Proceedings of the 5th International Conference on Advanced Science & Engineering ICOASE 2025, University of Zakho and Duhok Polytechnic University, 24th – 25th September 2025

(Special issue for Passer Journal of Basic and Applied Sciences)
Paper No. 04

Deep Learning for Facial Beauty Prediction: Integrating Transfer and Multi-Task Learning

Ali Hikmat Ibrahem [1], Adnan Mohsin Abdulazez [2]

[1] Department of IT, Technical College of Informatics - Akre, Akre University for Applied Sciences, Akre 42004, Kurdistan Region, Iraq.
[2] *Technical College of Engineering, Duhok Polytechnic University*, Duhok 42001, Kurdistan Region, Iraq.
Email: ali.hikmat@auas.edu.krd , adnan.mohsin@dpu.edu.krd

Abstract

Facial beauty prediction is a complex and subjective task with significant applications in cosmetic surgery, virtual makeovers, social media filtering, and personalized beauty recommendations. The perception of beauty varies across individuals and cultures, making it challenging for computational models to generalize effectively. conventional approaches based upon handcrafted features and statistical models often fail to capture the intricate patterns of facial aesthetics, limiting their adaptability to diverse populations. To address these challenges, this study proposes an advanced facial beauty model that integrates transfer learning with multi-task learning for improving this predictive accuracy and generalization**.** The EfficientNetV2B0 architecture is employed to leverage its superior feature extraction capabilities and memory efficiency, whereas multi-task learning is utilized to jointly predict beauty scores alongside auxiliary tasks such as gender and ethnicity classification. This joint learning enables the model to learn shared representations and capture subtle aesthetic features more effectively. Experiments on the SCUT-FBP5500 benchmark dataset show that the proposed approach outperforms single-task learning models, achieving higher Pearson correlation coefficients and lower mean absolute errors. multi-task model incorporating beauty score, ethnicity and gender prediction achieve a Pearson correlation coefficient of 0.9185, mean absolute error of 0.2068, and root mean squared error of 0.2762, surpassing existing state-of-the-art methods. These findings suggest that integrating transfer learning with multi-task learning significantly enhances facial beauty prediction, facilitating for more robust and generalizable computational aesthetics models with real-world applications in automated beauty assessment.

Keywords: Multi-Task Learning, Transfer Learning (TL), Beauty Score, Deep Learning, Convolution Neural Network.

1 Introduction

The prediction of facial beauty is considered a major and pragmatically important challenge, with wide-ranging applications across diverse domains such as cosmetic surgery planning, computer-aided virtual makeovers, online content filtering, and personalized recommendations for cosmetics and other beauty treatment [1, 2]. Beauty perception is subjective and shaped by cultural, social, and personal factors, making it a complex task for computational modeling. Traditional techniques to beauty assessment have mostly relied on handcrafted features extracted from facial symmetry, statistical analyses, and geometric proportions. However, the methods suffer from inherent limitations, involving predefined criteria dependence and an inability to generalize across diverse populations with changing aesthetic norms [3, 4]. With the advent of deep learning, automated feature extraction has demonstrated substantially stronger predictive accuracy in facial beauty analysis. TL [3], a powerful technique in computer vision, has gained prominence for its ability to leverage pre-trained models on large-scale datasets such as ImageNet and fine-tune them for domain-specific tasks, including facial beauty prediction (FBP) [5, 6]. This approach alleviates the challenge of acquiring extensive labeled datasets, which are often scarce in beauty-related research. However, current single-task transfer learning models may not fully capture the subtle interactions among various facial properties that contribute to aesthetic judgments, leading to potential limitations in predictive performance [7].

Despite significant advancements in deep learning-based FBP, most existing approaches predominantly rely on single-task learning structures that fail to capture the complex interrelationships among different facial attributes influencing beauty perception. TL, while effective, may overlook crucial auxiliary information that could further enhance predictive performance. Multi-task learning (MTL) offers a promising solution by jointly learning related tasks such as age estimation, gender classification, and facial expression recognition alongside facial beauty prediction. By sharing learned representations, MTL can improve generalization and robustness, ultimately leading to more accurate and adaptive beauty assessment models [8, 9].

In this study, the effectiveness of integrating TL with MTL strategies is investigated to achieve state-of-the-art performance in FBP. EfficientNetV2B0 architecture, known for its superior memory efficiency and feature extraction capabilities, is employed. This research systematically examines the impact of various auxiliary tasks and fine-tuning strategies on predictive accuracy and computational efficiency. By experimenting with different multi-task configurations, the study aims to identify optimal settings that maximize performance while minimizing complexity.

This work contributes to the development of robust and transferable facial attractiveness prediction models with practical applications in computational aesthetics. The remainder of this paper is organized as follows: Section 2 presents related work in FBP and TL; Section 3 describes the methodology, datasets, pre-trained models, and evaluation metrics used in this study; Section 4 discusses the experimental results and comparative analyses; and Section 5 concludes the paper with key findings and future research directions.

2. Related Work

Over the past decades, FBP has been an actively studied topic, with handcrafted features and statistical models serving as the earliest approaches. Nevertheless, these methods fail to fully capture the multifaceted aesthetic properties that shape human perceptions of beauty. With the emergence of deep learning techniques such as convolutional neural networks (CNNs), predictive accuracy in FBP has improved substantially. More recently, TL and MTL have been employed for this purpose, significantly enhancing model performance by enabling the joint learning of related tasks. This section reviews several influential studies in these domains and highlights the existing research gaps that the present study seeks to address.

2. 1 Facial Beauty Prediction Based on Multi-Task Learning

Recently, the use of MTL has attracted more attention as a technique to enhance the generalization and robustness of beauty prediction models based on auxiliary tasks. For example, Savchenko [10] suggested a lightweight CNN-based MTL framework for facial attribute recognition, showing competitive performance across multiple datasets. The face recognition was performed using the VGGFace2 pre-trained models and a compression technique was applied to enhance efficiency. Nevertheless, despite the excellent classification results of this model, the traditional classifiers such as support vector machines (SVMs) and random forests used are an indication for the limitations of the model in predictive power against contemporary deep learning methods.

