Effectiveness of Using Artificial Intelligence for Early Child Development Screening

Michael-Lian Gau[1] , Huong-Yong Ting[1]*, Teck-Hock Toh[2], Pui-Ying Wong[3], Pei-Jun Woo[3], Su-Woan Wo[3], Gek-Ling Tan[4]

[1]Design and Technology Centre, University of Technology Sarawak, No. 1, Jalan Universiti, 96000 Sibu, Sarawak, Malaysia
[2]Clinical Research Centre & Department of Paediatrics, Sibu Hospital, Ministry of Health Malaysia, Batu 5½, Jalan Ulu Oya, 96000 Sibu, Sarawak, Malaysia
[3]Department of Psychology, School of Medical Life Sciences, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia
[4]Spark Child Development Centre, B3, Jalan Teknologi 2/1B, Signature Park, 47810 Kota Damansara, Petaling Jaya, Selangor Darul Ehsan, Malaysia

*Correspondence: alan.ting@uts.edu.my

ABSTRACT: This study presents a novel approach to recognizing emotions in infants using machine learning models. To address the lack of infant-specific datasets, a custom dataset of infants' faces was created by extracting images from the AffectNet dataset. The dataset was then used to train various machine learning models with different parameters. The best-performing model was evaluated on the City Infant Faces dataset. The proposed deep learning model achieved an accuracy of 94.63% in recognizing positive, negative, and neutral facial expressions. These results provide a benchmark for the performance of machine learning models in infant emotion recognition and suggest potential applications in developing emotion-sensitive technologies for infants. This study fills a gap in the literature on emotion recognition, which has largely focused on adults or children and highlights the importance of developing infant-specific datasets and evaluating different parameters to achieve accurate results.

KEYWORDS: Facial expression recognition; machine learning; infants; emotional development; deep learning.

1. Introduction

The early detection of developmental and behavioral disabilities in children is crucial for effective intervention and improved outcomes. According to the Centers for Disease Control and Prevention (CDC), one in six three-to-17-year-olds has a behavioral or developmental disability, including autism, attention deficit hyperactivity disorder (ADHD), dyslexia, and mental health issues. Early intervention is key in reducing the adverse effects of these neurodevelopmental disorders [1]. To support the early detection of developmental issues, the American Academy of Pediatrics recommends screening children for developmental and behavioral issues at 9, 18, and 30 months [2]. However, self-report measures have limitations in screening young children's behaviors and development, including potential bias and limited coverage of behaviors and contexts [3]. Therefore, objective assessments or multiple data sources should be considered.

Computer vision (CV) and artificial intelligence (AI) can potentially improve medical diagnoses and treatment by analyzing past patient cases to predict future care. AI algorithms, such as neural networks and evolutionary algorithms, are used in various professions to tackle complex problems [4–6]. Specifically, computer vision, a subfield of AI, uses computer models to simulate human visual perception for tasks like object recognition and classification, with OpenCV being widely used in CV research [7]. Convolutional neural networks (CNN) require less preparation than conventional techniques and can be implemented in various ways, making them ideal for medical diagnosis and assessment. Recognizing emotions in infants is a crucial aspect of early development essential for adaptive social functioning. Recent advances in deep learning algorithms have significantly improved facial emotion recognition performance, a hot topic in CV and AI [8, 9]. However, recognizing emotions in infants presents additional challenges, including age-specific data collection and developing classifiers tailored to infant facial expressions [10–12]. Therefore, careful consideration of information extraction and classifier selection is required to achieve accurate and reliable results.

Several studies have focused on recognizing child facial expressions using the Child Affective Facial Expression (CAFE) dataset. One study achieved 56% accuracy using a Deep Boltzmann Machine with supervision [13]. Another study utilized transfer learning and a CNN network trained on the CK+ dataset to achieve 76% validation accuracy on a subset of the CAFE dataset [14]. A third study used a MobileNetV3 pre-trained network and achieved 65.78% accuracy [15]. Finally, a fourth study utilized multiple datasets to train the model and achieved a balanced accuracy of 66.9% on the CAFE dataset [16]. However, it is important to note that children's facial expressions differ from those of infants [17]. This study aimed to use facial expression detection algorithms on mobile technology for early child development screening in Asian cultures. The study evaluated the algorithm's accuracy and established a baseline for machine learning models' perception of infant emotions. The study results provide further information on developing culturally appropriate and effective screening tools and drive future research for improved approaches. Additionally, it contributes to the creation of infant-friendly, emotion-sensitive technologies to support young children's emotional well-being and development.

1.1. Machine learning algorithm.

Artificial intelligence (AI) is the oldest and broadest branch of computer science, encompassing all aspects of simulating cognitive capabilities for real-world problem-solving and developing systems that learn and think like humans. Therefore, it is often referred to as machine intelligence to differentiate it from human intelligence [18]. Recent research emphasizes that AI systems should not only address pattern recognition issues but also create causal models of the environment that allow explanation and comprehension [19]. Support Vector Machines (SVMs), which categorize data, are widely used in facial emotion recognition [20-22]. Research shows that SVMs can handle high-dimensional data and resist noise, making them suitable for facial expression recognition [23]. SVMs are based on finding the decision boundary that maximizes the margin, which is the distance between the boundary and the data points from different classes closest to it [24]. Several studies have used SVMs to examine different feature descriptors, including facial landmarks and the center of gravity [25], as well as SVM, KNN, LDA, Gabor, Haar, and LBP features for classification [26]. Other studies have used face geometry-based feature extraction to examine logistic regression, LDA, KNN, classification and regression trees, naive Bayes, and SVM [10]. Overall, SVMs have performed

well in these tasks, particularly when using face geometry-based feature extraction. However, the choice of feature descriptors and classification methods may depend on the specific application and dataset being used.

