A HYBRID APPROACH TO ADDRESS DATA SCARCITY AND IMBALANCE IN INFANT FACIAL EXPRESSION ANALYSIS

LIBRARY FTSM

WU TONG

UNIVERSITI KEBANGSAAN MALAYSIA

A HYBRID APPROACH TO ADDRESS DATA SCARCITY AND IMBALANCE IN INFANT FACIAL EXPRESSION ANALYSIS

LIBRARY FTSM

WU TONG

PROJECT SUBMITTED IN FULFILMENT FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF INFORMATION SCIENCE AND TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2025

PENDEKATAN HIBRID UNTUK MENANGANI KEKURANGAN DAN KETIDAKSEIMBANGAN DATA DALAM ANALISIS EKSPRESI WAJAH BAYI

WU TONG

LIBRARY FTSM

PROJEK YANG DIKEMUKAKAN UNTUK MEMPEROLEH IJAZAH SARJANA SAINS DATA

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2025

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

20 February 2025

WU TONG
P130105

LIBRARY FTSM

ACKNOWLEDGEMENTS

I am lucky to have the supervision of Dr. Elankovan A Sundararajan in my study. To Dr. Elankovan A Sundararajan, I would like to convey my sincere gratitude. He is a wonderful and patient supervisor who quickly responded and solved my questions during my research and dissertation writing.

To my family, I want to thank them sincerely for providing me the opportunity to study at UKM. Thanks for their care, assistance, support and understanding during my postgraduate education and thesis writing.

I want to appreciate all of the postgraduate DS students at UKM for their help and kindness during my time there.

LIBRARY FTSM

PENDEKATAN HIBRID UNTUK MENANGANI KEKURANGAN DAN KETIDAKSEIMBANGAN DATA DALAM ANALISIS EKSPRESI WAJAH BAYI

ABSTRAK

Generative Adversarial Networks (GANs) dapat mempelajari ciri taburan data dan menghasilkan sampel baharu yang sejajar dengan taburan data asal. Laporan ini meneroka penggunaan GANs sebagai teknik oversampling untuk meningkatkan prestasi klasifikasi set data yang tidak seimbang. Walau bagaimanapun, apabila berhadapan dengan set data yang sangat tidak seimbang, terutamanya dalam konteks analisis ekspresi wajah bayi berumur 0-2 tahun, bilangan sampel kelas minoriti yang terhad sering menghalang Conditional Generative Adversarial Networks (CGANs) daripada mempelajari taburan kelas minoriti sepenuhnya, yang membawa kepada penghasilan sampel sintetik berkualiti rendah. Untuk menangani batasan ini secara sistematik, laporan ini mencadangkan satu kaedah klasifikasi berkumpulan berdasarkan CGAN yang dipertingkatkan untuk set data tidak seimbang. Metodologi ini terdiri daripada tiga peringkat: Pertama, kaedah SMOTEENN (Synthetic Minority Oversampling Technique combined with Edited Nearest Neighbor) digunakan untuk mengembangkan set data ekspresi wajah bayi asal dengan menjana sampel kelas minoriti dan membuang sampel yang bising, memastikan keseimbangan dan kualiti data. Kemudian, CGAN dilatih pada set data yang telah diproses untuk menangkap sepenuhnya ciri taburan kelas minoriti dan menghasilkan imej ekspresi wajah bayi sintetik berkualiti tinggi untuk terus mengimbangi set data tersebut. Akhir sekali, pengklasifikasi Classification and Regression Tree (CART) yang digabungkan dengan algoritma Adaboost yang diubah suai digunakan untuk melatih dan menilai prestasi klasifikasi pada set data yang seimbang. Hasil eksperimen menunjukkan peningkatan yang signifikan dalam prestasi klasifikasi, terutamanya untuk kelas minoriti, berbanding dengan kaedah oversampling berasaskan CGAN asal, dengan skor F1 meningkat secara purata sebanyak 6.3%, G-mean sebanyak 2.3%, dan AUC sebanyak 5.6%, secara berkesan mengurangkan keterbatasan prestasi yang disebabkan oleh ketidakseimbangan data dalam analisis ekspresi wajah bayi berumur 0-2 tahun.

