Dissertation on

Infantza: Computer Vision and Deep Learning Enabled Infant Surveillance System

Submitted in partial fulfillment of the requirements for the award of degree of

Bachelor of Technology

in

Computer Science & Engineering

UE19CS390B – Capstone Project Phase - 2

Submitted by:

ANAGHA SURESH	**PES2UG19CS037**
IMMADISETTY SAI JAYANTH	**PES2UG19CS152**
JEEVAN ANIL	**PES2UG19CS166**
JITTA AMIT SAI	**PES2UG19CS169**

Under the guidance of

Dr. R Bharathi
Professor
PES University

June - Nov 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

Electronic City, Hosur Road, Bengaluru – 560 100, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

Infantza: Computer Vision and Deep Learning Enabled Infant Surveillance System

is a bonafide work carried out by

ANAGHA SURESH	**PES2UG19CS037**
IMMADISETTY SAI JAYANTH	**PES2UG19CS152**
JEEVAN ANIL	**PES2UG19CS166**
JITTA AMIT SAI	**PES2UG19CS169**

In partial fulfillment for the completion of seventh semester Capstone Project Phase - 2 (UE19CS390B) in the Program of Study -Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period June 2022 – Nov. 2022. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 7th semester academic requirements in respect of project work

.

Signature Dr. R Bharathi Professor	Signature Dr. Sandesh B J Chairperson **External Viva**	Signature Dr. B K Keshavan Dean of Faculty

Name of the Examiners **Signature with Date**

1. _____ _____

2. _____ _____

DECLARATION

We hereby declare that the Capstone Project Phase - 2 entitled **“Infantza:Computer Vision and Deep Learning enabled Infant Monitoring System”** has been carried out by us under the guidance of Dr. R Bharathi, Professor and submitted in partial fulfillment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester June – Nov. 2022. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2UG19CS037	**ANAGHA SURESH**
PES2UG19CS152	**IMMADISETTY SAI JAYANTH**
PES2UG19CS166	**JEEVAN ANIL**
PES2UG19CS169	**JITTA AMIT SAI**

ACKNOWLEDGEMENT

We would like to express my gratitude to Dr. R Bharathi, Department of Computer Science and Engineering, PES University, for her continuous guidance, assistance, and encouragement throughout the development of this UE19CS390B -Capstone Project Phase – 2.

We are grateful to the Capstone Project Coordinator, Dr. Sarasvathi V, Professor and Dr. Sudeepa Roy Dey, Associate Professor, for organizing, managing, and helping with the entire process.

We will take this opportunity to thank Dr. Sandesh B J, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support we have received from the department. We would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

We are deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement We have received from my family and friends.

ABSTRACT

With the rise in complexity of the job roles of today's parents and their hectic schedules, the need for infants to be observed frequently when left in the care of a caretaker to avoid any kind of injury and to constantly have an eye upon them all day becomes a tedious task. In today's world, small infants and children being subjected to abuse from caretakers and others has become a serious issue, due to which parents are unable to entrust them with the safety of their child.

We seek to provide a novel approach for infant monitoring that sends alerts to parents with respect to few cases of prospectively dangerous situations that an infant might be exposed to.To implement this a framework has been built.It consists of two phases. The first phase consists of the models for infant activity detection, harmful object detection and stranger detection. The second phase consists of the mobile application through which parents receive alerts if the infant is exposed to danger.

This would avoid the need for constant monitoring of the infant from the parent's side as the monitoring would mainly be required only when an alert has been issued indicating the occurrence of an abnormal situation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

With the rise in complexity of the job roles of today's parents and their hectic schedules, the need for infants to be observed frequently when left in the care of a caretaker to avoid any kind of injury and to constantly have an eye upon them all day becomes a tedious task. In today's world, small infants and children being subjected to abuse from caretakers and others has become a serious issue, due to which parents are unable to entrust them with the safety of their child.

A huge percentage of women end up leaving their professional dreams behind to take care of their infants because of concerns with respect to their safety. Parents shall not be able to spend all their time in their workspaces devoted to checking if their child is fine; hence, an efficient alert system is required in order to detect any kind of dangerous situation.

Through this project, we intend to develop an infant monitoring system that will provide alerts to parents in case the infant is subjected to any harm when left with the caretaker.

The cases for which the notification is to be sent are:

1. Hitting the infant
2. Infant choking
3. Infant crying
4. Any harmful objects in the vicinity of the infant.
5. Strangers' presence near the infant

Environment: The project environment shall be strictly constrained to that of a home environment within which the infant's room would be monitored.

Constraints:

1.This infant monitoring system shall be developed keeping in mind an infant below the age of 1 year.

2.The system shall be able to monitor a single infant at a time.The infant monitoring system can be used generically by anyone with the prerequisite of having to feed the photos of themselves prior to using the Application.

CHAPTER 2

PROBLEM STATEMENT

In the scenario when an infant is left with a caretaker, we seek to develop a model using computer vision and deep learning for infant monitoring in order to alert the parents when the infant is subjected to any harm or is in any unusual situation, thereby avoiding the need for 24/7 constant monitoring of the child. Some of the cases we intend to address include the infant crying, choking, being hit, having any harmful objects in its vicinity , and the presence of any stranger near the infant.

Environment:

The project environment shall be strictly constrained to that of a home environment within which the infant's room would be monitored.

Constraints:

1. This infant monitoring system will be developed keeping in mind an infant below the age of 1 year.
2. The system shall be able to monitor a single infant at a time. The infant monitoring system can be used generically by anyone with the prerequisite of having to feed the photos of themselves prior to using the Application.

CHAPTER 3

LITERATURE REVIEW

3.1 Real Time Crime Detection Using Deep Learning Algorithm

P.Sivakumar; Jayabalaguru.V;Ramsugumar. R; Kalaisriram.S [1]

3.1.1 Objective

Identifying abnormal situations with respect to crimes and sending alerts to police stations.

3.1.2 Technique:

1. Real time camera
2. Face Recognition
3. Comparing with criminal and weapon databases
4. Alert to nearby police station on match

3.1.3 Models:

1. Ybat annotation tool-dataset preparation
2. YOLO -Detection and classification of object detection
3. Darknet framework-Training neural networks for training YOLO.

3.1.4 Advantages:

1. The architecture is extremely fast and detects the criminals at 45 frames per second.
2. Can successfully detect criminals even in crowded areas.

3.1.5 Disadvantages:

1. The boxes for the Ybat annotation tool must already be manually annotated to ensure precision and regularity.
2. A high no of 2000 iterations must be done to ensure high efficiency during training.

3.2 Application of Deep Learning for Infant Vomiting and Crying Detection Chuan-Yu Chang, Fu-Ren Chen [2]

3.2.1 Objective

Detecting if the baby's mouth is covered with vomit or quilt.

3.2.2 Technique

Infant face detection->Vomit detection->Classification

1. Finding the mouth (gaussian filtering used to remove noise)
2. Determining the average pixel value in the mouth.
3. Difference value between previous and next frames is calculated.
4. If $r<0.5$ it is classified as vomit.