1.2. Deep learning algorithm.

Convolutional neural networks (CNNs) are a widely used subclass of artificial neural networks, especially in computer vision tasks [21]. Since their remarkable success in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012, CNNs have become the most established deep learning algorithm [12]. One of the main advantages of CNNs is their ability to learn automatically without human supervision, making them very popular [8]. However, their performance can be limited when training data is scarce or exclusive, especially with one-dimensional signals. Furthermore, deep CNNs face challenges such as training instability and feature duplication in facial expression recognition, especially when running on low-end devices [9, 27]. Nonetheless, researchers are developing new methods to enhance deep CNNs for facial expression recognition.

Most facial emotion recognition research has focused on adults, with CNNs yielding better results than customized methods. Several studies have achieved high accuracy rates for facial expression recognition using CNNs. For instance, Shaees et al. achieved 99.3% accuracy on thermal images using the Natural Visible and Infrared Expression (NVE) database and 98.3% accuracy on graphic illustrations using the Cohn-Kanade+ (CK+) database [28]. Similarly, Webb et al. achieved a state-of-the-art classification rate of 99.52% on a combined corpus of datasets [29]. Hassouneh et al. used virtual markers to gather six facial emotions from 55 college students and achieved a maximum recognition rate of 99.81% using CNN for emotion detection with facial landmarks [30].

Few studies have been conducted on infant facial expression recognition due to limited tagged datasets. However, some studies have proposed CNN-based techniques for infant facial expression recognition, achieving promising results. For example, one study achieved an accuracy of 87.16% and an F1-score of 62.54% using a CNN-based technique on the RCLA & NBH Smile Dataset. This study proposed a multi-label CNN for detecting infant facial action units, achieving 0.69 to 0.93 kappa with human coding [31].

Several methods can improve CNNs' emotion recognition performance. Weight initialization and data augmentation can improve CNNs' performance in emotion recognition [29, 32]. Ruiz-Garcia et al. used the stacked convolutional auto-encoder (SCAE) technique to pre-train each network layer as an unsupervised auto-encoder, and the results of this approach were compared with other methods [29]. Porcu et al. found that combining artificial images with other methods gave the best results [32]. Using a pre-trained visual object recognition network to recognize facial emotions and fine-tuning a pre-trained model with a cross-dataset strategy can also improve CNNs' performance [33, 34].

2. Materials and Methods

2.1. System overview.

The proposed system aimed to use machine learning models to recognize emotions in infants. Figure 1 presents an overview of the system, which involves a custom dataset of infant faces extracted from the AffectNet dataset, categorized into positive, neutral, and negative facial expressions. The dataset was then used to train a deep learning model that could recognize

infant emotions. Additionally, a mobile app was developed to collect facial image data of infants for this study. When the model was trained, it was integrated into the mobile app for infant development screening. This app can be used by caregivers, pediatricians, or other professionals to screen a child's socio-emotional development using emotion recognition in infants.
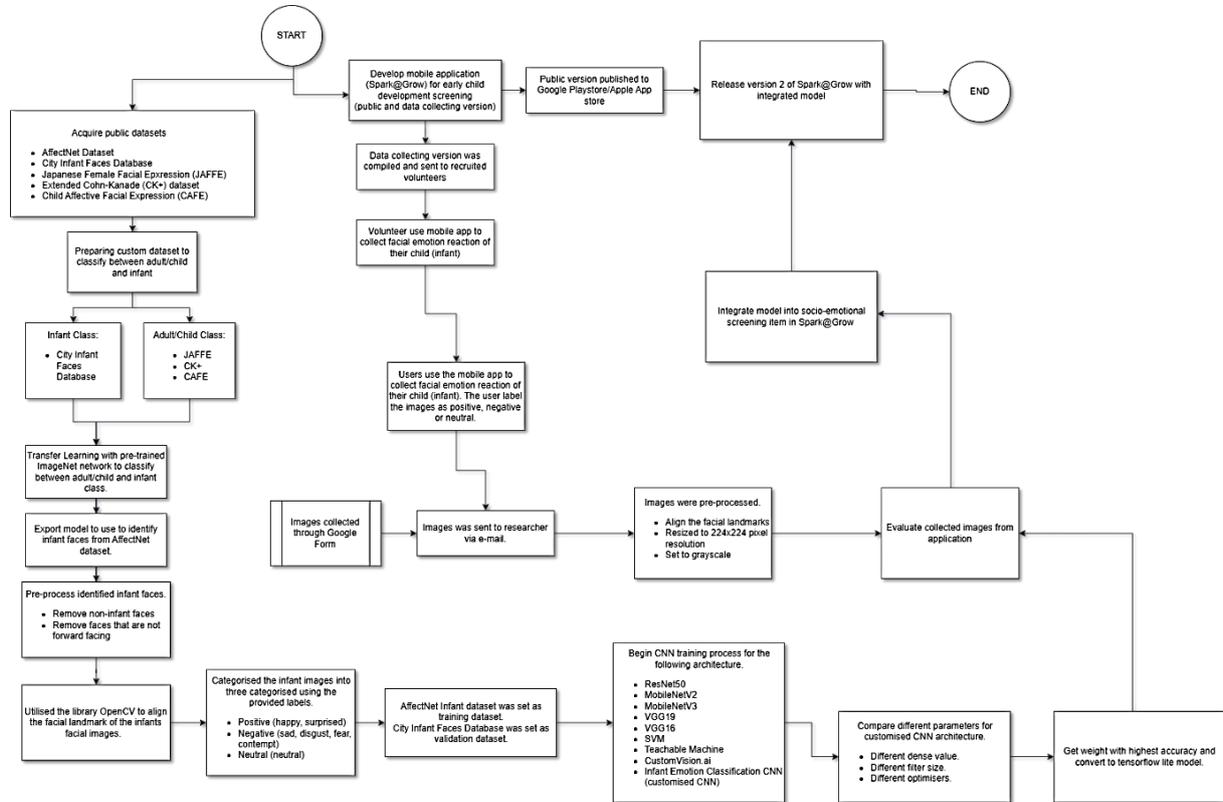

