Using the Deviation between Gaze Behaviour and Visual Saliency Maps to Classify Demographics

Njeri Ndungu

2635135

MSc Applied Data Science

Supervisors: Christoph Strauch

Alex J. Hoggerbrugge

Abstract

Human visual attention is a mechanism that has been highly researched with many studies focusing on demographic differences in visual attention. This analysis aims to examine these differences in visual attention across age and gender by investigating the differences in eye movements across these demographics. The dataset used in this analysis represents a unique opportunity to investigate these differences due to the quantity of data collected across a much wider range of demographics than is usually present in eye tracking studies, with a total of 534 participants included in the analysis. Saliency maps, predicted fixation locations, from 14 different saliency models were collected. The similarity between each participants fixation locations and these predicted fixation maps was calculated to create input features for classification. Logistic Regression, XGBoost and Neural Network algorithms were implemented to predict both age and gender. Classification results showed that none of these algorithms could predict either age or gender with an accuracy higher than a naïve baseline classifier. The saliency models and evaluation metrics implemented did not provide sufficient information to allow for the accurate classification of age or gender. This is a limitation of saliency models, they do not capture the demographic differences that are present in eye movements when viewing a visual scene, although this does not significantly impact the performance of these models.

Table of Contents

1. Introduction
    1.1. Human Visual Attention
    1.2. Saliency Maps as Models of Visual Attention
    1.3. Age Related Changes in Visual Attention
    1.4. Sex Related Changes in Visual Attention
2. Participants & Data Pre-processing
3. Methods
    3.1. Apparatus, Stimulus, & Procedure
    3.2. Saliency Maps
    3.3. Feature Extraction
    3.4. Machine Learning
        3.4.1. Logistic Regression
        3.4.2. XGBoost
        3.4.3. Artificial Neural Network
    3.5. Evaluation Metrics
4. Results
    4.1. Saliency
    4.2. Data Exploration
    4.3. Classification
5. Discussion
    5.1. Limitations
    5.2. Future Research
6. Conclusion

References

1. Introduction

1.1 Human Visual Attention

The human brain cannot process all elements of a visual scene at once. Visual attention is the process of selecting which objects or locations in the scene should receive more processing than others. There are a variety of factors that influence this and in general these can be separated two elements: bottom-up factors and top-down factors. Bottom-up attention is a largely involuntary process driven by the properties of the surrounding environment such as colour, border, brightness, shapes, and sizes. This type of attention is saliency driven (Theeuwes, 2010). It is assumed that top-down attention is a voluntary process similar to making an overt decision to focus attention on a particular object or location (Theeuwes, 2018). It is more complex and ambiguous than bottom-up attention and is influenced by factors including given task, emotions, and the observers' goals. Eye movements have been shown to be closely linked to visual attention (Groner & Groner, 1989). (Shepherd et al., 1986) have shown while attention can change without corresponding eye movements, it is not possible to make an eye movement without a corresponding shift in visual attention. As such eye tracking data is commonly used as a measurement of visual attention (Vehlen et al., 2021).

1.2 Saliency Maps as Models of Visual Attention

Visual saliency maps, computationally produced frequency maps, aim to predict human visual attention. These models were first proposed by Itti & Koch (Itti & Koch, 2001) with the aim of predicting how attention is distributed across a given visual scene. A saliency model takes a static image as its input and outputs a saliency map which contains the probability of each pixel grabbing viewers' attention. There are numerous methods that can be used to do this, and many models have been proposed over the last 20 years. Bottom-up models are features based, determining the salience of an area of an image by analysing the visual properties of the image. While each model is different, they do have some similarities. They extract mid and low-level image features such as orientation, texture, colour, faces, objects to feature maps. An image processing technique which attempts to model attention such as centre surround is then applied to each feature map and these feature maps are combined to create a saliency map (Riche & Mancas, 2016). While bottom-up maps focus on the features of an image to predict saliency, deep-learning saliency models are data-driven. These deep convolutional neural network models use existing networks that have been trained in object recognition to predict salient regions of an image which improves the quality of the predicted saliency maps (Kümmerer et al., 2015; Pan et al., 2016). Deep learning models have outperformed classic bottom-up saliency models however there is still a gap between the predicted saliency map and actual human gaze (Borji, 2021; Borji et al., 2019).

Evaluating the performance of saliency models is a widely researched area and at present there is no gold standard in visual saliency evaluation metrics. Most of the commonly used metrics measure the prediction ability of the saliency model while considering a different aspect of visual attention, for example Shuffled AUC excludes the effect of centre bias, whereas AUC includes centre bias. (Bylinskii et al., 2019) found that many of these metrics produce highly correlated outputs due to similar computation methods. They highlighted a so-called similarity cluster between NSS, CC and AUC. Thus, using a wide variety of metrics

to evaluate a saliency model provides more varied and detailed information on its performance.

