INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Decoding Innocence: Advancing Forensic Facial Discrimination through Comparative Analysis of Conventional CNN and Advanced Architectures

Mirza Jamal Ahmed [a], Nurul Azma Abdullah [a,*]

[a] *Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Jaya, Johor, Malaysia*
*Corresponding author: *azma@uthm.edu.my*

***Abstract*—The field of Digital Image Forensics (DIF) faces a critical issue in accurately identifying children in digital images, notably in cases involving the proliferation of child sexual abuse content. Existing techniques face hurdles due to model architecture limitations, dataset suitability concerns, and classification imbalance, impeding their ability to recognize children to deter pornographic images. Addressing this challenge, this study introduces Implicit Feature Extraction (IFE), a specialized approach for distinguishing child and adult images in object detection. Leveraging Convolutional Neural Networks (CNNs), the IFE method automates the extraction of discriminative facial features, surpassing the constraints of Explicit Feature Extraction (EFE) methods, which achieve an accuracy of around 70%. The research focuses on three core objectives introducing IFE for detailed face feature detection in DIF's child and adult image identification, implementing IFE with CNNs to enhance image classification, and conducting a thorough evaluation of the proposed technique's performance using key metrics like accuracy and balanced classification results and comparing the result with a basic CNN model's performance. This research's significance lies in its notable contributions to digital image forensics, particularly in combatting child exploitation. The fusion of IFE with CNNs showcases 92% accuracy in distinguishing child and adult images, promising advancements with practical implications in child protection and forensic investigations. The comprehensive evaluation using the UTKFace dataset underscores the proposed technique's efficacy, marking a substantial improvement in child image identification within digital image forensics.**

***Keywords*—Facial feature extraction; object detection; child identification; image forensics; digital image analysis.**

Manuscript received 27 Oct. 2023; revised 12 Dec. 2023; accepted 24 Feb. 2024. Date of publication 31 May. 2024.

I. Introduction

In the domain of Digital Image Forensics (DIF), the precise identification of children within digital images remains a challenge, notably worsened by the alarming proliferation of child sexual abuse content. The limitations inherent in current methodologies, characterized by constraints in model architecture, dataset suitability, and classification imbalance, significantly impede the accurate differentiation of child and adult images. Addressing this necessity, this study presents an innovative approach in digital image forensics, namely Implicit Feature Extraction (IFE), which is carefully engineered to enable accurate discrimination between child and adult images. Central to this novel approach is the utilization of convolutional neural networks (CNNs), which are renowned for their ability to extract visual features in the object detection domain. The IFE approach uses the power of CNNs to automate the extraction of discriminative facial features, effectively transcending the constraints that hampered traditional techniques like Haar-like feature, vola John, Hough, and neural networks. This research focuses on introducing and implementing a framework. A critical dimension of this study involves a comparative analysis of the effectiveness of conventional CNNs with advanced architectural models. The significance of this work lies in its potential to revolutionize digital image forensics, particularly in combatting child exploitation, promising substantial advancements in forensic investigations and child protection efforts. Through meticulous evaluation leveraging datasets like UTKFace, this research indicates a significant stride forward in child image identification within digital forensics.

II. Materials and Methods

Digital image forensics (DIF) focuses on examining digital images for authenticity, integrity, and origin determination; one of its subfields is image detection [1]. DIF procedures aid in the facilitation or advancement of reconstructing illicit events. Images depicting children in sexually explicit situations are considered child pornography, which documents child sexual abuse (CSA) [2]. DIF methods like manual inspection, hash sets, percentage of skin tones, faces, and edge detection have all been used to find illicit images. Manual DIF has been widely used in Child pornography investigations despite the substantial time and effort it requires and the inevitable human intervention that is needed. Message Digest 5 (MD5) and other hash sets have been used to spot CSA images [3]. Another method for identifying CSA content is to look for files with suspicious names, though this method does not correctly assess what is included within each file.

A critical aspect of DIF is refining the visual information to identify CSA content using a query-by-example approach employing content-based image retrieval (CBIR) techniques. File Hound, Web crawlers, X-Ways Forensics, Pornography Detection Stick, Advanced Digital Forensic Solutions, explicit picture detection, picture-Seeker, Adroit Photo Forensics, and RedLight are all examples of DIF technologies that can be used to detect illegally obtained images [4].

Manual methods, anthropometric data, checksum and hash set-based methods, child identification databases, and the delay between the generation of fresh child pornography and its inclusion in the hash set are all common problems to DIF as it is now practiced [5]. Children and adults can be identified by their unique eyes, nose, cheek, ear, lip, mouth, chin, hairline, and side profile traits.

In image processing, features are information about the picture's content and are often retrieved by retrieving sections of an image to obtain attributes [6]. Object detection, including the first real-time face detector, has used Haar-like features [7]. Features like the Haar descriptor benefit significantly from being quickly calculated and can even be computed in real-time. Calculating characteristics like the Haar transformation in real-time and conserving time and resources is possible. Facial features can be used for age estimation and classification using Haar, which has been implemented in methods like the Head-to-Body Ratio (HBR) and the Face to Iris Area Ratio (FIAR). However, they fall short when one's position and viewpoint limit one's ability to photograph the whole body. Using biometric ratios and wrinkle analysis, the Haar method divides facial photos into age groups [2].