Figure 1. System overview.

2.2. Mobile application, Spark@Grow.

In this study, a mobile application called Spark@Grow was developed for the purpose of early child development screening. The Xamarin Forms framework was used to develop the app, enabling its publication on both the Android and iOS platforms by the first two authors, MLG and HYT. The psychologists (PJW and SWW) and pediatricians (THT and GLT) in the study team provided the early child development screening items, which were implemented in the app. The initial version of the app had two versions, namely the public and data-collecting versions. The public version was published on Google PlayStore and Apple Appstore, while the data-collecting version was shared with recruited parents and volunteers to collect image data for building the locally collected infant image dataset.

2.3. Datasets.

2.3.1. AffectNet dataset.

This dataset contains over 1,000,000 facial photos gathered from the Internet by searching three major search engines with 1,250 emotion-related keywords in six languages. Approximately half of the photos were manually tagged for seven distinct facial expressions, strength of valence and arousal. Furthermore, a random subset of 30,000 images was used to prevent the occurrence of an out-of-memory error.

2.3.2. City infant faces database.

The City Infant dataset is a collection of photographs that depict infants' faces displaying positive, negative, and neutral facial expressions. The dataset was developed and validated by collecting photographs from parents of infants aged between 0 to 12 months and rating the images based on facial expression, intensity, clarity, genuineness, and valence. The final database comprises 154 images with rating agreements of at least 75%, including 60 photographs of positive infant faces, 54 of negative infant faces, and 40 of neutral infant faces. The images exhibit high criterion validity and good test-retest reliability, making the City Infant dataset a useful and valid tool for researchers studying the emotional expressions of infants [35].

2.3.3. Non-Infant dataset

The following datasets were used to build the custom dataset for classifying non-infant facial images: Japanese Female Facial Expression (JAFFE) dataset, which consists of 213 images of different facial expressions from ten Japanese female subjects [36]; the Extended Cohn-Kanade (CK+) dataset, which is a collection of 593 video sequences from 123 subjects of various ages, genders, and heritage [37]; and the Child Affective Facial Expression (CAFE) dataset, which is a collection of images of kids between the ages of two and [38]. Sample images from the CAFE dataset will not be displayed in this paper as it is copyright protected.

2.3.4. Locally collected dataset.

Figure 2 displays the sample images that were gathered from local communities through various online platforms and communication channels, such as Facebook, Instagram, WhatsApp, and Telegram. Parents were invited to complete a Google Form and upload an image of their child's facial expression to support this study, with the understanding that the child's emotional images would only be used to evaluate the classification algorithm and not be disclosed to anyone else. A total of 88 facial emotion images were categorized into positive, neutral, and negative categories. The images were classified based on the emotion that the participants perceived; if the parent believed that the image showed the child expressing joy, it was placed in the positive category, while images depicting sadness, crying, or anger were categorized as negative. The neutral category comprised images that did not display any intense emotion, such as joy, sadness, or anger.

Figure 2. Sample images from the locally collected dataset (with parental written consent).