3.2.3 Model

1. SSD Mobilenet network architecture (object detection and object prediction)
2. Tensorflow for infant vomit detection

3.2.4 Dataset

Public face dataset -WIDERFACE has been used.

3.2.5 Advantages

This approach can successfully recognise the baby's face in a variety of lighting conditions or complicated backgrounds.

3.3 Facial recognition using Haar cascade and LBP classifiers, Anirudha B Shetty,Bhoomika,Deeksha,Jeevan Rebeiro, Ramyashree [3]

3.3.1 Objective

Facial recognition, comparison of two face recognition techniques Haar Cascade and Local Binary Pattern edified for the classification.

3.3.2 Technique

1. Realtime camera
2. Face recognition
3. Detecting face
4. Detecting eyes

3.3.3 Advantages

1. Haar Cascade has a much higher level of accuracy and has the ability to detect a higher number of faces than the LBP classifier.
2. The LBP classifier is much faster than the Haar Cascade.

3.3.4 Disadvantages

1. LBP classifier has a lower level of accuracy when compared to the LBP classifier.
2. Haar Cascade is much slower than the LBP classifier.

3.4 Face recognition by support vector machines,Guodong Guo, S. Z. Li and Kapluk Chan [4]

3.4.1 Objective

Face recognition with a binary tree recognition strategy.

3.4.2 Technique

1. Image input
2. One against one strategy for facial recognition.

3.4.3 Model

1. SVM algorithm is used
2. Multi class recognition strategy of one against one

3.4.4 Dataset

Cambridge ORL face database(40 distinct persons)

3.4.5 Advantages

SVMs are a better algorithm than the nearest centre approach for facial recognition.

3.4.6 Disadvantages

1. Method is computationally intensive .
2. One against all is ambiguous.

3.5 Suspicious Activity Detection from Videos using YOLOv3,Nipunjita Bordoloi; Anjan Kumar Talukdar; Kandarpa Kumar Sarma [5]

3.5.1 Objective

Detection of any form of suspicious activities from the input video using YOLOv3

3.5.2 Technique

1. Input video
2. Extraction of frames

3. Extract region proposals
4. Classify regions
5. Detection

3.5.3 Model

1. The YoloV3 algorithm has been used
2. Dataset has been self prepared

3.5.4 Advantages

YOLOv3 outperforms Faster R-CNN

3.5.5 Disadvantages

1. The current feature extraction method gives accurate results only in controlled environments.
2. Data in training is extremely less

3.6 Face Detection and Recognition for Criminal Identification System,Sanika Tanmay Ratnaparkhi; Aamani Tandasi; Shipra Saraswat [6]

3.6.1 Objective

Performing facial recognition as well as detection for the identification of criminals.

3.6.2 Technique

1. Face detection:Finding the face in the image provided.
2. Normalising:Identifying facial landmarks.
3. Extract:Extracting features from face to make feature vectors
4. Facial Recognition:Verify and identify

5. the face.

3.6.3 Model

Multi Task Cascaded Convolutional Networks is used for facial detection and alignment in pictures. It is a 3 part CNN which can recognize landmarks on the face like nose, forehead,eyes etc.(Images loaded as numpy array)

1. Facenet has been used for verification and recognition of images.
2. Implementation has been done using python language in jupyter notebook.

3.6.4 Advantages

High accuracy in facial classification

3.6.5 Disadvantages

The dataset has been input in the form of 200 images and the model does not support a dynamic dataset.

3.7 Real Time Baby Facial Expression Recognition Using Deep Learning and IoT Edge Computing,Ramendra Pathak; Yaduvir Singh [7]

3.7.1 Objective

Using deep learning as well as iot edge computing techniques to do real time recognition of baby facial expressions.

3.7.2 Technique

1. Apply a face detection model to detect presence and location of the face.
2. Crop the face segment and compute 128d face embedding.
3. Model classification is done (happy,sad,sleeping).

3.7.3 Model

1. The DL face detector is trained using Caffe deep learning framework that is based on the Single Shot Detector framework with a ResNet base network.
2. Another DNN based model is used to merge the face into 128-D unit hypersphere that quantifies the face. The DNN model is based on a Deep Convolutional Neural Network (DCNN)
3. After training, the fine-tuned deep learning model is optimized as per the hardware and deployed for production on a low-cost Jetson Nano embedded device.
4. The deep learning model is deployed on the edge device. The deployed deep learning model is working as a web service where the image is sent through REST API to the webserver and in return, the model predicts the category of the image

3.7.4 Advantages

1. Average precision, recall, and f1-score of the proposed approach for happy, crying, and sleeping categories outperform the machine learning models.
2. All the processing can be performed on the edge device without using the internet

3.7.5 Disadvantages

1. The edge device has memory and computational constraints, so the size of the deep learning models should meet all constraints for functioning
2. Insights are sent to the cloud only if internet is available

3.8 A Novel Approach for Pose Invariant Face Recognition in Surveillance Videos,Manju.D. ,Radha. V. [8]

3.8.1 Objective

Pose invariant Facial recognition for surveillance videos using viola jones algorithm

3.8.2 Technique

1. Video input
2. Frame extraction
3. Integral image
4. Cascade structure

3.8.3 Model

1. Face detection using viola jones algorithm.
2. Adaboost

3.8.4 Advantages

1. The method improves the LBP code.
2. Highly accurate
3. Robust for facial recognition

3.8.5 Disadvantages

A little slower in execution than existing methods.

3.9 Fuzzy k-NN for choke infant detection,Muhammad Naufal Mansor; Shahryull Hi-Fi Syam Mohd Jamil; Mohd Nazri Rejab; Addzrull Hi-Fi Syam Mohd Jamil [9]

3.9.1 Objective

Using fuzzy-knn for detection of choking in infants

3.9.2 Technique

1. Image capturing
2. Face detection

3. Feature extraction
4. Classifier

3.9.3 Model

1. Fuzzy knn classifier
2. Svd
3. Fast fourier transform

3.9.4 Advantages

The results show that each of the module as well as the fused decision for each type of cry performs satisfactorily in the detection process.

3.9.5 Disadvantages

1. An improvement in the algorithms that detect the position of the eyebrows that are not very prominent in the case of infants, can improve the overall results considerably
2. Additional metrics could be fused to increase the robustness of the sound processing module.

3.10 Object Detection, Classification and Counting for Analysis of Visual Events,MyintSein; Khaing Suu Htet; Ken T. Murata; Somnuk Phon-A [10]

3.10.1 Objective

Detecting classifying and counting objects by analysing visual events.

3.10.2 Technique

1. Fast region proposal
2. Feature extraction

3. Segmentation

3.10.3 Model

1. Fast region proposal using CNN with hyperparameter optimisation
2. Yolo algorithm

3.10.4 Advantages

1. Robust to handle large datasets.
2. High performance

3.10.5 Disadvantages

Due to the illumination and shadow of the object, the detection and classification errors may occur.