1.3 Changes in Visual Attention with Age

Age related changes in visual attention are a well-researched area with different studies focusing on children, young adults, and the elderly (Takahashi et al., 2021; Walker et al., 2017). Most of these studies focus on overt visual attention and have shown that there are significant differences in scene viewing behaviour of observers across different age groups. (Egami et al., 2009) have shown that children from 3-6 exhibit less exploratory eye movements than 6–14-year-olds. The degree of centre bias, the tendency of the human gaze to be biased towards the centre of natural stimuli, has been shown by (Krishna et al., 2018) to decrease with age, with 4-year-olds having the highest bias among all age groups. (Açik et al., 2010) investigated developmental changes in scene viewing, particularly when free viewing natural images, and highlighted several differences. Firstly, local image features impact gaze allocation for children aged 7-9 more than for adults, which demonstrates that in children bottom-up factors drive scene exploration but as we age top-down factors become more influential. Secondly, older adults attend to narrower areas of a scene. Lastly, older adults displayed the highest number of fixations meaning they explore images more efficiently than other age groups.

1.4 Changes in Visual Attention across Sex

Similarly, the relationship between sex, eye movements and visual attention has been studied, although research into this area is more limited. (Mercer Moss et al., 2012) showed that there are differences in eye movements between men and women, with women exhibiting more exploratory eye movements and men making more frequent, shorter fixations. It has been shown that there are sex related differences in visual attention related to bottom-up processing of scenes, with women showing attention bias towards colour and men's attention being more biased towards spatial position (McGivern et al., 2019). Conversely, (Papavlasopoulou et al., 2020) found no significant difference in gaze between sexes and a second study found no sex related difference in the number of saccades produced when viewing an indoor scene, suggesting that the males and females attend to the same number of areas in a visual scene (Abdi Sargezeh et al., 2019). So, sex-based differences are a more disputed area in visual attention research.

These demographic differences in eye movements and visual attention suggest that while we all live in the same world, individuals with different demographics might view the world differently. It is well established that variations in human visual attention exist across both age and sex, it follows that our computational models of visual saliency should reflect these differences. Many of these models do not consider the impact that demographics have on the human visual attention mechanism in their design. As such it is possible that these models may not capture the variation in visual attention that has been detected in other studies using different methodologies.

The diagnostic value of eye movements and saliency maps has been demonstrated in research. (Rahman et al., 2020) proposed a framework in which the deviation between gaze behaviour and a predicted saliency map can successfully predict autism spectrum disorder and toddlers age. Developing on this premise, this study aims to assess whether there is a

difference in the way we view the world across demographics and additionally whether the current proposed saliency models capture these differences sufficiently. Several saliency maps are collected by implementing a variety of the saliency models that have been proposed in research. Actual gaze behaviour is used to generate a map of fixation locations which is compared to each of the predicted saliency maps using range of evaluation metrics. Classic machine learning algorithms and a deep learning algorithm are implemented to predict age and sex using these deviations between actual and predicted gaze behaviour. The data used in this study provides a unique opportunity to investigate these differences due to the number of participants and the range of demographics it includes.

2. Participants & Data Pre-Processing

In total there were 5604 participants, 4061 of which provided demographics. Since the experiment was not conducted in a controlled environment several steps were taken to assure the validity of the of the data that was collected. Free viewing is considered valid if there is no period of stable gaze greater than 500ms. Demographics are considered valid if there is no period of exactly stable gaze greater than 5s when selecting sex and year of birth. These periods of stable gaze likely indicate that the participant being tracked left the apparatus and the tracking was resumed on a new participant, as such these are not valid data points. Filtering in this manner returns 624 valid participants with valid free viewing and valid demographics. Due to outliers in the fixation points of some participants as a result of imprecise eye tracking or participants looking outside of the screen during free viewing, 39 data point were removed. The distribution of both age and sex can be seen in Fig. 1 (a-b). As is clear from Fig. 1 there is a significant outlier at year of birth 2000. A random sample 40 of the participants with year of birth 2000 was selected and these were the participants used for the analysis, with the remaining 36 data points being removed from the final dataset. Finally, 25 data points from participants with year of birth later than 2011 and before 1972 were removed due to the very low number of participants in these years. This resulted in a final dataset with 534 participants, the distribution of both age and sex after cleaning the data in this way is shown in Fig. 1 (c-d).

For classification, the age data was grouped into 4 separate categories of roughly equal size that were sufficiently large for analysis. Using equal width bins creates significantly imbalanced classes which would cause issues with prediction. Equal frequency binning offsets this issue and the resulting categories and their frequencies are shown in Table 1.