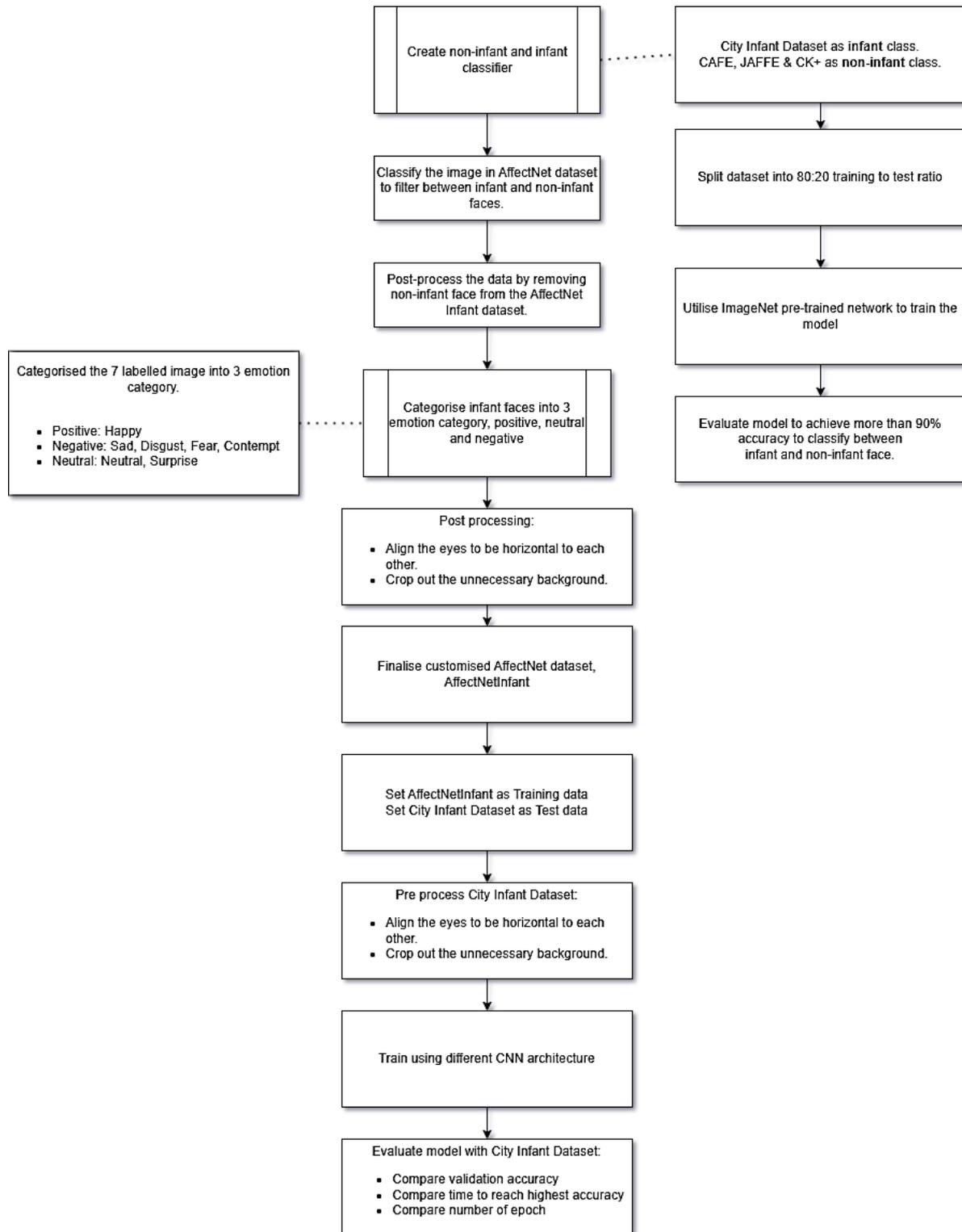*2.3.7. Preparing the custom dataset from AffectNet.*

Figure 3. Flowchart of preparing the custom dataset.

Figure 3 depicts the process of creating the custom infant dataset by extracting infant faces from the AffectNet dataset. A classifier was developed using transfer learning with a pre-trained network and images from the JAFFE, CK+, and CAFE datasets were used as the non-infant faces class, while the City Infant Dataset was used as the infant faces class. Around 6,000 infant face images were extracted, and five volunteers manually removed non-infant faces. The remaining images were categorized into positive, neutral, and negative emotion expression

categories, similar to the City Infant dataset. Finally, the custom dataset was utilized to train the machine learning models, and their performance was evaluated on the City Infant dataset for comparison purposes.

2.4. Hardware and software.

The deep learning model was developed and trained on an Nvidia RTX 2060 graphics card with 6GB VRAM. The model was built using the Keras API of the Tensorflow software library, which is widely used in machine learning and deep learning research. Python, NumPy, and OpenCV were also used to facilitate the training and evaluation of the model, including pre-processing the dataset, feature extraction, and performing data augmentation. OpenCV provides image and video processing tools, while NumPy supports large, multi-dimensional arrays and matrices. These tools helped to improve the model's performance in image classification tasks such as facial expression recognition.

2.5. Pre-process the datasets.

To pre-process the images, OpenCV was used to detect facial landmarks such as the eyes, nose, and mouth, which were used to align and center the facial images. The background was then cropped out, leaving only the aligned facial image, resized to a consistent size of 224 x 224 pixels. This process aimed to improve the machine learning models' performance by ensuring consistent and relevant features of the face were located and sized the same in all images. Resizing the images to a consistent size also aimed to reduce variability in the input data, potentially making it easier for the models to learn from the images.

2.6. Infant emotion classification CNN architecture.

Figure 4 shows the flowchart of the Infant Emotion Classification CNN model used in this study, specifically designed to recognize emotions in infant faces. The model had four blocks of convolutional layers, each followed by max pooling and dropout layers. The first block consisted of two convolutional layers with 32 filters and a kernel size of 3 x 3, which extracted features from the input image. The subsequent blocks increase the number of filters to 64, 128, and 256, respectively. The output was then passed through two fully connected (dense) layers with 64 units each and dropout layers, followed by a dense layer with the number of units equal to the number of classes and a softmax activation function. Finally, the model was trained using the Nadam optimization algorithm and evaluated on the validation data using the evaluate function.