The methods' performance differs between age groups and demonstrates unbalanced classification. Therefore, it's crucial to overcome its shortcomings in recognizing faces. Especially for real-time face detection, Haar-like characteristics have been a game-changer in object recognition [7]. Their feature extraction and orientation flexibility are, however, restricted. Improved feature extraction strategies like Vola Johns's face identification, Hough Transform, and Neural Networks' deep learning-based algorithms have been deployed to address these limitations. Although the Viola-Jones object detection framework's primary function is facing identification rather than recognition [8], it outperforms Haar-like features regarding detection and false-positive rates.

The Hough Transform is a feature extraction method utilized in digital image processing, computer vision, and image analysis [9]. Based on craniofacial growth and an investigation of skin wrinkles, the circular Hough has been employed for visual age classification and population counting. Across a range of ratio criteria, the average percentage of correctly identified samples is roughly 66.29 percent, suggesting room for advancement in this area [10].

In recent years, deep learning-based feature extraction (DLEF) has emerged as a game-changing tool for image processing and classification thanks to its ability to directly extract high-level abstract representations from pixel data [11]. Using convolutional layers and pooling techniques, CNNs can learn hierarchical representations of visual characteristics from input images, making them a state-of-the-art method for feature extraction in image classification [12].

Feature extraction detects and separates fundamental portions of digital images in object detection and classification. Extracting ideal features in a reduced order that can reflect the most relevant content of the images for image detection is still a challenging problem [13]. Very little research has paid attention to this problem. Automatic Implicit exploratory feature selection, dimension reduction, and Data Visualization are three reasons feature extraction is an essential problem in feature extraction techniques based on color, texture, and shape features.

Distinctive features are frequently chosen based on the perception that the features are considered the most crucial to include in a model [14]. Accordingly, the feature extraction methods are based on the predefined features extraction approach, Explicit Feature Extraction (EFE). EFE methods can usually effectively select good features that contain the essential discriminative information for image classification or detection. These methods shown in Fig. 1 do not use all the data points in an image, thus reducing processing efforts.

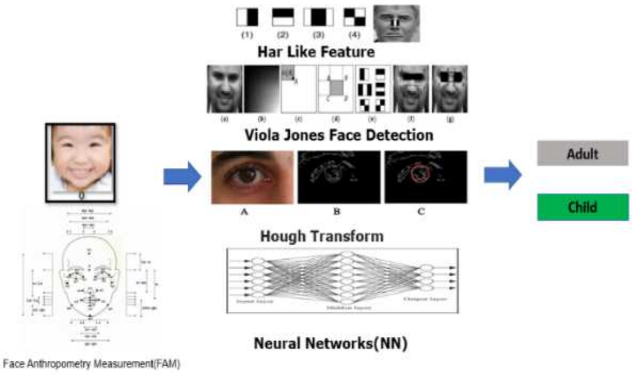


Fig. 1 EFE Methods

The research gap centers on the challenges of feature extraction in digital image analysis. Existing methods often rely on predefined features manually crafted by experts. However, this approach suffers from limitations such as inadequate representation of various image variations, time-consuming manual design for large datasets, and an inability to adapt to diverse tasks or image conditions. The fixed nature of predefined features restricts the capture of complex relationships within the data, keeping recognition accuracy at 70% [15]. Manual feature design takes a lot of time for large datasets and is not scalable. Their inability to adapt to various tasks and potential sensitivity to image conditions can reduce recognition accuracy. Complex features cannot be learned or captured because of their fixed nature [16]. The complex relationships between the features should have been captured.

The imbalance classification issue is prevalent in the previously used feature extraction techniques, leading to challenges in classification [17]. Biased models, reduced sensitivity, and misleading metrics are common problems. Addressing it requires resampling, class weighting, and improved methods to achieve more balanced and accurate predictions.

A more reliable and adaptive feature extraction technique, like deep learning, is required because the manual design approach is costly and limits scalability for large datasets. The required techniques should be fully automated type prediction models that can recognize the critical features in digital images [18]. Such automated methods are necessary primarily for two reasons. First and foremost, there are economic needs, such as processing large amounts of data quickly and with little manual oversight. Second, the issue and the data may be so novel that there is no prior knowledge of distinctive features to identify the image. Accordingly, a novel approach, Implicit Feature extraction (IFE), has been introduced to mitigate the problems. IFE involves leveraging Convolutional Neural Networks (CNNs) to automatically extract features directly from image pixel values to classify age groups. The research into IFE implementation for image classification could improve forensic analysis for age group categorization [19].

The carefully planned framework ensures a thorough and systematic approach throughout the investigation process, which serves as a methodological structure to direct the systematic execution and organization of research activities. The provided framework, Fig. 2, aims to overcome limitations in traditional feature extraction methods like Explicit Feature Extraction (EFE), which often requires human intervention for selecting features in image classification. This framework introduces Implicit Feature Extraction (IFE), leveraging Convolutional Neural Networks (CNNs) to automate feature extraction from images [20]. IFE integrates Digital Image Forensics (DIF) with Supervised Machine Learning (SML) for efficient image classification.