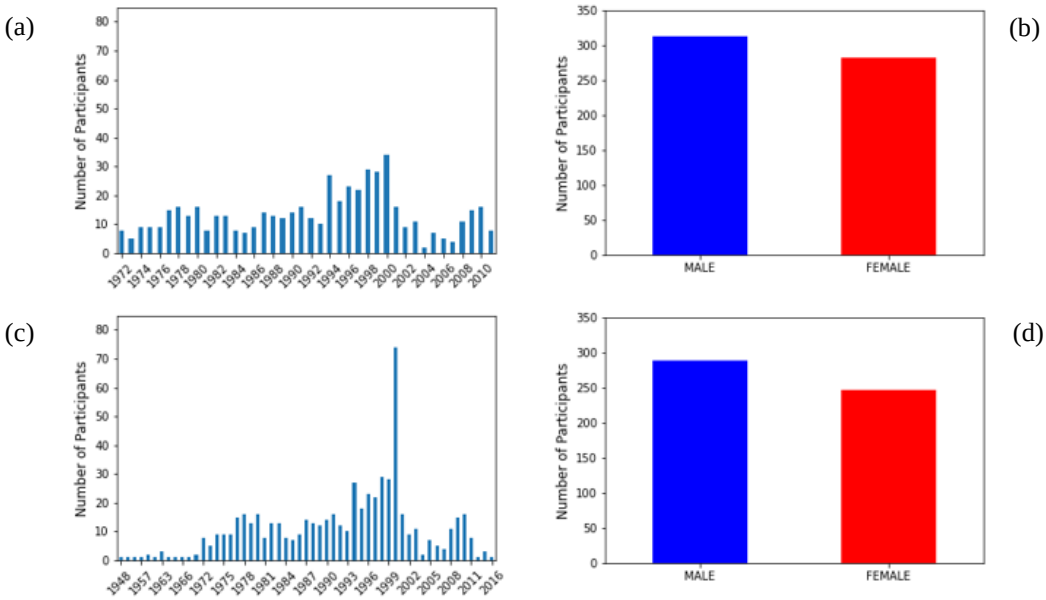

Fig. 1 Frequency of age and sex pre data cleaning, (a) & (b) respectively. Frequency of age and sex post data cleaning, (c) & (d).

Class	Date Range	Frequency
1	1972 – 1983	134
2	1984 – 1994	142
3	1995 – 2000	154
4	2001 - 2011	104

Table 1. Grouped age ranges and their frequencies

3. Methods

3.1 Apparatus, Stimulus, and Procedure

The data in this study was collected in the NEMO Science Museum in Amsterdam. The experiment is set up as part of an exhibition in the museum. Participants are shown one image, Fig. 2, for 10 seconds of free viewing. The experiment is unsupervised, and participants are given no specific instructions before viewing the image. Following the free viewing the participant is asked to provide their demographics, namely year of birth and sex, and permission is provided to allow the use of said data in further research. The experiment was conducted using a Tobii 4C eye tracker with sampling rate 60 Hz. The experiment had the following technical specifications: screen resolution 1920 x 1080 pixels, screen size 698 x 336 mm and viewing distance 800mm.

Fig. 2 Image used in eye tracking data collection

3.2 Saliency Maps & Evaluation Metrics

The following saliency models were implemented: AIM (Bruce & Tsotsos, 2009), RARE (Riche et al., 2013), QSS (Schauerte & Stiefelhagen, 2012), LDS (Fang et al., 2017), IMSIG (Hou et al., 2012), GBVS (Harel et al., 2007), FES (Rezazadegan Tavakoli et al., 2011), DVA (Wang & Shen, 2018), CVS (Erdem & Erdem, 2013), CAS (Goferman et al., 2012), Deepgaze I (Kümmerer et al., 2015), Deepgaze II (Kümmerer et al., 2015), Deepgaze IIE (Linardos et al., 2022) and ICF (Kummerer et al., 2017). This includes a mixture of both bottom-up saliency models and deep learning models. The bottom-up models require no previous input, however deep learning models need to be pre-trained on an image dataset. The models used in this study were pretrained on the ImageNet dataset. This resulted in a total of 14 predicted saliency maps.

In total 6 metrics were used to evaluate the performance of each of the predicted saliency maps: Normalised Scanpath Saliency (NSS), Area Under the ROC Curve (AUC), Information Gain (IG), Pearson's Correlation Coefficient (CC), Shuffled AUC (SAUC) and Similarity (SIM) (Bylinskii et al., 2019).

3.3 Feature Extraction

The spatial distribution of fixation locations was compared with all obtained saliency maps. The similarity between a participant's fixation map and each saliency map was measured using NSS, AUC, IG, SAUC, CC and SIM. Let F' be the fixation map of a given participant. Let X be the set of N collected saliency maps. The fixation map F' and the saliency maps X are compared using the T evaluation metrics discussed above to extract the necessary features for prediction. This process produces NxT p dimensional vectors, where p is the total number of participants in the study. This process is fully visualised in Fig. 3.

Data was scaled before the implementation of each classification algorithm. Each feature was standardised by removing the mean and scaling to unit variance.

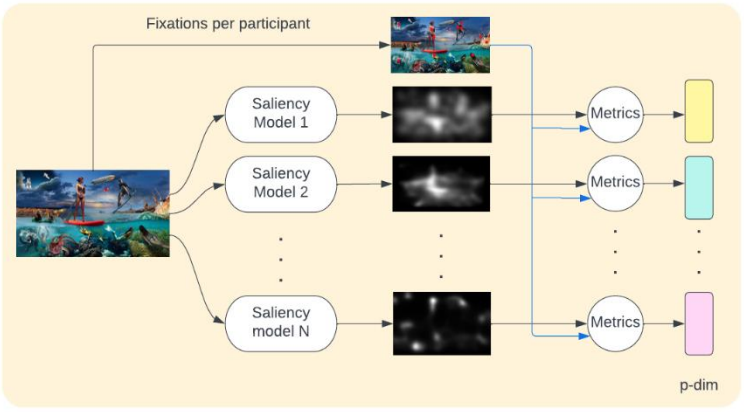

Fig. 3 Feature extraction process

3.4 Machine Learning

Since we are dealing with numerical predictors and labelled categorical response variables the ANOVA correlation coefficient was used to select the top k variables to predict both age and sex. Using this method, the features were ranked for both their relevance in predicting age and sex separately, and this ranking was used as a guide when inputting features into the classification algorithms used.