2.7. Comparison between algorithms and architectures.

To compare the effectiveness of different methods in recognizing emotions in infants' faces, nine methods were evaluated, including the Infant Emotion Classification CNN, ResNet50, MobileNetV2, MobileNetV3, VGG19, VGG16, Multiclass Support Vector Machine, Teachable Machine, and CustomVision.AI. The Infant Emotion Classification CNN is a model specifically designed for recognizing emotions in infants, while ResNet50 is a general-purpose CNN that has achieved success in various tasks. MobileNetV2 and MobileNetV3 are lightweight CNNs designed for efficient performance on devices with limited resources, and VGG19 and VGG16 are deeper CNNs that have achieved success in various tasks. The Multiclass Support Vector Machine is a traditional machine learning algorithm that does not

require large datasets for training, while Teachable Machine is a tool for creating machine learning models with a simple graphical interface. CustomVision.AI is a cloud-based deep learning service that allows users to train and deploy models using a web interface.

To compare these methods, a custom dataset of infant faces extracted from the AffectNet dataset was used to train and evaluate each model. The City Infant dataset was used as the validation set. The accuracy of each method was measured on the validation set, and the time required to train each method on the training set was recorded. Various machine learning models were used in the training process, including the TensorFlow library and Keras API for the CNN, ResNet50, MobileNetV2, MobileNetV3, VGG19, and VGG16 models. The Multiclass Support Vector Machine was trained using the scikit-learn library, and the default training procedure for this model was used. The Teachable Machine was trained by following the instructions on the Teachable Machine website, which involved uploading, labeling, and training the model using the provided interface. Similarly, the CustomVision.AI model was trained by following the instructions on the CustomVision.AI website, which involved creating a project, uploading, labeling, and training the model using the provided interface.

The accuracy of each method was measured by calculating the percentage of correctly classified samples in the validation set, and the training time for each method was measured by timing the execution of the training procedure. The performance of the methods was compared based on the accuracy and training time. Additionally, the relative simplicity and accessibility of each method were considered, as well as any additional resources or technical expertise required to use it. Based on the results, the most suitable method for implementing infant emotion recognition in a mobile application was selected.
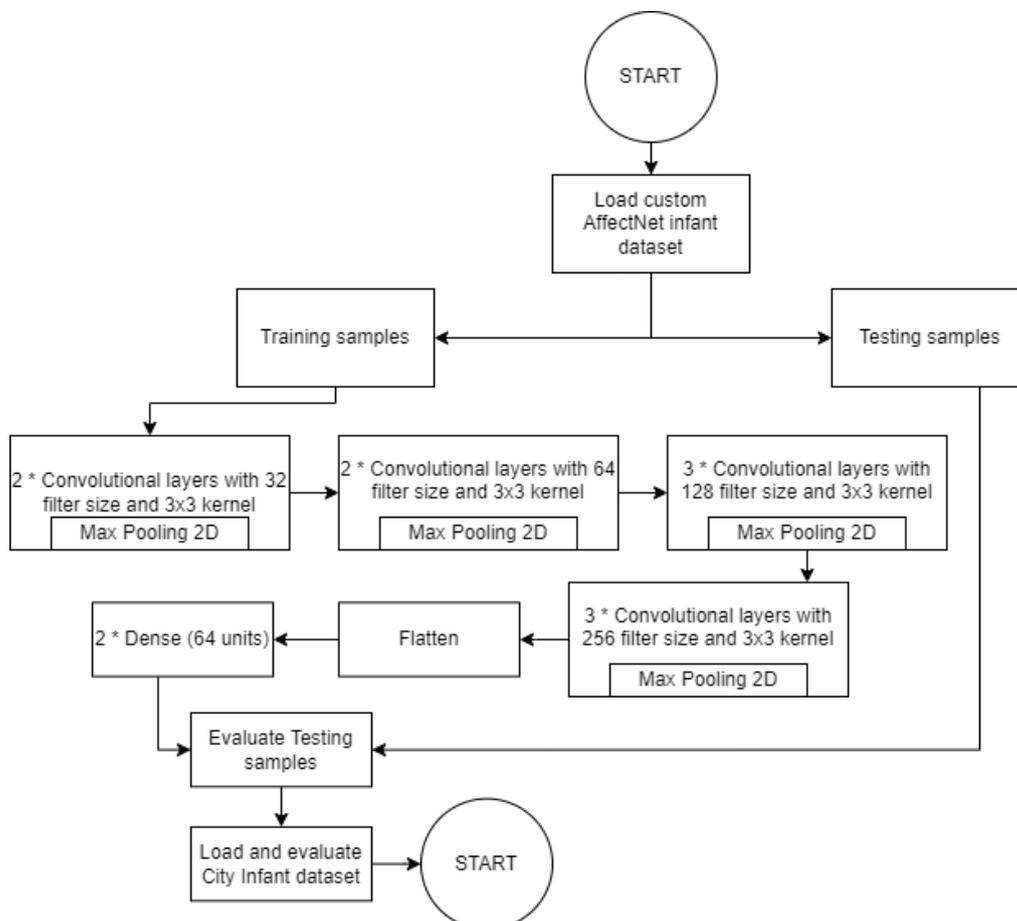

Figure 4. Flowchart of training infant emotion classification CNN model.