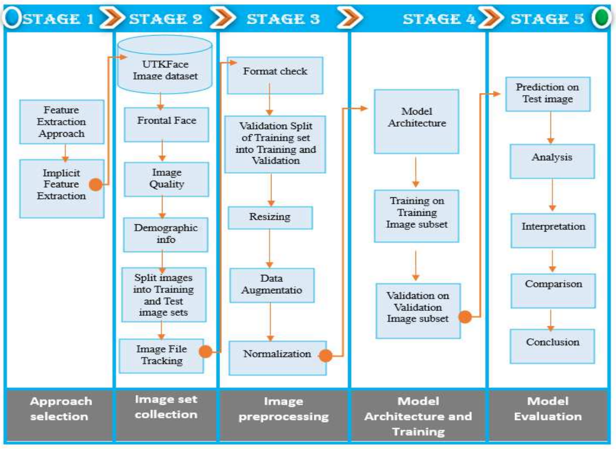


Fig. 2 Research Framework

The framework consists of carefully planned stages illustrated in Figure 2. It begins with feature selection, followed by criteria for image collection. Image processing steps are detailed to standardize image format, size, and pixel values for optimal input to the CNN model [21]. The framework's architecture includes an Improved Convolutional Neural Network (ICNN), which considers various architectural factors and layers for training. Finally, there's a comprehensive focus on evaluating the model's suitability and performance to ensure its effectiveness in image classification [16].

A. Approach Selection

Stage 1 of the Framework focuses on Approach Selection for image classification, assessing Implicit Feature Extraction (IFE) and Explicit Feature Extraction (EFE). EFE relies on predefined facial features like wrinkles and face ratios yet struggles with accuracy and adaptability due to its fixed nature. This approach suffers from limitations such as inadequate representation of various image variations, time-consuming manual design for large datasets, and an inability to adapt to diverse tasks or image conditions. The fixed nature of predefined features restricts the capture of complex relationships within the data, keeping recognition accuracy around 70%. This necessitated a need for a novel approach, namely IFE, which extracts all features directly from pixel values in images, making adaptability and abstract pattern recognition crucial for differentiating children and adults [22].

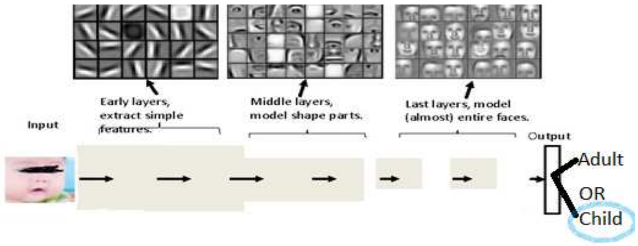

Input
Early layers, extract simple features.
Middle layers, model shape parts.
Last layers, model (almost) entire faces.
Output
Adult
OR
Child

Fig. 3 Implicit feature extraction approach

Fig. 3 illustrates IFE's process, starting with pixel extraction and progressively recognizing facial elements. IFE leverages CNN layers capabilities. IFE's adaptability extends to diverse computer vision tasks, offering automated and efficient prediction models, surpassing EFE's computational demands, and demonstrating enhanced efficiency in metrics. When applied to CNNs, IFE processes a labeled dataset of children and adults, enabling the network to extract informative features essential for classification. CNN's initial layers extract low-level features like edges and facial structures, advancing to higher-level abstract patterns distinguishing child and adult images [23]. The network learns these features iteratively through convolution and pooling operations, followed by classification in fully connected layers adjusted during training. Implementation involves dataset gathering, standardizing images, data augmentation, architecture design, feature extraction, CNN training, and model evaluation [24]. The IFE approach, leveraging CNNs, demonstrates adaptability and efficiency in classifying child and adult images, presenting a robust framework for digital image forensics and supervised machine learning.

B. Image Selection

Stage 2 of the framework centers on collecting an image set for child and adult classification.

1) UTKFace Dataset: This stage involves comprehensive steps, starting with the selection of JPEG photographs of children (13 years or younger) and adults (14 years or older) while ensuring the images are clear and single-subject front-facing shots [25]. The dataset is cleaned to remove irrelevant, duplicate, or incorrectly labeled images. The UTKFace dataset, containing around 20,000 images spanning ages from 0 to 116 years, provides diverse samples for age estimation [26]. This dataset's attributes encompass age, gender, and ethnicity information, vital for age-related tasks. The images reveal distinct age-related features like wrinkles, skin textures, and facial characteristics, enabling explicit feature extraction (EFE) for age classification purposes. Despite its advantages, the dataset presents limitations regarding image resolution, variability in image quality, and imbalanced gender and ethnicity distributions. This stage also highlights the specifics of image collection based on age groups, ensuring a comprehensive dataset for child and adult classification [27].

A tabular representation below demonstrates a balanced distribution of images across different ages for training and testing, totaling 4,990 images. These images are split equally for each age group, enabling robust machine learning model training and evaluation for age estimation or facial image-based classification tasks.

2) Frontal Face Image: Facial landmarks, like the mouth, nose, and eyes, are relatively easy to locate in frontal images. Accurate age classification depends on precise landmark localization, which is easier to achieve in frontal images [11]. Because the face is angled almost directly towards the camera, frontal images have slight variations in pose.

3) Image Quality: Image quality can significantly impact the model's performance. It involves identifying and removing images from the dataset that exhibit quality problems during the image set collection phase. This is done by visual verification that both child and adult images have sufficient resolution to capture fine facial details [28].

