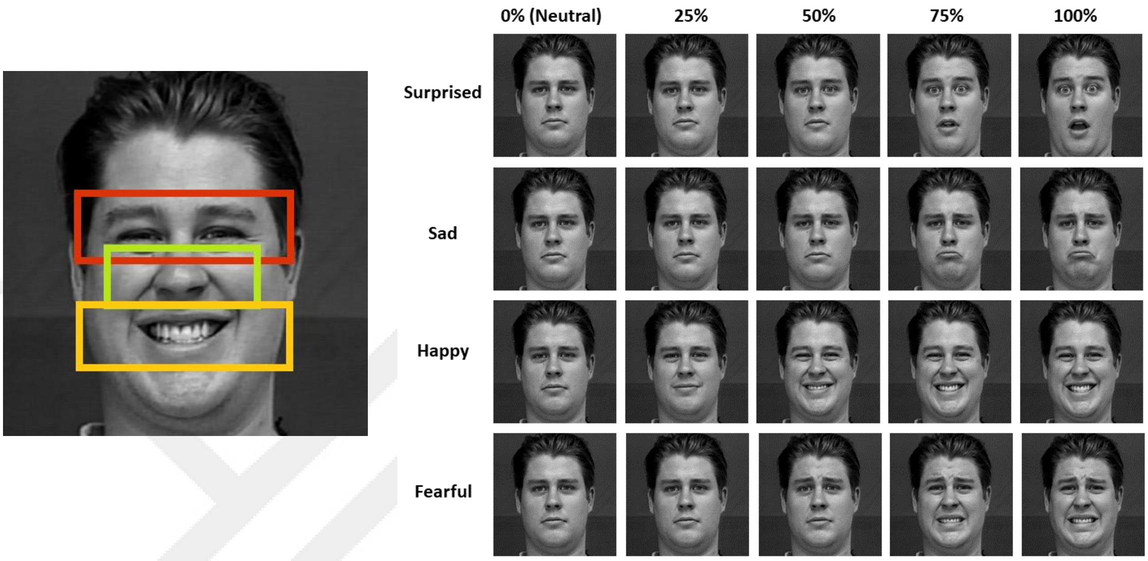
The inclusion criterion for the study was having a normal or corrected-to-normal vision. Twenty-seven of the participants received course credits for their participation. Current study was conducted with an approval from the Science-Ethics Committee of Yeditepe University.

2.2.Stimuli

Face images were selected from The Extended Cohn-Kanade (CK+) dataset (Lucey et al., 2010). CK+ dataset provides greyscale video frames for different subjects with different emotional expressions. These video frames capture the intensity change of expressions over time. We selected one female and one male middle-aged model expressing surprise, sadness, happiness and fear. Beginning of the frames were considered as neutral faces (0% intensity level) and the last frames with the peak of the emotion were considered as the highest intensity level (100%). Between these two frames, frame numbers equivalent to 25%, 50% and 75% intensity levels were selected (Figure 3b). A total of 40 images (8 neutral, 8 surprised, 8 sad, 8 happy and 8 fearful) were chosen for the experiment. Images were originally 640x490 pixels, but we cropped the images to 406x406 pixels, framing above the neck in order to have only faces in the images. In the experiment, we enlarged the images with a size of 640x640 pixels. We also adjusted the images such that the facial regions appear aligned when overlayed and created three areas of interest (AOIs): eyes, nose and mouth (Figure 3a).

Figure 3
Stimuli in the experiment and areas of interest

a. **b**.



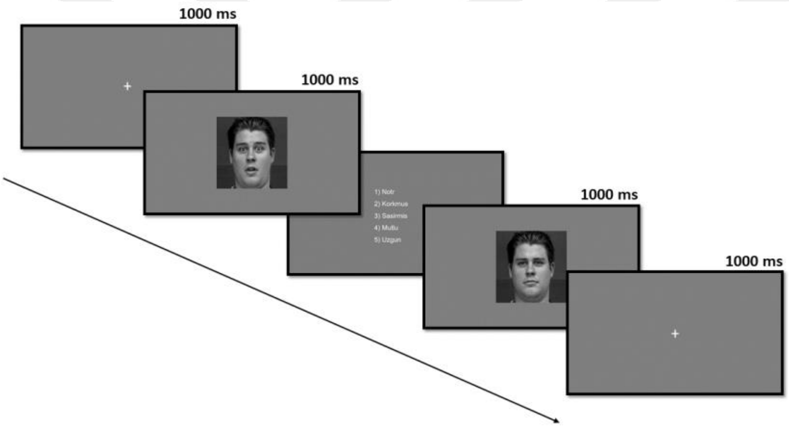
2.3.Procedure

All participants received an informed consent form before the experiment in accordance with the principles of the Declaration of Helsinki. Experiment was designed using MATLAB Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) and the stimuli were presented on a 27” ASUS monitor with a 120 Hz refresh rate and 1920x1080 resolution. Participants were seated approximately 70 cm from the computer screen in a dark room and they were asked to put their head on a chin rest in order to achieve a standard viewing angle and limit their head movements. Stimuli were presented on a gray screen.

2.3.1. Emotion recognition task

The emotion recognition task consisted of 4 emotion types, 5 intensity levels and 2 different subjects. Each condition (e.g., 50% angry male) repeated 3 times except for neutral images, which were presented one time, resulting in an experiment with 104 trials. First, participants were presented with an introduction screen. Then, every trial proceeds the following sequence: fixation point (1000 ms), face image (1000 ms), response screen (Figure 4). Response screen was presented until the participants responded. The task was to indicate the emotional expressions of the face they saw on the screen. Participants were presented with 5 options on the response screen in the following order for all the trials: neutral, fearful, surprised, happy, sad and they were asked to answer using the number keys (1, 2, 3, 4, 5) on the keyboard.

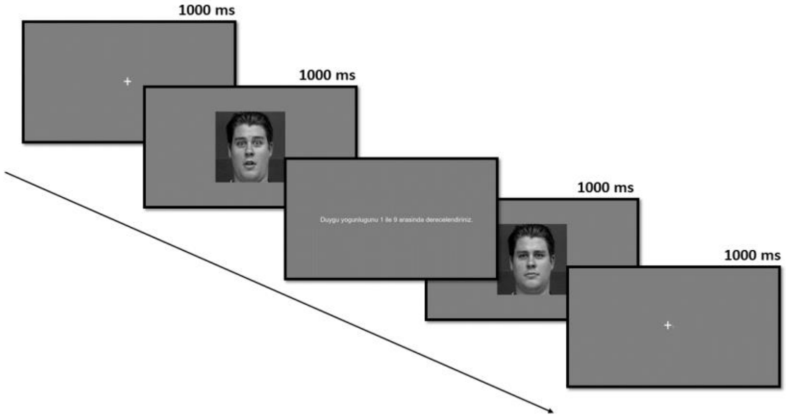
Figure 4
Sequence of the emotion recognition task



2.3.2. Intensity recognition task

Same number of trials and stimuli were used in the intensity recognition task. In this task, the only difference from the emotion recognition task was that we asked participants to decide on the intensity of the emotional expressions they saw on the screen. Sequences of the trials were the same except for the response screen (Figure 5). In the response screen, participants were asked to rate the emotional intensity of the face image on a scale of 1 to 9 by using the number keys on the keyboard.

Figure 5
Sequence of the intensity recognition task



2.3.3. Eye movement recording and preprocessing

Eye movements of the participants were recorded during the experiment with Gazepoint GP3 eye-tracking device (Vancouver, Canada). Participants were seated on an adjustable chair and an adjustable chin rest approximately 60 cm away from the eye-tracking device. Eye-tracker recorded the eye movements with a 60-Hz sampling rate and with an average 0.5-1 degree of visual angle accuracy. Before the experiment, a 5-point calibration and a 4-point validation were performed.

Data from eye tracker has been preprocessed before analyzing and feeding to machine learning models. Invalid (non-existent) pupil and fixation data were excluded, and participants with less than 70% of the total data were excluded from the eye tracking analysis. Fixation (≥ 80 ms) numbers were proportioned to the total fixation of the participants and we performed the analysis by calculating the percentage of fixations. For pupil size, a Butterworth low-pass filter with a 2 Hz cutoff frequency was applied to remove irrelevant data from the signal (Privitera et al., 2010). Pupil size of each participant was converted to *z*-score to create normalization and to avoid any data bias from pupil size variability.

2.4.Machine learning and CNN implementation

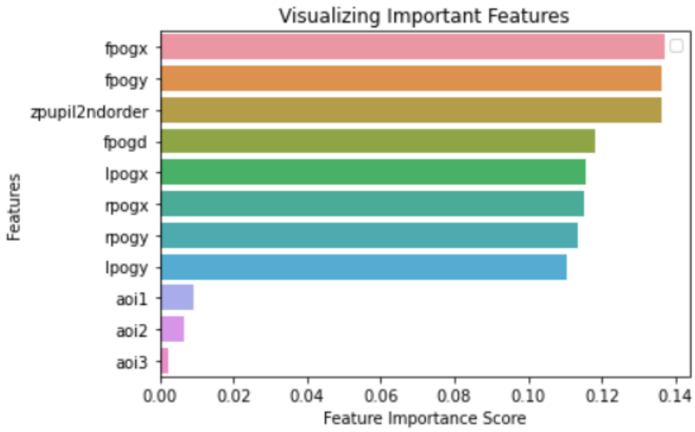
We fed the data we received from the eye tracking device to three different models to conduct a multiclass classification. k-Nearest Neighbors (k-NN) model searches for k number of closest samples for the test dataset from the training data set and selects the class that has the most k number of neighbors to the test dataset. Decision Tree model generally uses information entropy, passes the data through many rules and decides under which rule the classes can be. Finally, the Random Forest model is simply a combination of Decision Trees but avoids the tendency of overfitting observed in Decision Trees by choosing random features and bootstrapping.

When invalid data is excluded, each of the fixation-related parameters had 141.358 data points, pupil-related parameters had 188.326 data points and the combinations where we used both fixation- and pupil-related parameters had 125.217 data points. Test sample size was 25% and train sample size was 75% of the total sample. We optimized Decision Tree parameters as following: *criterion='entropy', splitter= 'best', max_depth=3, min_samples_leaf=2, max_features= 'log2', min_weight_fraction_leaf= 0.1, random_state = 42, max_leaf_nodes= 50* and we optimized Random Forest parameters as following: *n_estimators=1400, criterion='gini', max_depth=100, min_samples_split=2, min_samples_leaf=1, max_features='auto', bootstrap=True, oob_score=True.* For k-NN, we chose the number of neighbors with the lowest error rate among 40 neighbors.

Different eye tracker parameter combinations were created based on the feature importance scores (Figure 6). Models were fed with seven different combinations of eye tracking parameters: (1) fixation point of gaze (fpog), right eye fixation point of gaze (rfpog), left eye fixation point of gaze (lfpog), fixation duration (fpogd), left pupil size (lpupild), fixation numbers on AOIs, (2) fpog, rfpog, lfpog, fpogd, lpupild, (3) fpog, fpogd, lpupild, (4) fpog, lpupild, (5) fpog, (6) fpod, (7) lpupild. Vector sizes of these different combinations are as following: (1) 8 x 125.217, (2) 5 x 125.217, (3) 3 x 125.217, (4) 2 x 125.217, (5) 1 x 141.358, (6) 1 x 141.358, (7) 1 x 188.326.

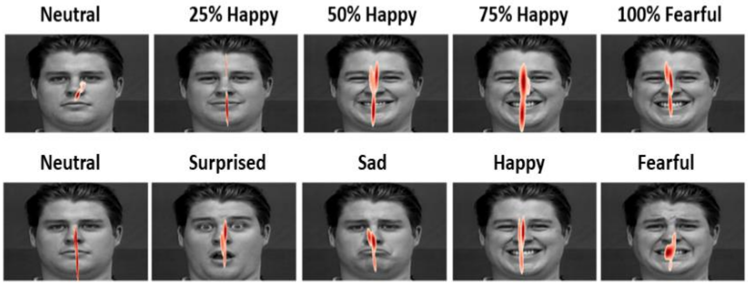
Figure 6

Feature importance scores



We also created heatmaps with the fixation points of the participants (Figure 7) and fed them to the ResNet50 convolutional neural network, which has 48 Convolution layers, 1 MaxPool and 1 Average Pool layer. The purpose of this step was to examine whether the fixation strategies around specific facial regions were different. The model is expected to capture features in the heatmap distribution and classify emotional expressions or intensities. There were a total of 2565 heatmaps that were projected on a white screen for analysis. We separated the data as 5% validation, 5% test and 90% train set.

Figure 7
Example of heatmaps of randomly chosen participant



2.5.Data Analysis

2.5.1. Behavioral data

Participants' total correct responses in the emotion recognition task were used to analyze their performance. Two-way ANOVA was conducted to see the effect of emotion type, intensity level and the interaction of the two on the performance for emotion recognition. Violation of homogeneity was ignored as the sample size for each condition was equal and the normality of the dependent variable was ignored as sample sizes were large enough according to our GPower (Faul et al., 2007) analysis. We run a one-way ANOVA to observe simple main effects of intensity level for each emotion type and the simple main effects of emotion type for each intensity level. Alpha level for significance was $p < .05$.

In the intensity recognition task, intensity levels of the presented stimuli ranged between 0 and 4. Participants rated these intensity levels on a scale of 1 to 9. Thus, we converted the ratings of participants proportional to the scale in the task and calculated the absolute difference between the real intensity level and the rating of the participants. We analyzed the error rates of the participants to observe the performance difference on the intensity recognition task for different emotion types. Two-way ANOVA was conducted to analyze the effect of emotion type, intensity level and the interaction effect between the emotion type and intensity level on the error rates for the intensity rating task.

2.5.2. Eye-tracking data

Total fixation percentage, fixation percentage on AOIs and pupil size were analyzed for emotion and intensity recognition tasks. Shapiro-Wilk normality assumption was violated. We conducted Kruskal-Wallis test to see the effect of emotion type and intensity level on recognition performance. Mann-Whitney test was applied for further pairwise comparisons. Alpha level for significance was $p < .05$.

3. RESULTS

3.1.Behavioral Results

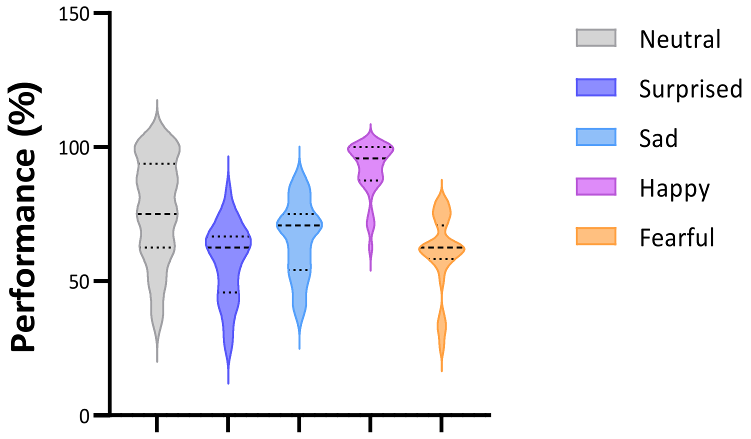
Two of the participants were not included in the analysis, as their overall performance was detected to be outliers in both emotion recognition task and intensity recognition task. Thus, we continued the study with 31 participants for the emotion recognition task (13 male, mean age= 24.27, *SD*= 4.02) and 31 participants for the intensity recognition task (12 male, mean age= 26.64, *SD*= 6.67).