LIBRARY FTSM

ABSTRACT

Generative Adversarial Networks (GANs) can learn the distribution characteristics of data and generate new samples that align with the original data distribution. This report explores the application of GANs as an oversampling technique to effectively improve the classification performance of imbalanced datasets. However, when dealing with extremely imbalanced datasets, particularly in the context of 0-2 year-old infant facial expression analysis, the limited number of minority class samples often prevents Conditional Generative Adversarial Networks (CGANs) from fully learning the minority class distribution, leading to poor-quality synthetic samples. To systematically address this limitation, this report proposes an ensemble classification method based on an improved CGAN for imbalanced datasets. The methodology consists of three stages: First, the SMOTEENN (Synthetic Minority Oversampling Technique combined with Edited Nearest Neighbor) method is applied to expand the original infant facial expression dataset by generating minority class samples and removing noisy samples, ensuring both data balance and quality. Then, a CGAN is trained on the preprocessed dataset to fully capture the distribution characteristics of the minority class and generate high-quality synthetic infant facial expression images to balance the dataset further. Finally, a Classification and Regression Tree (CART) classifier, combined with a modified Adaboost algorithm, is used to train and evaluate the classification performance on the balanced dataset. The experimental results demonstrated significant improvements in the classification performance, especially for the minority class, compared to the original CGAN-based oversampling method, with F1 score improved by an average of 6.3%, G-mean by 2.3%, and AUC by 5.6%, effectively mitigating the performance limitations caused by data imbalance in 0-2 year-old infant facial expression analysis.

LIBRARY FTSM

TABLE OF CONTENTS

		Page
DECLARATION		**iii**
ACKNOWLEDGEMENTS		**iv**
ABSTRAK		**v**
ABSTRACT		**vi**
TABLE OF CONTENTS		**ix**
LIST OF FIGURES		**x**
LIST OF TABLES		**xi**
LIST OF ABBREVIATIONS		**xii**
CHAPTER I	**INTRODUCTION**	
1.1	Introduction	1
1.2	Problem Background	3
1.3	Problem Statement	6
1.4	Research Objectives	7
1.5	Research Scope	8
1.6	Thesis Organization	8
CHAPTER II	**LITERATURE REVIEW**	
2.1	Introduction	10
2.2	Importance Of Infant Facial Expression Analysis	11
2.3	Existing Baby Facial Datasets	12
	2.3.1 City Infant Faces Datasets	13
	2.3.2 Babyexp	14
	2.3.3 The Child Emotion Facial Expression Set	15
	2.3.4 Tromso Infant Faces Datasets	15
	2.3.5 Rebel Dataset	15
	2.3.6 Emoreact	15
	2.3.7 The Multimodal Dyadic Behavior Dataset	16
	2.3.8 Child Affective Facial Expression Set	16
	2.3.9 NIMH-CHEFS	16

LIBRARY FTSM

2.4	Facial Expression Analysis Method	16
	2.4.1 Traditional Machine Learning Methods	19
	2.4.2 Deep Learning Methods	22
2.5	Dataset Construction Challenges And Methods	26
2.6	Applications Of Gans In Image Generation	28
	2.6.1 Standard Gan	30
	2.6.2 Conditional Gan	32
	2.6.3 Deep Convolutional Gan	34
2.7	Preprocessing Techniques	38
	2.7.1 Synthetic Minority Oversampling Technique	38
	2.7.2 Edited Nearest Neighbor	39
2.8	Summary	40
CHAPTER III	**METHODOLOGY**	
3.1	Introduction	42
3.2	The Research Framework	43
3.3	Hardware And Software Environments	44
3.4	Dataset	45
3.5	Data Preprocessing	46
	3.5.1 Synthetic Minority Oversampling Technique	47
	3.5.2 Edited Nearest Neighbor	48
	3.5.3 Preprocessing Results	49
3.6	Research Generative Modeling	50
	3.6.1 Cgan Architecture	50
	3.6.2 Training Process	52
	3.6.3 Sample Quality Evaluation Based on JS Divergence	53
3.7	Summary	55
CHAPTER IV	**RESULT AND DISCUSSION**	
4.1	Introduction	57
4.2	AdaBoost with F1-Based Weighting	57

LIBRARY FTSM

4.3	Experimental Results	59
4.4	Summary	61
CHAPTER V	**CONCLUSION AND FUTURE WORKS**	
5.1	Introduction	63
5.2	Research Summary	63
5.3	Strengths And Limitations	64
5.4	Future Works	65
REFERENCES		66

LIBRARY FTSM

LIST OF FIGURES

Figure No.		Page
1	Sample images from the City Infant Faces Database	14
2	Sample images from Babyexp	14
3	The new CNN model (Goumiri, S. et al., 2023).	25
4	Training process of CNN model for face emotion recognition (lu 2022.)	26
5	A graphical overview of the GAN architecture (Little, C. et al., 2023).	31
6	Balancing and Enhancing Classification Framework	44
7	Example images from our Collective datasets	46
8	Examples of Synthetic Infant Expressions	50