3. Results and Discussion

There were nine different models on the dataset of infant facial expressions: Infant Emotion Classification CNN, ResNet50, MobileNetV2, MobileNetV3, VGG19, VGG16, Multiclass Support Vector Machine (SVM), Teachable Machine, and CustomVision.AI. The results are shown in Table 1.

3.1. Validation accuracy and training speed.

Table 1. Validation accuracy and training speed of different infant facial emotion recognition models with the time to reach epoch and time to train.

Method	Validation Accuracy	Epoch with Highest Validation Accuracy	Time to reach epoch (s)	Time to train (s)
Infant Emotion Classification CNN	94.6%	17	2073	4894
Pre-trained with ImageNet: ResNet50	93.3%	34	4979	7176
Pre-trained with ImageNet: MobileNetV2	85.9%	7	176	1263
CustomVision.AI	85.9%	-	-	3584
Teachable Machine	81.9%	-	-	1800
Pre-trained with ImageNet: MobileNetV3	74.5%	14	216	772
Multiclass Support Vector Machine	53.0%	5	106	1061
Pre-trained with ImageNet: VGG19	39.6%	1	62	3102
Pre-trained with ImageNet: VGG16	33.5%	1	49	2482

Table 1 displays the results of experiments using different methods to train the AffectNet custom dataset and test on the City Infant dataset. Each method was tested three times. Among them, the Infant Emotion Classification CNN model achieved the highest validation accuracy, with 94.63%. The ResNet50 model also performed well with a validation accuracy of 93.3%, although it took longer to reach its accuracy compared to the Infant Emotion Classification CNN. The MobileNetV2 and CustomVision.AI models achieved similar accuracy, at 85.9%. The Teachable Machine and MobileNetV3 had intermediate accuracy, achieving 81.9% and 74.5% training accuracy, respectively, whereas the Multiclass SVM had relatively low accuracy, with only 53.0%. The VGG19 and VGG16 models had significantly lower accuracy, at 39% and 33%, respectively. These results suggest that the Infant Emotion Classification CNN and ResNet50 models were the most effective for infant facial emotion recognition, with validation accuracies of 94.63% and 93%, respectively. These models can learn complex features from the data and generalize well to unseen examples due to their deep architectures and advanced training techniques [39].

While the MobileNetV2 and MobileNetV3 models were designed to be efficient and lightweight, making them suitable for deployment on resource-constrained devices [40], their lower accuracy compared to the ResNet models suggests that they may not be as effective at learning complex features from the data. However, they were relatively fast in reaching their highest accuracy, taking 176 seconds with 85.9% accuracy and 216 seconds with 74.5% accuracy, respectively. The Multiclass SVM, Teachable Machine, and CustomVision.AI models had intermediate training speeds to reach their highest accuracy, taking 106 seconds with 53% accuracy, 1,800 seconds with 81.9% accuracy, and 3,584 seconds with 85.9% accuracy, respectively. The SVM model was less effective, and the CustomVision.AI model's performance may not be as good as models trained and optimized using more advanced techniques similar to the Teachable Machine model.

The lower accuracy of VGG19 and VGG16 may be due to their less suitability for the task because of the lack of data for training [41]. However, in terms of training speed, these models have relatively simple architectures and are, therefore, easier to train. Nevertheless, their lower accuracy compared to the ResNet models suggests that they may not be as effective at learning complex features from the data. The Infant Emotion Classification CNN and ResNet50 models had the slowest training speed to reach their highest accuracy, taking 582 seconds with 94.63% accuracy and 4,979 seconds with 93.3% accuracy, respectively. Although the accuracy difference was only 1.33%, the Infant Emotion Classification CNN model achieved its highest accuracy with about 58% lower training time. Therefore, these models have more complex architectures and are more computationally intensive to train. However, their higher accuracy compared to the other models suggests that they can learn more complex features from the data and generalize better to unseen examples, making them more effective at capturing the nuances of infant facial expressions.

The results of the experiment are significant for the use of machine learning in the field of infant facial emotion recognition. The high accuracy of the Infant Emotion Classification CNN and ResNet50 models indicates that these models could be valuable for accurately detecting and classifying infants' emotions based on their facial expressions. Conversely, the lower performance of the VGG19 and VGG16 models and the intermediate performance of the MobileNetV2 and MobileNetV3 models suggest that these models might not be as suitable for this task. These findings underscore the importance of carefully selecting an appropriate machine learning model for the task at hand, considering both accuracy and training speed. For infant facial emotion recognition, the Infant Emotion Classification CNN and ResNet50 models seem to be the most effective, with high accuracy and relatively fast training speed.

3.2. Evaluation of locally collected dataset.

Table 2 demonstrates that the classifier had the highest accuracy in detecting neutral emotion, with 82.3% of the images correctly classified. This was followed by positive emotion, with 78.8% accuracy, and negative emotion, with 69.8% accuracy. Overall, the classifier performed relatively well in detecting positive and neutral emotions, with accuracy rates above 70%. However, the accuracy for detecting negative emotion was lower, with almost a third of the images being misclassified. This might be due to the subtle nature of negative expressions, which could be more difficult to distinguish than more intense positive and neutral expressions. Finally, it is worth noting that the accuracy rates for positive and neutral emotions were similar, but the classifier had a higher number of correct detections for neutral emotions due to the smaller number of images in this category. This highlights the importance of having sufficient examples in each category to obtain reliable results.