3.1.1. Behavioral results of emotion recognition task

Participants showed an average of 69.38% performance for the emotion recognition task (*SD*= 6.843). Results showed a main effect of emotion type on the recognition performance, $F\ (4, 3203) = 7.424$, $p < .001$, $np^2 = .009$ (Figure 8). The highest recognition performance was observed for happy faces (*M*= 22) and it was significantly higher than the performance for all other emotion types and neutral faces, $p < .001$. The lowest recognition performance was found for surprised faces (*M*=14). Results showed that recognition performance was significantly higher for sad faces (*M*= 16) compared to surprised faces, $p < .001$, but not for fearful faces (*M*=14.94), $p = .136$. There was also no significant difference between the recognition performance for surprised faces and fearful faces, $p = .482$.

Figure 8

Recognition performance on emotional expressions

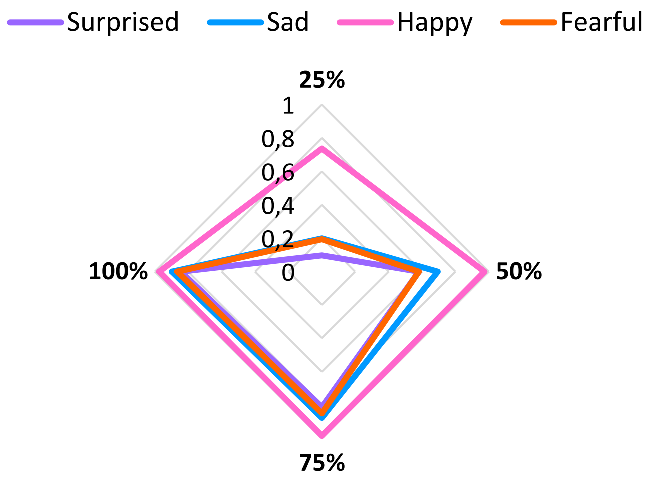


A main effect of intensity level on the emotion recognition performance was found, *F* (4, 3203) = 42.025, *p* = < .001, np^2= .281. Further post-hoc comparisons revealed that recognition performance for faces with %0 intensity level (*M*= .77) was significantly higher than the faces with %25 intensity level (*M*= .31), *p* < .001 and it was lower compared to the faces with %50 (*M*= .71), p = .001, %75 (*M*= .88) and %100 intensity level (*M*= .90), *p* < .001. Recognition performance was significantly lower for the faces with %25 emotional intensity than all other intensity levels, *p* < .001. Emotions with %50 intensity level resulted in lower recognition performance compared to the emotions with %75 and %100 intensity level, *p* <.001. However, there were no significant performance differences between faces with %75 intensity and %100 intensity, *p*= 1.

Results showed an interaction effect between the emotion type and intensity level for emotion recognition performance, *F* (12, 3203) = 13.843, *p* < .001, np^2= .048. We run a one-way ANOVA to observe simple main effects of intensity level for each emotion type.

Figure 9

Recognition performance on emotion types by intensity levels



Surprised, sad and fearful faces resulted in similar performance difference and only happy faces revealed different results (Figure 9). Recognition performance was significantly higher for surprised faces with 75% intensity level (*M*= .81) compared to surprised faces with 50% intensity (*M*= .56), p < .001, on sad faces with 75% intensity level (*M*= .87) compared to sad faces with 50% intensity (*M*= .67), p < .001 and fearful faces with 75% intensity level (*M*= .82) compared to fearful faces with 50% intensity (*M*= .57), p < .001. When the level of intensity was %25 for surprised (*M*= .09), sad (*M*= .19) and fearful faces (*M*= .18), the performance was found significantly lower compared to their higher emotional intensity, p < .001. Finally, we did not observe a significant difference for the recognition performance between surprised faces with %75 (*M*= .81) and %100 (*M*=.81) intensity level, p= 1. Recognition performance for happy faces did not differ between the faces with %50 (*M*= .97), %75 (*M*= .98) and %100 (*M*= .97) intensity level, p=1. However, happy faces with %25 intensity level (*M*= .74) resulted in significantly lower recognition performance compared to %50 (*M*= .97), %75 (*M*= .98), and %100 level of intensity (*M*=.97), p < .001.

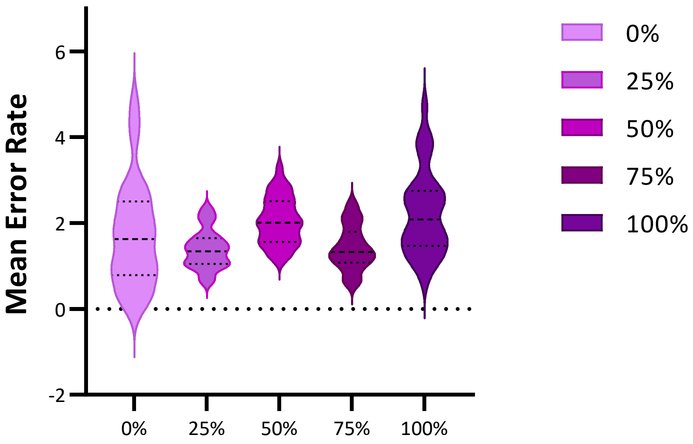
3.1.2. Behavioral result of intensity recognition task

Participants showed an average of 1.81 (*SD*= 1.48) error rate for the intensity recognition task. Two-way ANOVA results revealed a main effect of emotion types on the error rates for intensity recognition task, F (3, 3360) = 20.253, p < .001, np^2= .018. The highest error rate for rating the intensity levels was observed for sad faces (*M*= 2.18), while participants showed the lowest error rate for surprised faces (*M*= 1.58). Post-hoc results showed that error rates for sad faces were significantly greater compared to other three emotion types, p < .001. Additionally, error rates for surprised faces were significantly lower than error rates for happy faces (*M*= 1.75), p= .002. There was no significant difference between the error rates for happy faces and fearful faces (*M*= 1.71).

There was also a main effect of intensity level of emotions on the error rates, F (4, 3360) = 61.509, p < .001, np^2= .068. Lowest error rate was observed for emotional faces with 25% level of intensity (*M*= 1.37) and the highest was observed for 100% level of intensity (*M*= 2.29) (Figure 10). The error rate for %100 intensity level was significantly higher than neutral faces (*M*= 1.69), p < .001, faces with 25% intensity, p < .001, faces with 50% intensity (*M*= 2.05), p = .002 and faces with 75% intensity (*M*=1.49), p < .001. Emotional faces with %50 intensity (*M*= 2.05) level resulted in significantly higher error rates on the intensity recognition task compared to the faces with %25, p < .001 and %75 (*M*= 1.49) intensity level, p < .001 and neutral faces, p = .007. Finally, it was found that error rates to neutral faces (%0 intensity level) (*M*= 1.69) were significantly higher compared to faces with %25 emotional intensity, p = .005.

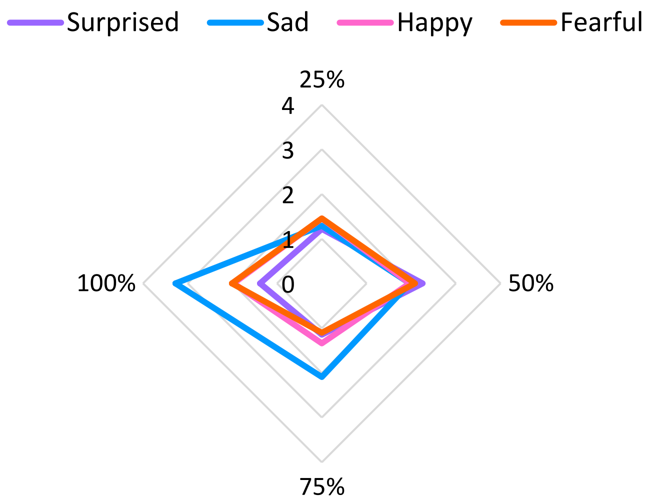
Figure 10

Error rates by intensity levels on intensity rating



Two-way ANOVA results revealed an interaction effect between emotion type and intensity level on the error rates of intensity recognition task, F (12, 3360) = 17.544, p < .001. We conducted a one-way ANOVA to observe the simple main effects of intensity levels on the error rates for each emotion type. Greater error rates were found for surprised faces with %50 intensity level (M= 2.26) when compared to the error rates for surprised faces with %25 (M= 1.22, p <.001), %75 (M= 1.5, p < .001), and %100 intensity levels (M= 1.38, p <.001) (Figure 11). There was no difference between the error rates for surprised faces with %100 and %75 emotional intensity, p= .615 and between %100 and %25 emotional intensity, p= .583. For sad faces, error rates were significantly higher for %100 emotional intensity (M= 3.46) compared to the error rates to %25 (M= 1.33), p < .001, %50 (M= 1.99), p < .001, %75 emotional intensity (M=2.15), p <.001 and neutral faces (M=1.52), p <.001.

Figure 11
Error rates on emotion types by intensity levels



When emotional intensity was %75, significantly higher error rates was found compared to the sad faces with %25 intensity levels, *p* < .001 and neutral faces, *p*= .021. Error rates were also greater for sad faces with %50 intensity level compared to the error rates for sad faces with %25 emotional intensity, *p* < .001. There was also no significant difference between the error rates for neutral faces and sad faces with %50 intensity level, *p*= .194. Error rates to happy faces with %100 intensity level were higher compared to error rates for %75 intensity level, *p* < .001 and %25 intensity level, p < .001. Error rates for happy faces with %75 intensity level (*M*= 1.43) was significantly lower than the error rates for %50 (*M*= 1.91) intensity level, *p*= .011. Also, more error rates were observed for happy faces with %50 intensity level compared to %25 intensity level (*M*=1.48), *p* = .036. However, there was no difference between happy faces with %25 intensity level and both neutral faces and happy faces with %75 intensity level, *p*=1. Finally, results revealed a lower error rate for fearful faces with %0 (*M*= 1.59), *p*= .026, %25 (*M*= 1.44), p< .001 and %75 intensity level (*M*= 1.22, p < .001 compared to fearful faces with 100% emotional intensity (*M*= 2.16). Fearful faces with %50 intensity levels revealed a greater error rate for both %25 ($p < .001$) and %75 intensity level ($p < .001$).

3.2.Eye movement results

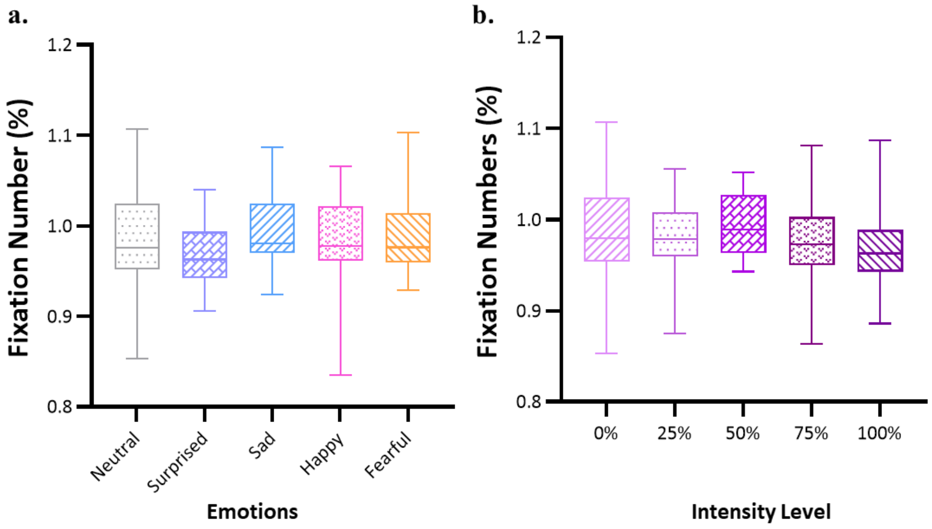
3.2.1. Number of fixations

3.2.1.1.Emotion recognition task.

Kruskal-Wallis test showed a significant effect of emotion type on the number of fixations on the faces, *H* (4) = 11.746, *p*= .019 (Figure 12a). Number of fixations were significantly lower for surprised faces (*Mdn*= 0.97, *n*= 672) compared to neutral faces (*Mdn*= 0.98, *n*= 216), *U*= 64171, *z*= -2.563, *p*= .010, compared to sad faces (*Mdn*= 0.98, *n*= 672), *U*= 205648, *z*= -2.831, *p*=.005, compared to happy faces (*Mdn*= 0.98, n= 674), *U*= 209691, *z*= -2.352, *p*= .019 and compared to fearful faces (*Mdn*= .98, n= 674), *U*= 209944, *z*= -2.317, *p*=.021. There was also an effect of intensity level on the total fixation numbers on the faces, *H* (4) = 12.979, *p*= .011 (Figure 11b). Higher fixation numbers were found for faces with %0 intensity level (*Mdn*= 0.98, *n*=224) when compared to %100 intensity level (*Mdn*= 0.97, n=671), *U*= 67581, *z*= -2.260, *p*= .024. Number of fixations for faces with %100 intensity level was also significantly lower compared to %25 intensity level (*Mdn*= 0.97, *n*=672), *U*= 208441, *z*= -2.394, *p*= .017. Faces with %50 intensity level (*Mdn*= 0.98, *n*=670) resulted in higher fixation numbers compared to the faces with %75 intensity level (*Mdn*= 0.97, *n*= 671), *U*= 208986, *z*= -2.228, *p*= .026 and compared to faces with %100 intensity level level, *U*= 205382, *z*= -2.736, *p*= .006.

Figure 12

Fixation numbers (emotion recognition task)

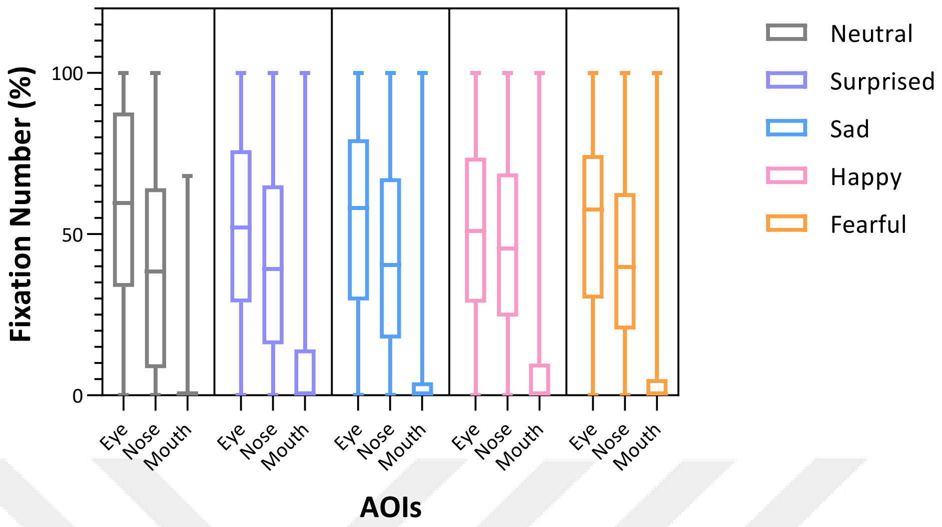


Note. (a) Fixation numbers by emotion type, (b) Fixation numbers by intensity level

Another Kruskal-Wallis test showed an effect of emotion type on the fixation numbers over AOIs, on the nose area, H (4) = 14.251, p= .007 and mouth area, H (4) = 22.732, $p < .001$ (Figure 13). Number of fixations towards the mouth was significantly lower for neutral faces (*Mdn*= 0, *n*= 216) compared to other emotional expressions. Mouth region was more looked at for surprised faces (*Mdn*= 0, *n*= 672) compared to both fearful faces (*Mdn*= 0, *n*= 674), *U*= 209272, *z*= -2.927, *p*= .003 and sad faces (*Mdn*= 0, *n*= 672), *U*= 209359, *z*= -2.805, *p*= .005. Also, more fixation towards nose region was found for happy faces (*Mdn*= 0, *n*= 674) compared to fearful faces, *U*= 214052, *z*= 3.046, *p*= .002, sad faces, *U*= 211608, *z*= -2.087, *p*= .037 and surprised faces, *U*= 205921, *z*= -2.886, *p*= .004.