LIBRARY FTSM

LIST OF TABLES

Table No.		Page
1	Overview of the existing facial expression datasets.	13
2	Emotion Category Before and After SMOTE-ENN Processing	49
3	JS Divergence for Synthetic Samples by Test Iteration	55
4	Proposed Method Performance on Various Datasets	60

LIBRARY FTSM

LIST OF ABBREVIATIONS

GANs	Generative Adversarial Networks
CGAN	Conditional Generative Adversarial Network
DCGAN	Deep Convolutional Generative Adversarial Network
CNN	Convolutional Neural Networks
RNN	Recurrent Neural Networks
HMM	Hidden Markov Model
DBN	Deep Belief Networks
PSO	Particle Swarm Optimization
SMOTE	Synthetic Minority Oversampling Technique
ENN	Edited Nearest Neighbor
CART	Classification and Regression Tree
AUC	Area Under the Curve
G-Mean	Geometric Mean
JS Divergence	Jensen-Shannon Divergence

LIBRARY FTSM

CHAPTER I

INTRODUCTION

1.1 INTRODUCTION

Generative Adversarial Networks (GANs) have fundamentally changed the field of artificial intelligence and machine learning since they were proposed by Ian J. Goodfellow et al. in 2014 (Goodfellow et al., 2014). It is composed of two main parts: generator (G) and discriminator (D). Generative Adversarial Networks (GANs) can learn and generate complex high-dimensional patterns. For instance, various 3D GAN models for lung tissue modeling in pulmonary CT were evaluated by Sam Ellis et al. in 2022 (Ellis et al., 2022), highlighting both their capabilities and limitations in generating data that resembles real distributions. The influence of GANs has spanned various fields, among which in art they have been used to create realistic images, music, text, and video (Ali et al., 2021). In the field of computer vision, GANs have been shown to be effective in generating high-quality images, which can effectively mitigate the shortcomings of traditional algorithms in processing high-dimensional data (Wang et al., 2019). In the medical field, especially in disease detection and surgical simulation, GANs can assist in making diagnoses and treatment plans by generating realistic medical images (Jin et al., 2020). At the same time, their wide applicability and advantages have also been demonstrated in the fields of weather forecasting models and materials science (Leinonen et al., 2019; Yang et al., 2019). One of the main advantages of GANs is their ability to model large datasets efficiently, opening up new possibilities for data representation and analysis. By combining GANs with unsupervised learning algorithms, potential prediction problems in the network can be shown, while object detection can also be performed, and other applications. In addition, GANs are now being used to enhance data security, particularly in tasks such

LIBRARY FTSM

as botnet detection and image authentication (Randhawa et al., 2021; Poredi, 2024). These networks operate on the principles of game theory, where the generator is responsible for generating adversarial samples and the discriminator is responsible for distinguishing between real and generated samples (Randhawa et al., 2021). GANs are a transformative technology with profound implications for artificial intelligence and machine learning. They have good universality and effectiveness in data generation, distribution modeling, and complex problem-solving, and have important significance in promoting research and application in various fields.

The analysis of babies' facial expressions is a new and challenging area of artificial intelligence research, with implications for fields as diverse as child psychology, medicine, and computer science. Understanding and being able to interpret a baby's facial expressions can provide valuable insights into their emotional state, cognitive development, and overall health. However, a major challenge in this area is the small number of relevant datasets, as it would be labor-intensive and ethically sensitive to collect large datasets on infant facial expressions (Messinger et al., 2009). In this context, GANs offer a promising solution to the lack of datasets on infant facial expressions. With increasing attention to the application of GANs in facial expression synthesis and analysis tasks, they have shown a good ability to generate realistic facial expressions from limited data (Yan et al., 2020).

This study proposes a comprehensive data augmentation pipeline to address the limitations of existing infant facial expression datasets. Initially, the dataset is balanced using SMOTE (Synthetic Minority Oversampling Technique) to generate synthetic samples for minority emotion categories (Chawla et al., 2002). To ensure data quality, ENN (Edited Nearest Neighbor) is applied to remove noisy or redundant samples (Wilson, 1972). These preprocessing steps result in a cleaner and more balanced dataset that serves as input to Conditional GANs (CGAN). CGAN is then employed to generate high-quality synthetic facial expression images for underrepresented categories (Mirza & Osindero, 2014). Finally, the augmented dataset is validated using Classification and Regression Tree (CART), which classify the facial expressions and demonstrate the effectiveness of the proposed pipeline in improving recognition performance, especially for minority classes.