3.11 Facial Emotion Detection Using Neural Network,Rhad AliMehenag Khatun Nakib Aman Turzo [11]

3.11.1 Objective

Using neural networks to detect facial emotions.

3.11.2 Technique

1. Dataset has been self prepared.
2. Feature extraction has been done for the input.

3.11.3 Model

1. CNN along with keras,tensorflow and pretraining concepts .
2. Viola jones algorithm to detect eye and lips region
3. ML ,DL ,NN algorithms can be used for emotion recognition

3.11.4 Advantages

1. Accuracy is high and has been determined using decision trees.
2. 7 emotions have been detected and classified

3.11.5 Disadvantages

A large quantity of test data and keywords are needed if it wants to get greater accuracy

3.12 Face verification based on convolutional neural network and deep learning,A.Lebedev;V.Khryashchev;A.Priorov;O.Stepanova [12]

3.12.1 Objective

Using deep learning and convolutional neural networks for face verification.

3.12.2 Technique

The algorithm produces face feature vectors, distance between these vectors allows to determine whether images are from the same class.

3.12.3 Model

Deep Convolutional Neural Network has been used as the model.

3.12.4 Advantages

1. Modern face recognition algorithm
2. Testing was carried out under unsupervised learning

3.12.5 Disadvantages

Preprocessing has not been carried out but this can be done to enhance the AUC value.

3.13 Video Analytics for Face Detection and Tracking,Vaibhavi Kulkarni; Kiran Talele [13]

3.13.1 Objective

Detection of faces and objects from videos using Viola Jones.

3.13.2 Technique

1. Video Input
2. Preprocessing
3. Image Segmentation
4. Detect and recognize objects
5. Track detected objects
6. Data Fusion

3.13.3 Model

1. Detection and cropping is done using the Viola-Jones Algorithm
2. Tracking continuous feature points is done using the KLT algorithm

3.13.4 Advantages

1. The algorithm is robust even when noise and clutter is present.
2. Selecting the facial frame from the real-time surveillance videos and analyzing at the edge reduces human effort and also eliminates human errors

3.13.5 Disadvantages

1. Viola Jones algorithm has a very slow training time.
2. It is mainly effective only when the face is in frontal view

3.14 Moving object detection and tracking Using Convolutional Neural Networks,Shraddha Mane; Supriya Mangale [14]

3.14.1 Objective

Using convolutional neural networks for moving object detection and tracking

3.14.2 Technique

1. Input the video
2. Frame Extraction
3. Object detection
4. Get object location
5. Object tracking

3.14.3 Model

1. Tensorflow based object detection API has been used for the purpose of object detection.
2. The object tracking has been done using the CNN architecture.

3.14.4 Advantages

1. The object detection module robustly detects the object.
2. Object tracking requires a huge number of features,using CNN for imageclassification improves the performance significantly as it is trained in millions of classes.

3.14.5 Disadvantages

1. An accuracy of 90% has been determined.which can be improved.
2. CNN is extremely expensive to train.

3.15 Facial Emotion Detection Using Deep Learning,Akriti Jaiswal; A. Krishnama Raju; Suman Deb [15]

3.15.1 Objective

Using deep learning for facial emotion detection .

3.15.2 Technique

1. Face detection
2. Feature extraction
3. Emotion classification

3.15.3 Model

Convolutional neural network based deep learning model.

3.15.4 Advantages

1. This paper presents the design of an artificial intelligence (AI) system capable of emotion detection through facial expressions.
2. Results of the experiment show that the model proposed is better in terms of the results of emotion detection

3.15.5 Disadvantages

Higher accuracy can be obtained in terms of FERC dataset

CHAPTER 4

PROJECT REQUIREMENT SPECIFICATION

4.1 Product Perspective

With the rise in complexity in the job roles of today's parents and their hectic schedules, the need for infants to be observed frequently to avoid any kind of injury and to constantly have an eye upon them all day becomes a tedious task.

In today's world, infants being subjected to abuse from caretakers and others have become a serious issue, due to which parents are unable to entrust them with the safety of their child. A huge percentage of women end up leaving their professional dreams behind to take care of their infants purely out of concerns with respect to their safety .

Parents shall not be able to spend all their time in their workspaces devoted to checking if their child is fine; hence, an efficient alert system is required in order to detect any kind of prospectively dangerous situations.

4.1.1 Product Features

Through this project we intend to develop an infant monitoring system that will provide alerts to parents in case of the infant being subjected to any harm when left with the caretaker.

The functionalities that shall be provided include:

1. Conversion of the input real time video into frames.
2. Extraction of features.
3. Classification of normal and abnormal behaviour by the model.
4. Notification to parents on harm detection.
5. Facial recognition to recognize and distinguish strangers from members of the house and notification on identification of strangers.

4.1.2 User Classes and Characteristics

The end users are going to be:

1. Parents
2. Relatives
3. Caretaker

No prior technical knowledge/expertise is required to use this product.

4.1.3 Operating Environment

The System can be operated in all the major platforms with a good GPU like Nvidia.

4.1.4 General Constraints, Assumptions and Dependencies

Assumptions

1. This infant monitoring system shall be developed keeping in mind a child between the age of 3 months to 1 year.
2. The project environment shall be strictly constrained to that of a home environment within which the room in which the infant would be in is monitored.
3. The system shall be able to monitor a single infant at a time.

Dependencies

1. Android studio
2. Android Emulator
3. Ngrok
4. Flutter
5. Flask
6. Jupyter notebook
7. Visual studio code

4.1.5 Risks

1. Mispredictions i.e alert messages being sent when the environment is completely normal.
2. Alert messages not being sent when an abnormal activity is happening in the environment.

4.2 Functional Requirements

1. Conversion of input video into frames.
2. Extraction of features.
3. Classification of normal and abnormal behaviour by the model.
4. Notification to parents on harm detection.
5. Facial recognition to recognize and distinguish strangers from members of the house and notification on identification of strangers.

4.3 External Interface Requirements

4.3.1 User Interfaces

The UI is a mobile application.The alerts for the use cases will be displayed to the user in the application.

4.3.2 Hardware Requirements

High Resolution CCTV Surveillance Camera for the purpose of real time capturing.

GPU(provided by Google Collaboratory)

4.3.3 Software Requirements

1. Python libraries.
2. Jupyter Notebook
3. Python 3.7

4. Tensorflow
5. Computer Vision
6. Flask
7. Ngrok
8. Flutter

4.4 Non Functional Requirements

4.4.1 Performance Requirement

1. **Higher Accuracy:** The model will have the ability to correctly assess and comprehend the genuine positives and negatives of various events and will classify with high accuracy such that alerts can be sent during abnormal or emergency situations.
2. **Fast In Real time**: The machine learning model must be able to detect abnormal activity and send notifications within minimal time in real time.
3. **Efficient And Reliable System**: Our baby monitoring system must make the end user feel safe, secure, and dependable. The user must believe that this system is more efficient than the already available product on the market.
4. **Highly Available**: We will make our application available to our users 24 hours a day, seven days a week, extremely accessible at all times.