4) Demographic Information: Recording age information is critical in ensuring the reliability and credibility of the dataset used for training and evaluating the model. Gender recording for each image, Male:0 and Female:1, highlights the significance of recording gender information within the dataset and its impact on the model's performance and fairness, which can be evaluated. The ethical considerations related to data privacy and the potential re-identification of individuals from the outputs of the model have been addressed by applying manual occlusion on the faces shown in this research [13].

5) Image Split: The dataset has been divided into distinct subsets for training and testing purposes. The training dataset has been divided automatically into training and validation datasets with a ratio of 80% to 20% [29]. Age, gender, ethnicity recording, and ethical considerations regarding data privacy and re-identification risks are thoroughly discussed and managed.

TABLE I
IMAGE SET DISTRIBUTION-AGE-GENDER

		Age		Gender	
	Total	**Child**	**Adult**	**Male**	**Female**
Training	2495	605	1890	1103	1392
Testing	2495	605	1890	1103	1392
Total	4990	1210	3780	2206	2784

Table 1 shows the image distribution set by gender and age group (adults and children). With 2,495 instances divided into age and gender categories in each set, the training and testing datasets appear to have the same distribution.

TABLE II
IMAGE SET DISTRIBUTION-ETHNICITY

		Ethnicity				
	Total	**White**	**Black**	**Asian**	**Indian**	**other**
Training	2495	1476	84	341	371	223
Testing	2495	1476	84	341	371	223
Total	4990	2952	168	682	742	446

Table 2 shows the image dataset of various ethnic groups: White, Black, Asian, Indian, and Other. The training and testing datasets exhibit the same distribution, with 2,495 instances in each set across these ethnic categories. This information outlines the distribution of image sets among different ethnicities within the dataset used for training and testing.

6) Image File Tracking: Furthermore, properly implemented image file tracking contributes to the overall quality and reliability of the dataset and streamlines the entire machine learning prediction analysis and interpretation.

This detailed approach to image collection ensures a comprehensive and representative dataset, vital for training and testing machine learning models for child and adult image classification tasks. The dataset's diverse attributes and balanced distribution contribute to its effectiveness in addressing age-related classification challenges.

C. Image Processing

In Stage 3, essential preprocessing steps are undertaken to prepare the image dataset for robust training of a child and adult image classification model using Convolutional Neural Networks (CNNs) [30]. This phase focuses on critical details of several key preprocessing steps. Selection of JPEG format photographs of children and adults to align with research objectives and common image standards. Uniform resizing of all images to a consistent size (180x180) for compatibility with CNN models, optimizing memory usage, and ensuring uniformity in model inputs [31].

Applying random rotation and horizontal flipping increases data diversity and enhances the model's ability to generalize to real-world image variations. Adjusting pixel values of images to a standardized range ([0, 1]) for computational efficiency, stability in algorithms, and ensuring consistent input values for the CNN model [14]. Utilizing visual inspection to gain insights into the dataset, understand label distributions, identify outliers, and assess data quality before model training. Each step contributes significantly to refining and aligning the dataset with the CNN architecture's requirements, ensuring a high-quality dataset for training a child and adult image classification model.

Hardware and software components were used to simulate the CNN-based models for retraining. The procedure used TensorFlow and Python in conjunction with Anaconda's environment within Jupyter Notebook. The system used for this simulation had the following hardware specifications: it was an Apple computer running a 64-bit version of Windows 10 Enterprise. 8 GB of RAM, and an x64-based processor driven by an Intel Core i7-8550U CPU running at 1.80GHz with a maximum frequency of 1.99 GHz, make up the system.

D. Model Architecture and Training

In stage 4, a deep learning model is created from scratch to implement a feature extraction approach for classifying child and adult images. The approach, called IFE, directly extracts image features without external information. It relies on a formula where a CNN model performs classification based on preprocessed images, extracting features implicitly through its layers. An improved CNN (ICNN) model Fig.4 is customized better to identify features specific to child and adult characteristics, aiming for higher classification accuracy.