Gan et al. [11] proposed 2M BeautyNet, a multi-input, multi-task (MI-MT) model designed to predict facial beauty and gender. In this study, an adaptive feature-sharing policy was employed to facilitate multi-task learning. To address the imbalance between different predictive objectives and prevent one task from dominating the others, a multi-task loss function was implemented. Evaluation on the LSFBD dataset demonstrated improved predictive accuracy, with an FBP accuracy of 68.23% achieved on the SCUT-FBP5000 dataset. However, the study focused primarily on gender as an auxiliary task, limiting its generalizability to other demographic factors such as ethnicity.

Gao et al. [12] suggested a deep multi-task learning framework that predicts beauty scores and localizes facial landmarks at the same time for capturing geometric-based and texture-based features. Whereas traditional methods treat individual factors independently, this approach combines them into a unified learning framework, achieving a Pearson correlation coefficient (PC) of 0.92 on the SCUT-FBP dataset. These results indicate that incorporating information about facial structure improves predictive performance. On the other hand, transfer learning methods that can increase the performance of feature extraction for beauty predictions are not investigated in the study.

2. 2 Transfer Learning for Facial Beauty Prediction

Transfer learning has been extensively addressed in beauty prediction tasks to overcome the data scarcity problem. Lebedeva et al. applied a pre-trained FaceNet model for extracting high-level features; the high-level grouping was fused with a Support Vector Regression (SVR) [13] and attractiveness scores were predicted. The method was evaluated on the SCUT-FBP5500 dataset in both constrained (lab-controlled) and unconstrained (real-world) settings. Although the regression performance was competitive, its shallow nature potential is limited to learning more complex hierarchical representations.

Similarly, Boukhari et al. [14] proposed an ensemble learning method by fusing three deep convolutional neural network architecture: InceptionV3, MobileNetV2 and a manually designed light-weight CNN (S-CNN). This ensemble model utilized a range of representational capacities of distinct architectures, producing a Pearson correlation coefficient (PC) of 0.935 on the SCUT-FBP5500 dataset. While this strategy succeeds, there is a computational complexity increase using ensemble methods, making these models less suitable for real-time applications.

2. 3 Identified Research Gaps and Contributions

Although previous mentioned works confirm that both MTL and TL can improve FBP, they have been explored mostly as separate entities. Previous works have mainly concentrated on using pre-trained models or auxiliary tasks, but they have not seen a comprehensive framework to incorporate both (the path of least resistance)." Furthermore, since most of studies adopted single-task beauty prediction frameworks, they could not effectively utilize the interrelations among facial attributes including gender, ethnicity and aesthetic perception as an open set.

In this work, the first deep learning-based approach that uses the EfficientNetV2B0 framework and a MTL strategy for finer predictions of facial beauty is presented. The proposed model suggests how to predict beauty scores together with gender and ethnicity classification, enabling it to generalize better via shared feature representation. Strategies for fine-tuning are also evaluated systematically to help optimize the performance within available computing power while keeping accurate predictions. When the SCUT-FBP5500 benchmark dataset is implemented, the proposed model achieved superior performance in comparison with state-of-the-art single-task and multi-task models, significantly advancing the state-of-the-art in facial beauty assessment. MTL and TL are both effective for Facial Beauty Prediction, but they differ in performance and efficiency. TL leverages pre-trained models to improve feature extraction, reducing the need for large datasets. However, it may not fully capture the complexity of beauty perception. MTL enhances predictive accuracy by incorporating related tasks, for example gender, ethnicity, leading to richer feature representations and improved generalization.

Building upon these conceptual differences, this study proposes a new and unified framework that systematically integrates TL and MTL for Facial Beauty Prediction. The originality of the proposed work can be summarized as follows:

1. Integration of EfficientNetV2B0 within an MTL framework: To the best of our knowledge, this is the first study to employ the EfficientNetV2B0 architecture recognized for its outstanding balance between accuracy, parameter efficiency, and computational speed in a multi-task beauty prediction setting.

2. Systematic evaluation of auxiliary tasks: Four configurations were investigated (beauty-only, beauty + gender, beauty+ethnicity, and beauty+gender+ethnicity), providing new insights into how demographic attributes influence shared feature learning and predictive performance.

3. Improved generalization and interpretability: The proposed TL–MTL integration not only outperforms single-task and existing multi-task models on the SCUT-FBP5500 dataset but also demonstrates robustness and adaptability, paving the way for fairer and more inclusive computational aesthetics.

These contributions distinguish the present study from prior incremental approaches by combining architectural efficiency, methodological depth, and interpretability within a single, coherent framework. Table 1 below summarizes their key differences:

Table 1: Comparison of TL and MTL for Facial Beauty Prediction

Feature	Transfer Learning	Multi-Task Learning
Performance	Strong feature extraction but limited in capturing multi-faceted beauty aspects.	Higher accuracy by learning shared features across multiple tasks.
Generalization	Relies on pre-trained models but may not fully adapt to beauty-related complexities.	Improves generalization by leveraging auxiliary tasks.
Computational Efficiency	More efficient, as it only fine-tunes pre-trained weights.	Higher computational cost due to multiple tasks but optimizes learning.
Best Use Case	When labelled beauty data is limited and efficiency is a priority.	When maximizing accuracy and capturing beauty perception complexity is essential.

Overall, the combination of MTL with TL produces enhance accuracy, but at a marginally higher computational cost, making it ideal for robust FBP models. TL alone is preferable when efficiency is an important concern.

Various studies have explored FBP using TL and MTL. TL-based models are pre-trained to improve feature extraction, during MTL enhances performance by incorporating auxiliary tasks such as gender or ethnicity classification. Table 2 summarizes key studies, comparing their methodologies, models, performance metrics, strengths, and limitations. Building on the insights identified in the related literature, the following section describes the dataset, experimental setup, and the proposed TL–MTL framework used for model development and evaluation.

Table 2: Comparison of Different FBP Approaches Using TL and MTL.