Figure 13

Fixation numbers on areas of interest (emotion recognition task)

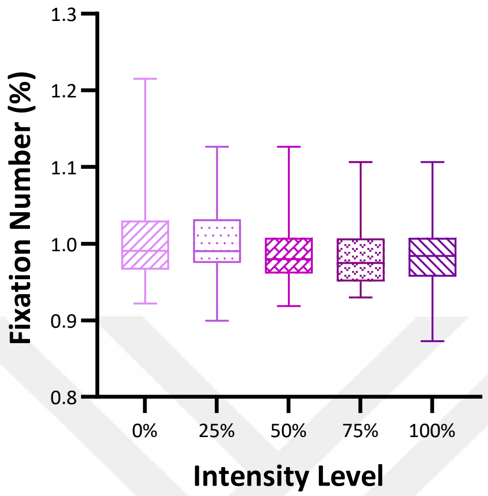


3.2.1.2.Intensity recognition task.

Level of intensity of emotional faces significantly affected the number of fixations of participants, *H* (4) =10.757, *p*=.029 (Figure 14). Number of fixations were higher for faces with %0 intensity level (*Mdn*= .99, *n*=215) compared to faces with %75 intensity level (*Mdn*= .97, *n*=646), *U*= 61824, *z*= -2.413, *p*= .016 and compared to %100 intensity level (*Mdn*= .97, *n*=643), *U*= 60706, *z*= -2.706, *p*= .007. A lower fixation number was observed for faces with %100 emotional intensity compared to faces with %25 emotional intensity (*Mdn*= .99, *n*= 648), *U*= 193887.5, *z*= -2.202, *p*= .028. There was no effect of emotion type on the number of fixations to face images, *H* (3) = 13.771, *p*= .775. However, effect of emotion revealed itself on the number of fixations over mouth regions of faces, *H* (3) =13.771, *p*= .003. More fixations over mouth region of surprised faces (*Mdn* =0, *n*=700) compared to sad faces (*Mdn*= 0, *n*=699), *U*= 224796, *z*= -3.273, *p*= .001 and compared to fearful faces (*Mdn*= 0, *n*= 700), *U*= 228472, *z*= -2.711, *p*= .007. Also, fixations over mouth regions were significantly higher for happy faces (*Mdn*=, 0 *n*= 700) compared to sad faces, *U*= 231247, *z*= -2.248, *p*= .025.

Figure 14

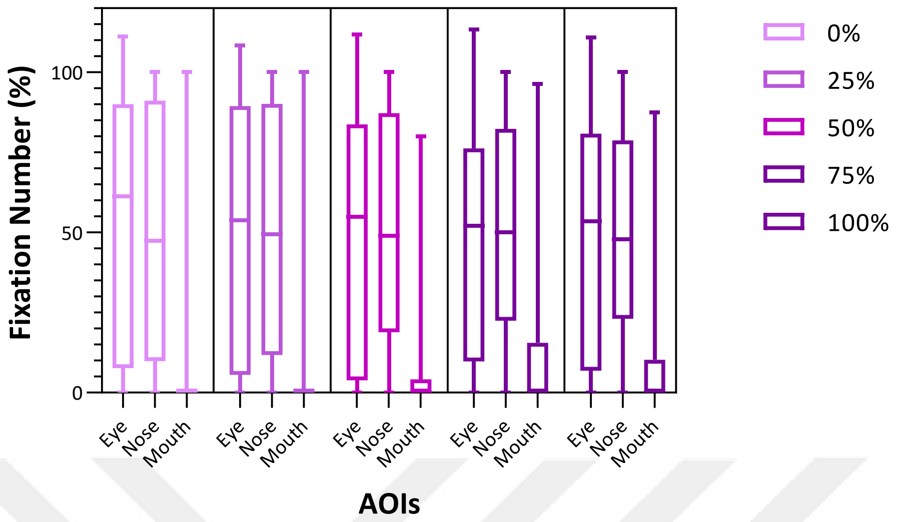
Fixation numbers on faces by emotional intensity (intensity recognition task)



There was also an effect of intensity level the number of fixations over mouth region, *H* (4) = 28.207, *p* < .001 (Figure 15). Number of fixations over mouth regions was lower for faces with %0 intensity level (*Mdn*= 0, n= 215) compared to faces with %75 intensity level (*Mdn* = 0, *n*= 646), *U*= 60075, *z*= -3.642, *p*< .001 and compared to faces with % 100 intensity level (*Mdn*= 0, *n*= 643), *U*= 60826, *z*= -3.268, *p*= .001. Fixations over mouth regions were also lower for faces with %25 emotional intensity (*Mdn*= 0, *n*= 648) compared to faces with %75 emotional intensity, *U*= 187761.5, *z*= -4.016, *p* < .001 and compared to faces with %100 emotional intensity, *U*= 190014.5, *z*= -3.449, *p*= .001. Faces with %50 emotional intensity (*Mdn*= 0, *n*= 647) resulted in lower fixations over mouth regions compared to faces with %75 emotional intensity, *U*= 194448.5, *z*= -2.659, *p* = .008 and compared to faces with %100 emotional intensity, *U*= 196763, *z*= -2.078, *p*= .038.

Figure 15

Fixation numbers on areas of interest by intensity level (intensity recognition task)



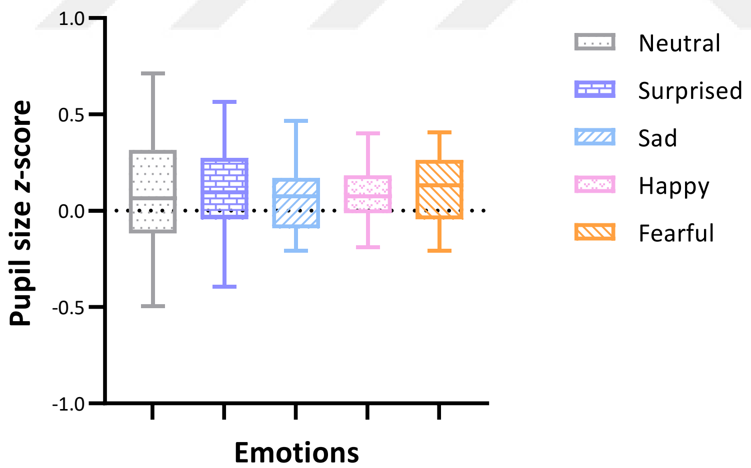
3.2.2. Pupil size

3.2.2.1.Emotion recognition task.

There was a significant effect of emotion type on the pupil size of participants, *H* (3) = 74.615, *p* < .001 (Figure 16). Pupil size was significantly larger when participants viewed fearful faces (*Mdn*= .14, *n*= 43336) when compared to the surprised faces (*Mdn*= .09, *n*= 43543), *U*= 914998463, *z*= -7.708, *p* < .001, compared to sad faces (*Mdn*= .99, *n*= 43522), *U*= 916835843, *z*= -7.091, *p* < .001 and compared to happy faces (*Mdn*= .09, *n*= 43812), *U*= 928069258, *z*= -5.722, *p*< .001. A larger pupil size is also observed for happy faces compared to surprised faces, *U*= 945854753.5, *z*= -2.146, *p* = .032.

An effect of intensity level on pupil size was also observed, *H* (4) = 78.374, *p* < .001. Pupil size for faces with %0 emotional intensity (*Mdn*= .09, *n*= 14577) was significantly larger compared to faces with 50% intensity level (*Mdn*= 0.9, *n*=43537), *U*=311457893.5, *z*= -3.343, *p*=.001 and compared to the faces with 100% intensity level (*Mdn*= .10, *n*= 43370), *U*= 313045965.5, *z*= -1.981, *p*= .048. However, pupil size was larger for faces with 75% intensity level (*Mdn*=, *n*= 43370) compared to 0% intensity level, *U*= 311990544, *z*= -2.353, *p*= .019. Pupil size for 25% emotional intensity (*Mdn*= .09, *z*= 43414) was significantly larger compared to the faces with 50% emotional intensity, *U*= 937201743.5, *z*= -2.123, *p*= .034 but, it was significantly smaller compared to pupil size for 75% intensity level, *U*= 918920460, *z*= -6.101, *p* <.001. Pupil size were also smaller for with 50% emotional intensity compared to the faces with 75% intensity level, *U*= 913135237, *z*= -8.373, *p* < .001 and compared to the faces with 100% intensity level, *U*= 937586957, *z*= -2.089, *p*= .37. Finally, a smaller pupil size was observed for faces with 100% emotional intensity compared to faces with 75% emotional intensity, *U*= 919027597, *z*= -6.141, *p* < .001.

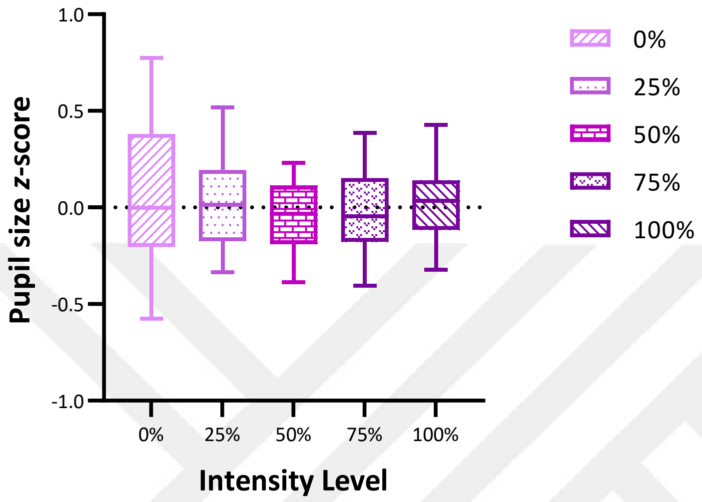
Figure 16
Pupil size for emotional expressions (emotion recognition task)



3.2.2.2.Intensity recognition task.

A significant effect of emotion type on the pupil size was found, *H* (3) = 738.480, *p* < .001. Pupil size for surprised faces (*Mdn*= .04, *n*= 27935) was significantly larger compared to pupil size for sad faces (*Mdn*= -.16, *n*= 27797), *U*= 345134940, *z*= -22.706, *p* < .001. Pupil size for sad faces were observed to be smaller compared to pupil size for happy faces (*Mdn*= .05, *n*= 27793), *U*= 345095929, *z*= -21.770, *p* < .001 and compared to pupil size for fearful faces (*Mdn*= .04, *n*= 27822), *U*= 344968844.5, *z*= -22.033, *p* < .001.

Figure 17
Pupil size for intensity levels (intensity recognition task)



An effect of emotional intensity level was also observed on pupil size of participants, *H* (4) = 90.602, *p*= < .001 (Figure 17). Pupil size for faces with 0% emotional intensity (*Mdn*= .05, *n*= 8698) was found to be larger compared to 25% (*Mdn*= .01, *n*= 25560), *U*= 109029005.5, *z*= -2.675, *p* = .007, 50% (*Mdn*= -.02, *n*= 25855), *U*= 109021287, *z*= -4.253, *p* < .001, and 75% emotional intensity (*Mdn*= -.06, *n*= 25762), *U*= 106599893, *z*= -6.780, *p* < .001. Expressions with 25% intensity level results in larger pupil size compared to expressions with 50% intensity level, *U*= 326420921.5, *z*= -2.381, *p*= .017 and compared to expressions with 75% intensity level, *U*= 319148794, *z*= -6.012, *p* <.001. Faces with 100% emotional intensity (*Mdn*= .04, *n*= 25472) resulted in larger pupil size compared to faces with 25% emotional intensity, *U*= 321948533.5, *z*= -2.154, *p*= 0.31, compared to 50% emotional intensity, *U*= 321559480, *z*= -4.606, *p* < .001 and compared to pupil size for faces with 75% intensity level, *U*= 314298858.5, *z*= -8.248, *p* < .001. Finally, emotional expressions with 50% intensity level also resulted in larger pupil size than expressions with 75% intensity level, U= 327111737, z= -3.501, p < .001.

3.3.Machine Learning Results

We applied three different algorithms (k-NN, Decision Tree, Random Forest) with seven combinations of eye tracking parameters. Random Forest model achieved the highest performance on all classifications. Below, the results from Random Forest will be given in detail. Then, results from other models will be briefly presented.

3.3.1. Emotion recognition task

The RF model achieved the highest accuracy rates with the following parameters: (1) fpog, rfpog, lfpog, fpogd, lpupild, fixation numbers on AOIs, (2) fpog, rfpog, lfpog, fpogd, lpupild, (3) fpog, fpogd, lpupild. Model was able to predict emotional expressions with an average accuracy rate of 86.06% with the first combination, 85.03% with the second combination and 73.31% with the third combination (Table 1).

Even though 1st combination (fpog, rfpog, lfpog, fpogd, lpupild fixation numbers on AOIs) resulted in the highest classification performance, we considered 2nd combination (fpog, rfpog, lfpog, fpogd, lpupild) as the most successful as the difference is very small and fixations on areas of interest are secondary parameters. The 2nd combination resulted in 99% precision for neutral face classification with 67% recall which is pointing at a weaker classification (Figure 18a). Accuracy for surprised faces was slightly higher (86%) compared to other emotions and sad, fearful and happy faces were predicted with similar values.