The features extracted from the saliency maps were forwarded to a supervised machine learning algorithm. Both classical algorithms, Multinomial Logistic Regression, XG-Boost and a deep learning algorithm, Artificial Neural Network were tested for performance.

3.4.1 Logistic Regression

A logistic regression model was fitted to the data and used to predict category membership based on the features extracted. A simple binary logistic regression was applied to predict gender and a multinomial model, an extension of the binary model, (Kmenta et al., 1988; Scott et al., 1991), was used to predict age. This algorithm was selected for this data since its assumptions are less restrictive than other more powerful classification algorithms such as discriminant function analysis. Logistic regression does not assume normality, linearity, or homoscedasticity, which is well suited to this dataset (Stoltzfus, 2011).

Further feature selection was performed using sequential forward selection. In this way the most informative features are selected at each stage and the number of features required to perform classification while maintaining classification results (Rückstieß et al., 2011). After feature selection the models' parameters were optimised using the Grid Search (GS) algorithm (Bergstra et al., 2011). The following parameters were considered: solver, penalty and C value.

3.4.2 XGBoost

Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016) is a gradient boosting technique used in both classification and regression modelling problems. This algorithm has

been shown to be one of the most efficient machine learning classifiers which is why it was implemented in this research (Bentéjac et al., 2019).

Additional feature selection was performed based on feature importance weights calculated by the XGBoost algorithm. Extensive hyperparameter tuning was performed using the GS algorithm since incorrect tuning can lead to poor model performance (Bentéjac et al., 2019). The hyperparameters considered were max depth, min child weight, gamma, subsample, colsample by tree and regularisation.

3.4.3 Artificial Neural Network

Artificial neural networks (ANN) are a powerful prediction tool with strong performance at classification tasks, outperforming white box models such as logistic regression (Dreiseitl & Ohno-Machado, 2002). These models make no distributional assumptions about the data and can detect more complex, non-linear relationships in the data; thus it was selected as a suitable model for classification.

Model architecture and hyperparameter selection can significantly impact the performance of ANN's. There is no industry standard for the number of layers that a model should have or the number of neurons per layer. Some studies suggesting one hidden layer is sufficient while others state that 3 hidden layers is optimal (Uzair & Jamil, 2020) (Gaurang Panchal et al., 2011). Ultimately, model architecture, number of hidden layers and number of neurons, was decided empirically using k-fold cross validation. The hyperparameters were tuned for the data using Bayesian optimisation (Eggensperger et al., 2013). Bayesian optimisation has a quicker performance speed than Grid Search, which is necessary given the increased training time for neural networks versus Logistic Regression and XGBoost. The hyperparameters considered were activation function, optimiser, learning rate, batch size and epochs. Deep learning doesn't require advanced feature selection techniques since any redundancies in the data will be automatically disregarded by the algorithm.

3.4 Evaluation Metrics

Accuracy was used as evaluation metric for the performance of each model (relating predicted outcomes to possible outcomes). A 5x2cv paired t test was used as an additional test for model comparison (Dietterich, 1998).

4. Results

4.1 Saliency

Fig. 4 shows a sample of the saliency maps collected. The brightness of the pixels corresponds to the level of saliency the model has predicted for that area of the input image.

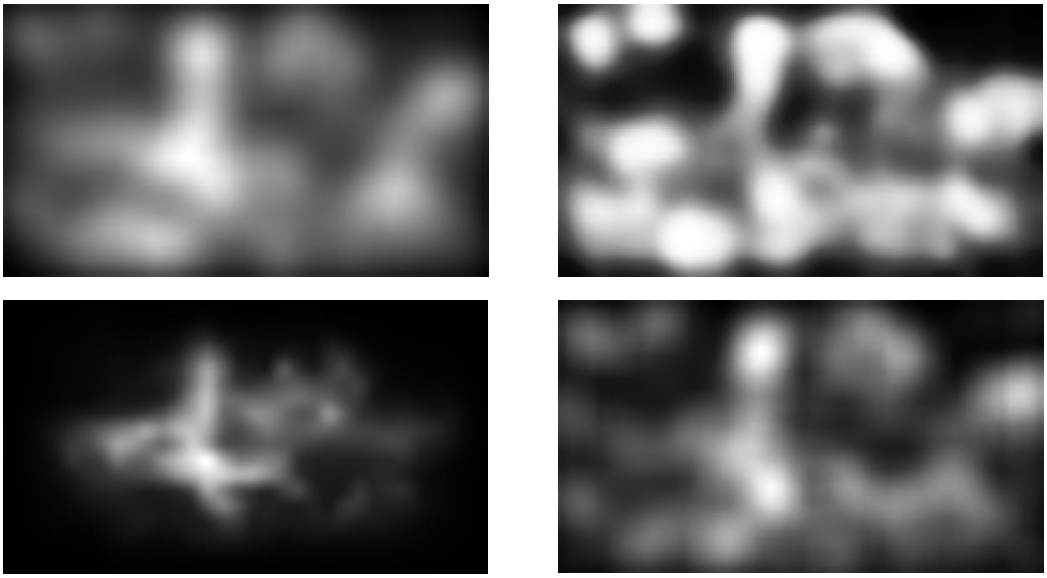

Fig. 4 Predicted saliency maps from the following models IMSIG (top left), DGII (top right), LDS (bottom left), QSS (bottom right)

4.2 Data Exploration

The feature extraction process resulted in a total of 84 features. Initial data exploration indicates minimal variation in the data for both age and sex. Fig. 5, a tSNE (van der Maaten & Hinton, 2008) plot of all features split by age and sex, shows that features of different classes are not well separated.