Study	Methodology	Learning Model	Results (PC, MAE, RMSE)	Strengths	Weaknesses
Lebedeva et al. (2021) [13]	TL	FaceNet + SVR	PC = 0.88	Uses a pre-trained model, reducing data needs.	Does not exploit relationships between other aesthetic features.
Gan et al. (2020) [11]	TL with MTL	2M BeautyNet	PC = 0.6823	Uses multi-task loss balancing to improve beauty score predictions.	Limited demographic attributes reduce generalizability.
Gao et al. (2018) [12]	TL with MTL	CNN + Facial Landmarks	PC = 0.92	Captures both texture and geometric features for FBP.	Limited by dataset constraints, affecting model generalization.
Boukhari et al. (2023) [14]	TL + Ensemble	InceptionV3 + MobileNetV2 + S-CNN	PC = 0.9350	Ensemble two model leverage diverse feature extractions for better accuracy.	Computationally expensive due to multiple models.
Our Study	TL + MTL	EfficientNetV2 B0	PC = 0.9185, MAE = 0.2068, RMSE = 0.2762	MTL enhances generalization and predictive power.	Increased complexity and computational cost.

3 Methodology

3. 1 Dataset Overview and Partitioning Strategy

To build the model, the SCUT-FBP5500 dataset was used [15], the standard for facial beauty prediction research. The data comprises 5,500 facial photographs, each annotated with beauty scores, as shown in Figure 1. It is intended for facial beauty prediction tasks (regression with beauty scores from 1 to 5, classification with gender (male, female) and ethnicity (Caucasia male, Caucasia female); The ASIAN Female 2000 dataset which contains 2000 each for Asian Female, Asian Male, Caucasian Female and Caucasian Male 750. Our dataset SCUT-FBP5500 is more challenging because it contains diverse real-world facial images with variations in facial expressions, lighting conditions, head poses and occlusions. These real-world complexities are reflected in ensuring that models are being evaluated on data near realistic application. The dataset also consists of a diverse set of subjects with different genders, ethnicity, and ages which increases its usefulness for analyzing robustness and generalization [16]. SCUT-FBP5500 is a highly practical and reliable data set that contains 5,500 annotated facial images for FBP models. It has gender, facial expressions, lighting conditions and head poses variations, making it generalizable. The model was trained on an increasing number of images to evaluate dataset adequacy, performance metrics (PC, MAE, RMSE) settling after 4000 images, sufficient to comprise a stable data size. Furthermore, transfer learning with the utilization of EfficientNetV2B0 and ImageNet pre-training also reduces data scarcity, making it possible to capitalize on previous knowledge from a wide-ranging and expansive image dataset. Consequently, the dataset is reasonably leveraged with transfer learning for the planned generalization and proper performance of the model.

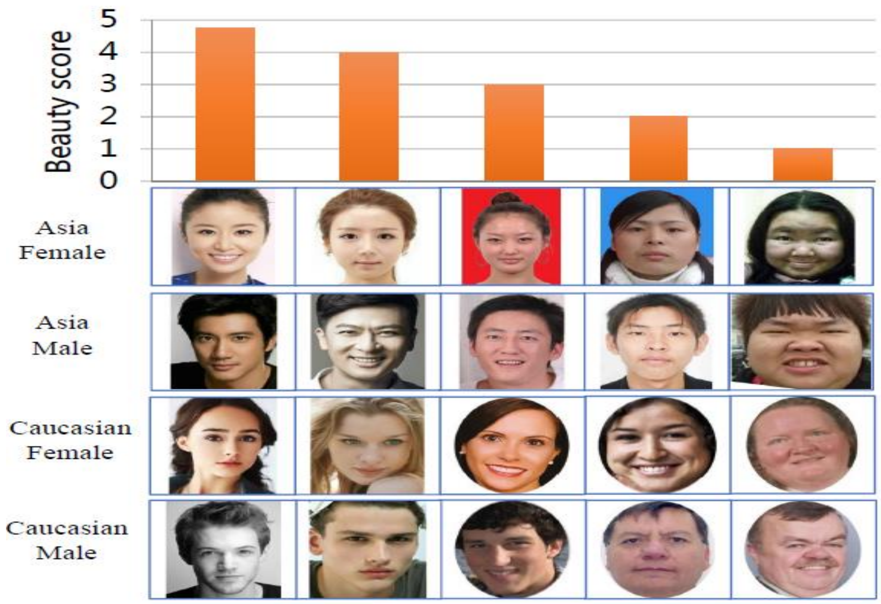

Figure 1: Sample images from the SCUT-FBP5500 dataset showing diversity in gender, ethnicity, and beauty scores [15].

For training and testing preparation, minimal preprocessing was performed to maintain the raw form of the data. The image size of all images was normalized to 384x384 pixels using the default image size used by the EfficientNetV2B0 model used in this research. Input size control through resize steps to meet the model architecture requirements. Built-in preprocessing functions from the TensorFlow/Keras library was used for EfficientNetV2B0 to normalize and scale channels for the pre-trained weights. These preprocessing functions, for example, rescale pixel values by default to the range of [-1, 1], which is often the expected range.

SCUT-FBP5500 Dataset Partitioning In particular, 4,400 images were used for training, 550 for validation, and 550 for testing. It was instructed on the training set so that the model can learn underlying relationships in the data. The validation set was used for hyperparameter tuning and early stopping to avoid overfitting and tune on the model performance during training. The test was then used to evaluate the final performance of the model, giving an unbiased estimation for the ability of the model to generalize invisible data.

3. 2 Pre-trained Models Used in this Research

This study assessed the EfficientNetV2B0 model performance for the FBP tasks within a multi-task framework, predicting beauty scores, gender, and ethnicity. An efficient architecture from the EfficientNet family, the EfficientNetV2B0 is slight and optimized for deployment, achieving similar accuracy to the larger EfficientNet/3 architecture, but with a lower computational cost and fewer parameters. EfficientNet models use combination scaling to find the right depth, width, and resolution, allowing the model to be optimized for both accuracy and speed. EfficentNetV2B0 was we the backbone and adapted model for multi-task FBP by introducing several modifications to the original architecture. The base of the pretrained network and replaced the output layer with fully connected layers were used to do various prediction tasks. As shown in figure 2, three distinct output layers were added: one predicting beauty score, another predicting gender and the last predicting ethnicity. The output layer is appropriate to the expected output type: continuous value for beauty score, and categorical for gender and ethnicity.

Figure 2 describes the key steps of the facial beauty prediction through the SCUT-FBP5500 dataset. data is split into a training and a testing set. The regression and classification applied methods the EfficientNetV2B0 model. The regression task is to predict a beauty score from 1 indicate more unattractiveness to 5 is indicate more attractiveness, and the classification tasks identify gender (Male or Female) and ethnicity (Caucasian or else Asian). This full multi-task is beneficial for model generalization with related tasks leading to improved predictive performance.