Table 1

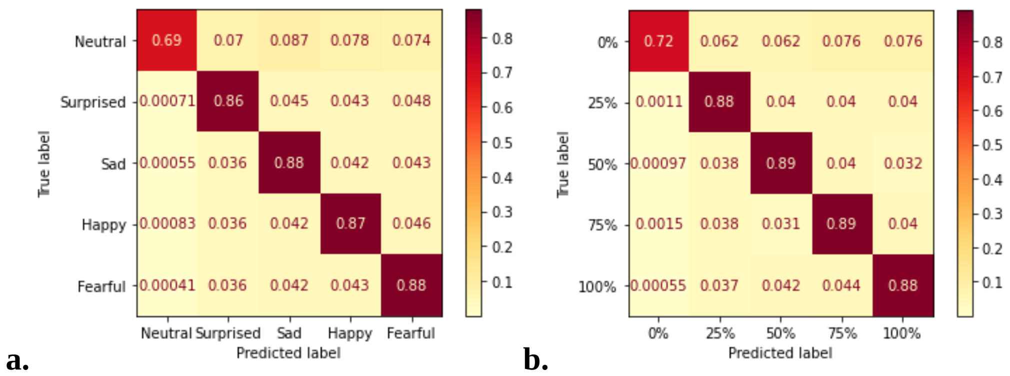
Random Forest outputs for the combinations of parameter on emotion and intensity classification on emotion recognition task

Parameters	Emotion	*Precision*	*Recall*	*F1-score*	Intensity	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral	.99	.72	.83	0%	.99	.72	.83
	Surprised	.87	.87	.87	25%	.86	.89	.88
	Sad	.85	.87	.86	50%	.87	.88	.87
	Happy	.85	.87	.86	75%	.86	.89	.87
	Fearful	.84	.88	.86	100%	.87	.88	.87
Fpog Lpog Rpog Fpogd Pupild	Neutral	.99	.69	.81	0%	.98	.71	.83
	Surprised	.87	.86	.86	25%	.86	.89	.87
	Sad	.85	.88	.86	50%	.87	.88	.87
	Happy	.85	.87	.86	75%	.84	.87	.86
	Fearful	.85	.88	.86	100%	.86	.87	.87
Fpog Fpogd Pupild	Neutral	.81	.58	.68	0%	.83	.57	.67
	Surprised	.75	.74	.74	25%	.72	.75	.74
	Sad	.72	.74	.73	50%	.73	.75	.74
	Happy	.73	.75	.74	75%	.72	.74	.73
	Fearful	.72	.75	.73	100%	.73	.74	.73
Fpog Pupild	Neutral	.60	.37	.46	0%	.58	.36	.45
	Surprised	.59	.58	.58	25%	.56	.57	.57
	Sad	.55	.59	.57	50%	.54	.58	.56
	Happy	.56	.59	.58	75%	.56	.57	.57
	Fearful	.57	.57	.57	100%	.55	.56	.56
Fpog	Neutral	.24	.16	.19	0%	.24	.16	.19
	Surprised	.36	.37	.37	25%	.36	.37	.36
	Sad	.34	.35	.35	50%	.35	.37	.36
	Happy	.35	.36	.36	75%	.35	.36	.35
	Fearful	.35	.36	.36	100%	.34	.35	.35
Fpogd	Neutral	.11	.03	.04	0%	.10	.03	.04
	Surprised	.24	.19	.21	25%	.24	.22	.23
	Sad	.24	.30	.27	50%	.24	.24	.24
	Happy	.23	.23	.23	75%	.23	.26	.24
	Fearful	.23	.27	.25	100%	.23	.27	.25
Pupild	Neutral	.08	.07	.08	0%	.07	.07	.07
	Surprised	.25	.25	.25	25%	.24	.24	.24
	Sad	.23	.24	.24	50%	.24	.24	.24
	Happy	.24	.24	.24	75%	.23	.23	.23
	Fearful	.23	.23	.23	100%	.24	.24	.24

Training accuracy of the model was 100% for the first five parameter combinations, 40% for fpogd and 85% for lpupild. Since training accuracy is already expected to be 100% or very high with this algorithm and with such a large data, we also examined out of bag scores to validate the learning of the model. Out of bag error (OOB error) is basically averaging errors based on the predictions from the trees. However, their own bootstrap samples are not included in these trees. Thus, OOB error is a good way to see if our predictions are biased or not for this specific algorithm. It is argued that predicting training accuracy is done by using OOB samples and the closer the OOB score (1- OOB error) is to the testing accuracy, the better the model learns (Ramosaj & Pauly, 2019). We observed that the OOB score of the emotion classification accuracy values in the emotion recognition task is always less than 0.01 (See Appendix A).

Figure 18

Confusion matrix for emotion and intensity classification on emotion recognition task



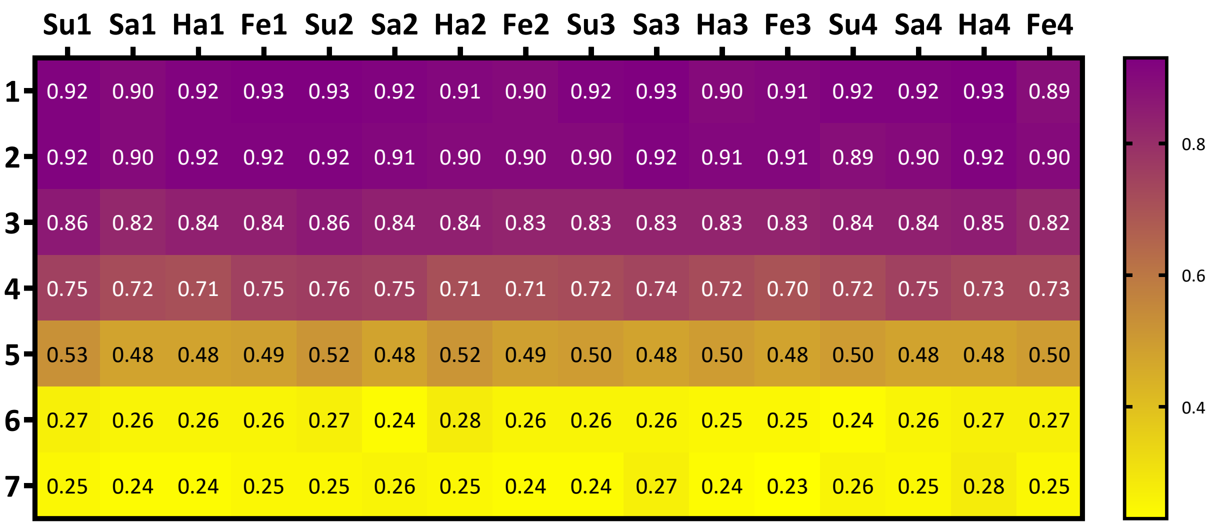
Note. *(a) Confusion matrix of emotion classification, (b) confusion matrix for intensity classification.*

Algorithm achieved 87.11% accuracy for intensity classification with the 1st combination, 86.55% accuracy with the 2nd combination and 72.93% accuracy with the 3rd combination. Neutral faces again had a high precision in intensity classification (98%) but again a lower recall value (77%) with the second combination (Figure 18b).

The highest classification performance was observed for stimuli with 50% intensity level, even though it is a minor difference. OOB score were again very close to testing accuracy (< .01) indicating an unbiased and reliable (See Appendix A).

Classification performance of the model increased when the emotional intensity was set to a certain level. However, we did not observe an eminent difference between emotion types or intensity levels for emotion classification (Figure 19). Classification for each emotion type followed the pattern of the algorithm which shows an increase in the accuracy when fed with more information. First two combinations reached the highest performance while the last two combinations reached the lowest performance for emotion classification. Same pattern was observed for intensity classification within each emotion type on emotion recognition task as well (See Appendix B).

Figure 19
Accuracy heatmap of emotion classification by different eye tracking parameters



Note. 1= fpog, rfpog, lfpog, fpogd, lpupild, fixation numbers on AOIs, (2) fpog, rfpog, lfpog, fpogd, lpupild, (3) fpog, fpogd, lpupild, (4) fpog, lpupild, (5) fpog, (6) fpod, (7) lpupild. Su= surprised, sa = sad, ha= happy, fe= fearful.

3.3.2. Intensity Recognition Task

Model was able to predict emotional intensity with an average accuracy rate of 89.9% with the first combination, 89.42% with the second combination and 72.29% with the third combination from the eye movement data on intensity recognition task (Table 2).

Table 2

Random Forest outputs for the combinations of parameters on emotion and intensity classification on intensity classification task

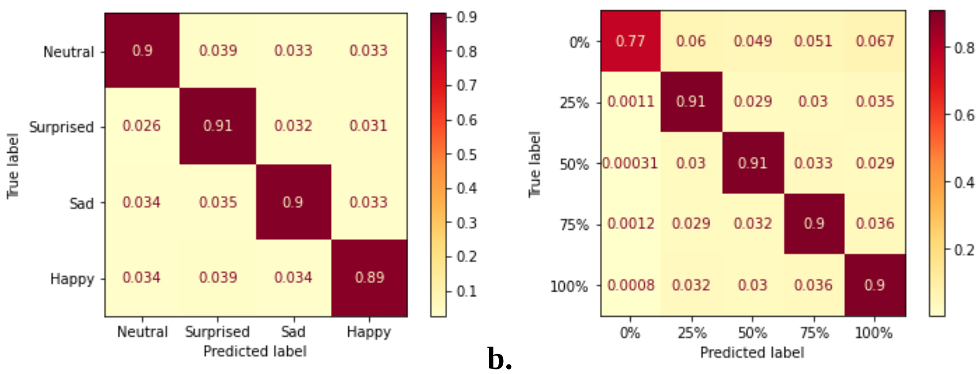
Parameters	**Emotion**	Precision	Recall	F1-score	**Intensity**	Precision	Recall	F1-score
Fpog Lpog	Neutral	NaN	NaN	NaN	0%	.99	.80	.88
Rpog	Surprised	.90	.89	.90	25%	.90	.91	.91
Fpogd	Sad	.90	.90	.90	50%	.89	.91	.90
Pupild	Happy	.90	.90	.90	75%	.89	.91	.90
Fix. On AOIs	Fearful	.90	.89	.90	100%	.89	.90	.90
Fpog	Neutral	NaN	NaN	NaN	0%	.99	.77	.87
Lpog	Surprised	.90	.89	.89	25%	.89	.91	.90
Rpog	Sad	.89	.90	.89	50%	.90	.91	.90
Fpogd	Happy	.89	.90	.89	75%	.89	.90	.89
Pupild	Fearful	.89	.89	.89	100%	.88	.90	.89
Fpog	Neutral	NaN	NaN	NaN	0%	.87	.72	.79
Fpogd	Surprised	.81	.79	.80	25%	.79	.80	.79
Pupild	Sad	.79	.79	.79	50%	.79	.80	.79
	Happy	.78	.79	.79	75%	.79	.80	.79
	Fearful	.79	.79	.79	100%	.78	.80	.79
Fpog	Neutral	NaN	NaN	NaN	0%	.68	.44	.54
Pupild	Surprised	.61	.60	.61	25%	.61	.61	.61
	Sad	.61	.62	.61	50%	.60	.62	.61
	Happy	.61	.60	.61	75%	.60	.62	.61
	Fearful	.60	.61	.61	100%	.59	.63	.61
	Neutral	NaN	NaN	NaN	0%	.28	.19	.23
	Surprised	.40	.38	.39	25%	.37	.38	.37
Fpog	Sad	.38	.39	.38	50%	.37	.38	.38
	Happy	.39	.39	.39	75%	.37	.39	.38
	Fearful	.39	.40	.40	100%	.36	.38	.37
	Neutral	NaN	NaN	NaN	0%	.07	.02	.04
	Surprised	.25	.24	.25	25%	.23	.20	.21
Fpogd	Sad	.25	.24	.24	50%	.22	.24	.23
	Happy	.24	.25	.24	75%	.23	.27	.25
	Fearful	.25	.26	.26	100%	.23	.24	.23

	Neutral	NaN	NaN	NaN	0%	.08	.08	.08
	Surprised	.26	.26	.26	25%	.23	.23	.23
Pupild	Sad	.26	.26	.26	50%	.23	.24	.23
	Happy	.26	.25	.25	75%	.24	.24	.24
	Fearful	.25	.26	.26	100%	.23	.24	.23

Model achieved 90% percent accuracy for surprised faces 89% for sad, happy and fearful faces with high recall values with 2nd combination of parameters (Figure 20a). As there cannot be a different neutrality of each expression, neutral faces were not included in this classification. However, the previous classification pattern revealed itself in intensity classification as well. Model showed highest precision for neutral faces with lowest recall and other intensity levels showed similar precision and recall values (Figure 20b). OOB scores for intensity classification and emotion classification on intensity recognition task were close to the testing accuracy (< .01) (See Appendix A).

Figure 20

Confusion matrix for emotion and intensity classification on intensity recognition task

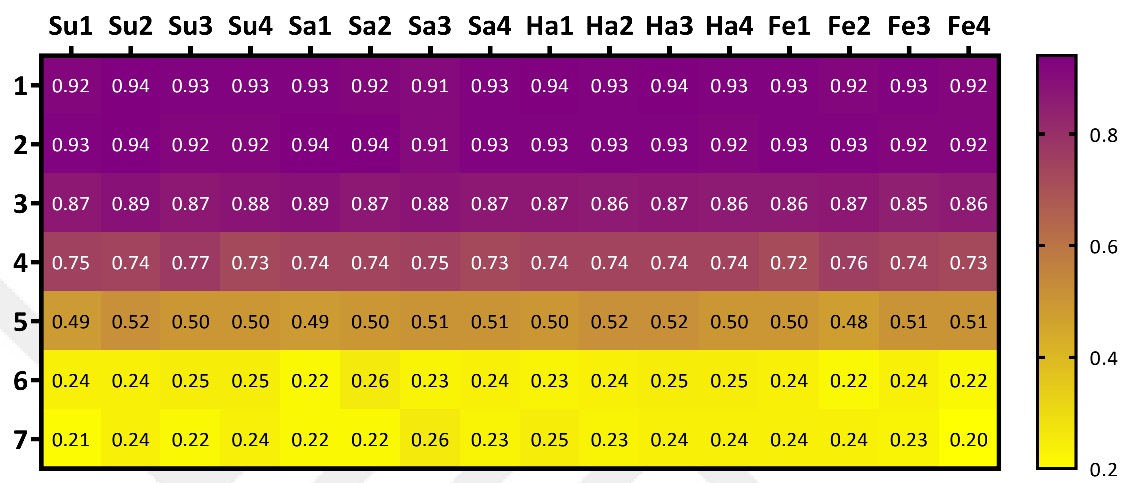


a. **b.**

Note. (a) Confusion matrix of emotion classification, (b) confusion matrix for intensity classification.

Each intensity level was also predicted within each emotion type and we observed an increase in the model's classification performance (Figure 21). Algorithm did not show any superior performance towards one intensity level. We achieved highest performance with the 1st and 2nd combinations and lowest performance with 6th and 7th combinations. Model followed a similar pattern for emotion classification on intensity recognition task (See Appendix B).

Figure 21
Accuracy heatmap of intensity classification by different eye tracking parameters



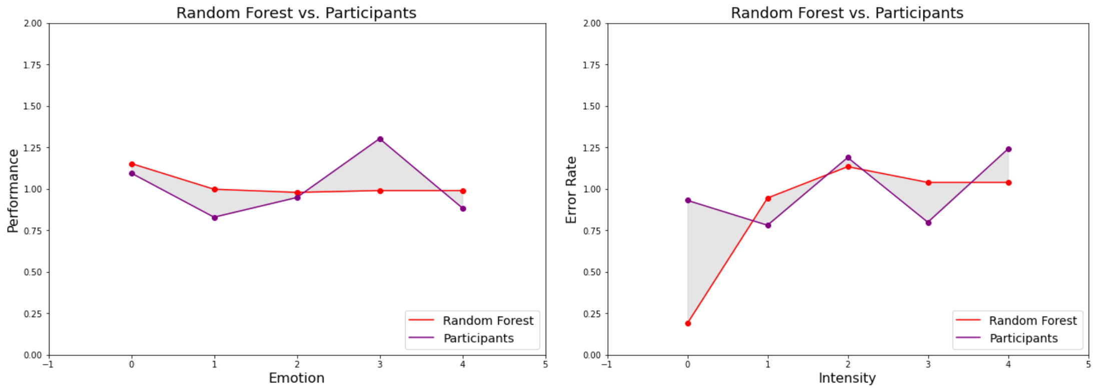
3.3.3. Random Forest Model vs. Participants

We also wanted to compare the emotion recognition performance of the participants with the emotion recognition performance of the model. We divided the performance of the Random Forest and humans for each emotion by their overall performance, thus leveling the performance of the two (Figure 22a). We calculated the polygonal area to find the performance difference. We found that the people and model showed a .10 difference in emotion recognition performance. Neutral, surprised, sad, and fearful faces did not differ greatly between the model and the participants, but we found that the model was more successful than the participants in classifying happy faces.