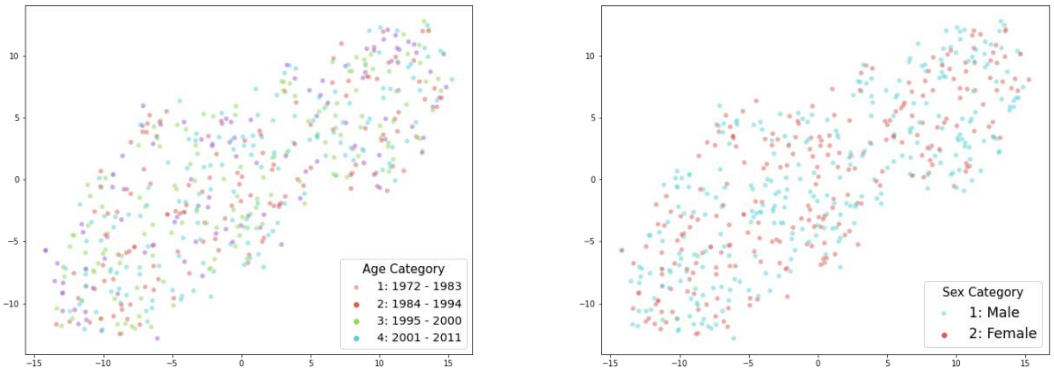

Fig. 5 2D tSNE visualisation of extracted features for age (right) and sex (left)

After feature selection the following features were used in each of the machine learning models:

	Logistic Regression	XGBoost
Age	IMSIG – SIM RARE – IG GBVS – SIM CVS – NSS CVS – CC QSS – AUC QSS – SAUC QSS – SIM	ICF – CC ICF – NSS QSS – CC QSS – AUC QSS – SAUC QSS – IG QSS - NSS
Sex	IMSIG – SAUC DVA – CC DVA – NSS DGIIE – NSS LDS – CC LDS – NSS QSS – SAUC	IMSIG – CC IMSIG – AUC DGI – IG DVA – SAUC

Table 2. Input features to MLR and XGBoost for age and sex prediction after feature selection

4.3 Classification

The results of each machine learning algorithm for age and sex are shown in Fig. 6, Fig. 7 respectively. Each model was trained on the same data and the training/test data split was 80% to 20% for each model.

Fig. 6 shows that Logistic Regression is the most efficient algorithm at predicting age when compared to XGBoost and ANN, with a test accuracy of 39%. However, the baseline accuracy when predicting majority class only is 31%. A 5x2cv paired t test shows that the Logistic Regression model does not predict age significantly better than the naïve baseline model. Both Logistic Regression and Neural Networks predict sex with similar efficiency on both the training and test set (Fig. 7). A 5x2cv paired t test shows that this accuracy is not significantly different from the naïve baseline classifier which predict sex with an accuracy of 50%.

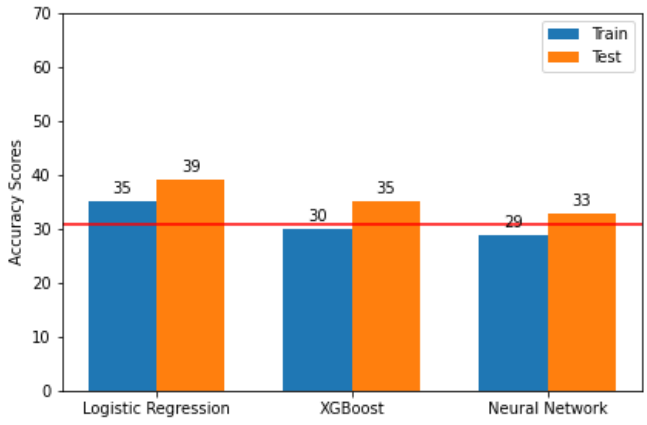

Fig. 6. Train and test accuracy for age prediction

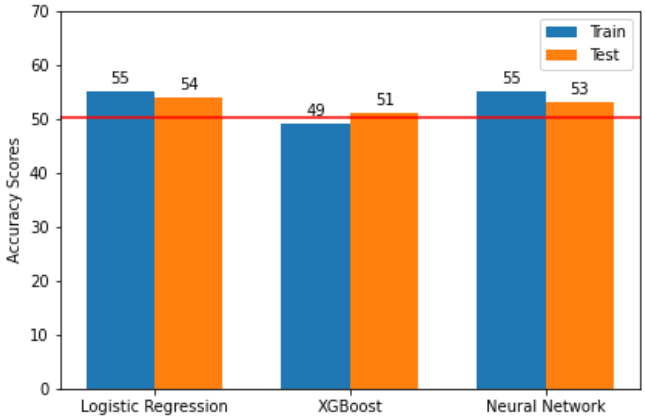

Fig. 7. Train and test accuracy for sex prediction

5. Discussion

The aim of this thesis was to analyse whether there is a difference in the way we view the world by investigating differences in gaze behaviour cross demographics. This analysis was performed by attempting to predict an individual's demographics based on a match to a variety of saliency maps. The results of the machine learning classification algorithms show that there is no clear relationship between a individuals match to a saliency map and their age and sex. This suggests that there is no distinguishable difference in eye movements across demographics and as such we do not differ in the way we move our eye across a visual scene.