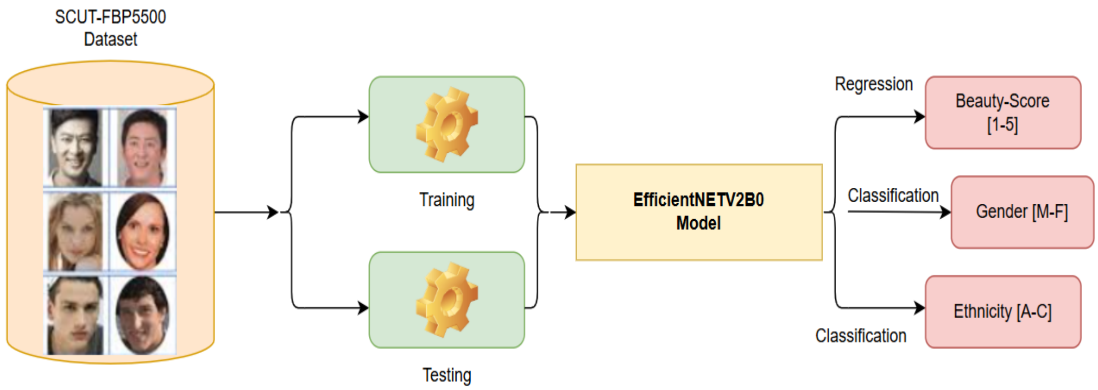

Figure 2: Architecture of the proposed multi-task model based on EfficientNetV2B0 for beauty, gender, and ethnicity prediction.

To better explain the TL process, figure 3 illustrates how the EfficientNetV2B0 Model is fine-tuned with the FBP task. The image describes how the pre-trained model, which was initialized with weights from ImageNet, tailored itself to the task: predicting beauty scores, gender, and ethnicity. This allows the model to generalize its representation of features that are learned from ImageNet and exploit general awareness of facial beauty [17, 18].

The EfficientNetV2B0 model was adopted as the baseline because of its superior feature extraction capabilities, computational optimization, and memory efficiency. It balances accuracy and speed, making it appropriate for real-time applications and edge devices. Its architecture, which scales depth, width, as well as resolution, allows it to achieve high performance with reduced parameters and less computational costs compared to larger models. Pre-trained on large data sets like ImageNet, it leverages TL to reduce the demand for extensive labeled data, which is particularly beneficial for FBP. Furthermore, its adaptability supports MTL, enabling the model to predict beauty scores alongside auxiliary tasks such as gender classification, enhancing generalization and predictive accuracy. EfficientNetV2B0 was selected for its efficiency, adaptability, and robust performance in combining transfer and MTL for FBP.

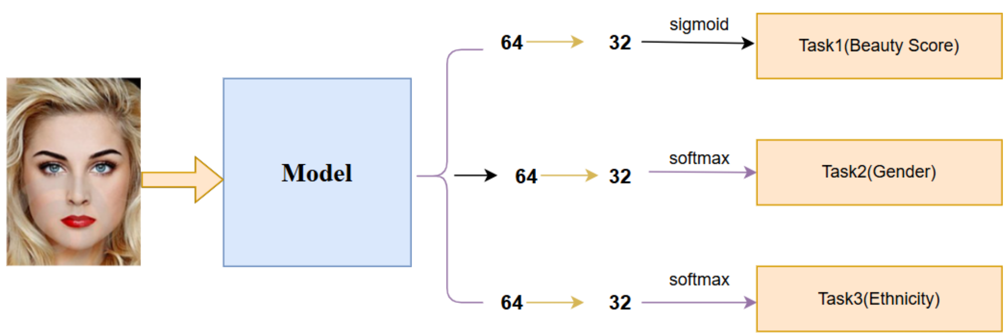

Figure 3: Transfer learning workflow used to fine-tune EfficientNetV2B0 on the SCUT-FBP5500 dataset.

The ImageNet weights with network was first initialized, then all layers on the database were fine-tuned. The approach enables the model to employ the general feature extraction learned by using ImageNet but adapted to the Facial beauty model. This allows the model to be MTL by stacking the output layers into different task-related layers, where the outputs are the predicted beauty score, gender and ethnicity respectively.

In addition, the training process for the suggested FBP involved fine-tuning the EfficientNetV2B0 model architecture, pre-trained on ImageNet, using the SCUT-FBP5500 dataset. It contains 5,500 facial images annotated with beauty scores, gender, and ethnicity, and was separated into 4,400 training, 550 validation, and 550 test images. Images were resized to 384x384 pixels and normalized to the range [-1, 1]. The model was trained for 60 epochs using the Nadam optimizer with a triangular cyclical learning rate (min: 1e-6, max: 1e-3), a batch size of 32, and other training settings are summarized in Table 3. A MTL approach was employed, where the model simultaneously predicted beauty scores (regression), gender (binary classification), and ethnicity (multi-class classification). A multi-task loss function balanced the contributions of each task, using MSE for beauty scores and categorical cross-entropy for gender and ethnicity. All layers in Fine-tuning allowed the model to adapt to FBP while retaining general feature extraction techniques. The EfficientNetV2B0 model achieved a PC of 0.9185, MAE of 0.2068, and RMSE of 0.2762, exceed single-task models and state-of-the-art methods. This integration of TL and MTL enhanced feature generalization and predictive accuracy, demonstrating the effectiveness of the approach for robust FBP.

Table 3: Training Hyperparameters.

Parameter	Value	Justification
Input Size	384×384×3	Default for EfficientNetV2B0
Optimizer	Nadam	Combines Adam with Nesterov momentum
LR Range	1e-6–1e-3 (Cyclical)	Ensures stable convergence
Epochs	60	Prevents overfitting with early stopping
Batch Size	32	Balanced memory/computation
Loss Functions	MSE + Cross Entropy	Suitable for regression and classification tasks

This setup ensured that the model learns task-specific features and thereby achieves optimal performance on FBP. Even though the MTL problem is inherently complex, it is necessary to applied on architectures that are computationally efficient, which is suitable for real-time applications as well as devices at the edge [19, 20] and to perform that, the EfficientNetV2B0 efficient architecture was utilized.

The complexity of various deep CNN architectures can be assessed through two primary metrics: the number of parameters and the Training time per patch. These metrics are critical as they directly impact the model's performance, computational efficiency, and its applicability in real-time scenarios. The details of these models are summarized in Table 4.