LIBRARY FTSM

By leveraging Generative Adversarial Networks (GANs), researchers can create synthetic datasets of babies' facial expressions, which can then be used to effectively train facial expression recognition models. In addition, GANs can train deep learning models by generating realistic synthetic faces, thereby helping to improve the accuracy of facial expression analysis (Bozorgtabar et al., 2019). At the same time, the synthetic data generated by GANs also helps to improve the performance of the expression recognition system, which is especially useful when the real training images are limited. Furthermore, GANs can be used to transfer facial expressions between different faces, increasing the diversity of datasets and facilitating better model generalization for expression analysis (Fan et al., 2023). By integrating SMOTE, ENN, CGAN, and CART, this study provides a novel and scalable approach to overcoming the challenges of imbalanced and insufficient datasets for infant facial expressions. This systematic framework not only enhances the recognition accuracy for underrepresented emotion categories but also offers valuable insights into the emotional and developmental analysis of infants.

LIBRARY FTSM

1.2 PROBLEM BACKGROUND

Analyzing an infant's facial expressions can provide important insights into an infant's emotional state, cognitive development, and overall health (Parsons et al., 2010). Babies primarily express their needs, emotions, and experiences through facial expressions, so accurate understanding and interpretation of these expressions is critical for caregivers, parents, and healthcare professionals. Changes in a baby's facial expression can indicate discomfort, pain, or underlying health problems, and continuous monitoring can help detect distress or illness at an early stage, leading to timely intervention and improved care for the child. The ability of infants to express facial expressions by distinguishing between positive and negative emotions is related to their emotional development and parents' interaction with them (Messinger, 2002). Studying how babies respond to different facial expressions can enhance understanding of their emotional processing abilities and social cognition. Analyzing a baby's facial expressions is also important to identify risk factors or early signs of developmental disorders or mental health problems. For example, mothers with postpartum depression may have difficulty accurately interpreting negative facial expressions expressed by their infants (Stein et al., 2010). Understanding how caregivers perceive and respond to infants' facial expressions sheds light on a potential link between parental mental health and infant health.

In this context, generative adversarial networks (GANs) offer a promising approach by addressing data scarcity and improving the accuracy of facial expression recognition models (Pumarola et al., 2018). By using GANs to generate simulated datasets of infant facial expressions, researchers can enhance data that can be used for training and validation, resulting in more robust analytical results. Similarly, GANs can help researchers create realistic facial expressions for training deep learning models, thereby improving the performance of facial expression recognition systems. However, traditional GAN-based data augmentation methods have limitations when applied to highly imbalanced and scarce datasets, such as those available for infant facial expressions. Standard GANs often struggle to generate diverse samples, particularly for minority classes, due to the lack of sufficient training examples from these categories. This results in synthetic data that is either overly similar to the original samples or does not accurately reflect the underlying distribution of minority classes (Yi et al., 2019). Moreover, traditional GANs are not equipped to handle noise or redundant data present in the initial dataset, which can propagate through the generative process and negatively impact the quality of the augmented data. These limitations hinder the ability of traditional GANs to fully address the challenges of dataset imbalance and insufficient diversity.

To overcome these limitations, this study adopts a multi-step pipeline that integrates SMOTE (Synthetic Minority Oversampling Technique), ENN (Edited Nearest Neighbor), and Conditional GANs (CGANs). First, SMOTE is used to generate initial synthetic samples for minority emotion categories, effectively balancing the dataset and mitigating class imbalance at the outset (Chawla et al., 2002). Next, ENN is applied to remove noise and redundant samples, ensuring a cleaner and higher-quality dataset for the generative model (Wilson, 1972). Finally, CGANs are employed to further augment the dataset, focusing on generating realistic and diverse

LIBRARY FTSM

samples for underrepresented categories (Mirza & Osindero, 2014). This combination of techniques allows for a more targeted and effective data augmentation process that ensures both quality and diversity in the final dataset.

The scarcity of infant facial expression datasets is a major challenge, as it limits the variety and number of expressions that can be analyzed. Limited datasets may not adequately capture the entire aspect of infant facial expressions, leading to gaps in the understanding of infant emotional cues and responses (Peltola et al., 2007). The lack of diversity in existing infant facial expression datasets may hinder the development of accurate and robust facial expression recognition systems. If the dataset contains a small number of expressions or a large number of repetitive samples, it may not be possible to capture the variability of infant facial expressions in different backgrounds, emotions, or individuals (Haviland & Lelwica, 1987). This lack of diversity can limit the generalization and effectiveness of machine learning models.