Figure 22

Performance of Random Forest and participants

a. **b.**



Note. *(a) Emotion recognition performance of participants and the model, 0 = neutral, 1 = surprised, 2 = sad, 3 = happy, 4 = fearful. (b) Error rates on intensity classification of participants and the model, 0 = 0%, 1 = 25%, 2 = 50%, 3= 75%, 4 = 100%.*

The model's error for classifying intensity levels (1 - accuracy = 1 - (number of classified correct) / (number of classified total)) was compared to the error of participants on intensity level ratings. Again, performance difference on intensity recognition between model and people was found by calculating the polygonal area between the lines (Figure 22b). A difference of .12 in rating was observed. Model and participants recognized with similar error rates for 25% and 50% intensities. The model was found to classify 0% and 100% intensity levels better than humans, and 75% intensity worse than humans.

3.3.4. Other Models

Although the k-NN and Decision Tree models performed above the chance level in emotion and intensity classification, they lagged far behind the performance of Random Forest. In particular, the Decision Tree performance was only slightly above the chance level. While Random Forest achieved the highest accuracy rate with the 1st parameter combination, the k-NN algorithm was able to classify emotions correctly with an average of 41.6% (k= 38) with the emotion recognition task and Decision Tree showed a performance of 26% in the same situation (Table 3). While small differences were observed between the training accuracy and testing accuracy of the Decision Tree, this difference was higher for k-NN (See Appendix E and F). It can be suggested that even though k-NN resulted in higher classification performance, it has slightly lower reliability.

Table 3
Emotion classification of k-NN and Decision Tree on emotion recognition task

	k-NN Emotion Classification					Decision Tree Emotion Classification		
Parameters	**Emotion**	*k*	*Precision*	*Recall*	*F1-score*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral	38	.41	.11	.17	.00	.00	.00
	Surprised		.42	.43	.43	.26	.24	.25
	Sad		.41	.46	.43	.25	.41	.31
	Happy		.42	.45	.43	.27	.29	.28
	Fearful		.42	.42	.42	.27	.18	.21

K-NN and Decision Tree, intensity recognition performance on intensity recognition task was also similar to emotion classification performance with the 1st parameter combination (Table 4). Additionally, both parameters followed a similar pattern as Random Forest and showed an increase in their accuracy with the increase of the number of parameters (See appendix C and D). Both algorithms continued to show the pattern in training and testing accuracy differences for intensity recognition task as well (See Appendix E and F).

Table 4

Intensity classification of k-NN and Decision Tree on intensity recognition task

	k-NN Intensity Classification					Decision Tree Intensity Classification		
Parameters	*Intensity*	*k*	*Precision*	*Recall*	*F1-score*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	0%	38	.43	.14	.22	.00	.00	.00
	25%		.43	.49	.46	.25	.20	.22
	50%		.43	.47	.45	.24	.48	.32
	75%		.43	.44	.43	.25	.38	.30
	100%		.44	.42	.43	.00	.00	.00

Finally, when we put the heatmaps of the participants' eye movements on the images into image classification with CNN, we saw that the model did not provide sufficient learning. The model did not increase its prediction accuracy higher than 24.91% so, the best accuracy rate we got from the CNN model was 24.91%.

4. DISCUSSION

We examined which eye movement parameters are informative for recognition processes of emotion and intensity of facial expressions. The eye movements behavior across tasks, emotions and intensities and the role of different parameters on these variables were investigated. In this section, behavioral and eye tracking results are discussed first. Then, eye tracking parameters that increase the classification accuracy of models are discussed with regard to eye tracking and emotional expressions literature. Finally, the difference between the performance of the model and the performance of humans is examined.

4.1.Recognition Performance and Eye Movements

Behavioral and eye movement analyses provided insights about the recognition processes of emotional expressions and the expression intensity. While an effect of emotion type on fixation number was observed in the emotion recognition task, this effect was not pursued in the intensity recognition task. Although we cannot claim that the participants changed their scanning patterns due to their task based on these findings, it can be suggested that we may need more detailed information in general when recognizing emotional intensity. The level of emotional intensity, on the other hand, showed its effect in both tasks. We expected that emotion recognition would become easier and the number of fixations would decrease with the increase in the intensity level (Guo, 2012; Montagne et al., 2007). Participants showed greater fixation for the expressions with 0% and 25% intensity compared the expressions with 100% level of intensity in both tasks. Our behavioral results showed that participants were better at recognizing neutral faces than expressions with 25% emotional intensity. These results may suggest that participants may have made their scanning patterns more thorough to understand whether neutral faces have emotion. However, an emotional intensity as low as 25% may not be successfully recognized even with a detailed search. Considering that the recognition threshold of emotions starts from the lowest 40% level of intensity, these findings were expected (Calvo et al., 2016). Opposite results were revealed in the intensity task. We observed that participants rate the expressions with 25% and 75% emotional intensity better, and they were worse at intensities at extreme levels of intensity such as 0% or 100% which was in line with the study by Vaidya et al. (2014). These results may also be due to the fact that base and the peak of the expressions are expected differently in natural settings. Finally, our eye tracking results showed the same pattern for the AOIs in both the emotion task and the intensity task. As previous studies have shown, the eyes were looked at most, followed by the nose and the mouth is the least (Guo, 2012; Scheller et al. 2012).

4.2.Informative Parameters on Emotion and Intensity Recognition

In this chapter, results of Random Forest will be discussed as the algorithm achieved the highest accuracy rates when fed with different eye tracking parameters. Other models are discussed in Chapter 4. 4.. Random Forest achieved the best performance with the information from the following parameters: fixation point of gaze, fixation point of gaze (right eye), fixation point of gaze (left eye), fixation duration, pupil size. Adding the fixation count on AOIs increased the accuracy averagely .4% but, we will focus on the former parameters as AOIs are secondary parameters which are calculated later. The model achieved an average accuracy of 87.54% (average of both tasks) for emotion classification and an average accuracy of 87.98% for intensity classification with the second parameter combination. Considering these results, combination of different eye tracking parameters has a more predictive value for both emotion and intensity classification. As we increase the number of parameters fed to the model, we see that the accuracy also increases. However, the explanatory value of fixation point of gaze was higher compared to pupil size. These results were expected, as the fixation point of gaze gives information about the regions on which viewers focus, as well as the order of gaze orientation. It has also been shown by many studies that if people do not adapt specific fixation patterns for recognizing emotional facial expressions, their recognition performance will decrease (Carbon, 2020; Eisenbarth & Alpers, 2011; Jack et al., Kashihara, 2013; 2009; Scurgin et al., 2014). Strategies to examine emotional expressions have also been demonstrated by previous behavioral and eye tracking studies. For example, when examining threat-related emotional expressions such as angry or fearful, it was observed that there was a longer distance between fixations and fewer but longer fixations were performed (Calvo & Nummenmaa, 2008; Green et al., 2003; Guo, 2012; Horley et al., 2005). However, much less total fixation number, duration and even first fixation duration was observed for happy expressions (Calvo & Nummenmaa, 2008; Guo, 2012). Some studies have argued that people fixate less on threatful facial expressions such as fearful and angry, that is, they show less fixation and a holistic scanning pattern is followed (Hunnius et al., 2011). Importance of fixation-related strategies has also been seen in studies comparing the performance of emotion recognition across cultures suggesting distribution of the gaze across the face rather than fixating mostly to the eyes improves performance (Jack et al., 2009). Still, people generally follow a scanning pattern according to diagnostic features that reflect emotion (Smith et al., 2005; Yitzhak et al., 2020). People direct their gaze more to the eyes when recognizing sad and angry faces, and more to the mouth area for happy and disgusted faces (Beaudry et al., 2014; Eisenbarth & Alpers, 2011; Schurgin et al., 2014). Eye area was also found to be important for recognition of fearful and surprised faces (Guo, 2012). In addition to these, there are also studies suggesting that a more holistic strategy is required for surprised and fearful faces compared to other facial expressions (Meaux & Vuilleumier, 2012; Tanaka et al., 2012). Although these studies cannot reach an agreement on the suggested pattern or diagnostic feature for each specific emotion, they point out the importance of gaze allocation, fixation patterns and duration in emotion identification. The need for information from diagnostic features, especially for emotions, remains for low or high emotional intensity as well, even though fixation numbers change with the intensity of emotions (Eisenbarth & Alpers, 2011). As the intensity of emotional expression decreases, we may need more information from faces, longer exploration time, and more fixation in order to make a correct judgment. It seems logical for these parameters to be explanatory in the same way for the intensity of emotion (Gao & Maurer, 2010; Herba et al., 2006; Vaidya et al., 2014). Contrary to fixation patterns, our model could barely pass the chance level when only pupil size values were fed. It can be argued that an autonomous response from a pupil is not a strong parameter in predicting the expression of observed expression. It has long been known that the pupil is sensitive to emotional stimuli (Hess & Polt, 1960). With the advancement of technology, this sensitivity was expected to be a very useful feedback sensor, especially in HCI studies by providing information about people’s emotional and cognitive states (Jacob & Karn, 2003). In fact, the pupil was expected to be a powerful source of information that could inform us about the person's internal state or arousal level, as many studies have shown that the pupil is responsive to emotional stimuli. Larger pupil size as a response to both negative and positive stimuli have been shown repeatedly (Arriaga et al., 2015; Bradley et al., 2008; Geangu et al., 2011; Henderson et al., 2014; Bradley & Lang, 2015; Partala & Surakka, 2003; Babiker et al., 2015; Steenberg et al., 2011;

Snowden et al., 2016). However, it was also found that the pupil dilates only with negatively charged emotional stimuli (Burley et al., 2017) and even especially with fearful (Laeng et al., 2013) or angry faces (Kret et al., 2013). Our eye tracking analyzes also showed dilation towards fearful faces compared to other faces but, there was no significant difference between the pupillary responses for neutral faces and fearful and happy faces. Pupil size for emotional intensity was also larger for both low intensity levels and high intensity levels. Furthermore, pupil size is also known to be sensitive to decision-making processes (Hess, 1972). Participants had to use decision-making mechanisms due to the nature of the task in the current study. These uninformative results and possible interference of decision-making might have provided less discriminative information for emotion and intensity classification. Finally, when pupil size and fixation point of gaze are combined, the model gets more powerful and the information from gaze position increases the model's accuracy while classifying within a certain emotion or a certain intensity. However, no such effect of pupil size was observed which shows that the strategy used by viewers when examining facial expressions provides more distinctive information than autonomous responses such as pupillary responses.

4.3.Random Forest Model vs. Participants

RF model classified the emotions of the faces with 86.06% accuracy for the emotion recognition task. The difference between the classification accuracies among emotions were quite small, so performance differences of participants on the emotion recognition were not observable for the model. Emotion classification was slightly higher within specific intensity levels for both tasks. This finding was expected as we limited the variability of the emotions by setting a certain intensity level so the model could classify the emotion types more precisely. Even though previous literature on expression recognition did not agree upon the diagnostic features for each expression, it seems like we follow a distinctive sequential gaze pattern through faces for each expression and this pattern continues to reveal itself with different emotional intensities as well. This suggests that even though the emotional intensity of an expression is low, we continue to adapt the same strategy for extracting the necessary information from the face. Vaidya et al. (2014) also investigated the fixation patterns for subtle and extreme expression and showed results contradictory to ours. Their model reached a better prediction for subtle expressions but, based on the fixation patterns towards the eye. Examining the whole scanning pattern instead of focusing on certain features seems to be more informative for understanding emotion recognition. Even though we did not statistically analyze the classification performance for each participants' data, when we compared the emotion classification performance of RF with the performance of participants, we observed a total of 0.10 difference between the two. Both humans and the machine showed a similar performance for classification of neutral, surprised and fearful expressions and almost equal performance for sad expressions. However, participants classified happy expressions much better than the model. This might be due to the saliency of a smiling mouth (Calvo et al., 2018; Elsherif et al., 2017). Participants may have used the peripheral information especially for happy expressions with high intensity. Thus, the machine may not have learned and recognized as well as the participants as they used the information coming outside of their direct vision. Model classified intensity levels as successful as emotion classification with 89.9% accuracy for intensity recognition tasks. We compared the error rate of the model and the error rate of participants for intensity classification and calculated a difference of 0.12 between their error rates. The model made much less errors for faces with 0% emotional intensity compared to the participants. Our eye tracking results also showed higher fixations to neutral faces in intensity recognition tasks, we also already know from our previous study that lower emotional intensities are poorly rated by participants (Türk et al., 2021). Thus, participants' eye movements seem to provide more information for the model to learn while exploring and judging the neutrality of the face. Finally, one of the reasons why intensity classification was as successful as emotion classification may be because the participants judge the intensity of the emotions while still acknowledging the expressed emotion. This is also supported by the fact that the classification accuracy remained similar, even though the task was changed which suggests a scanning pattern for emotional expressions independent from the task. Thus, it can be suggested that we follow a certain strategy when we encounter emotional facial expressions varying in intensity regardless of the task. Studies using emotion or gender discrimination tasks suggested an effect, as people tend to direct their attention towards diagnostic features for the discrimination (Schyns, Bonnar, & Gosselin, 2002). However, Vaidya et al. (2014) asked participants to either categorize an expression or search for a certain expression; people did not change their fixation patterns. They interpreted these findings as we adapt scanning strategies with respect to the emotional value of the faces rather than the task at hand.

4.4.Other Models

It is important that we are able to conduct higher-level analysis with methods such as machine learning, since the statistical analyzes we perform with eye tracking data cannot be performed on raw data, and we need to convert them to events such as fixation numbers, saccades or other secondary parameters. We fed the data we received from the eye tracker to k-NN, Decision Tree and Random Forest algorithms and we got the best results from the Random Forest model by far. High performance of Random Forest has been previously shown by Zemblys (2017) which compared 10 different machine learning algorithms (including k-NN, Decision Tree and Random Forest) with the eye movement data and found that the Random Forest algorithm is the best model for eye tracking event detection among other 9 models. In support of these findings, Random Forest has been shown to be successful in eye movement event detection (Fikri et al., 2021), prediction of facial expressions (Khan et al., 2012), and autistic spectrum traits detection (Lin et al., 2021). Also, the Decision Tree algorithm is known to perform better with binary classification (Yan et al., 2020). If we applied one vs. rest approach, we could have reached better results with the algorithm. Combining k-NN with methods such as Support Vector Machine (SVM) has widely applied in previous studies and it has been shown that it gives better classification results with large and multiclass data (Çomak & Arslan, 2008; Yuan et al., 2008; Zhang et al., 2006). Finally, the low accuracy we got from the CNN model supports that, as we discussed earlier, eye movements that lead to the recognition process of emotions are based mainly on information from eye movement sequence and duration rather than information on diagnostic features. As we created the heatmaps based on the x-y points of gaze, the information was limited to the fixation numbers on areas of interest. This model might also not have worked as the focal points on the faces are spatially very close to each other and CNN summarizes the image in the intermediate layers and loses detail. Further, we did not tune any parameters for the CNN model or manipulate the layers of the model. In summary, we could have obtained different results from other models other than Random Forest, if they were applied with different approaches.