Table 4: Performance analysis of the four pretrained models in terms of complexity.

Method	Parameters	Training time(minute)
Single-Task: beauty score. (STL)	Total params: 6003409 (22.90 MB)	43
Multi-Task: beauty score and ethnicity	Total params: 6087539 (23.22 MB)	46
Multi-Task: beauty score and Gender	Total params: 6087539 (23.22 MB)	44
Multi-Task: beauty score, Ethnicity and Gender	Total params: 6171669 (23.54 MB)	47

The above table presented outlines the total parameter counts for each model configuration utilized in this research. As auxiliary tasks like gender and ethnicity classification are incorporated, the model's complexity and parameter count increase. STL, dedicated to predicting beauty scores alone, possesses the fewest parameters at 6,003,409. On the other hand, the multi-task model, which predicts beauty scores alongside gender and ethnicity, has the largest parameter count at 6,171,669. This rise in parameters signifies the added layers and complexity necessary to manage multiple tasks concurrently. Following the implementation of the proposed TL and MTL models and the definition of the experimental settings, the next section reports the experimental results and offers a detailed comparative analysis to assess the effectiveness and performance of the proposed approaches.

4 Experimental Results

4. 1 Experimental Setup

To address the accurately predicting beauty scores challenges, this study explores the integration of TL and MTL to enhance the predictive accuracy of deep learning models for FBP. The proposed framework leverages pre-trained deep neural networks to extract high-level representations during usage MTL strategies to improve feature generalization by incorporating auxiliary tasks such as gender and ethnicity classification. The SCUT-FBP5500 dataset, an often-used benchmark for FBP research, serves as the foundation for training and evaluating the models.

The experiments were conducted in Kaggle Notebooks which provided a cloud environment with a Tesla P100 GPU. Python 3.10.12, TensorFlow 2.15, Keras 2.15 were utilized for building and training the deep learning models and training the network for 60 epochs with the Nadam optimizer and a triangular cyclical learning rate (which can be implemented using TensorFlow Addons). The parameters of the learning rate scheduler were set with a rate of 1e-6, maximal and minimal learning rates of 1e-3 and 1e-6, respectively, a step size of len(train_data)//32 (one cycle per epoch), and a scale mode of "cycle". They were trained with a batch size of 32, restoring the model to weights giving maximum validation accuracy.

Moreover, the tesla P100 GPU was selected for this study because of its very high computational abilities that could solve deep learning models especially ones that required large datasets and complex architectures like EfficientNetV2B0 Model. These characteristics make it an ideal candidate to handle the computational burden of TL and MTL frameworks employed in this study. The fact that it is widely used in research settings, such as Kaggle Notebooks, proves how reliable and impactful it is for deep learning work. Moreover, no performance-related issues were observed during the entirety of training and evaluation process, meaning the computational power of the GPU was sufficient to meet the demands of the study, thus allowing a successful model training and validation on the SCUT-FBP5500 dataset.

The SCUT-FBP5500 dataset used in this study to test the performance of the proposed method, which includes 5500 face images of different gender (Male and Female) and face races (Asian and Caucasian). Some face images of the SCUT-FBP5500 dataset are shown in Figure 1. For each face image, human raters would rate the attractiveness on a scale of 1–5, where 1 is most unattractive and 5 is extremely attractive. The average of the 60 ratings equals the ground-truth score. For more solid results, Experiments were conducted with an initial learning rate of 1e-6, maximal_learning_rate = 1e-3 triangularCyclicalLearningRate and select 80% of facial images as the training set and 20% of facial images as the testing set. In this study, various prediction metrics PC, RMSE and MAE [21] are used to evaluate the performance of the automatic rater. Indeed, accurate predictions yield larger PC values, as well as smaller errors. These are PC (Eq. 1), RMSE (Eq. 2), and MAE (Eq. 3). The three-prediction metrics are as follows.

$$\mathrm{PC} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(r_i - \bar{r})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(r_i - \bar{r})^2}} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - r_i)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - r_i| \tag{3}$$

The Pearson Correlation of (*y1, y2,..., yn*) and *(r1, r2,..., rn)* is calculated using (1), where *y1, y2,..., yn* are the ground-truth scores and *r1, r2,..., rn* are the scores predicted by our model. Where ‾y and ‾p are the mean of *(y1, y2,..., yn)* and *(r1, r2, ..., rn*); respectively. Importantly, “PC” is between -1 to 1, as the significant positive linear association is signified by “1”.

4.2 Experiments on the SCUT-FBP5500 Benchmark Dataset

Experiments on the SCUT-FBP5500 benchmark dataset were conducted to evaluate the performance of STL and MTL models for facial beauty prediction. The models were assessed using key metrics, including PC, RMSE, and MAE (Metrics 1 – 3). The STL model, which focused solely on beauty score prediction, achieved a PC of 0.9130, an MAE of 0.2089, and an RMSE of 0.2808. In contrast, MTL approaches, which incorporated auxiliary tasks such as ethnicity and gender classification, demonstrated improved performance. The best-performing model, which simultaneously predicted beauty score and gender, achieved a PC of 0.9175, an MAE of 0.2038, and an RMSE of 0.2731, outperforming the single-task approach. Another multi-task model that included beauty score, gender, and ethnicity prediction achieved a slightly higher PC of 0.9185 but exhibited a marginal increase in MAE (0.2068) and RMSE (0.2762). As illustrated in table 5, These findings suggest that integrating additional tasks can enhance predictive accuracy by leveraging shared feature representations.