Compared to traditional single-step GAN augmentation, the proposed pipeline addresses these pain points by combining oversampling, noise removal, and generative modeling into a cohesive process. This ensures that the generated data not only balances the dataset but also represents the diverse range of infant facial expressions. Unlike traditional GANs, which rely solely on the generative process, this multi-step approach leverages the strengths of preprocessing and advanced generative techniques to produce high-quality synthetic datasets. As a result, the enhanced dataset better supports the training of machine learning models, ultimately improving the accuracy and robustness of infant facial expression recognition systems.

To address these limitations, efforts are needed to improve the quality and diversity of infant facial expression datasets. Collecting data from a larger, more diverse group of infants, a step that includes infants from different cultural backgrounds, age groups, and emotional states, can help create more representative datasets. This study, by combining SMOTE, ENN, and CGANs, seeks to enhance the comprehensiveness of facial expression recognition systems through an innovative pipeline. By generating synthetic data and ensuring diversity while preserving data quality, the proposed approach addresses the scarcity and lack of variability in infant

LIBRARY FTSM

facial expression datasets. This will ultimately improve the performance of machine learning models, advancing the understanding and analysis of infant emotions across psychology, medicine, and computer science.

1.3 PROBLEM STATEMENT

The scarcity of infant facial expression data presents a significant challenge for the development of accurate and robust facial expression recognition systems. Collecting infant facial expression data is extremely challenging in practice due to the difficulty of collecting data from infants on a large scale and continuously, coupled with ethical sensitivities regarding the privacy and rights of minors. Many parents are reluctant to share photos of their children, making traditional methods of data collection impractical. Existing datasets are often limited in size, diversity, and representation, making them inadequate to train complex machine learning models. These limitations lead to significant gaps in the ability of machine learning algorithms to generalize and perform effectively in analyzing infant facial expressions.

To address these issues, traditional approaches often attempt to increase the amount of data collected or simplify machine learning models to accommodate limited data. However, these solutions are either cost-prohibitive or fail to leverage advanced machine learning algorithms to improve recognition performance and analytical accuracy. Instead, a promising alternative is to employ generative adversarial networks (GANs) to generate large quantities of synthetic infant facial expression images. GANs can simulate a wide variety of complex infant facial expressions, rapidly increasing the size and diversity of datasets in a cost-effective manner. This approach has the potential to solve the technical and application bottlenecks caused by insufficient data, allowing machine learning models to train more deeply and perform more accurately.

While GANs offer significant promise, generating high-quality synthetic data alone is insufficient to address all the challenges associated with dataset scarcity and diversity. The quality, representativeness, and balance of the generated data must also be ensured to effectively replace real data in training machine learning models. To

LIBRARY FTSM

achieve this, this study proposes a systematic pipeline that integrates SMOTE (Synthetic Minority Oversampling Technique), ENN (Edited Nearest Neighbor), and Conditional GANs (CGANs). Specifically, SMOTE is employed to balance the dataset by generating synthetic samples for minority emotion categories, addressing the issue of class imbalance at the outset (Chawla et al., 2002). Following this, ENN is used to remove noise and redundant samples, ensuring a cleaner and higher-quality dataset for the generative process (Wilson, 1972). Finally, CGANs are applied to generate high-quality and diverse synthetic images for underrepresented emotion categories, further enhancing the variability and fidelity of the dataset (Mirza & Osindero, 2014). By combining these techniques, the proposed pipeline ensures that the augmented data is not only well-balanced and diverse but also of sufficient quality to effectively support the training of machine learning models.

In addition, this study aims to evaluate the practical application of this pipeline by training a Classification and Regression Tree (CART) on the augmented dataset. The CART model will serve as a validation tool to assess the performance improvements brought about by the enhanced dataset, particularly for minority emotion categories. Through this comprehensive approach, the proposed pipeline aims to address technical challenges such as data scarcity, quality assurance, and diversity, while advancing the accuracy and applicability of infant facial expression recognition systems in real-world scenarios.

1.4 RESEARCH OBJECTIVES

The following are the objectives of this study:

1. Balance and Preprocess the Dataset Using SMOTEENN.

2. Train CGAN to Generate Minority Class Samples and Balanced Datasets.

3. Evaluate Classification Performance Using CART with Improved Adaboost.

4. Establish an Enhanced Framework for Imbalanced Data Handling.

LIBRARY FTSM

1.5 RESEARCH SCOPE

This study investigates the generation and augmentation of infant facial expression datasets for ages 0-2, with a focus on addressing class imbalance through a hybrid data augmentation approach. Six publicly available datasets were employed, including UTK and B3FD from Kaggle, along with four additional datasets sourced from previously published peer-reviewed research papers, as detailed in Chapter 2. These datasets were selected for their relevance to infant facial expression analysis and their inherent data imbalance challenges, which limit the effectiveness of traditional machine learning models.