4.4.2 Safety Requirements

1. **Application security**:Security in terms of allowing access by setting usage limitations.
2. **Authentication**: Giving the user access and ensuring that the users are legitimate. The first step in a successful identity and access management strategy is authentication.
3. **Authorization**: What resources a user has access to is determined by authorization such as giving a certain user extra privileges.

4.4.3 Security Requirements

1. **Authentication**: Giving the user access and ensuring that the users are legitimate. The first step in a successful identity and access management strategy is authentication.

2. **Authorization**: What resources a user has access to is determined by authorization such as giving a certain user extra privileges.

CHAPTER 5

SYSTEM DESIGN

5.1 Design Goals

Newly proposed system:

In the scenario when an infant below the age of 1 year is left with a caretaker, we seek to develop a model using computer vision and deep learning for infant monitoring with an alert system in order to alert the parents when the infant is subjected to any harm or is in any unusual situation, thereby avoiding the need for 24/7 constant monitoring of the infant.The monitoring would mainly be required only when an alert has been issued indicating the occurrence of an abnormal situation. Some of the cases we intend to address include the infant crying, being hit, choking or any harmful objects in its vicinity.We will also include a facial recognition system that will help us to detect the presence of any stranger in the infant's vicinity and an alert will be sent to the parents in all such situations.

Difference between new and existing system:

The tasks that have been handled in the existing system for infant monitoring involve only the detection of harm occurring to an infant in a particular situation where the categories of harmful activities that are being handled are limited to a few.

The newly proposed system that we seek to implement will cover a broader range of dangerous situations that an infant might be exposed to which include the infant crying, being hit, choking or any harmful objects in its vicinity.We will also include a facial recognition system that will help us to detect the presence of any stranger in the infant's vicinity and an alert will be sent to the parents in such a situation.

5.1.1 Features of the System

Availability:

We intend to make our application available to our users 24 hours a day, seven days a week, extremely accessible at all times.

Security

1. Security in terms of allowing access by setting usage limitations.
2. Facial Recognition to enhance application security .
3. Usage Limitations: The usage limitation can be imposed on the application by setting a limit of allowing a maximum of 4 people to access the CCTV footage using a single login.

Privacy:

- **Authentication:** Giving the user access and ensuring that the users are legitimate.

Speed:

- **Fast In Real time:** The model must be able to detect abnormal activity and send notifications within minimal time in real time.

5.2 Architecture Diagram

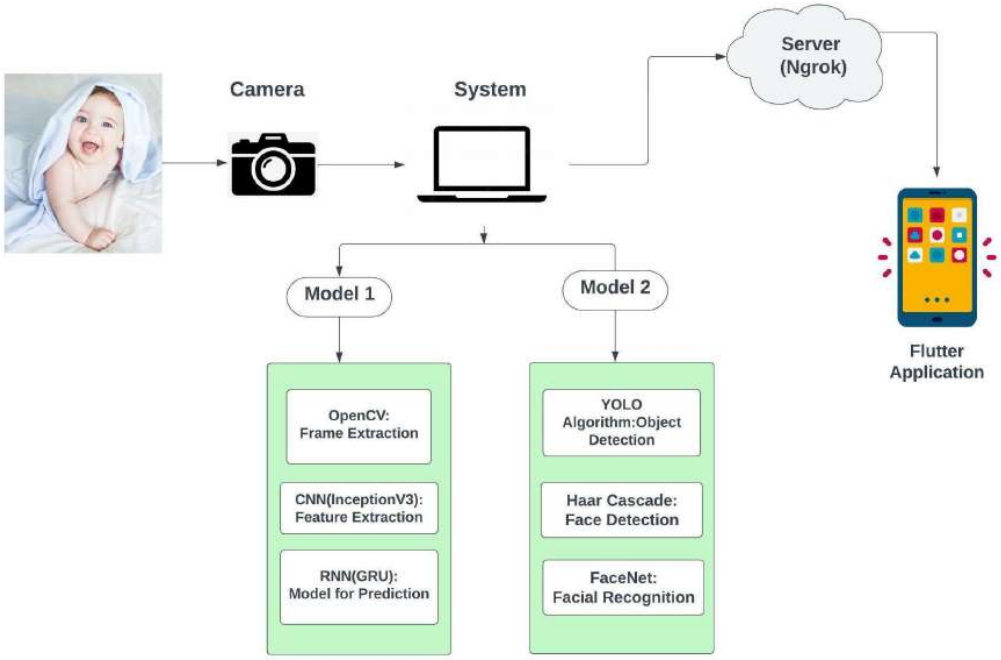


Fig No: 5.1 Architecture Diagram

5.3 Constraints, Assumptions and Dependencies

Assumptions:

1. This infant monitoring system shall be developed keeping in mind an infant i.e below the age of 1 year.
2. The project environment shall be strictly constrained to that of a home environment within which the room in which the infant would be in is monitored.
3. The system shall be able to monitor a single infant at a time.

Hardware Requirements:

1. High Resolution CCTV Surveillance Camera for the purpose of real time capturing.
2. GPU(provided by Google Collaboratory)

Software Requirements:

1. Python libraries
2. Jupyter Notebook
3. Python 3.7
4. Tensorflow
5. Computer Vision
6. Flask
7. Ngrok
8. Flutter

5.4 High Level System Design

5.4.1 Use Case Diagram

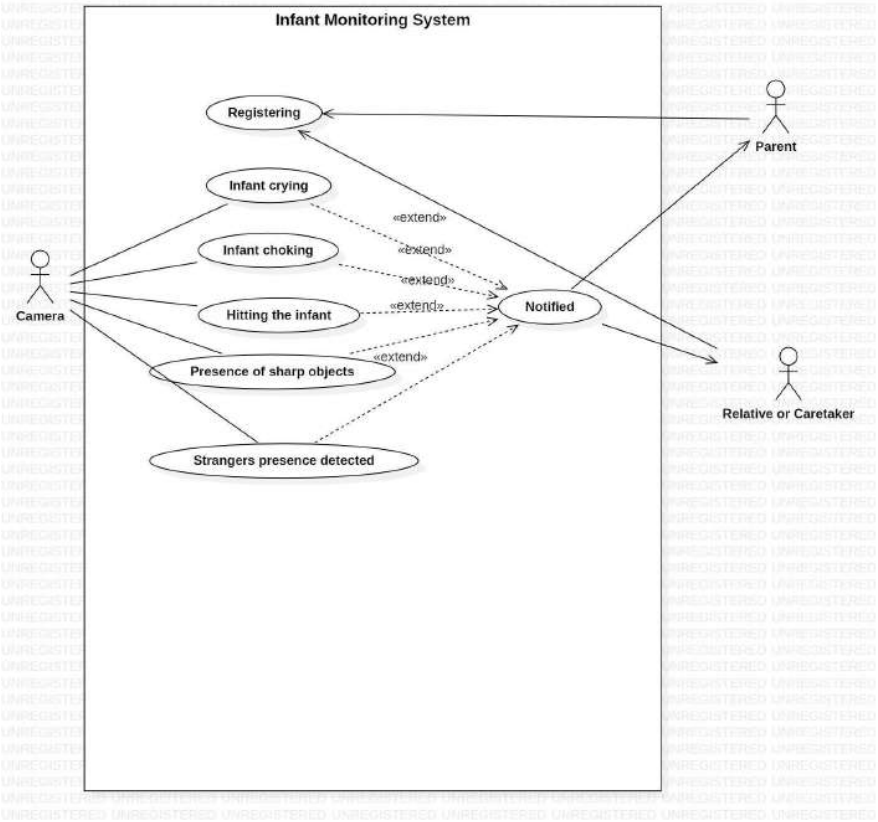


Fig No: 5.2 Use Case Diagram

5.4.2 User Interface Diagram

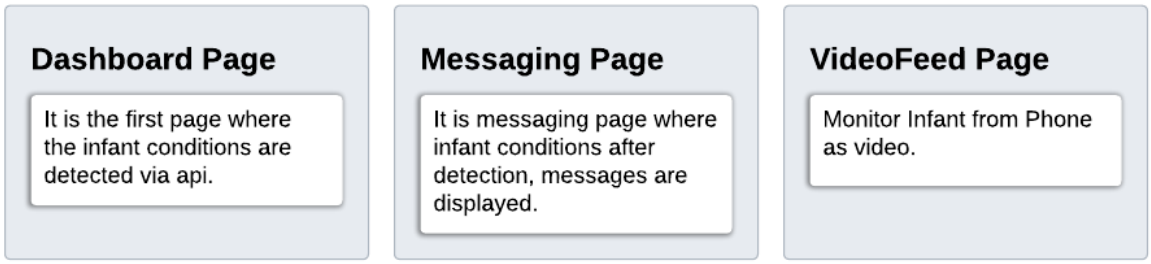


Fig No: 5.3 User Interface Diagram

5.4.3 External Interface Diagram

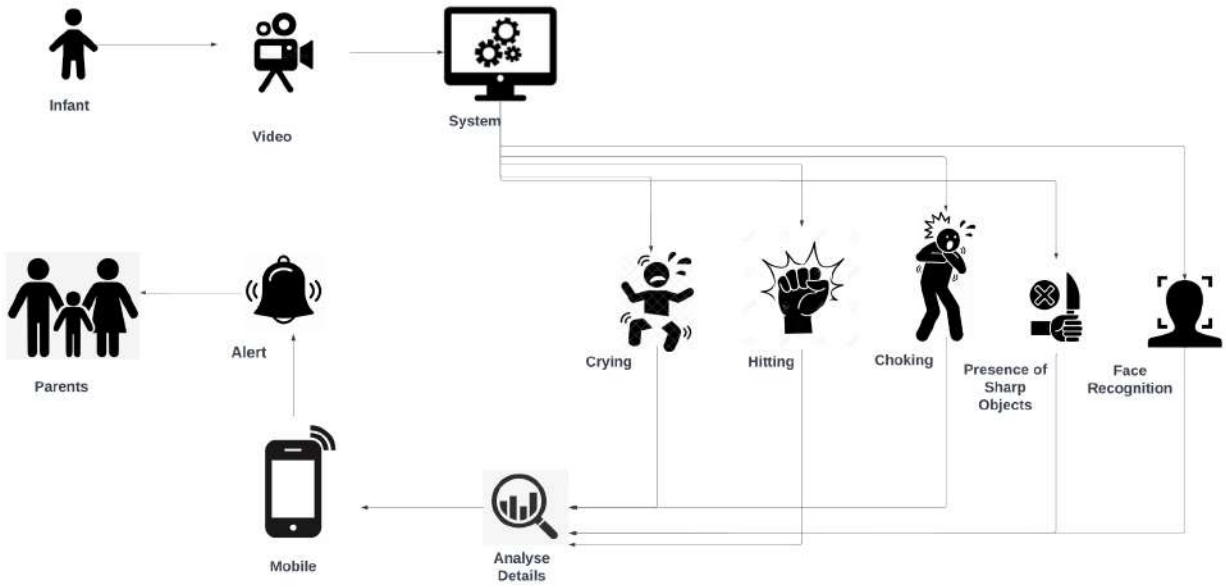


Fig No: 5.4 External Interface Diagram

5.5 Low Level Design

5.5.1 Methods of Model 1 and Model 2

Table no:5.1

Sl No	Method Name	Method Description
1	load_video	Conversion of input videos to frames
2	crop_center_square	The frames that are obtained using load_video are resized using this method.
3	build_feature_extractor	Consists of the InceptionV3 pre-trained network that performs feature extraction.
4	prepare_all_videos	This method converts every video in the dataset to frames as well as performs padding and feature extraction.
5	get_sequence_model	Consists of the GRU model that is used for prediction.
6	run_experiment	Trains the model based on provided hyperparameters.
7	prepare_single_video	Used to convert videos into frames while testing.
8	sequence_prediction	This method is used to predict the class.
9	face_match	Used to train the FaceNet model