Table 5: Performance Comparison of STL and MTL Models for FBP

Method	↑ PC	↓ MAE	↓ RMSE
Single Task: Beauty Score	0.9130	0.2089	0.2808
Multitask: Beauty Score and Ethnicity	0.9163	0.2121	0.2787
Multitask: Beauty Score and Gender	0.9175	0.2038	0.2731
Multitask: Beauty Score, gender and Ethnicity	0.9185	0.2068	0.2762

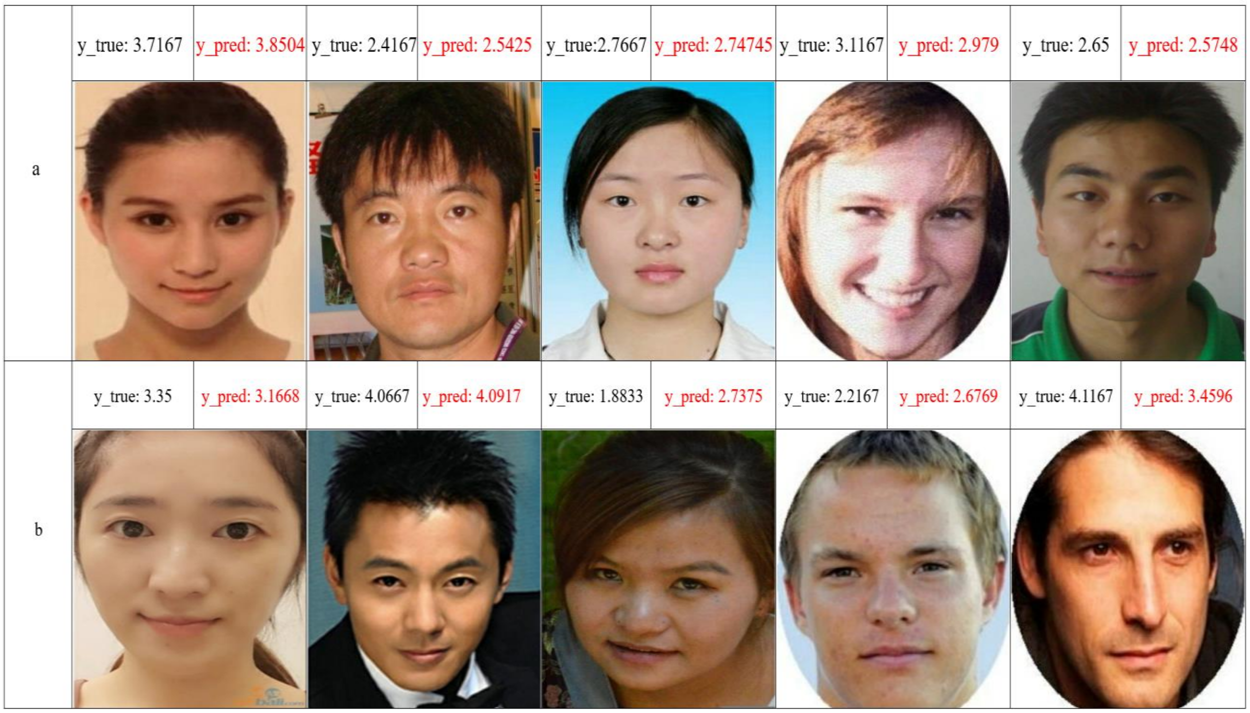

Figure 4: Examples of ground-truth beauty scores (y_true) and predicted scores (y_pred): (a) samples with accurate predictions, (b) samples with noticeable prediction errors.

As shown in figure 4, a visual comparison of true beauty scores (y_true) and predicted scores (y_pred) for various samples using the proposed model is provided. The figure is classified into two parts: (a) represents the examples with higher prediction accuracy, where the model's predictions closely align with the true beauty scores. (b) represents the examples with noticeable prediction deviations, where the model struggles to accurately predict beauty scores. These deviations may occur in cases where facial features are ambiguous or when beauty scores are extreme.

To have a better visual representation, the actual y_true and the predicted beauty scores y_pred for the SCUT-FBP5500 database using multi-task model plotted in Figure 5. The figure further demonstrates the robust alignment between the two in several cases. This visual study highlights the effectiveness of the model in capturing the essential patterns of FBP whilst additionally emphasizing areas where further improvements could be made, by way of illustration in handling ambiguous or random cases. Points closer to the ideal line indicate accurate predictions, while variances highlight areas where the model may struggle due to factors, for example lighting conditions, facial expressions, or database biases. This visualization adds to the quantitative metrics PC, MAE, and RMSE discussed in the previous parts, and provides a more intuitive understanding about the model effectiveness.

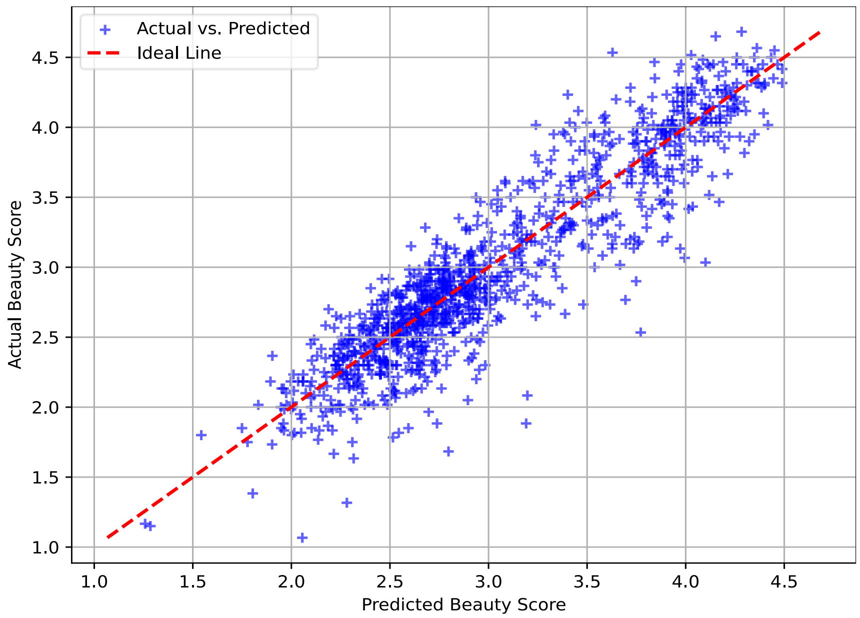

Figure 5: Scatter plot of actual (y_true) versus predicted (y_pred) beauty scores; points aligned with the diagonal indicate accurate predictions.

Overall, the experiments demonstrate the potential of MTL to enhance FBP by leveraging additional facial attributes, facilitating more accurate and robust FBP systems in the future.