To mitigate this imbalance, a three-stage augmentation strategy was implemented, integrating Synthetic Minority Oversampling Technique (SMOTE) for oversampling minority classes, Edited Nearest Neighbor (ENN) for noise reduction, and Conditional Generative Adversarial Networks (CGAN) for generating additional synthetic data. This hybrid approach was designed to balance the datasets while minimizing overfitting and preserving the diversity of facial expressions within the infant category.

The scope of this research is limited to static 2D facial images of infants aged 0-2 years. It does not cover real-time video analysis, other age groups, or demographic variables such as ethnicity and health conditions. The synthetic images generated were evaluated using a CART model combined with an F1-based Adaboost algorithm, with performance assessed through standard metrics, including F1 score, G-mean, and AUC. While the results demonstrate promising improvements in class balance and classification accuracy, the findings are constrained to the six datasets used and may require further validation for broader applicability.

1.6 THESIS ORGANIZATION

The objective of this research is to validate whether GANs can compensate for the scarcity of infant facial expression datasets. Below is a description of the remaining portions of this report.

LIBRARY FTSM

Chapter2 reviews foundational research on infant facial expression analysis and the challenges of data scarcity and imbalance in datasets. It explores existing datasets and preprocessing methods, such as SMOTE and ENN, used for handling imbalanced data. Additionally, the chapter discusses the application of GANs, particularly CGANs, for generating synthetic data, highlighting their ability to capture and replicate complex data distributions. The review focuses on the importance of constructing balanced datasets to enhance the performance of downstream classification models.

Chapter 3 describes the research methodology in detail, starting with dataset preparation, including balancing and cleaning the dataset using SMOTE and ENN. It outlines the design and training of the CGAN model, which is used to generate high-quality synthetic samples that align with the original data distribution. The chapter further explains the use of Jensen-Shannon Divergence (JS Divergence) to evaluate the similarity between the distributions of the generated and original data. These steps culminate in the construction of a balanced dataset for subsequent classification tasks.

Chapter 4 presents a comparative analysis of the classification performance of datasets processed with different methods: original, SMOTEENN-enhanced, and SMOTEENN + CGAN-enhanced. The evaluation metrics include F1 score, G-mean, and AUC, which provide a comprehensive understanding of the impact of the proposed pipeline on improving classification accuracy, particularly for minority emotion categories. The CART classifier, combined with an F1-based Adaboost method, is employed to demonstrate the advantages of the enhanced datasets in addressing class imbalance. Results are analyzed to validate the proposed framework's effectiveness.

Chapter 5 concludes the study by summarizing the main contributions and findings. It highlights the effectiveness of the hybrid SMOTEENN + CGAN approach in addressing data imbalance and improving classification performance. The chapter also discusses the limitations of the proposed method, such as computational complexity and dataset constraints, and suggests future research directions. These include exploring alternative generative models, expanding dataset diversity, and extending the framework's applications to other domains with imbalanced data challenges.

LIBRARY FTSM

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

In the rapidly developing field of artificial intelligence and machine learning, the analysis of facial expressions stands out as an important research topic, particularly in infants, whose ability to express emotions is still in a nascent stage. The study of infant facial expressions has significant applications in psychology, medicine, and human-computer interaction. Babies' facial expressions serve as key indicators of their emotions and needs, offering critical insights into their emotional state, cognitive development, and overall health. However, due to the complexity, subtlety, and variability of infant facial expressions, accurately recognizing and analyzing these expressions remains a significant challenge.

This chapter aims to review relevant research on infant facial expression analysis, introduce existing datasets and analysis methods, and explore the application of generative adversarial networks (GANs) in image generation. Particular emphasis is placed on Conditional Generative Adversarial Networks (CGANs) and their role in addressing the scarcity and imbalance of infant facial expression datasets. Additionally, this chapter discusses preprocessing techniques, including SMOTE (Synthetic Minority Oversampling Technique) and ENN (Edited Nearest Neighbor), which are used to enhance dataset quality and balance prior to image generation.