4.5.Future Applications

Promising and predictive results of the model can propose different applications for various fields. We showed that eye movements can predict the emotional value of the face seen by the viewer. Results of the current study can be applied both in behavioral and eye tracking studies examining emotional face perception, in abnormal psychology and also in the field of sensor technologies. Model showed that recognition of emotions and emotional intensity on faces depends more on information in fixation and gaze patterns rather than information on facial features. Thus, visual perception is more complicated than attentional mechanisms directed to certain salient regions. Considering the eye tracking parameters which model is successful or failed at classifying emotions and intensities, future eye tracking studies can acknowledge which parameters are important to understand emotional face processing and further investigate the gaze patterns rather than the distinctive and diagnostic features on faces. Since the model provides important clues about reliable eye movements for emotion and intensity recognition, scanning strategies should be investigated based on these eye tracking parameters in depth. Eye tracking data is also frequently used in early diagnosis of autism spectrum disorder (Wan et al., 2019), and children with bipolar disorder (Deveney et al., 2012, Kim et al., 2013) by using emotion labeling tasks as people with these disorders show different scanning patterns compared to the healthy population (Dadds et al., 2008; Pelphrey et al., 2002; Spezio et al., 2007). The explanatory eye tracking features we showed for emotion recognition could be used in these kinds of researches as well and could reduce the computational cost and the time on models for diagnosis. Having said that, eye tracking devices output huge and complicated data. Especially devices with higher resolution can produce 1000 data points per second and software of most devices offers simple, visualization-oriented tool boxes such as heat map creation or describing area of interests. Results of the current study may offer the opportunity to improve the toolboxes by adding another tool for a deeper analysis of eye movements which can predict the content of the observed stimulus.

4.6.Limitations

This study showed that fixation information is important in the recognition of emotion and intensity. Although we have suggested that the fixation point of gaze has an important role in these processes, fixation sequences for emotions and emotional intensities can be examined in depth in future studies. Moreover, the study can be fed to more models by collecting data with different emotional expression datasets in order to implement the proposed application areas.

4.7.Conclusion

The main aim of the current study was to explore the responsible eye movement parameters for recognition of facial expressions and emotional intensity. We implemented machine learning algorithms to find out the informative value of different parameters. Our analysis on eye tracking data showed that both emotional expressions and the emotional intensity have an effect on fixation numbers and pupil size of people during recognition tasks. The results we obtained from machine learning models showed that especially fixation related primary eye movement parameters have an important role in classifying emotion types and intensity levels. We concluded that people follow a certain scan path for each emotional expression varying intensity levels and each emotional intensity.

REFERENCES

Abdallah, A. S., Elliott, L. J., & Donley, D. (2020, August). *Toward Smart Internet of Things (IoT) Devices: Exploring the Regions of Interest for Recognition of Facial Expressions using Eye-gaze Tracking*. In 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-4). IEEE.

Alshehri, M., & Alghowinem, S. (2013, October). *An exploratory study of detecting emotion states using eye-tracking technology*. In 2013 Science and Information Conference (pp. 428-433). IEEE.

Arriaga, P., Adrião, J., Madeira, F., Cavaleiro, I., Maia e Silva, A., Barahona, I., & Esteves, F. (2015). A "dry eye" for victims of violence: Effects of playing a violent video game on pupillary dilation to victims and on aggressive behavior. *Psychology of Violence, 5*(2), 199.

Babiker, A., Faye, I., Prehn, K., & Malik, A. (2015). Machine learning to differentiate between positive and negative emotions using pupil diameter. *Frontiers in psychology*, *6*, 1921.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition and Emotion*, *28*, 416 – 432. http://dx.doi.org/10.1080/ 02699931.2013.833500

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, *17*(7), 1107.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838.

Bradley, M. M., & Lang, P. J. (2015). Memory, emotion, and pupil diameter: Repetition of natural scenes. *Psychophysiology*, *52*(9), 1186-1193.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607.

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology, 46*(1), 1-11.

Burley, D. T., Gray, N. S., & Snowden, R. J. (2017). As far as the eye can see: Relationship between psychopathic traits and pupil response to affective stimuli. *PLoS One*, *12*(1), e0167436.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research, 41*(9), 1179-1208.

Calvo, M. G., & Beltrán, D. (2013). Recognition advantage of happy faces: tracing the neurocognitive processes. *Neuropsychologia, 51*(11), 2051-2061.

Calvo, M. G., Avero, P., Fernández-Martín, A., & Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion, 16*(8), 1186.

Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., & Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Scientific reports, 8*(1), 1-10.

Calvo, M. G., Gutiérrez-García, A., & Del Líbano, M. (2018). What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychological Research*, *82*(2), 296-309.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior research methods*, *40*(1), 109-115.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, *137*(3), 471.

Cao, H., & Elliott, F. (2021, September). *Analysis of Eye Fixations During Emotion Recognition in Talking Faces*. In 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 1-7). IEEE.

Carbon, C. C. (2020). Wearing face masks strongly confuses counterparts in reading emotions. *Frontiers in Psychology*, 2526.

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72*(1), 164–176. https://doi.org/10.1037/0022-3514.72.1.164

Ramosaj, B., & Pauly, M. (2019). Consistent estimation of residual variance with random forest Out-Of-Bag errors. *Statistics & Probability Letters*, 151, 49-57.

Çomak, E. & Arslan, A. (2008). A new training method for support vector machines: Clustering k-NN support vector machines. *Expert Systems with Applications, 35*, 564- 568.

Dadds, M.R., El Masry, Y., Wimalaweera S., Guastella, A.J. (2008) Reduced eye gaze explains “fear blindness” in childhood psychopathic traits. *J Am Acad Child Adolesc Psychiatry, 47*, 455-63.

Deveney, C. M., Brotman, M. A., Decker, A. M., Pine, D. S., & Leibenluft, E. (2012). Affective prosody labeling in youths with bipolar disorder or severe mood dysregulation. *Journal of child psychology and psychiatry, and allied disciplines*, *53*(3), 262–270. https://doi.org/10.1111/j.1469-7610.2011.02482.x

Duchowski, A.T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods 34*(4), 455–470.

Duchowski, A. (2007). Eye tracking techniques. *In Eye tracking methodology* (pp. 51 59). Springer, London.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, *11*(4), 860.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384

Ekman, P. (1992). An argument for basic emotions, *Cognition & Emotion*, *6*(3-4), 169–200.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology, 17*(2), 124.

Elsherif, M. M., Saban, M. I., & Rotshtein, P. (2017). The perceptual saliency of fearful eyes and smiles: A signal detection study. *PloS one*, *12*(3), e0173199

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Fenn, W. O., & Hursh, J. B. (1936). Movements of the eyes when the lids are closed. *American Journal of Physiology-Legacy Content*, *118*(1), 8-14.

Fikri, M. A., Santosa, P. I., & Wibirama, S. (2021, October). A Review on Opportunities and Challenges of Machine Learning and Deep Learning for Eye Movements Classification. In *2021 IEEE International Biomedical Instrumentation and Technology Conference (IBITeC)* (pp. 65-70). IEEE.

Gao, X., & Maurer, D. (2010). A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of experimental child psychology*, *107*(2), 67-86.

Geangu, E., Hauf, P., Bhardwaj, R., & Bentz, W. (2011). Infant pupil diameter changes in response to others' positive and negative emotions. *PloS one*, *6*(11), e27132.

Grahlow, M., Rupp, C. I., & Derntl, B. (2022). The impact of face masks on emotion recognition performance and perception of threat. *PLoS one*, *17*(2), e0262840. https://doi.org/10.1371/journal.pone.0262840

Green, M. J., Williams, L. M., & Davidson, D. (2003). In the face of danger: Specific viewing strategies for facial expressions of threat? *Cognition and Emotion*, *17*, 779-786.

Guo, K. (2012). Holistic gaze strategy to categorize facial expression of varying intensities. *PloS one*, *7*(8): e42585. https://doi.org/10.1371/journal.pone.0042585

Guo, J. J., Zhou, R., Zhao, L. M., & Lu, B. L. (2019, July). Multimodal emotion recognition from eye image, eye movement and EEG using deep neural networks. *In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3071-3074). IEEE

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2014). Modulation of the initial light reflex during affective picture viewing. *Psychophysiology*, *51*(9), 815 818.

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry*, *47*(11), 1098-1106.

Hess, E.H. (1972). Pupillometrics. In: Greenfield, N.S., Sternbach, R.A. (Eds.), *Handbook of Psychophysiology*. Holt, Richard & Winston, New York, NY, pp. 491–531.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*(3423), 349-350.

Horley, K., Williams, L., Gonsalvez, C., & Gordon, E. (2004). Face to face: Visual scan path evidence for abnormal processing of facial expressions in social phobia. *Psychiatry Research*, *127*, 43-53.

Horng, W. B., Chen, C. Y., Chang, Y., & Fan, C. H. (2004, March). Driver fatigue detection based on eye tracking and dynamic template matching. *In IEEE International Conference on Networking, Sensing and Control*, (Vol. 1, pp. 7 12). IEEE.

Hunnius, S., de Wit, T. C., Vrins, S., & von Hofsten, C. (2011). Facing threat: Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and fearful emotional expressions. *Cognition and Emotion*, *25*(2), 193-205.

Itti, L. & Koch, C. (2001). Computational Modelling of Visual Attention. *Nature Reviews Neuroscience, 2*, 194–203.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology, 19,* 1.

Jessen, S., Altvater-Mackensen, N., & Grossmann, T. (2016). Pupillary responses reveal infants’ discrimination of facial emotions independent of conscious perception. *Cognition, 150,* 163-169.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye* (pp. 573 605). North-Holland.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583-1585.

Kahneman, D., Tursky, B., Shapiro, D., Crider, A. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79*(1p1), 164

Kashihara, K., Okanoya, K., & Kawai, N. (2014). Emotional attention modulates microsaccadic rate and direction. *Psychological Research, 78*(2), 166-179.

Khan, R. A., Meyer, A., Konik, H., & Bouakaz, S. (2012, June). Exploring human visual system: study to aid the development of automatic facial expression recognition framework. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 49-54). IEEE.

Kim, P., Arizpe, J., Rosen, B. H., Razdan, V., Haring, C. T., Jenkins, S. E et al. (2013). Impaired fixation to eyes during facial emotion labelling in children with bipolar disorder or severe mood dysregulation. *Journal of Psychiatry and Neuroscience, 38*(6), 407-416.

Kleberg, J. L., Hanqvist, C., Serlachius, E., & Högström, J. (2019). Pupil dilation to emotional expressions in adolescent social anxiety disorder is related to treatment outcome. *Journal of Anxiety Disorders, 65*, 26-33.

Kret, M. E., Roelofs, K., Stekelenburg, J., & de Gelder, B. (2013). Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience, 7*, 810.

Laeng, B., Sæther, L., Holmlund, T., Wang, C. E. A., Eisemann, M., & Halvorsen, M. (2013). Invisible emotional expressions influence social judgments and pupillary responses of both depressed and non-depressed individuals. *Frontiers in Psychology, 4,* 291.

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of neuroscience methods*, *201*(1), 196-203.

Lang, P. J., & Bradley, M. M. (2010). Emotion and the motivational brain. *Biological psychology*, *84*(3), 437-450.

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological research*, *69*(1), 22-29.

Le Meur, O., Le Callet, P., Barba, D. & Thoreau, D. (May 2006). A Coherent Computational Approach to Model Bottom-up Visual Attention, *IEEE Trans. Pattern Analysis and Machine Intelligence, 28*(5), 802-817.

Lin, Y., Gu, Y., Xu, Y., Hou, S., Ding, R., & Ni, S. (2022). Autistic spectrum traits detection and early screening: A machine learning based eye movement study. *Journal of Child and Adolescent Psychiatric Nursing*, *35*(1), 83-92.

Lu, Y., Zheng, W. L., Li, B., & Lu, B. L. (2015, June). Combining eye movements and EEG to enhance emotion recognition. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 ieee computer society* *conference on computer vision and pattern recognition-workshops* (pp. 94 101). IEEE

Meaux, E. & Vuilleumier, P. (2016). Facing mixed emotions: Analytic and holistic perception of facial emotion expressions engages separate brain networks. *NeuroImage 141*, 154–173.

Montagne, B., Kessels, R. P., De Haan, E. H., & Perrett, D. I. (2007). The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and motor skills, 104*(2), 589 598.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers, 36*(4), 634-638.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies, 59*(1-2), 185 198.

Pelphrey, K.A., Sasson, N., Reznick, J.S., Paul, G., Goldman, B.D. & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 3,* 249-61.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research*, *15*(1), 31-45.

Peterson, M.F., Eckstein, M.P. (2012). Looking just below the eyes is optimal across face recognition tasks. In: *Proceedings of the National Academy of Sciences of the United States of America.* 109. pp. E3314–E3323. https://doi.org/10.1073/pnas. 1214269109 48

Porges, S. W. (2003). Social engagement and attachment: a phylogenetic perspective. *Annals of the New York Academy of Sciences*, *1008*(1), 31-47.

Prabha, A.J., & Bhargavi, R. (2019). Prediction of dyslexia from eye movements using machine learning. *IETE Journal of Research*, 1-10.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, *10*(3).

Prunty, J. E., Keemink, J. R., & Kelly, D. J. (2021). Infants show pupil dilatory responses to happy and angry facial expressions. *Developmental science*, *25*(2), e13182.

Raatikainen, P., Hautala, J., Loberg, O., Kärkkäinen, T., Leppänen, P., & Nieminen, P. (2021). Detection of developmental dyslexia with machine learning using eye movement data. *Array, 12,* 100087.

Recio, G., Schacht, A., & Sommer, W. (2013). Classification of dynamic facial expressions of emotion presented briefly. *Cognition & Emotion*, *27*(8), 1486 1494.

Recio, G., Schacht, A., & Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: Specificity and intensity effects in event-related brain potentials. *Biological Psychology, 96*, 111-125.

Rensink, R.A. (2000). The dynamic representation of scenes. *Visual Cognition, 7* (1 3), 17-42.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102.

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic features of emotional expressions are processed preferentially. *PloS one*, *7*(7), e41792.

Schmidhuber, J. (2015). Deep learning in neural networks: An Overview. *Neural Networks, 61,* 85–117

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision, 14*(13), 14-14.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402-409.

Siegle, G. J., Steinhauer, S. R., Carter, C. S., Ramel, W., & Thase, M. E. (2003). Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. *Cognitive Therapy and Research, 27*(3), 365-382.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184-189.

Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology, 53*(8), 1217 1223.

Spezio, M. L., Adolphs, R., Hurley, R. S., & Piven, J. (2007). Analysis of face gaze in autism using “Bubbles”. *Neuropsychologia, 45*(1), 144-151.

Sun, L., Ren, J., & He, W. (2017). Neural correlates of facial expression processing during a detection task: An ERP study. *Plos one, 12*(3), e0174016.

Tamuly, S., Jyotsna, C., & Amudha, J. (2019, July). Tracking eye movements to predict the valence of a scene. In *2019 10th International Conference on*

Computing, Communication and Networking Technologies (ICCCNT) (1-7). IEEE.

Tanaka, J. W., Kaiser, M. D., Butler, S. & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion, 26*(6), 961-977.