Many of the features selected to train the classification models in this analysis were extracted from bottom-up features-based maps such as QSS and CVS. This aligns with previous research which has shown that bottom-up factors are more influential when a child is viewing a scene as compared to adults (Açik et al., 2010). Interestingly, the features extracted from deep learning maps were not as shown to be relevant when predicting age, but features extracted from deep learning maps, Deepgaze I and Deepgaze IIE in particular, were shown to relevant in prediction of sex. Deep learning maps have been claimed to bring a significant advancement in predicting gaze behaviour from saliency maps and have improved prediction over classic saliency models. According to the MIT/Tuebingen saliency benchmark (Borji et al., 2019) these deep learning models are the state-of-the-art saliency models. However, these models are generally tested on much smaller datasets than the data used in this analysis and in general the participants are adults only. It is therefore possible that their performance would vary on a dataset with a wider range of demographics. It is also worth noting that since this study only focuses on the saliency of one image, models pretrained on the ImageNet dataset were used to predict the saliency of the image (Deng et al., 2010). The images in this dataset are mostly from natural scenes whereas the image used in this study is more artificial with a variety of different objects and people merged into one scene. This may adversely impact the performance of the deep learning models as they have not been trained to classify images of this nature.

Previous research into eye movements and human visual attention has shown that there are differences in how our eye move across a visual scene as we age (Egami et al., 2009; Helo et

al., 2014; Takahashi et al., 2021; Walker et al., 2017). There have been studies that show there are sex-based differences in visual attention (Abdi Sargezeh et al., 2019; Hwang & Lee, 2018; Miyahira et al., 2000), but this is a less researched area of the human visual attention system (Mercer Moss et al., 2012). The conflicting evidence between the analysis in this research and prior research into demographic differences in attention could suggest that the saliency models used in this analysis do not capture demographic differences in saliency across gender age that have been shown to exist in other studies. This is a limitation in the current research into predicting visual saliency. Much of the current research into visual saliency focuses on the prediction of a universal saliency map but these maps are insufficient in detecting the variation in salient regions across demographics. Some studies have suggested the integration of demographics into saliency prediction. This has been addressed by (Krishna & Aizawa, 2017) who developed a saliency model that is age-adapted and outperformed other saliency models that did not consider age. Research into sex adapted saliency models is more limited but could also lead to improvements in saliency prediction. However, given the results of this research any expected gains in prediction performance from integration of age and sex into saliency models would be limited. To address this limitation, Zaib & Yamamura propose the idea of a personalised saliency map which considers gender, age and personal preference when predicting a saliency map for an individual (Zaib & Yamamura, 2022). Deep learning maps predict equally well across all age groups and are currently the best models at prediction. But saliency prediction will always be limited by the fact that it excludes the variation in salient regions across demographics. Personalised saliency maps provide an opportunity to address this limitation although research is currently limited, and they require much more data for prediction.

5.1 Limitations

The nature of the data collection process in this study is the most significant limitation. The setting of the eye tracking experiment in a busy environment like the NEMO Science Museum can lead to a lot of noise in the data. Additionally, since the experiment is entirely unsupervised there is no way to confirm that a participant's eye movements and demographics were collected correctly, except for the necessity to continuously fixate a central circle to start the experiment, which ensured minimal data quality. The method of collecting demographics via the gaze interactive interface made it slightly difficult for participants to enter the exact correct age. Incorrect entries in the data will adversely impact any classification algorithm, preventing the detection of patterns in the data and limiting the accuracy of any predictions. Lastly the image used in the experiment is a cluttered image with many different objects and people. Saliency models tend to perform more poorly on these images, this poor performance may limit the predictive ability of the classification algorithms (Zaib & Yamamura, 2022).

5.2 Further Research

Improvements to the data collection process may improve prediction accuracy. Improving user input of demographics would reduce noisy entries in year of birth or gender. The inclusion of more than one image in the experimental set up would provide further information on an individual's fixation patterns. The inclusion of more varied images in the

dataset would provide more data to input into prediction algorithms which could lead to increased classification accuracy.

Helo and colleagues have shown that bottom-up processing takes place only in the early stages of scene viewing (Helo et al., 2014). Further attempts to classify demographics based on match to saliency maps may be improved by considering only the first 3-5 seconds of free viewing when extracting features as any fixations after this time may be driven by top-down factors which the saliency maps do not consider.

Prediction efforts, particularly for age which has been shown to be linked to top-down factors, may be improved by the inclusion of saliency models that include the impact of top-down factors in the prediction of saliency maps. Saliency models such as those proposed by (Mahdi et al., 2020; Tanner & Itti, 2019; Zhang & Zakir, 2019) where top-down factors that influence visual attention such as goal relevance and prior knowledge are implemented in model definition may capture more of the variation in visual attention across demographics that have been recorded in other research.