4. 3 Comparison with the State of the Art on the SCUT-FBP5500 Benchmark Dataset

Our proposed MTL model demonstrates superior performance compared to the available state-of-the-art methods on the SCUT-FBP5500 benchmark dataset. As illustrated in table 6, traditional approaches such as Geometric Features combined with Support Vector Regression (SVR) yielded a Pearson Correlation Coefficient (PC) of 0.6668, with a relatively high Mean Absolute Error (MAE) of 0.3898 and Root Mean Squared Error (RMSE) of 0.5132. Deep learning methods, including HMTNet and AlexNet, showed significant improvements, achieving PCs of 0.8783 and 0.9141, respectively. However, their MAE and RMSE values remained higher than those of our model. For instance, HMTNet achieved an MAE of 0.2501 and an RMSE of 0.3263, while AlexNet recorded an MAE of 0.2448 and an RMSE of 0.3097. The most recent methods, such as the ResNet-18 based AaNet and the VGGFace2+ Ensemble Stack, further improved performance, achieving PCs of 0.9055 and 0.9112, respectively, with MAEs of 0.2236 and 0.2304, and RMSEs of 0.2954 and 0.2951. In contrast, our comprehensive multi-task model, which simultaneously predicts beauty scores, gender, and ethnicity, achieved the highest PC of 0.9185, surpassing all previous methods. Additionally, our model recorded the lowest MAE of 0.2068 and an RMSE of 0.2762, indicating both higher prediction accuracy and reduced error margins.

Table 6: Performance comparison with FBP state-of-the-arts.

Method	↑ PC	↓ MAE	↓ RMSE
Geometric Features + SVR [15]	0.6668	0.3898	0.5132
HMTNet [22]	0.8783	0.2501	0.3263
AlexNet [23]	0.9141	0.2448	0.3097
ResNet- 18 based AaNet [24]	0.9055	0.2236	0.2954
VGGFace2 + Ensemble Stack [25]	0.9112	0.2304	0.2951
ours	**0.9185**	**0.2068**	**0.2762**

The obtained results indicate the effectiveness of the integration of MTL and TL, enabling our model to better recognize the inherent complexity of beauty perception. The global scheme not only demonstrates improved predictive performance but also improved generalization across facial features, setting a new benchmark in FBP research.

4. 4 Ethical Considerations, Fairness, and Interpretability

Facial beauty prediction involves critical ethical considerations due to its sensitivity and potential societal impact. Beauty perception is inherently subjective and influenced by cultural, gender-based, and socio-demographic factors, meaning computational models may unintentionally embed or amplify these biases. Although the SCUT-FBP5500 dataset includes both Asian and Caucasian subjects, it remains demographically imbalanced, which may lead to performance disparities across subgroups and reduce fairness when models are deployed in real-world settings. The proposed multi-task learning strategy partially mitigates this risk by jointly predicting gender and ethnicity, encouraging more balanced shared feature representations; however, such auxiliary tasks may themselves introduce hidden correlations that need careful monitoring.

From an ethical perspective, automated attractiveness scoring systems may reinforce normative beauty standards, influence self-perception, and be misused in hiring, security, or social filtering applications. To address this, future work should incorporate fairness-aware loss functions, demographic parity evaluation, and cross-cultural benchmarking to assess equitable performance across diverse populations. Moreover, enhancing interpretability through explainable AI techniques (e.g., Grad-CAM, saliency attribution, concept activation vectors) can reveal which facial regions drive attractive predictions, enabling transparency and identification of stereotypical or non-semantic cues. Establishing clear deployment guidelines, privacy protections, and user consent mechanisms will be essential to ensure the ethical reliability and reduce potential harm. Expanding upon the preceding discussion of ethical considerations, fairness, and interpretability in facial beauty prediction, the ensuing section presents the experimental evaluation that substantiates the efficacy of integrating TL and MTL. Furthermore, it consolidates the principal findings of this research and articulates potential avenues for future investigation aimed at advancing model performance, transparency, and ethical accountability.

5 Conclusion and Future Work

The results show that by combining TL and MTL, the so-called TL-MTL, achieves state of the art performance at a PC of (0.9185), MAE of (0.2068) and a RMSE of (0.2762) on the SCUT-FBP5500 dataset. These results reinforce the effectiveness of using EfficientNetV2B0 for high-quality feature extraction and using auxiliary tasks, specifically predicting gender and ethnicity during training, to improve generalization performance. The proposed framework effectively improves the predictive accuracy over both baseline single-task approaches and current state-of-the-art techniques. Facial beauty prediction remains inherently subjective and influenced by cultural diversity. Although the SCUT-FBP5500 dataset includes Asian and Caucasian subjects, demographic imbalance may still introduce bias. The proposed MTL framework, which jointly predicts beauty, gender, and ethnicity, aims to promote fairer and more balanced feature learning. Future work will focus on fairness-aware learning strategies and cross-cultural datasets to enhance inclusiveness and ethical reliability. Additionally, by creating more inclusive datasets such as CelebA and FairFace that represent a broader spectrum of demographics, can build more equitable, automated beauty grading algorithms. To put it in a nutshell, this work contributes to the development of computational aesthetics based on deep learning, paving the way for applications in personalized recommendations, aesthetic evaluations and human-computer interaction, among others. Future studies of this field can be considered in incorporated with supplementary auxiliary tasks (for example, facial expression recognition and age estimation) that would enhance the predictive capabilities of the model. Moreover, research on the performance of advanced fine-tuning techniques and different MTL setups may provide additional enhancements. Using explainable Artificial Intelligence techniques to increase model interpretability, while reducing potential biases in beauty assessment, represents another critical area of future work.

Acknowledgements

The authors would like to thank the department of Information Technology at the University of Akre for applied science for the assistance in the academic performance.

Authors contribution

Ali Hikmat Ibrahem: Conceptualization; Methodology; Software Development; Data Curation; Experimentation; Formal Analysis; Visualization; Writing, Original Draft Preparation.

Adnan Mohsin Abdulazez: Supervision; Conceptualization; Validation; Review & Editing; Project Administration; Resources; Funding Acquisition.

Conflict of interests

The authors of this research confirm that they have no conflict of interest concerning the publication of the study.

Data Availability Statement

SCUT-FBP5500 dataset analyzed during the present study is on hand in the GitHub repository, https://github.com/HCIILAB/SCUT-FBP5500-Database-Release (accessed on 15 June 2023).

References

[1] Gan, J., Luo, H., Xiong, J., Xie, X., Li, H. & Liu, J. Facial beauty prediction combined with multi-task learning of adaptive sharing policy and attentional feature fusion. *Electronics* **13**, 179 doi:10.3390/electronics13010179 (2024).