Table 2. Evaluating locally collected dataset.

Emotion	Correct Detect / Total Image	Accuracy
Negative	30 / 43	69.8%
Neutral	14 / 17	82.3%
Positive	41 / 52	78.8%

4. Conclusion

The proposed Infant Emotion Classification CNN model achieved the highest accuracy and quickest training time, demonstrating the effectiveness of machine learning techniques in identifying infant emotions. The model's high accuracy was due to the unique dataset created

by excluding baby photos from the AffectNet dataset. The study's methodology involved testing several parameters. The Nadam optimizer, a feature set with a beginning value of 32 and a dense parameter of 64, produced the best results. These findings offer a useful benchmark for future research on identifying newborn emotions and suggest potential applications for developing emotion-sensitive devices for infants. However, there is still room for improvement in the model's performance. Further research could explore using more extensive and diverse datasets and advanced machine-learning models. Machine learning for infant emotion recognition has the potential to support the emotional well-being of infants by enabling the development of emotion-sensitive technologies. These results are helpful for other researchers working on similar tasks and provide a benchmark for evaluating various machine learning models' performance on infant emotion recognition. Further research can build upon these findings and explore the use of machine learning in other applications related to infant development and care.

Acknowledgments

References

[1]   Hadders-Algra, M. (2021). Early diagnostics and early intervention in neurodevelopmental disorders—age-dependent challenges and opportunities. *Journal of Clinical Medicine, 10*, 861. https://doi.org/10.3390/jcm10040861.

[2]   Lipkin, P.H.; Macias, M.M.; Norwood, K.W.; Brei, T.J.; Davidson, L.F.; Davis, B.E.; Voigt, R.G. (2020). Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics, 145,* e20193449. https://doi.org/10.1542/peds.2019-3449.

[3]   Silver, H.C.; Zinsser, K.M. (2020). The interplay among early childhood teachers' social and emotional well-being, mental health consultation, and preschool expulsion. *Early Education and Development, 31,* 1133−1150. https://doi.org/10.1080/10409289.2020.1785267.

[4]   Khan, A.A.; Shaikh, A.A.; Shaikh, Z.A.; Laghari, A.A.; Karim, S. (2022). IPM-Model: AI and metaheuristic-enabled face recognition using image partial matching for multimedia forensics investigation with genetic algorithm. *Multimedia Tools and Applications, 81*, 23533−23549. http://doi.org/10.1007/s11042-022-12398-x.

[5]   Oreški, D.; Androček, D. (2020). Genetic algorithm and artificial neural network for network forensic analytics. 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 1200−1205. https://doi.org/10.23919/MIPRO48935.2020.9245140.

[6]   Bonizzoni, P.; Costantini, M.; De Felice, C.; Petescia, A.; Pirola, Y.; Previtali, M.; Zizza, R. (2022). Numeric Lyndon-based feature embedding of sequencing reads for machine learning approaches. *Information Sciences, 607*, 458−476. https://doi.org/10.1016/j.ins.2022.06.005.

[7]   Das, S.; Sett, S.; Saha, S.; Saha, S.; Roychoudhury, S.; Pal, K. (2021). A Novel Face Detection Technique Using OpenCV. *International Journal of Research in Engineering, Science and Management, 4,* 121−124.

[8] Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*, 1−74. https://doi.org/10.1186%2Fs40537-021-00444-8.

[9] Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing, 151*, 107398. https://doi.org/10.1016/j.ymssp.2020.107398.

[10] Akhand, M.A.H.; Roy, S.; Siddique, N.; Kamal, M.A.S.; Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. *Electronics, 10*, 1036. https://doi.org/10.1109/ICoICT52021.2021.9527512.

[11] Joseph, A.; Geetha, P. (2020). Facial emotion detection using modified eyemap-mouthmap algorithm on an enhanced image and classification with tensorflow. *The Visual Computer, 36*, 529−539. https://doi.org/10.1007/s00371-019-01628-3.

[12] Cengiz, E.; Yilmaz, C.; Kahraman, H. (2021). Classification of human and vehicles with the deep learning based on transfer learning method. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 9*, 215−225. http://doi.org/10.29130/dubited.842394.

[13] Hassouneh, A.; Mutawa, A.M.; Murugappan, M. (2020). Development of a Real-Time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked, 20*, 100372. https://doi.org/10.1016/j.imu.2020.100372.

[14] Nagpal, S.; Singh, M.; Vatsa, M.; Singh, R.; Noore, A. (2019). Expression classification in children using mean supervised deep Boltzmann machine. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

[15] Witherow, M.A.; Samad, M.D.; Iftekharuddin, K.M. (2019). Transfer learning approach to multiclass classification of child facial expressions. *Applications of Machine Learning, 11139*, 275−282. http://doi.org/10.1117/12.2530397.

[16] Banerjee, A.; Mutlu, O.C.; Kline, A.; Surabhi, S.; Washington, P.; Wall, D.P. (2022). Training and profiling a pediatric facial expression classifier for children on mobile devices: Machine learning study. *Journal of Medical Internet Research Formative Research, 7,* e39917. https://doi.org/10.2196/39917.