Tarnowski, P., Kołodziej, M., Majkowski, A., & Rak, R. J. (2020). Eye-tracking analysis for emotion recognition. *Computational Intelligence and Neuroscience.*

Tavakoli H., Atyabi, A., Rantanen, A., Laukka, S.J., Nefti-Meziani, S., Heikkilä, J. (2015) Correction: Predicting the Valence of a Scene from Observers' Eye Movements. *PloS one, 10*(10), https://doi.org/10.1371/journal.pone.0141174

Tian, Y.L., Kanade, T. & Cohn, J. F. (2005). Facial expression analysis. in *Handbook of face recognition*. Springer, 247–275.

Torre, F. D. L., & Cohn, J. F. (2011). *Facial expression analysis. Visual analysis of humans*, 377-409.

Torro-Alves, N., Bezerra, I. A. D. O., Claudino, R. G. E., & Pereira, T. C. L. (2013). Influences of sex, type and intensity of emotion in the recognition of static and dynamic facial expressions. *Avances en Psicología Latinoamericana, 31*(1), 192-199.

Türk, D.D., Eskikurt, G., Tasliyurt, S. & Yildirim, F. (2021). Neural Underpinnings of Negative Emotional Expressions Evoked by Audiovisual Stimuli, *in 43rd European Conference on Visual Perception (ECVP) 2021 Online.* https://doi.org/10.1177/03010066211059887

Türk, D. D., & Yildirim, F. (2022, February 3). Attention towards Facial Regions Varies between Emotion Types but not between Attachment Styles. PsyArXiv, https://doi.org/10.31234/osf.io/268vw

Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition, 133*(2), 443-456.

Van Steenbergen, H., Band, G. P., & Hommel, B. (2011). Threat but not arousal narrows attention: evidence from pupil dilation and saccade control. *Frontiers in Psychology, 2,* 281.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron, 30*(3), 829-841.

Walther, D. & Koch, C. (2006). Modeling Attention to Salient Proto-Objects, *Neural Networks, 19*(9), 1395-1407.

Wan, G., Kong, X., Sun, B., Yu, S., Tu, Y., Park, J., Lang, C., Koh, M., Wie, Z., Feng, Z., Lin, Y. & Kong, J. (2019). Applying eye tracking to identify autism spectrum disorder in children. *Journal of autism and developmental disorders*, *49*(1), 209-215.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS one, 12*(5), e0177239.

Yan, J., Zhang., Z., Lin, K., Yang, F. & Luo, X. (2020). A hybrid scheme-based one vs-all decision trees for multi-class classification tasks, *Knowledge-Based Systems*, *198,* https://doi.org/10.1016/j.knosys.2020.105922.

Yitzhak, N., Pertzov, Y., Guy, N., & Aviezer, H. (2020). Many ways to see your feelings: Successful facial expression recognition occurs with diverse patterns of fixation distributions. *Emotion (Washington, D.C.)*, 10.1037/emo0000812. Advance online publication. https://doi.org/10.1037/emo0000812

Yuan, P., Chen, Y., Jin, H. & Huang, L. (2008). MSVM-kNN: Combining SVM and k-NN for Multi-Class Text Classification. *International Workshop on Semantic Computing and Systems.* doi: 10.1109/WSCS.2008.36.

Zemblys, R., Niehorster, D. C., Komogortsev, O. & Holmqvist, K. (2018). Using machine learning to detect events in eye-tracking data. *Behavior Research Methods*, *50*(1), 160-181, 2018.

Zemblys, R. (2017). Eye-movement event detection meets machine learning. *BioMed 2017*

Zhang, H., Berg, A.C., maire, M. & Malik, J. (2006). SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. *Computer Society Conference on Computer Vision and Pattern Recognition.*

APPENDIX A: Training Accuracy Scores and Out of Bag Scores of Random Forest

	RF Emotion Classification (Emotion Recognition Task)			RF Intensity Classification (Emotion Recognition Task)		
Parameters	*OOB Score*	*Training Accuracy*	*Testing Accuracy*	*OOB Score*	*Training Accuracy*	*Testing Accuracy*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	,86255	1,00	,86065	,86963	1,00	,87161
Fpog Lpog Rpog Fpogd Pupild	,86081	1,00	,85998	,86746	1,00	,87100
Fpog Fpogd Pupild	,72980	1,00	,73313	,72951	1,00	,72939
Fpog Pupild	,57306	1,00	,57213	,55500	1,00	,55478
Fpog	,34661	1,00	,34644	,34393	1,00	,34602
Fpogd	,22769	,40	,23186	,22800	,39	,23118
Pupild	,22520	,85	,22709	,22287	,85	,22418

	RF Emotion Classification (Intensity Recognition Task)			RF Intensity Classification (Intensity Recognition Task)		
Parameters	*OOB Score*	*Training Accuracy*	*Testing Accuracy*	*OOB Score*	*Training Accuracy*	*Testing Accuracy*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	.89555	1.00	.89707	.89477	1.00	.89370
Fpog Lpog Rpog Fpogd Pupild	.89916	1.00	.89079	.89327	1.00	.89310
Fpog Fpogd Pupild	.78974	1.00	.79103	.78848	1.00	.79290
Fpog Pupild	.61174	1.00	.60850	.60213	1.00	.60627
Fpog	.38848	1.00	.39070	.36249	1.00	.36494
Fpogd	.24508	.49	.24805	.22956	.47	.22168
Pupild	.25666	.92	.25724	.21996	.91	.22215

APPENDIX B: Additional Outputs of Random Forest

Figure 1

Accuracy heatmap of intensity classification within each emotion type in emotion recognition task

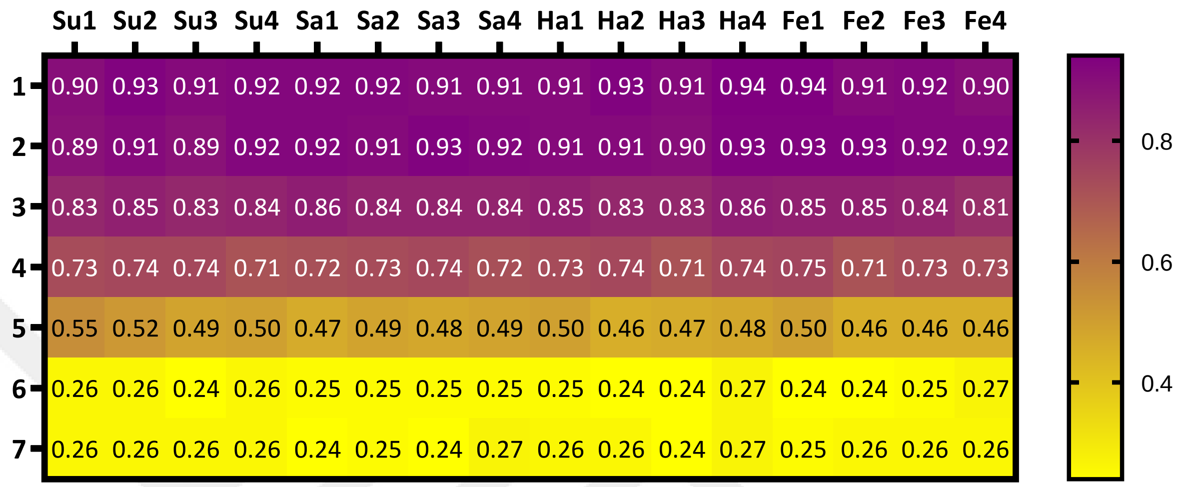
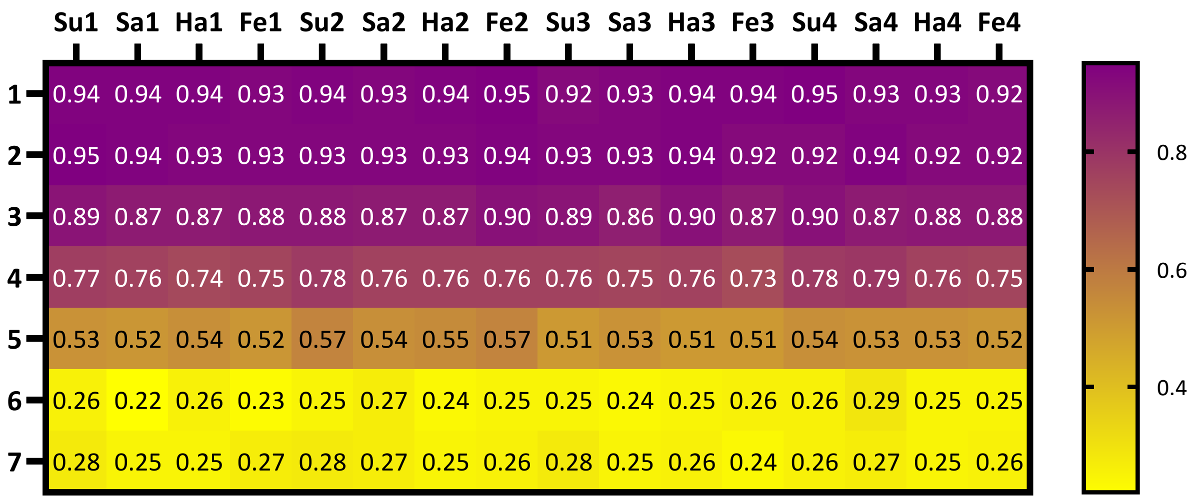


Figure 2

Accuracy heatmap of emotion classification within each intensity level in intensity recognition task



Note. (1) fpog, rfpog, lfpog, fpogd, lpupild, fixation numbers on AOIs, (2) fpog, rfpog, lfpog, fpogd, lpupild, (3) fpog, fpogd, lpupild, (4) fpog, lpupild, (5) fpog, (6) fpod, (7) lpupild.

APPENDIX C: k-NN Outputs

	k-NN Emotion Classification (Emotion Recognition Task)					k-NN Intensity Classification (Emotion Recognition Task)				
Parameters	*Emotion*	*k*	*Precision*	*Recall*	*F1-score*	*Intensity*	*k*	*Precision*	*Recall*	*F1-score*
Fpog Lpog	Neutral	38	.41	.11	.17	0%	40	.39	.09	.15
Rpog	Surprised		.42	.43	.43	25%		.40	.47	.43
Fpogd	Sad		.41	.46	.43	50%		.40	.46	.43
Pupild	Happy		.42	.45	.43	75%		.41	.41	.41
Fix. On AOIs	Fearful		.42	.42	.42	100%		.41	.38	.40
Fpog	Neutral	38	.44	.10	.17	0%	38	.41	.10	.16
Lpog	Surprised		.43	.43	.43	25%		.40	.48	.44
Rpog	Sad		.40	.47	.44	50%		.40	.45	.43
Fpogd	Happy		.42	.45	.43	75%		.40	.41	.41
Pupild	Fearful		.42	.42	.42	100%		.41	.38	.39
Fpog	Neutral	38	.35	.06	.10	0%	38	.36	.06	.10
Fpogd	Surprised		.36	.38	.37	25%		.35	.41	.38
Pupild	Sad		.35	.40	.37	50%		.34	.39	.37
	Happy		.36	.38	.37	75%		.34	.35	.35
	Fearful		.35	.36	.36	100%		.36	.33	.35
Fpog	Neutral	38	.35	.05	.09	0%	38	.35	.05	.09
Pupild	Surprised		.34	.37	.35	25%		.33	.40	.36
	Sad		.33	.37	.35	50%		.32	.36	.34
	Happy		.33	.36	.35	75%		.33	.33	.33
	Fearful		.34	.33	.34	100%		.34	.32	.33
Fpog	Neutral	38	.31	.09	.14	0%	38	.31	.10	.15
	Surprised		.38	.40	.39	25%		.36	.41	.38
	Sad		.35	.38	.36	50%		.36	.39	.38
	Happy		.36	.39	.38	75%		.36	.37	.36
	Fearful		.37	.36	.37	100%		.36	.34	.35
Fpogd	Neutral	38	.03	.00	.00	0%	38	.03	.00	.00
	Surprised		.22	.26	.24	25%		.24	.29	.26
	Sad		.23	.25	.24	50%		.23	.28	.25
	Happy		.24	.26	.25	75%		.23	.22	.22
	Fearful		.22	.22	.22	100%		.24	.23	.23
Pupild	Neutral	38	.12	.00	.00	0%	37	.15	.00	.01
	Surprised		.24	.30	.26	25%		.24	.30	.27
	Sad		.24	.26	.25	50%		.24	.27	.25
	Happy		.24	.25	.25	75%		.23	.24	.24
	Fearful		.24	.22	.23	100%		.24	.22	.23

	k-NN Emotion Classification (Intensity Recognition Task)					k-NN Intensity Classification (Intensity Recognition Task)				
Parameters	*Emotion*	*k*	*Precision*	*Recall*	*F1-score*	*Intensity*	*k*	*Precision*	*Recall*	*F1-score*
Fpog Lpog	Neutral		NaN	NaN	NaN	0%		.43	.14	.22
Rpog	Surprised		.45	.46	.45	25%		.43	.49	.46
Fpogd	Sad	40	.44	.47	.46	50%	38	.43	.47	.45
Pupild	Happy		.45	.44	.45	75%		.43	.44	.43
Fix. On AOIs	Fearful		.46	.43	.44	100%		.44	.42	.43
Fpog	Neutral		NaN	NaN	NaN	0%		.45	.13	.20
Lpog	Surprised		.46	.47	.46	25%		.42	.50	.46
Rpog	Sad	38	.44	.48	.46	50%	38	.43	.45	.44
Fpogd	Happy		.46	.44	.45	75%		.43	.44	.44
Pupild	Fearful		.46	.43	.45	100%		.44	.42	.43
Fpog	Neutral		NaN	NaN	NaN	0%		.32	.05	.09
Fpogd	Surprised		.39	.42	.40	25%		.34	.40	.37
Pupild	Sad	38	.39	.42	.41	50%		.33	.36	.35
	Happy		.39	.37	.38	75%		.34	.35	.34
	Fearful		.39	.35	.37	100%		.34	.32	.33
Fpog	Neutral		NaN	NaN	NaN	0%		.32	.05	.09
Pupild	Surprised		.36	.38	.37	25%		.34	.40	.37
	Sad	38	.37	.39	.38	50%	38	.33	.36	.35
	Happy		.37	.36	.37	75%		.34	.35	.34
	Fearful		.36	.32	.34	100%		.34	.32	.33
	Neutral		NaN	NaN	NaN	0%		.35	.12	.18
	Surprised		.39	.41	.40	25%		.37	.42	.40
Fpog	Sad	38	.40	.43	.41	50%	40	.37	.39	.38
	Happy		.40	.39	.40	75%		.37	.39	.38
	Fearful		.41	.37	.39	100%		.37	.36	.36
	Neutral		NaN	NaN	NaN	0%		.02	.00	.00
	Surprised		.25	.31	.28	25%		.23	.29	.26
Fpogd	Sad	38	.26	.27	.27	50%	38	.23	.26	.24
	Happy		.24	.22	.23	75%		.23	.23	.23
	Fearful		.25	.22	.23	100%		.23	.21	.22
	Neutral		NaN	NaN	NaN	0%		.16	.01	.01
	Surprised		.27	.31	.29	25%		.23	.29	.26
Pupild	Sad	38	.28	.30	.29	50%	38	.24	.26	.25
	Happy		.27	.25	.26	75%		.24	.24	.24
	Fearful		.27	.22	.24	100%		.24	.22	.23