LIBRARY FTSM

2.2 IMPORTANCE OF INFANT FACIAL EXPRESSION ANALYSIS

Newborns are extremely limited in the early stage of environmental cognition. Their learning is mainly derived from observing the surrounding environment and imitating the behavior of others. The core strength lies in the skill of parsing facial expressions. Research has shown that babies have the ability to discriminate between different movements and sounds, which allows them to interact with other people's emotional signals. For example, seeing a smile, they experience happiness; Hearing a cry may cause anxiety. Although seemingly basic, this ability has a decisive impact on the development of early learning and social interaction in infants.

Babies' amazing facial recognition ability accurately matches known faces with their corresponding sounds, new research shows. They are extremely sensitive to emotions and can recognize and understand multiple emotions similar to adults (Vaillant-Molina et al., 2013). Newborns have an acute sense of context beyond adult cognition, which gives them unusual understanding in social interactions. Recent studies strongly confirm that newborns are highly sensitive and can accurately capture and quickly respond to the emotional state of those around them (Kahana-Kalman , Walker-Andrews, 2001).

The instinct for emotion recognition is not only an innate advantage for babies, but also a key factor in their long-term social development. It can improve their ability to sense and interpret the emotions of the environment, thus developing effective communication and interaction skills in future social interactions. The initial emotional cognitive ability will evolve over time and further develop into fine emotion recognition and regulation skills, which is very important for them to build a stable emotional cornerstone in the socialization process. Therefore, parents and caregivers should take advantage of these god-given advantages to promote the overall development of emotional and social skills through positive interaction and loving care.

The latest research focuses on how the mother's emotional expression and behavior patterns profoundly affect the emotional formation of the newborn and the early social interaction with the mother, which has been deeply analyzed (Moore et al.,

LIBRARY FTSM

2009). Within the focus of our study, researchers revealed a key finding: when mothers respond to their newborns' discomfort with appropriate expressions, infants more accurately identify and reflect their pain (Hu et al., 2017). The key stage of children's mental health development and emotional formation is to establish a stable attachment relationship. Intensive research has found that babies have an early talent for recognizing different emotional expressions. The ability of infants to express facial expressions by distinguishing between positive and negative emotions is related to their emotional development and parents' interaction with them (Lamb, 1977). So is a baby's ability to quickly recognize and reflect pain by looking at facial expressions. Existing studies have focused on how facial features assist infants to perceive and recognize pain, and revealed the specific efficacy of these features in mother-infant communication (Zhi et al., 2018).

Studies have shown that the ability of newborns to distinguish emotions does not depend on a single way, but results from the dual effects of facial expressions and acoustic signals (Caron et al., 1988). This interaction suggests that without verbal communication, infants can form an understanding concept of their surroundings based solely on the combination of non-verbal signals such as facial expressions and voices.

These findings reveal important new insights into neonatal emotion recognition and attachment development. They highlight the critical role of mothers in caregiving during infancy, a period that has a profound impact on children's emotional development and social resilience. Future research is expected to dig deeper into these areas to deepen the understanding of the whole process of early human emotion development.

LIBRARY FTSM

2.3 EXISTING BABY FACIAL DATASETS

In the field of neonatal facial expression analysis, there is a dearth of data sets created specifically for infant facial expressions, which greatly hinders further research. The general adult-adolescent related emotion detection dataset does not adequately address the subtle and complex expressions unique to newborns (aldahoud,ugail, 2018). So a dedicated dataset that focuses on newborn expressions is crucial to further understanding how babies communicate emotions and understanding what babies do when they interact. Such a dataset would enable accurate collection and analysis of the facial movements of newborns in different situations, which is crucial for measuring their emotional and developmental status (Webb et al., 2018; LoBue & Thrasher, 2015). Projects like InfAnFace, which focuses on improving the estimation of facial features in newborns, are critical because they provide the precise data necessary to build complex machine learning models and extend our grasp of developmental psychology (Wan et al., 2021).Table1 presents an overview of the existing baby face dataset.

Table 1 Overview of the existing facial expression datasets.

Datasets	**JAFFE**	**PIE**	**MMI**	**BU-3DFE**	**CK+**	**FER**	**CAFE**	**SFEW**	**RAF-DB**
Images	213	40,000	740	2500	593	35,887	1192	1766	29,672
Subjects	10	68	25	100	137	-	100	-	-
Class	7	4	-	7	7	7	7	7	7
Size	256x256	-	720x576	-	640x480	48x48	Square	720x576	-
Age	-	-	19-62	18-70	18-50	-	2-8	-	0-70
Gender	Female	Both	Both	Both	Both	Both	Both	Both	Both
Year	1988	2000	2005	2006	2010	2013	2014	2015	2017

2.3.1 City Infant Faces Datasets

The database contained 195 baby faces, including 40 images of neutral expressions, 54 images of negative expressions, and 60 images of positive expressions. The image has high criterion validity and good retest reliability. The database contains 154 portrait images in both color and black and white formats (Figure 1).