5.5.2 Use Case Diagram

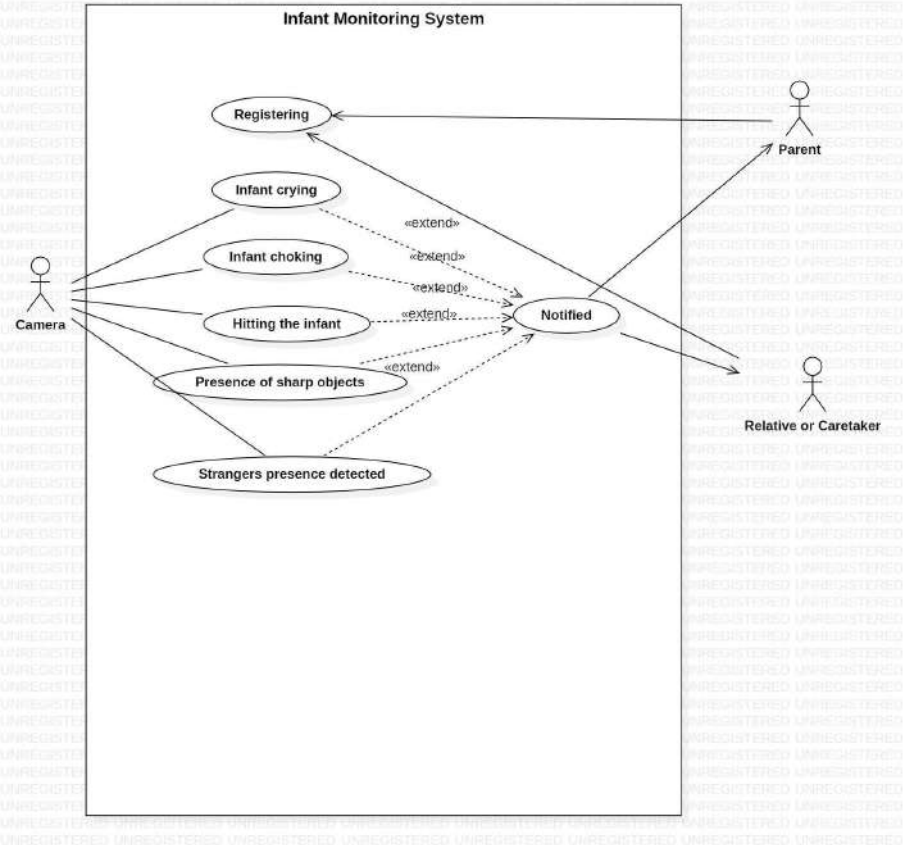


Fig No: 5.5 Use Case Diagram

Actors:

1. Camera
2. Parent
3. Relative/Caretaker

5.5.3 Master Class Diagram

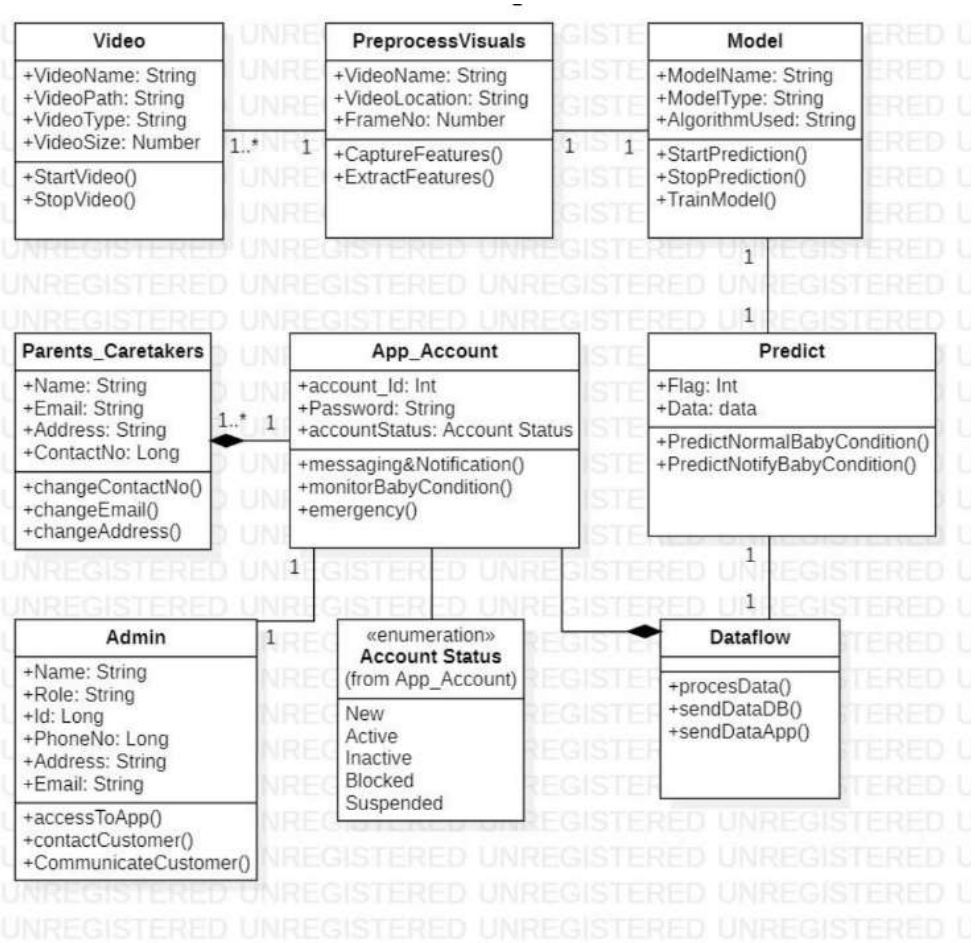


Fig No: 5.6 Master Class Diagram

CHAPTER 6

PROPOSED METHODOLOGY

Our project is composed of two models .The first one is for the purpose of infant activity detection and the second is for harmful object and stranger detection.The two models work in parallel.

6.1 Infant Activity Detection

The first model is a sequential model that is used for infant activity detection.The three cases for which alerts are generated are - infant crying ,infant choking and infant being hit. The sequential model is composed of the InceptionV3 pretrained network for the purpose of feature extraction followed by GRU for prediction. On completion of the literature review, it has been understood that GRU is much faster due to the presence of less number of gates when compared to LSTM.The results computed by the sequential model are hosted using Ngrok as an API. Flask has been used as the web framework for the API. The API is used by the application to fetch the results which are displayed as alerts on the Flutter application if any of the three cases are encountered.

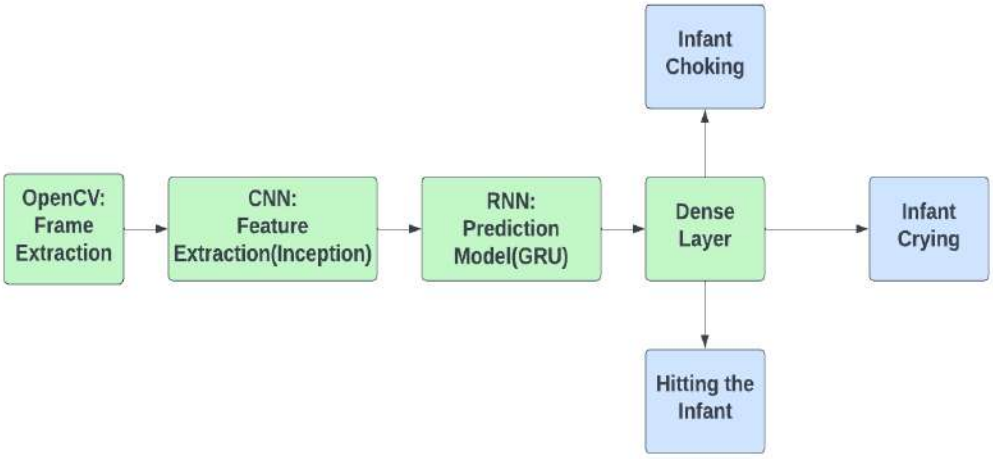


Fig No: 6.1 Model 1

6.2 Harmful Object and Stranger Detection

The second model covers the cases of harmful objects and stranger detection.The YOLO based framework has been used for object detection. It is used to detect harmful objects like knives, scissors , and forks near the infant.It is also used to detect the presence of any stranger in the infant's room.On detection of a face by the classifier, the FaceNet model is launched to identify faces by comparing them to the database of photographs initially fed into the application by the user.The results computed by the sequential model are hosted using Ngrok as an API. Flask has been used as the web framework for the API. The API is used by the application to fetch the results which are displayed as alerts on the Flutter application if any of three cases are encountered.