APPENDIX D: Decision Tree Outputs

	DT Emotion Classification (Emotion Recognition Task)				DT Intensity Classification (Emotion Recognition Task)			
Parameters	*Emotion*	*Precision*	*Recall*	*F1-score*	*Intensity*	*Precision*	*Recall*	*F1-score*
Fpog Lpog	Neutral	.00	.00	.00	0%	.00	.00	.00
Rpog	Surprised	.26	.24	.25	25%	.23	.49	.32
Fpogd	Sad	.25	.41	.31	50%	.25	.19	.22
Pupild	Happy	.27	.29	.28	75%	.24	.37	.29
Fix. On AOIs	Fearful	.27	.18	.21	100%	.00	.00	.00
Fpog	Neutral	.00	.00	.00	0%	.00	.00	.00
Lpog	Surprised	.00	.00	.00	25%	.26	.27	.26
Rpog	Sad	.26	.21	.23	50%	.24	.61	.34
Fpogd	Happy	.26	.45	.33	75%	.00	.00	.00
Pupild	Fearful	.24	.43	.31	100%	.24	.18	.21
Fpog	Neutral	.00	.00	.00	0%	.00	.00	.00
Fpogd	Surprised	.29	.13	.18	25%	.25	.20	.22
Pupild	Sad	.00	.00	.00	50%	.23	.48	.31
	Happy	.26	.54	.35	75%	.26	.23	.24
	Fearful	.25	.44	.31	100%	.24	.14	.18
Fpog	Neutral	.00	.00	.00	0%	.00	.00	.00
Pupild	Surprised	.26	.31	.29	25%	.25	.27	.26
	Sad	.24	.25	.25	50%	.24	.46	.32
	Happy	.26	.43	.32	75%	.25	.33	.28
	Fearful	.26	.11	.15	100%	.00	.00	.00
	Neutral	.00	.00	.00	0%	.00	.00	.00
	Surprised	.26	.27	.27	25%	.24	.23	.24
Fpog	Sad	.00	.00	.00	50%	.24	.26	.25
	Happy	.26	.13	.18	75%	.25	.43	.32
	Fearful	.24	.67	.35	100%	.24	.14	.18
	Neutral	.00	.00	.00	0%	.00	.00	.00
	Surprised	.23	.10	.14	25%	.25	.32	.28
Fpogd	Sad	.00	.00	.00	50%	.24	.13	.17
	Happy	.24	.36	.29	75%	.24	.60	.34
	Fearful	.24	.57	.34	100%	.00	.00	.00
	Neutral	.00	.00	.00	0%	.00	.00	.00
	Surprised	.24	.30	.27	25%	.24	.23	.24
Pupild	Sad	.24	.22	.23	50%	.25	.29	.27
	Happy	.23	.41	.30	75%	.25	.28	.26
	Fearful	.25	.11	.15	100%	.24	.26	.25

	DT Emotion Classification (Intensity Recognition Task)				DT Intensity Classification (Intensity Recognition Task)			
Parameters	*Emotion*	*Precision*	*Recall*	*F1-score*	*Intensity*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.27	.27	.27	25%	.25	.20	.22
	Sad	.26	.49	.34	50%	.24	.48	.32
	Happy	.28	.20	.23	75%	.25	.38	.30
	Fearful	.27	.11	.16	100%	.00	.00	.00
Fpog Lpog Rpog Fpogd Pupild	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.27	.38	.32	25%	.24	.76	.36
	Sad	.30	.35	.32	50%	.25	.15	.19
	Happy	.29	.11	.16	75%	.27	.16	.20
	Fearful	.27	.29	.28	100%	.00	.00	.00
Fpog Fpogd Pupild	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.27	.30	.29	25%	.25	.33	.29
	Sad	.29	.52	.37	50%	.24	.59	.34
	Happy	.00	.00	.00	75%	.28	.14	.19
	Fearful	.28	.30	.29	100%	.00	.00	.00
Fpog Pupild	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.26	.63	.37	25%	.26	.38	.31
	Sad	.29	.14	.19	50%	.23	.54	.33
	Happy	.26	.29	.27	75%	.28	.15	.20
	Fearful	.00	.00	.00	100%	.00	.00	.00
Fpog	Neutral	.00	.00	.00	0%	.00	.00	.00
	Surprised	.27	.45	.33	25%	.25	.37	.30
	Sad	.26	.10	.15	50%	.24	.56	.33
	Happy	.28	.12	.17	75%	.28	.12	.17
	Fearful	.26	.39	.31	100%	.00	.00	.00
Fpogd	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.25	.36	.30	25%	.00	.00	.00
	Sad	.26	.33	.29	50%	.24	.36	.29
	Happy	.00	.00	.00	75%	.24	.56	.34
	Fearful	.26	.34	.30	100%	.23	.11	.15
Pupild	Neutral	NaN	NaN	NaN	0%	.00	.00	.00
	Surprised	.26	.37	.31	25%	.24	.42	.30
	Sad	.29	.45	.35	50%	.24	.47	.32
	Happy	.27	.28	.27	75%	.26	.15	.19
	Fearful	.00	.00	.00	100%	.00	.00	.00

APPENDIX E: k-NN Training Accuracy Scores

	k-NN Emotion Classification (Emotion Recognition Task)					k-NN Intensity Classification (Emotion Recognition Task)				
Parameters	*Emotion*	*k*	*Precision*	*Recall*	*F1-score*	*Intensity*	*k*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral		.54	.14	.22	0%		.52	.14	.22
	Surprised		.47	.49	.48	25%		.45	.53	.49
	Sad	38	.46	.53	.49	50%	40	.46	.52	.49
	Happy		.47	.50	.49	75%		.47	.47	.47
	Fearful		.48	.48	.48	100%		.47	.44	.46
Fpog Lpog Rpog Fpogd Pupild	Neutral		.55	.13	.20	0%		.55	.14	.22
	Surprised		.48	.49	.48	25%		.46	.55	.50
	Sad	38	.46	.53	.49	50%	38	.46	.53	.49
	Happy		.48	.50	.49	75%		.47	.48	.48
	Fearful		.48	.49	.48	100%		.49	.45	.47
Fpog Fpogd Pupild	Neutral		.48	.07	.13	0%		.46	.08	.14
	Surprised		.42	.45	.43	25%		.41	.48	.44
	Sad	38	.41	.47	.44	50%	38	.41	.47	.44
	Happy		.42	.44	.43	75%		.41	.43	.42
	Fearful		.43	.43	.43	100%		.43	.39	.41
Fpog Pupild	Neutral		.46	.07	.12	0%		.45	.08	.13
	Surprised		.40	.43	.41	25%		.38	.47	.42
	Sad	38	.39	.44	.41	50%	38	.39	.43	.41
	Happy		.40	.43	.41	75%		.39	.40	.39
	Fearful		.41	.39	.40	100%		.41	.38	.39
	Neutral		.38	.11	.18	0%		.39	.13	.19
	Surprised		.41	.45	.43	25%		.40	.47	.43
Fpog	Sad	38	.40	.44	.42	50%	38	.40	.44	.42
	Happy		.41	.45	.43	75%		.41	.41	.41
	Fearful		.42	.41	.41	100%		.41	.40	.41
	Neutral		.28	.00	.01	0%		.22	.00	.01
	Surprised		.29	.34	.31	25%		.29	.36	.32
Fpogd	Sad	38	.28	.35	.31	50%	38	.29	.30	.30
	Happy		.29	.30	.30	75%		.29	.32	.30
	Fearful		.30	.27	.28	100%		.30	.28	.29
	Neutral		.29	.01	.01	0%		.27	.01	.02
	Surprised		.32	.39	.35	25%		.31	.38	.34
Pupild	Sad	38	.32	.35	.33	50%	37	.32	.36	.34
	Happy		.32	.34	.33	75%		.32	.33	.32
	Fearful		.32	.30	.31	100%		.33	.31	.32

	k-NN Emotion Classification (Intensity Recognition Task)					k-NN Intensity Classification (Intensity Recognition Task)				
Parameters	*Emotion*	*k*	*Precision*	*Recall*	*F1-score*	*Intensity*	*k*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral	40	NaN	NaN	NaN	0%	38	.54	.18	.26
	Surprised		.51	.51	.51	25%		.48	.55	.51
	Sad		.50	.53	.51	50%		.48	.53	.50
	Happy		.50	.51	.51	75%		.48	.49	.49
	Fearful		.51	.48	.49	100%		.50	.48	.49
Fpog Lpog Rpog Fpogd Pupild	Neutral	38	NaN	NaN	NaN	0%	38	.56	.16	.25
	Surprised		.52	.52	.52	25%		.48	.56	.51
	Sad		.50	.53	.52	50%		.49	.52	.50
	Happy		.51	.51	.51	75%		.48	.50	.49
	Fearful		.52	.49	.50	100%		.50	.49	.49
Fpog Fpogd Pupild	Neutral	38	NaN	NaN	NaN	0%		.50	.09	.16
	Surprised		.45	.47	.46	25%		.42	.48	.45
	Sad		.45	.48	.47	50%		.43	.47	.45
	Happy		.45	.44	.45	75%		.42	.45	.43
	Fearful		.46	.42	.44	100%		.44	.43	.43
Fpog Pupild	Neutral	38	NaN	NaN	NaN	0%	38	.45	.09	.16
	Surprised		.43	.45	.44	25%		.39	.46	.42
	Sad		.42	.45	.44	50%		.40	.43	.41
	Happy		.43	.42	.43	75%		.40	.42	.41
	Fearful		.43	.38	.41	100%		.41	.38	.40
Fpog	Neutral	38	NaN	NaN	NaN	0%	40	.40	.14	.21
	Surprised		.44	.47	.45	25%		.42	.47	.44
	Sad		.44	.46	.45	50%		.42	.44	.43
	Happy		.45	.44	.44	75%		.42	.44	.43
	Fearful		.46	.42	.44	100%		.43	.42	.42
Fpogd	Neutral	38	NaN	NaN	NaN	0%	38	.24	.00	.01
	Surprised		.32	.38	.34	25%		.30	.36	.32
	Sad		.32	.33	.33	50%		.30	.35	.32
	Happy		.32	.29	.30	75%		.30	.33	.31
	Fearful		.33	.29	.31	100%		.31	.26	.28
Pupild	Neutral	38	NaN	NaN	NaN	0%	38	.26	.01	.01
	Surprised		.34	.38	.36	25%		.31	.39	.34
	Sad		.35	.37	.36	50%		.32	.36	.34
	Happy		.34	.33	.34	75%		.32	.33	.33
	Fearful		.35	.29	.32	100%		.33	.30	.31

APPENDIX F: Decision Tree Training Accuracy Scores

	DT Emotion Classification (Emotion Recognition Task)				DT Intensity Classification (Emotion Recognition Task)			
Parameters	*Emotion*	*Precision*	*Recall*	*F1-score*	*Intensity*	*Precision*	*Recall*	*F1-score*
Fpog Lpog Rpog Fpogd Pupild Fix. On AOIs	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,30	,31	,31	25%	,24	,51	,33
	Sad	,27	,59	,37	50%	,26	,20	,23
	Happy	,00	,00	,00	75%	,25	,37	,30
	Fearful	,30	,24	,27	100%	,00	,00	,00
Fpog Lpog Rpog Fpogd Pupild	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,00	,00	,00	25%	,26	,26	,26
	Sad	,26	,20	,23	50%	,24	,62	,35
	Happy	,26	,45	,33	75%	,00	,00	,00
	Fearful	,25	,45	,32	100%	,24	,19	,21
Fpog Fpogd Pupild	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,29	,13	,18	25%	,26	,21	,23
	Sad	,00	,00	,00	50%	,25	,50	,33
	Happy	,26	,53	,35	75%	,26	,24	,25
	Fearful	,25	,45	,32	100%	,24	,15	,18
Fpog Pupild	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,27	,32	,29	25%	,25	,27	,26
	Sad	,25	,27	,26	50%	,24	,47	,32
	Happy	,26	,42	,32	75%	,24	,32	,28
	Fearful	,26	,11	,16	100%	,00	,00	,00
Fpog	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,27	,28	,27	25%	,24	,23	,24
	Sad	,00	,00	,00	50%	,24	,26	,25
	Happy	,27	,13	,18	75%	,25	,43	,32
	Fearful	,24	,67	,35	100%	,24	,14	,18
Fpogd	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,24	,11	,15	25%	,25	,32	,28
	Sad	,00	,00	,00	50%	,24	,13	,17
	Happy	,25	,38	,30	75%	,24	,60	,34
	Fearful	,24	,56	,34	100%	,00	,00	,00
Pupild	Neutral	,00	,00	,00	0%	,00	,00	,00
	Surprised	,24	,29	,26	25%	,25	,23	,24
	Sad	,25	,22	,23	50%	,25	,29	,27
	Happy	,24	,42	,30	75%	,25	,29	,27
	Fearful	,25	,11	,15	100%	,24	,26	,25

	DT Emotion Classification (Intensity Recognition Task)				DT Intensity Classification (Intensity Recognition Task)			
Parameters	*Emotion*	*Precision*	*Recall*	*F1-score*	*Intensity*	*Precision*	*Recall*	*F1-score*
Fpog Lpog	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
Rpog	Surprised	,27	,27	,27	25%	,25	,20	,22
Fpogd	Sad	,27	,50	,35	50%	,24	,48	,32
Pupild	Happy	,28	,19	,23	75%	,25	,38	,30
Fix, On AOIs	Fearful	,28	,11	,16	100%	,00	,00	,00
Fpog	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
Lpog	Surprised	,28	,39	,33	25%	,24	,76	,36
Rpog	Sad	,30	,35	,33	50%	,25	,15	,19
Fpogd	Happy	,28	,11	,16	75%	,27	,15	,20
Pupild	Fearful	,27	,28	,28	100%	,00	,00	,00
Fpog	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
Fpogd	Surprised	,28	,31	,29	25%	,25	,33	,29
Pupild	Sad	,29	,54	,38	50%	,24	,60	,34
	Happy	,00	,00	,00	75%	,28	,14	,19
	Fearful	,28	,30	,29	100%	,00	,00	,00
Fpog	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
Pupild	Surprised	,27	,64	,38	25%	,26	,38	,31
	Sad	,30	,15	,20	50%	,24	,54	,33
	Happy	,26	,29	,27	75%	,28	,15	,20
	Fearful	,00	,00	,00	100%	,00	,00	,00
Fpog	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
	Surprised	,27	,46	,34	25%	,25	,38	,30
	Sad	,27	,11	,16	50%	,24	,57	,34
	Happy	,29	,12	,17	75%	,28	,13	,17
	Fearful	,27	,40	,32	100%	,00	,00	,00
Fpogd	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
	Surprised	,25	,36	,30	25%	,00	,00	,00
	Sad	,26	,34	,30	50%	,24	,37	,29
	Happy	,00	,00	,00	75%	,24	,57	,34
	Fearful	,27	,35	,30	100%	,24	,11	,15
Pupild	Neutral	NaN	NaN	NaN	0%	,00	,00	,00
	Surprised	,28	,39	,32	25%	,24	,42	,30
	Sad	,29	,46	,36	50%	,24	,48	,32
	Happy	,27	,28	,27	75%	,26	,15	,19
	Fearful	,00	,00	,00	100%	,00	,00	,00