LIBRARY FTSM

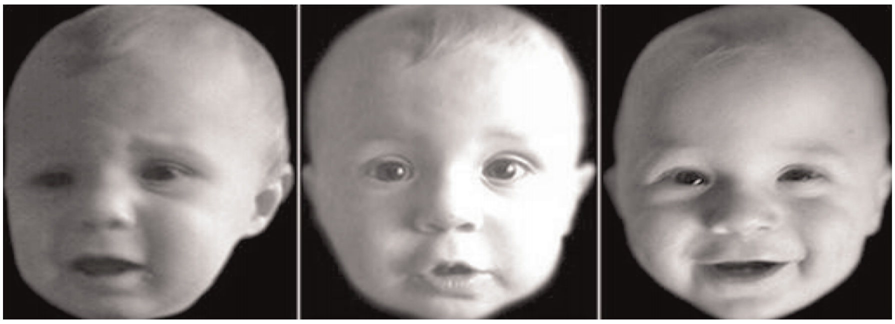

Figure 1 Sample images from the City Infant Faces Database

2.3.2 Babyexp

The dataset contains 5400 photos displaying three different forms of infant's face facial expressions: cry, laugh, and neutral. To accurately characterize more universal face expressions, these three expressions must first be identified. For that process, initially, the photographs of infant actions such as crying and laughing are gathered from the private database termed infant action database. The database contains film of children performing various acts from which photos of various actions have been taken. The photographs of neutral activity and some images of laughter were collected from the internet (Figure 2)

Figure 2 Sample images from Babyexp

LIBRARY FTSM

2.3.3 The Child Emotion Facial Expression Set

The seven induced and posed universal emotions as well as a neutral expression were applied to develop a video and image library of 4- to 6-year-old youngsters.Participants were involved in video and image sessions meant to induce certain emotions, and the resulting photos were then reviewed in two rounds by impartial judges. For each emotion, there were 87 stimuli for neutrality, 363 stimuli for joy, 170 stimuli for disgust, 104 stimuli for surprise, 152 stimuli for fear, 144 stimuli for sadness, 157 stimuli for anger, and 183 stimuli for contempt (Gioia et al. 2021).

2.3.4 Tromso Infant Faces Datasets

In addition to grading the images' intensity, clarity, and valence, over 700 adult images grouped them into 7 emotion categories: glad, sad, disgusted, furious, terrified, astonished, and neutral.

2.3.5 Rebel Dataset

It comprises of 50 movies of infants aged 6–10 months that were acquired from the University of Nevada, Las Vegas' Department of Psychology (UNLV). There are a lot of unlabeled movies of infants in the Rebel collection that need to be categorized (Huguet Cabot and Navigli 2021).

2.3.6 Emoreact

Children between the ages of 4 and 14 make up this multimodal emotion dataset. The collection includes 1102 audio-visual clips with annotations for 17 different emotional states, including 9 difficult emotions like irritation, skepticism, and curiosity, as well as neutral and valence (Nojavanasghari et al. 2016).

LIBRARY FTSM

2.3.7 The Multimodal Dyadic Behavior Dataset

A solitary collection of multimodal (video, audio, and physiological) recordings of infants and toddlers'social and communicative behavior that was captured during a semi-structured play engagement with an adult. According to an IRB process endorsed by the university, the sessions were videotaped in the Georgia Tech Child Study Lab (CSL).

2.3.8 Child Affective Facial Expression Set

This research project brings together a vast collection of 1,192 images of children of all ethnic and cultural backgrounds, ranging in age from 2 to 8 years old, who vividly convey six basic emotions: anger, fear, sadness, joy, surprise, and disgust (Vanessa and Cat,2015).

2.3.9 NIMH-CHEFS

There are 482 photographs in the collection of child faces in 2 different gaze states direct stare and averted gaze including those who are terrified, furious, joyful, sad,and neutral (Egger et al. 2011).

2.4 FACIAL EXPRESSION ANALYSIS METHOD

Facial expression analysis, which is the frontier of the intersection of computer vision and affective computing, has developed into a key technology for scientific and technological innovation and improving the efficiency of human-computer interaction. In the rapid progress of artificial intelligence, multiple cutting-edge technologies have played a role in this field. Feature extraction plays a central role in facial expression recognition, which is a key step to analyze and express expression characteristics. LBP local binary pattern (LBP), as a prominent feature extraction technique, is unique in that it can effectively

LIBRARY FTSM