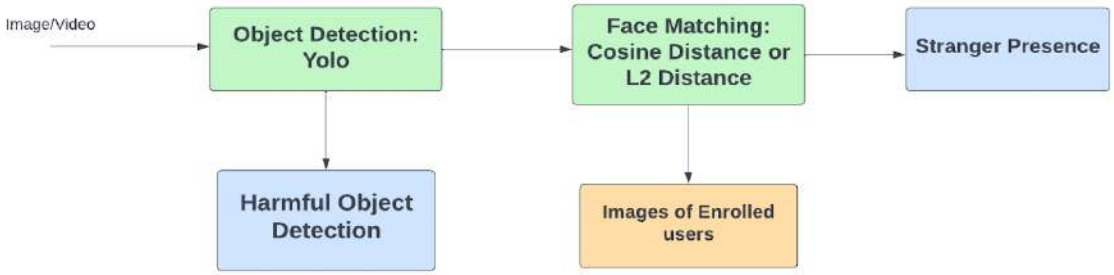


Fig no:6.2 Model 2

6.3 Dataset

The dataset for the sequential model has been synthesized and several videos have been collected for infant crying, infant choking and infant hitting respectively. Live streamed videos captured via webcam are provided as input to the model for harmful object and stranger detection.

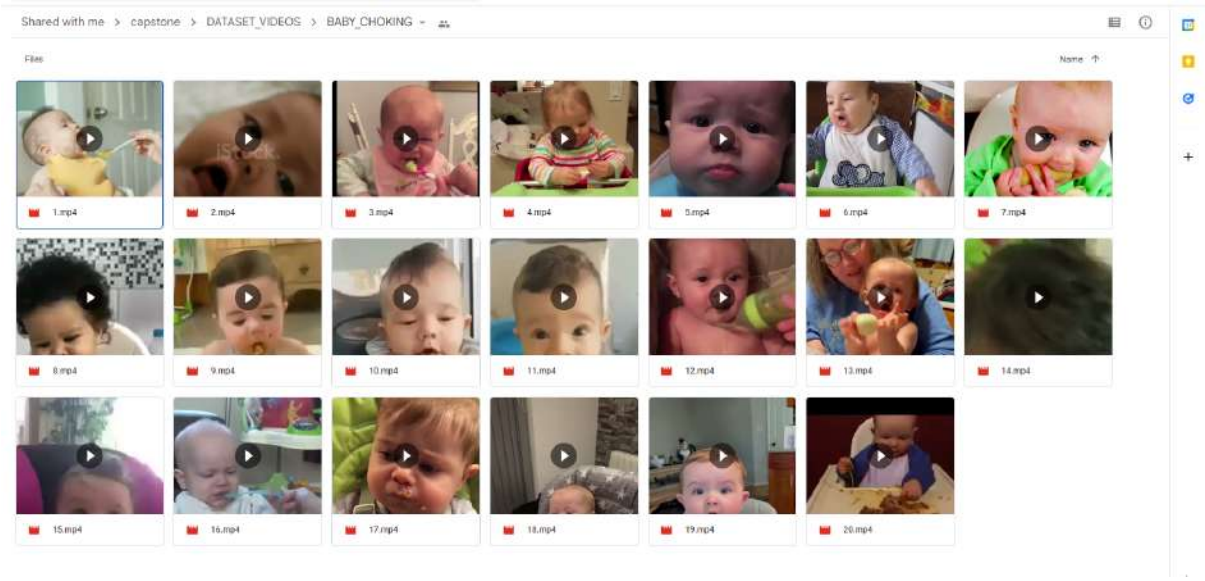


Fig No: 6.3 The dataset for infant choking

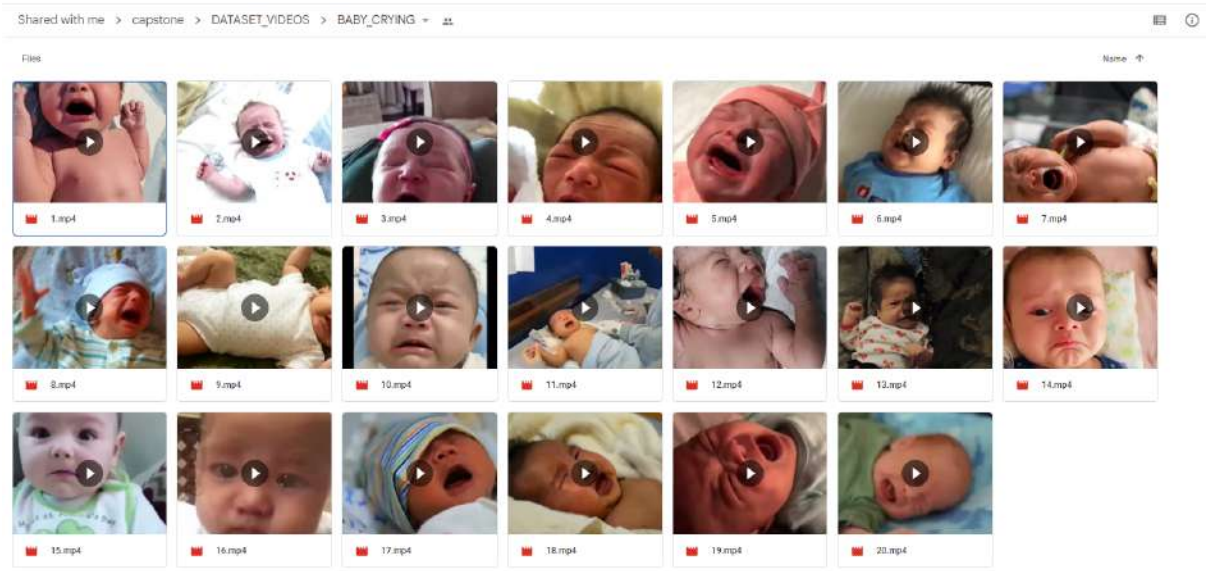


Fig no:6.4 The Dataset for infant crying

Fig no:6.5 The dataset for infant hitting

Dataset Pre-Processing:

Resizing of video frames to a size of 224 x 244

Manual data augmentation to the dataset for sequential model.

Padding of the frames of the video.

The collected videos have been trimmed in order to ensure that only the necessary parts with respect to the infant activity are retained.