6. Conclusion

The main goal of this analysis was to investigate demographic differences in human visual attention. This was implemented by attempting to predict these demographics using the match between participants fixations and saliency maps. These saliency maps were collected from a variety of saliency models both classical and deep learning.

The results of the classification showed that neither age nor sex could be accurately predicted using this method. The classification models proposed failed to predict with an error rate that was statistically different from a naive classifier. The saliency maps implemented in this analysis did not provide sufficient information to allow for demographic prediction. These results suggest that these maps do not capture demographic differences in gaze behaviour although this does not significantly impact the performance of saliency models.

References

Abdi Sargezeh, B., Tavakoli, N., & Daliri, M. R. (2019). Gender-based eye movement differences in passive indoor picture viewing: An eye-tracking study. *Physiology and Behavior*, *206*. https://doi.org/10.1016/j.physbeh.2019.03.023

Açik, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: Bottomup and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, *1*(NOV). https://doi.org/10.3389/fpsyg.2010.00207

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2019). *A Comparative Analysis of XGBoost*. https://doi.org/10.1007/s10462-020-09896-5

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*.

Borji, A. (2021). Saliency Prediction in the Deep Learning Era: Successes and Limitations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(2). https://doi.org/10.1109/TPAMI.2019.2935715

Borji, A., Bylinskii, Z., Durand, F., Itti, L., Judd, T., Kümmerer, M., Oliva, A., & Torralba, A. (2019). *Mit/tübingen saliency benchmark.* Https://Saliency.Tuebingen.Ai/.

Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency, attention and visual search: An information theoretic approach. *Journal of Vision*, *9*(3). https://doi.org/10.1167/9.3.5

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation Metrics Tell Us about Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3). https://doi.org/10.1109/TPAMI.2018.2815601

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*. https://doi.org/10.1145/2939672.2939785

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2010). *ImageNet: A large-scale hierarchical image database*. https://doi.org/10.1109/cvpr.2009.5206848

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7). https://doi.org/10.1162/089976698300017197

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, *35*(5–6). https://doi.org/10.1016/S1532-0464(03)00034-0

Egami, C., Morita, K., Ohya, T., Ishii, Y., Yamashita, Y., & Matsuishi, T. (2009). Developmental characteristics of visual cognitive function during childhood according to exploratory eye movements. *Brain and Development*, *31*(10). https://doi.org/10.1016/j.braindev.2008.12.002

Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H. H., & Leyton-Brown, K. (2013). Towards an Empirical Foundation for Assessing Bayesian Optimization of Hyperparameters. *BayesOpt Workshop @ NeurIPS*.

Erdem, E., & Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, *13*(4). https://doi.org/10.1167/13.4.11

Fang, S., Li, J., Tian, Y., Huang, T., & Chen, X. (2017). Learning Discriminative Subspaces on Random Contrasts for Image Saliency Analysis. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(5). https://doi.org/10.1109/TNNLS.2016.2522440

Gaurang Panchal, Amit Ganatra, Y P Kosta, & Devyani Panchal. (2011). Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers. *International Journal of Computer Theory and Engineering*, *3*(2).

Goferman, S., Zelnik-Manor, L., & Tal, A. (2012). Context-aware saliency detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(10). https://doi.org/10.1109/TPAMI.2011.272

Groner, R., & Groner, M. T. (1989). Attention and eye movement control: An overview. *European Archives of Psychiatry and Neurological Sciences*, *239*(1). https://doi.org/10.1007/BF01739737

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*. https://doi.org/10.7551/mitpress/7503.003.0073

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*. https://doi.org/10.1016/j.visres.2014.08.006

Hou, X., Harel, J., & Koch, C. (2012). Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1). https://doi.org/10.1109/TPAMI.2011.146

Hwang, Y. M., & Lee, K. C. (2018). Using an Eye-Tracking Approach to Explore Gender Differences in Visual Attention and Shopping Attitudes in an Online Shopping Environment. *International Journal of Human-Computer Interaction*, *34*(1). https://doi.org/10.1080/10447318.2017.1314611

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3). https://doi.org/10.1038/35058500

Kmenta, J., Aldrich, J. H., Nelson, F. D., Newbold, P., & Bos, T. (1988). Linear Probability, Logit, and Probit Models. *Journal of the American Statistical Association*, *83*(401). https://doi.org/10.2307/2288960

Krishna, O., & Aizawa, K. (2017). Age-adapted saliency model with depth bias. *Proceedings - SAP 2017, ACM Symposium on Applied Perception*. https://doi.org/10.1145/3119881.3119885

Krishna, O., Helo, A., Rämä, P., & Aizawa, K. (2018). Gaze distribution analysis and saliency prediction across age groups. *PLoS ONE*, *13*(2). https://doi.org/10.1371/journal.pone.0193149

Kümmerer, M., Theis, L., & Bethge, M. (2015). Deep Gaze I: Boosting saliency prediction with feature maps trained on ImageNet. *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*.