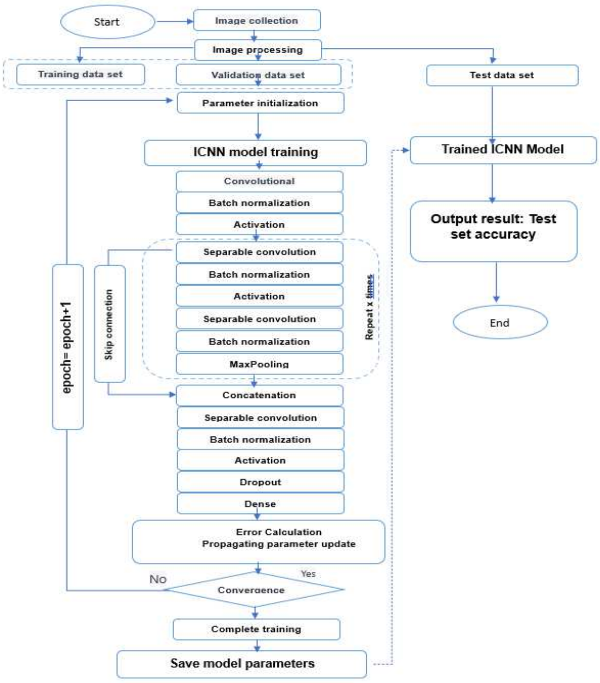


Fig. 4 Improved CNN model architecture

Multiple Convolutional 2D layers are used in Fig. 4. 1st Convolutional 2D layers with 32 filters. 2nd Convolutional 2D layer with 64 filters. For loop with 4 sets of Convolutional 2D layers: [128, 256, 512, 728] filters each—final Separable Convolutional 2D layer with 1024 filters. Batch Normalization layers are utilized after each Convolutional 2D and Separable Convolutional 2D layer.1 Batch Normalization after the 1st Convolutional 2D layer.1 Batch Normalization after the 2nd Convolutional 2D layer. Multiple Batch Normalization layers after each set of Separable Convolutional 2D layers within the for loop.1 Batch Normalization after the final Separable Convolutional 2D layer. Activation Layer (ReLU): ReLU activation layers are present after each Convolutional 2D and Separable Convolutional 2D layer. 1 ReLU activation after the 1st Convolutional 2D layer. 1 ReLU activation after the 2nd Convolutional 2D layer. Multiple ReLU activations within each set of Separable Convolutional 2D layers within the for loops.

1 ReLU activation after the final Separable Convolutional 2D layer. Separable Convolution Layer: Separable Convolutional 2D layers are used after each ReLU activation within the FOR loop—8 sets of Separable Convolutional 2D layers within the for loops. One Max Pooling 2D layer is employed after each set of Separable Convolutional 2D layers within the for loop, resulting in 4 Max Pooling 2D layers within the for loop. Concatenation Layer (Skip Connection): The addition (skip connection) is used inside the for loop. 4 addition operations (layers. add). Finally, Dropout Layer: There's a single Dropout layer with a rate of 0.5 applied before the final Dense layer.

The improved CNN (ICNN) model integrates advanced CNN technologies like separable convolutions [20] and residual blocks [32], overcoming issues of vanishing gradients and overfitting. Adaptability and dynamic adjustments in activation functions and output units yield significantly higher balance accuracy (92%) than the basic CNN model (67%), ensuring symmetry in identifying child and adult images. ICNN's sophistication and diverse architectural components enhance generalization and performance, making it superior to the basic CNN model. For training, important hyperparameter values opted for are Learning Rate: 0.001, Batch Size: 32, Number of Epochs: 50, Dropout: 0.2, Optimizer: Adam.

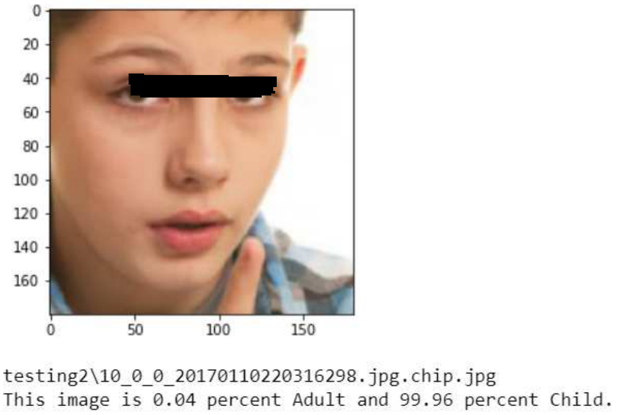
In the training process, the epoch number signifies the training stage. Epoch time in seconds illustrates the duration for each epoch. Training loss, starting at 0.4316 and decreasing to 0.0457, reflects improved model effectiveness with the training dataset. The training accuracy begins at 0.8148 and increases to 0.9848, depicting the model's learning advancement. Validation loss showcases the model's performance on unseen validation data, displaying occasional spikes (from 1.7551 to 1.5511) but generally trending downward. Validation accuracy, fluctuating between 0.8 and 0.9, signifies the accuracy of predictions on the validation set. For comparison purposes, a basic CNN model has been designed. The basic CNN model comprises an input layer defining image shapes, followed by rescaling to standardize pixel values [33].

The basic CNN model includes three sequential convolutional blocks. The first block features a 2-dimensional convolutional layer with 32 filters, size 3x3 with stride 2, same padding, and ReLU activation. Subsequent blocks follow a similar structure but with 64 and 128 filters, respectively [16]. Max Pooling 2D down samples spatial dimensions. The Global Average Pooling Layer condenses spatial information, and a final dense layer with sigmoid activation enables binary classification [34]. This architecture ensures progressive feature extraction through convolutional layers, spatial down-sampling via pooling, and information condensation before the classification step.

When distinguishing between images of children and adults, ICNN performs noticeably better than Basic CNN (92% vs. 67%). To address problems like vanishing gradients and overfitting, ICNN makes use of cutting-edge CNN technologies. Separable convolutions are a more sophisticated type that breaks down the standard convolution into depth-wise and pointwise convolutions to lower computational costs. This can result in improved efficiency and generalization. The architecture of ICNN includes these blocks, which enable the network to discover residual mappings. They aid in improving performance and assist in resolving the vanishing gradient issue during the training of deeper networks. Since these technologies can enhance network efficiency, learning capacity, and performance metrics like accuracy and generalization, they are regarded as advanced in the context of CNN architectures.