[17] Washington, P.; Kalantarian, H.; Kent, J.; Husic, A.; Kline, A.; Leblanc, E.; Wall, D.P. (2022). Improved digital therapy for developmental pediatrics using domain-specific artificial intelligence: Machine learning study. *Journal of Medical Internet Research Pediatrics and Parenting, 5,* e26760. https://doi.org/10.2196/26760.

[18] Guzman, A.L.; Lewis, S.C. (2020). Artificial intelligence and communication: A Human-Machine Communication research agenda. *New Media & Society, 22*, 70−86. http://doi.org/10.1177/1461444819858691.

[19] Holzinger, A.; Müller, H. (2021). Toward human-AI interfaces to support explainability and causability in medical AI. *Computer, 54*, 78−86.

[20] Pan, H.; Xie, L.; Lv, Z.; Li, J.; Wang, Z. (2020). Hierarchical support vector machine for facial micro-expression recognition. *Multimedia Tools and Applications, 79*, 31451−31465. https://doi.org/10.1007/s11042-020-09475-4.

[21] Sharma, S.; Bhatt, M.; Sharma, P. (2020). Face recognition system using machine learning algorithm. 5th International Conference on Communication and Electronics Systems (ICCES), 116.

[22] Mohan, L.; Pant, J.; Suyal, P.; Kumar, A. (2020). Support vector machine accuracy improvement with classification. 12th International Conference on Computational Intelligence and Communication Networks (CICN), 477−481.

[23] Sun, Z.; Chiong, R.; Hu, Z.P. (2020). Self-adaptive feature learning based on a priori knowledge for facial expression recognition. *Knowledge-Based Systems, 204,* 106124. https://doi.org/10.1016/j.knosys.2020.106124.

[24] Wu, J.; Xia, Z.; Feng, X. (2022). Improving Adversarial Robustness of CNNs via Maximum Margin. *Applied Sciences, 12*, 7927. https://doi.org/10.3390/app12157927.

[25] Chouhayebi, H.; Riffi, J.; Mahraz, M.A.; Yahyaouy, A.; Tairi, H.; Alioua, N. (2020, June). Facial expression recognition based on geometric features. International Conference on Intelligent Systems and Computer Vision (ISCV), 1−6.

[26] Punuri, S.B.; Kuanar, S.K.; Kolhar, M.; Mishra, T.K.; Alameen, A.; Mohapatra, H.; Mishra, S.R. (2023). Efficient Net-XGBoost: An Implementation for Facial Emotion Recognition Using Transfer Learning. *Mathematics, 11*, 776. https://doi.org/10.3390/math11030776.

[27] Ma, H.; Celik, T.; Li, H. C. (2021). Lightweight attention convolutional neural network through network slimming for robust facial expression recognition. *Signal, Image and Video Processing, 15*, 1507−1515. https://link.springer.com/article/10.1007/s11760-021-01883-9.

[28] Tannugi, D.C.; Britto, A.S.; Koerich, A.L. (2019). Memory integrity of CNNs for cross-dataset facial expression recognition. IEEE International Conference on Systems, Man and Cybernetics (SMC), 3826−3831.

[29] Pons, G.; Masip, D. (2017). Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis. *IEEE Transactions on Affective Computing*, *9,* 343−350. https://doi.org/10.1109/TAFFC.2017.2753235.

[30] Shaees, S.; Naeem, H.; Arslan, M.; Naeem, M. R.; Ali, S. H.; Aldabbas, H. (2020). Facial emotion recognition using transfer learning. International Conference on Computing and Information Technology (ICCIT-1441). https://doi.org/10.1109/iccit-144147971.2020.9213757.

[31] Chronaki, G.; Hadwin, J.A.; Garner, M.; Maurage, P.; Sonuga-Barke, E.J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33*, 218−236. http://doi.org/10.1111/bjdp.12075.

[32] Webb, N.; Ruiz-Garcia, A.; Elshaw, M.; Palade, V. (2020). Emotion recognition from face images in an unconstrained environment for usage on Social robots. International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/ijcnn48605.2020.9207494.

[33] Akamatsu, M.J.S.L.; Kamachi, M.; Gyoba, J.; Budynek, J (1998). The Japanese female facial expression (JAFFE) database. Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14−16.

[34] Lucey, P.; Cohn, J. F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 94-101.

[35] LoBue, V.; Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology, 5,* 1532. https://doi.org/10.3389/fpsyg.2014.01532.

[36] Alshamrani, H. A.; Rashid, M.; Alshamrani, S. S.; Alshehri, A. H. (2023). Osteo-NeT: An Automated System for Predicting Knee Osteoarthritis from X-ray Images Using Transfer-Learning-Based Neural Networks Approach. *Healthcare, 11*, 1206. http://doi.org/10.3390/healthcare11091206.

[37] Cui, C.; Gao, T.; Wei, S.; Du, Y.; Guo, R.; Dong, S.; Lu, B.; Zhou, Y.; Lv, X.; Liu, Q.; Hu, Z.; Yu, D.; Ma, Y. (2021). PP-LCNet: A lightweight CPU convolutional neural network. https://doi.org/10.48550/arXiv.2109.15099.