Kummerer, M., Wallis, T. S. A., Gatys, L. A., & Bethge, M. (2017). Understanding Low- and High-Level Contributions to Fixation Prediction. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*. https://doi.org/10.1109/ICCV.2017.513

Linardos, A., Kummerer, M., Press, O., & Bethge, M. (2022). *DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling*. https://doi.org/10.1109/iccv48922.2021.01268

Mahdi, A., Qin, J., & Crosby, G. (2020). DeepFeat: A bottom-up and top-down saliency model based on deep features of convolutional neural networks. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(1). https://doi.org/10.1109/TCDS.2019.2894561

McGivern, R. F., Mosso, M., Freudenberg, A., & Handa, R. J. (2019). Sex related biases for attending to object color versus object position are reflected in reaction time and accuracy. *PLoS ONE*, *14*(1). https://doi.org/10.1371/journal.pone.0210272

Mercer Moss, F. J., Baddeley, R., & Canagarajah, N. (2012). Eye Movements to Natural Images as a Function of Sex and Personality. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0047870

Miyahira, A., Morita, K., Yamaguchi, H., Morita, Y., & Maeda, H. (2000). Gender differences and reproducibility in exploratory eye movements of normal subjects. *Psychiatry and Clinical Neurosciences*, *54*(1). https://doi.org/10.1046/j.1440-1819.2000.00632.x

Pan, J., Sayrol, E., Giro-I-Nieto, X., McGuinness, K., & O'connor, N. E. (2016). Shallow and deep convolutional networks for saliency prediction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*. https://doi.org/10.1109/CVPR.2016.71

Papavlasopoulou, S., Sharma, K., & Giannakos, M. N. (2020). Coding activities for children: Coupling eye-tracking with qualitative data to investigate gender differences. *Computers in Human Behavior*, *105*. https://doi.org/10.1016/j.chb.2019.03.003

Rahman, S., Rahman, S., Shahid, O., Abdullah, M. T., & Sourov, J. A. (2020). Classifying eye-tracking data using saliency maps. *Proceedings - International Conference on Pattern Recognition*. https://doi.org/10.1109/ICPR48806.2021.9412308

Rezazadegan Tavakoli, H., Rahtu, E., & Heikkilä, J. (2011). Fast and efficient saliency detection using sparse sampling and kernel density estimation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6688 LNCS*. https://doi.org/10.1007/978-3-642-21227-7_62

Riche, N., & Mancas, M. (2016). *Bottom-Up Saliency Models for Still Images: A Practical Review*. https://doi.org/10.1007/978-1-4939-3435-5_9

Riche, N., Mancas, M., Duvinage, M., Mibulumukini, M., Gosselin, B., & Dutoit, T. (2013). RARE2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, *28*(6). https://doi.org/10.1016/j.image.2013.03.009

Rückstieß, T., Osendorfer, C., & van der Smagt, P. (2011). Sequential feature selection for classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7106 LNAI*. https://doi.org/10.1007/978-3-642-25832-9_14

Schauerte, B., & Stiefelhagen, R. (2012). Quaternion-based spectral saliency detection for eye fixation prediction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7573 LNCS*(PART 2). https://doi.org/10.1007/978-3-642-33709-3_9

Scott, A. J., Hosmer, D. W., & Lemeshow, S. (1991). Applied Logistic Regression. *Biometrics*, *47*(4). https://doi.org/10.2307/2532419

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The Relationship between Eye Movements and Spatial Attention. *The Quarterly Journal of Experimental Psychology Section A*, *38*(3). https://doi.org/10.1080/14640748608401609

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine*, *18*(10). https://doi.org/10.1111/j.1553-2712.2011.01185.x

Takahashi, J., Miura, K., Morita, K., Fujimoto, M., Miyata, S., Okazaki, K., Matsumoto, J., Hasegawa, N., Hirano, Y., Yamamori, H., Yasuda, Y., Makinodan, M., Kasai, K., Ozaki, N., Onitsuka, T., & Hashimoto, R. (2021). Effects of age and sex on eye movement characteristics. *Neuropsychopharmacology Reports*, *41*(2). https://doi.org/10.1002/npr2.12163

Tanner, J., & Itti, L. (2019). A top-down saliency model with goal relevance. *Journal of Vision*, *19*(1). https://doi.org/10.1167/19.1.11

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2). https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2018). Visual selection: Usually fast and automatic; Seldom slow and volitional; A reply to commentaries. In *Journal of Cognition* (Vol. 1, Issue 1). https://doi.org/10.5334/joc.32

Uzair, M., & Jamil, N. (2020). Effects of Hidden Layers on the Efficiency of Neural networks. *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*. https://doi.org/10.1109/INMIC50486.2020.9318195

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*.

Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., & Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-81987-x

Walker, F., Bucker, B., Anderson, N. C., Schreij, D., & Theeuwes, J. (2017). Looking at paintings in the Vincent Van Gogh Museum: Eye movement patterns of children and adults. *PLoS ONE*, *12*(6). https://doi.org/10.1371/journal.pone.0178912

Wang, W., & Shen, J. (2018). Deep Visual Attention Prediction. *IEEE Transactions on Image Processing*, *27*(5). https://doi.org/10.1109/TIP.2017.2787612

Zaib, S. E., & Yamamura, M. (2022). Personalized saliency prediction using color spaces. *Multimedia Tools and Applications*, *81*(13). https://doi.org/10.1007/s11042-022-12341-0

Zhang, D., & Zakir, A. (2019). Top-Down Saliency Detection Based on Deep-Learned Features. *International Journal of Computational Intelligence and Applications*, *18*(2). https://doi.org/10.1142/S1469026819500093