E. Model Evaluation

Stage 5 involves evaluating the Improved Convolutional Neural Network (ICNN) model through several procedures. Loading the trained ICNN model for classifying test images into child or adult categories. Resize and normalize images before feeding [35] them to the model.



```
testing2\10_0_0_20170110220316298.jpg.chip.jpg
This image is 0.04 percent Adult and 99.96 percent Child.
```

Fig. 5 ICNN model prediction

Fig. 5 showcases the model's predictions with confidence percentages for each class (Adult and Child) on test images. Analysis and interpretation have been made utilizing evaluation metrics such as accuracy, precision, recall, and F1 score, including the confusion matrix, to assess model performance [36]. Interpretation of these metrics to understand model behavior in classifying adults and children, including the calculation and interpretation of various measures. A comparison with the Basic CNN Model has been provided. Moreover, the analysis extends to employing statistical techniques to understand better predictions, including descriptive statistics to evaluate differences between age groups, genders, and races. Eventually, this stage encompasses thorough model evaluation, statistical analysis, and comparison with a simpler CNN model to assess performance and model behavior in classifying child and adult images.

III. Results and Discussion

Table 3 provides the experiment results. It classifies adult and child images by implying the new CNN-based model. This classification also has two categories: adult (0) and Child (1).

TABLE III
CONFUSION TABLE FOR IMPROVED CCN MODEL

		Total	**(Test outcome Negative) Predicted: No/0**		**(Test outcome Positive) Predicted: Yes/1**	
Adult (0)	Actual No	1890	TN	1726	FP	164
Child (1)	Actual Yes	605	FN	43	TP	562
	Total	2495	Predicted No	1769	Predicted Yes	726

Table 3 compares the actual outcomes (Actual No/Actual Yes) with the predicted outcomes (Predicted No/Predicted Yes) based on the test results.

TABLE IV
CONFUSION TABLE FOR BASIC CCN MODEL

		Total	**(Test outcome Negative) Predicted: No/0**		**(Test outcome Positive) Predicted: Yes/1**	
Adult (0)	Actual No	1890	TN	1812	FP	78
Child (1)	Actual Yes	605	FN	367	TP	238
	Total	2495	Predicted No	2179	Predicted Yes	316

The following performance matrix has been calculated based on the confusion matrix shown in Table 4.

TABLE V
PERFORMANCE MATRICES

Matrices	**Formula**	**ICNN**	**Basic CNN**
Accuracy	= (TP+TN)/Total	91.70%	82.20%
Error Rate	= (FP+FN)/Total	8.30%	17.80%
Recall	= TP/Actual yes	92.89%	39.30%
False Positive Rate	=FP/Actual no	8.68%	4.10%
Specificity	=TN/Actual no	91.32%	95.90%
Precision	=TP/Predicted yes	77.41%	75.30%
Prevalence	=Actual yes/Total	24.25%	24.25 %
Null Error Rate	= Actual no/Total SAMPLE	75.75%	75.75%
F Score	= (2*Precision* Recall)/(Precision+ Recall)	84.45%	51.90%
Balanced accuracy (BA)	= TPR + TNR/2	92.11%	67.60%

As per Table 5, the ICNN outperforms Basic CNN by almost 10% in accuracy. ICNN has a significantly lower misclassification rate compared to Basic CNN. ICNN has a much higher ability to correctly identify positive cases (92.89% vs. 39.3% in Basic CNN. Basic CNN has slightly higher specificity compared to ICNN. ICNN has a marginally higher precision than Basic CNN.ICNN has a notably higher F Score, indicating a better balance between precision and recall than Basic CNN. ICNN's balanced accuracy is substantially higher, reflecting a better balance between sensitivity and specificity.

For the most part, the ICNN model performs better than the Basic CNN model in terms of accuracy, sensitivity, balanced accuracy, and F Score. Higher accuracy, sensitivity, and balanced accuracy show that ICNN performs better overall and can detect positive instances with greater accuracy. Compared to ICNN, basic CNN exhibits lower false positive rates and higher specificity, but it is less sensitive and accurate overall. This shows the ICNN model performs better across a range of evaluation metrics, appearing to be more resilient and prosperous in capturing both positive and negative instances.

Further statistical analysis has been applied to evaluate the performance of the ICNN features of the study results.

TABLE VI
GENDER GROUP STATISTICS

Gender	N	Mean	Std. Dev.	Std. Error Mean
Male	1103	92.88	21.88	0.6588148
Female	1392	89.24	26.46	0.7093114

The group statistics for the gender-based performance are shown in Table 6. Male and female data are divided into two groups. These group statistics cover the average performance scores, score variability, and accuracy of the mean estimate for the Male and Female groups. The male group appears to have a slightly higher mean performance score than the female group, but the female group displays a marginally higher performance score variability.

TABLE VII
ETHNIC GROUP PERFORMANCE

	N	Mean	Std. Dev.	Std. Error	95% Confidence Interval for Mean
					Lower Bound
White	1476	89.45%	26.56%	0.69	88.09
Black	84	93.83%	18.72%	2.04	89.77
Asian	341	90.00%	24.88%	1.35	87.36
Indian	371	95.93%	14.98%	0.78	94.40
other	223	91.82%	24.50%	1.64	88.59
Total	2495	90.85%	24.61%	0.49	89.88

The descriptive statistics in Table 7 shed light on the range of performance scores for various ethnic groups and the average performance scores, variability, precision of the mean estimates, and confidence intervals. The Black group appears to have the lowest mean performance score, whereas the Indian group appears to have the highest. Different ethnic groups have different standard deviations and confidence intervals, indicating variations in the Test of Homogeneity of Variances results for the variable "Performance," which are shown in the table. The overall mean prediction and its confidence interval indicate the model's aggregated performance across these ethnic categories, showing that the model can identify children of different ethnic groups.