[2] Dornaika, F. & Moujahid, A. Multi-view graph fusion for semi-supervised learning: application to image-based face beauty prediction. *Algorithms* **15**, 207 doi:10.3390/a15060207 (2022).

[3] Espejo-Garcia, B., Malounas, I., Mylonas, N., Kasimati, A. & Fountas, S. Using EfficientNet and transfer learning for image-based diagnosis of nutrient deficiencies. *Computers and Electronics in Agriculture* **196**, 106868 doi:10.1016/j.compag.2022.106868 (2022).

[4] Bougourzi, F., Dornaika, F., Barrena, N., Distante, C. & Taleb-Ahmed, A. CNN based facial aesthetics analysis through dynamic robust losses and ensemble regression. *Applied Intelligence* **53**, 10825–10842 doi:10.1007/s10489-022-03943-0 (2023).

[5] Xie, Y., et al. Cross-dataset transfer learning for motor imagery signal classification via multi-task learning and pre-training. *Journal of Neural Engineering* **20**, 056037 doi:10.1088/1741-2552/acf6c0 (2023).

[6] Zhao, J., Zhang, M., He, C., Xie, X. & Li, J. A novel facial attractiveness evaluation system based on face shape, facial structure features and skin. *Cognitive Neurodynamics* **14**, 643–656 doi:10.1007/s11571-020-09608-4 (2020).

[7] Gan, J., Jiang, K., Tan, H. & He, G. Facial beauty prediction based on lighted deep convolution neural network with feature extraction strengthened. *Chinese Journal of Electronics* **29**, 312–321 doi:10.1049/cje.2020.01.009 (2020).

[8] Lin, Y., Gao, Z., Zhang, S., Li, L. & Huang, L. Multi-Task Network and Optimization for Face Detection and Attribute Analysis. In: IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 1–5. doi:10.1109/ICAICA53447.2021.00005 (2021).

[9] Lu, H., Xu, S. & Wang, J. Multi-dataset fusion for multi-task learning on face attribute recognition. *Pattern Recognition Letters* **173**, 72–78 doi:10.1016/j.patrec.2022.10.003 (2023).

[10] Savchenko, A. V. Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. In: IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY), 119–124 doi:10.1109/SISY52806.2021.9605947 (2021).

[11] Gan, J., et al. 2M BeautyNet: facial beauty prediction based on multi-task transfer learning. *IEEE Access* **8**, 20245–20256 doi:10.1109/ACCESS.2020.2966012 (2020).

[12] Gao, L., Li, W., Huang, Z., Huang, D. & Wang, Y. Automatic facial attractiveness prediction by deep multi-task learning. In: 2018 24th International Conference on Pattern Recognition (ICPR), 3592–3597 doi:10.1109/ICPR.2018.8545516 (2018).

[13] Lebedeva, I., Guo, Y. & Ying, F. Transfer learning adaptive facial attractiveness assessment. *Journal of Physics: Conference Series* **1922**, 012004. doi:10.1088/1742-6596/1922/012004 (2021).

[14] Boukhari, D. E., Chemsa, A., Taleb-Ahmed, A., Ajgou, R. & Bouzaher, M. T. Facial Beauty Prediction Using an Ensemble of Deep Convolutional Neural Networks. *Engineering Proceedings* **56**, 125. doi:10.3390/engproc2023056125 (2023).

[15] Liang, L., Lin, L., Jin, L., Xie, D. & Li, M. SCUT-FBP5500: a diverse benchmark dataset for multi-paradigm facial beauty prediction. In: 2018 24th International Conference on Pattern Recognition (ICPR), 1598–160. doi:10.1109/ICPR.2018.8545548 (2018).

[16] Ibrahem, A. H. & Abdulazeez, A. M. A comprehensive review of facial beauty prediction using multi-task learning and facial attributes. *ARO-The Scientific Journal of Koya University* **13**, 10–21. doi:10.14500/aro.5471 (2025).

[17] Dikmen, M. Investigating transfer learning performances of deep learning models for classification of GPR B-Scan images. *Traitement du Signal* **39**, doi:10.18280/ts.390501 (2022).

[18] Kantaria, N., Khokhar, D. & Goel, P. A lightweight deep learning approach for Indian fruit classification using EfficientNetV2B2. In: 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), 1063–1069, doi:10.1109/ICESC60888.2024.1013456 (2024).

[19] Lo, L.-J., Yang, C.-T., Chiang, W.-C. & Lin, H.-H. A quantitative method for the assessment of facial attractiveness based on transfer learning with fine-grained image classification. *Pattern Recognition* **145**, 109970, doi:10.1016/j.patcog.2023.109970 (2024).

[20] Radha, H. M., Abdul Hassan, A. K. & Al-Timemy, A. H. Classification of different shoulder girdle motions for prosthesis control using a time-domain feature extraction technique. *ARO-The Scientific Journal of Koya University* **10**, 73–81, doi:10.14500/aro.3522 (2022).
[21] Kukharev, G. A. & Kaziyeva, N. Digital facial anthropometry: application and implementation. *Pattern Recognition and Image Analysis* **30**, 496–511, doi:10.1134/s1054661820030141 (2020).
[22] Xu, L., Fan, H. & Xiang, J. Hierarchical multi-task network for race, gender and facial attractiveness recognition. In: 2019 IEEE International Conference on Image Processing (ICIP), 3861–3865, doi:10.1109/ICIP.2019.8803664 (2019).
[23] Saeed, J. N., Abdulazeez, A. M. & Ibrahim, D. A. Automatic facial aesthetic prediction based on deep learning with loss ensembles. *Applied Sciences* **13**, 9728, doi:10.3390/app13179728 (2023).
[24] Lin, L., Liang, L., Jin, L. & Chen, W. Attribute-aware convolutional neural networks for facial beauty prediction. In: IJCAI 2019, pp.847–853, doi:10.24963/ijcai.2019/120 (2019).
[25] Vahdati, E. & Suen, C. Y. Female facial beauty analysis using transfer learning and stacking ensemble model. In: *Image Analysis and Recognition: 16th International Conference, ICIAR 2019*, Waterloo, ON, Canada, August 27–29, 2019, *Proceedings, Part II*, pp.255–268, doi:10.1007/978-3-030-19141-0_21 (2019).