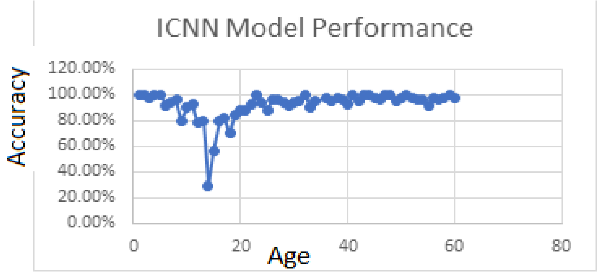


Fig. 6 ICNN Moel performance

The model has shown the ability to predict images of various age groups as shown in Fig. 6. Depending on the age group, the model performs differently, with some age groups achieving greater accuracy than others. In contrast to other age groups, such as 14 and 15, which have lower accuracy percentages, age groups 1, 2, 4, and 5 have perfect accuracy of 100%. Overall, the model exhibits very high degrees of age-group accuracy, demonstrating its capacity to categorize images into child or adult.

IV. CONCLUSION

The study's conclusion signifies a breakthrough in digital image forensics, particularly in the detection of child sexual abuse content. By developing an Implicit Feature Extraction (IFE) method and utilizing improved convolutional neural networks (ICNNs), the research achieved a remarkable 92% classification accuracy, surpassing previous methods, which averaged around 70%. These advancements hold promising implications for law enforcement, child exploitation prevention, and child protection efforts.

Furthermore, integrating supervised machine learning with Digital Image Forensics (DIF) introduces a novel approach to enhancing digital image analysis and forensic processes. The study emphasizes the need for more adaptive and automated techniques, highlighting the limitations of traditional explicit feature extraction (EFE) methods.

Evaluation of the ICNN model demonstrates its superior performance to the basic CNN model, showcasing higher accuracy, sensitivity, balanced accuracy, and F Score. The model's efficacy in accurately distinguishing between child and adult images underscores its potential in real-world scenarios. Additionally, the study explores the impact of demographic factors such as gender and ethnicity on model performance, revealing variations in performance scores across different demographic groups.

In conclusion, the research significantly contributes to advancing digital image forensics by introducing innovative methodologies for efficient and accurate detection of child sexual abuse content. The proposed framework, coupled with advanced feature extraction techniques and supervised machine learning, offers promising avenues for enhancing forensic analysis and protecting vulnerable populations.

ACKNOWLEDGMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vote Q414).

REFERENCES

[1] H. K. Tu, L. T. Thuong, H. V. Uyen Synh, H. T. San, and H. Van Khoa, “Develop an algorithm for image forensics using feature comparison and sharpness estimation,” 2017 International Conference on Recent Advances in Signal Processing, Telecommunications &amp; Computing (SigTelCom), Jan. 2017, doi:10.1109/sigtelcom.2017.7849800.

[2] V. Chatzis, F. Panagiotopoulos, and V. Mardiris, “Face to Iris Area Ratio as a feature for children detection in digital forensics applications,” 2016 Digital Media Industry &amp; Academic Forum (DMIAF), Jul. 2016, doi: 10.1109/dmiaf.2016.7574915.

[3] A. Ulges and A. Stahl, “Automatic detection of child pornography using color visual words,” 2011 IEEE International Conference on Multimedia and Expo, Jul. 2011, doi: 10.1109/icme.2011.6011977.

[4] C. Chen, S. Member, and J. Ni, “Blind Forensics of Successive Geometric Transformations in Digital Images Using Spectral Method : Theory and Applications,” vol. 26, no. 6, pp. 2811–2824, 2017.

[5] P. Kruchten, “Certification 1, 2, 3,” IEEE Software, vol. 27, no. 3, pp. 92–94, May 2010, doi: 10.1109/ms.2010.70.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/tpami.2016.2577031.

[7] O. F. Ince, I. F. Ince, J. S. Park, J. K. Song, and B. W. Yoon, "Child and adult classification using biometric features based on video analytics," *ICIC Express Lett. Part B Appl.*, vol. 8, no. 5, pp. 819–825, 2017, doi: 10.5281/zenodo.890713.

[8] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," 2017.

[9] C. Raghavachari, V. Aparna, S. Chithira, and V. Balasubramanian, "A Comparative Study of Vision Based Human Detection Techniques in People Counting Applications," Procedia Computer Science, vol. 58, pp. 461–469, 2015, doi: 10.1016/j.procs.2015.08.064.

[10] R. Hussin, M. R. Juhari, N. W. Kang, R. C. Ismail, and A. Kamarudin, "Digital Image Processing Techniques for Object Detection From Complex Background Image," Procedia Engineering, vol. 41, pp. 340–344, 2012, doi: 10.1016/j.proeng.2012.07.182.

[11] C. R. Kumar, S. N, M. Priyadharshini, D. G. E, and K. R. M, "Face recognition using CNN and siamese network," Measurement: Sensors, vol. 27, p. 100800, Jun. 2023, doi: 10.1016/j.measen.2023.100800.

[12] S. Biswas, D. Chambers, W. D. Hairston, and S. Bhattacharya, "Head pose classification for passenger with CNN," Transportation Engineering, vol. 11, p. 100157, Mar. 2023, doi:10.1016/j.treng.2022.100157.

[13] A. A. Solanke, "Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models," Forensic Science International: Digital Investigation, vol. 42, p. 301403, Jul. 2022, doi: 10.1016/j.fsidi.2022.301403.

[14] X. Wang, F. Cheng, S. Wang, H. Sun, G. Liu, and C. Zhou, "Adult Image Classification," *2018 25th IEEE Int. Conf. Image Process.*, pp. 2989–2993, 2018.

[15] H. Amroun, Mh. H. Temkit, and M. Ammi, "Best Feature for CNN Classification of Human Activity Using IOT Network," 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Jun. 2017, doi: 10.1109/ithings-greencom-cpscom-smartdata.2017.145.

[16] O. Uparkar, J. Bharti, R. K. Pateriya, R. K. Gupta, and A. Sharma, "Vision Transformer Outperforms Deep Convolutional Neural Network-based Model in Classifying X-ray Images," Procedia Computer Science, vol. 218, pp. 2338–2349, 2023, doi:10.1016/j.procs.2023.01.209.

[17] W.-H. Yun, D. Lee, C. Park, J. Kim, and J. Kim, "Automatic Recognition of Children Engagement from Facial Video Using Convolutional Neural Networks," IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 696–707, Oct. 2020, doi:10.1109/taffc.2018.2834350.

[18] B. Liu, C.-M. Pun, and X.-C. Yuan, "Digital Image Forgery Detection Using JPEG Features and Local Noise Discrepancies," The Scientific World Journal, vol. 2014, pp. 1–12, 2014, doi: 10.1155/2014/230425.

[19] J. Waleed and T. M. Hasan, "Techniques (AFTs) Based Compressed Image," no. March, pp. 7–9, 2017.

[20] V. Kate and P. Shukla, "A 3 Tier CNN model with deep discriminative feature extraction for discovering malignant growth in multi-scale histopathology images," Informatics in Medicine Unlocked, vol. 24, p. 100616, 2021, doi: 10.1016/j.imu.2021.100616.

[21] P. Kuppusamy and V. C. Bharathi, "Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance – A survey," Measurement: Sensors, vol. 24, p. 100510, Dec. 2022, doi:10.1016/j.measen.2022.100510.

[22] A. Shah et al., "A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN)," Clinical eHealth, vol. 6, pp. 76–84, Dec. 2023, doi:10.1016/j.ceh.2023.08.002.

[23] A. Kamel, B. Sheng, P. Yang, P. Li, R. Shen, and D. D. Feng, "Deep Convolutional Neural Networks for Human Action Recognition Using Depth Maps and Postures," pp. 1–14, 2018.

[24] Y. Zhang, J. Gao, and H. Zhou, "Breeds Classification with Deep Convolutional Neural Network," Proceedings of the 2020 12th International Conference on Machine Learning and Computing, Feb. 2020, doi: 10.1145/3383972.3383975.

[25] R. Chikkala, S. Edara, and P. Bhima, "Human facial image age group classification based on third order four pixel pattern (TOFP) of wavelet image," *Int. Arab J. Inf. Technol.*, vol. 16, no. 1, pp. 30–40, 2019.

[26] K. Kärkkäinen and J. Joo, "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age," 2019, [Online]. Available: http://arxiv.org/abs/1908.04913.

[27] V. Mirjalili, S. Raschka, and A. Ross, "PrivacyNet: Semi-Adversarial Networks for Multi-Attribute Face Privacy," IEEE Transactions on Image Processing, vol. 29, pp. 9400–9412, 2020, doi:10.1109/tip.2020.3024026.

[28] B. Johnston and P. de Chazal, "A review of image-based automatic facial landmark identification techniques," EURASIP Journal on Image and Video Processing, vol. 2018, no. 1, Sep. 2018, doi:10.1186/s13640-018-0324-4.

[29] A. Nurhadiyatna, S. Cahyadi, F. Damatraseta, and Y. Rianto, "Adult content classification through deep convolution neural network," 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Oct. 2017, doi: 10.1109/ic3ina.2017.8251749.

[30] C. Gautam and K. R. Seeja, "Facial emotion recognition using Handcrafted features and CNN," Procedia Computer Science, vol. 218, pp. 1295–1303, 2023, doi: 10.1016/j.procs.2023.01.108.

[31] L. F. de J. Silva, O. A. C. Cortes, and J. O. B. Diniz, "A novel ensemble CNN model for COVID-19 classification in computerized tomography scans," Results in Control and Optimization, vol. 11, p. 100215, Jun. 2023, doi: 10.1016/j.rico.2023.100215.

[32] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification," vol. 6, 2018.

[33] Y. Nie, S. Xia, and Y. Wu, "Wheel classification using convolutional neural networks," 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), May 2018, doi:10.1109/yac.2018.8406429.

[34] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-Stream Neural Networks for Tampered Face Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jul. 2017, doi: 10.1109/cvprw.2017.229.

[35] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic Estimation from Face Images: Human vs. Machine Performance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 6, pp. 1148–1161, Jun. 2015, doi: 10.1109/tpami.2014.2362759.

[36] N. N. Prakash, V. Rajesh, D. L. Namakhwa, S. Dwarkanath Pande, and S. H. Ahammad, "A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis," Scientific African, vol. 20, p. e01629, Jul. 2023, doi: 10.1016/j.sciaf.2023.e01629.