CHAPTER 7

IMPLEMENTATION AND PSEUDOCODE

7.1 Architecture diagram

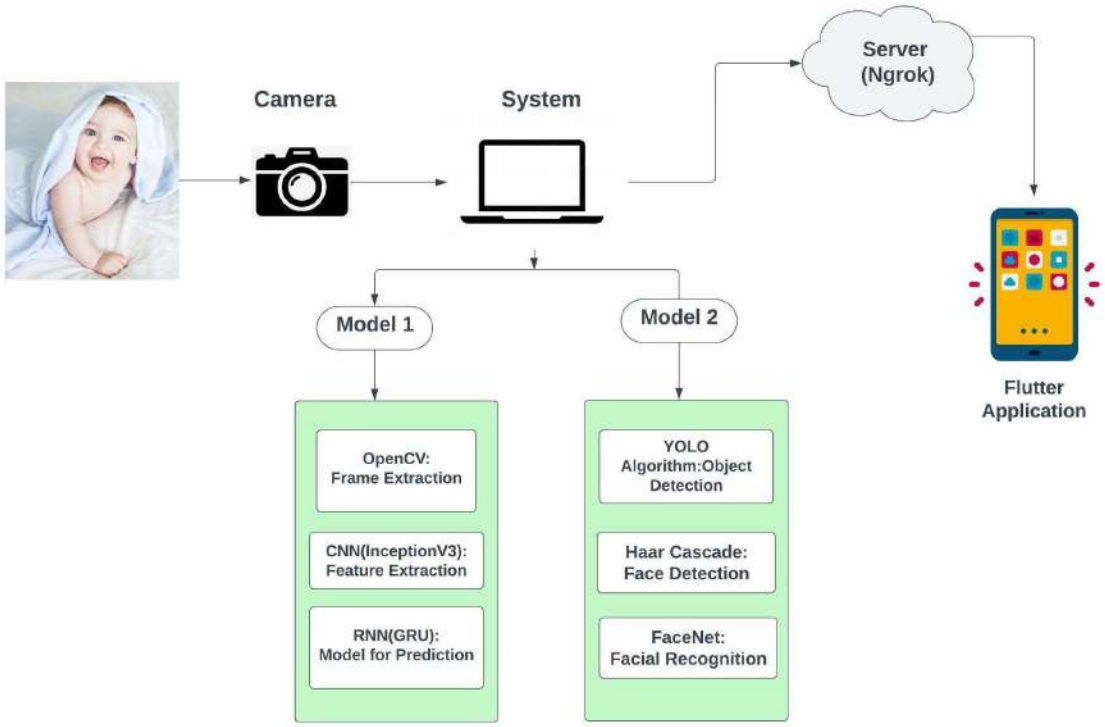


Fig No: 7.1 System Architecture Diagram

7.2 Modules of the project

7.2.1 Module 1

Module Name: Sequential Module

Technology used:CNN and GRU

This module uses the CNN-GRU model in order to capture the sequence of actions to send alerts or infant choking,infant crying or infant being hit.

7.2.2 Module 2

Module Name: Object Detection Module

Technology used:YOLO V3

1. In this module we are doing sharp object and person detection.
2. The sharp objects include knife,scissors and fork.

7.2.3 Module 3

Module Name: Face Recognition Module

Technology used:FaceNet

The face recognition module is used after person detection in order to identify any unknown faces for stranger detection.

7.2.4 Module 4

Module Name: Application Module

Technology Used:Flutter

The app will let the parents know of the infant's current circumstances/situation as we fetch the results from the api and display them here.

7.3 Pseudocode

```
Load Models
A:Sequential model  =>Infant hitting,crying and choking
B:Yolo Model          =>Object Detection-> Person, Sharp Objects like => Knife, Spoon, Fork
C:Face net            =>Face detection

A:Sequential model
 For each frame from the frames(cv2.capture())
        InceptionV3 for feature extraction(CNN)
        Frames are given as input to sequential model(GRU) which  predicts

B:Yolo Object Detection model
For each frame from the frames(cv2.capture())
        do yolo object detection
                find the class id's of Sharp Object


C:Do FaceMatching for the Database of Images and learn the their cosine distances for each
Person
Test it real time on cv2(computer vision)
Check each frame person's cosine distance
and check person's cosine distance and detect if the person is known/unknown

cv2.capture(Computer Vision)
Video Input:
  Images/frames
  Use A: detection of Infant hitting   => res:results
  Use A: detection of Infant crying    => res:results
  Use A: detection of Infant choking => res:results
  Use B: detect knife  => res:results
  Use B: detect spoon=> res:results
  Use B: detect fork    => res:results
  Use B:detect Person=>
  Use C: do face matching   => res:results
res:host  on server using python
      ngrok web hosting service ,flask as web framework for API
fetch Results in Flutter Application
Notify parents/caretaker  in the Application
```

CHAPTER 8

RESULTS AND DISCUSSION

The project is composed of two models.The proposed sequential model has been trained on several videos. The model hyperparameters consist of sparse categorical cross entropy as the loss function and Adam as the optimizer.The model has been trained for 100 epochs with a batch size of 64.

```
In [65]: print(classification_report(labels1,y_predict))
```

	precision	recall	f1-score	support
0	0.97	0.87	0.92	101
1	0.90	0.96	0.93	84
2	0.95	1.00	0.97	73
accuracy			0.94	258
macro avg	0.94	0.95	0.94	258
weighted avg	0.94	0.94	0.94	258

Fig No:8.1 Classification report for training data of sequential model

```
from sklearn import metrics
confusion_matrix = metrics.confusion_matrix(labels1, y_predict)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix, display_labels = ["choking","crying
cm_display.plot()
plt.show()
```

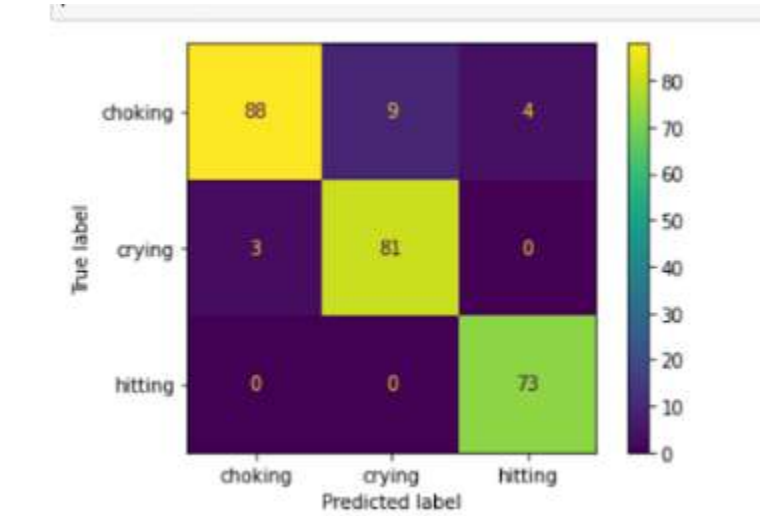


Fig No:8.2 Confusion matrix for training data of sequential model

```
: print(classification_report(labels1,y_predict))
```

	precision	recall	f1-score	support
0	0.88	0.64	0.74	11
1	0.88	0.70	0.78	10
2	0.67	1.00	0.80	10
accuracy			0.77	31
macro avg	0.81	0.78	0.77	31
weighted avg	0.81	0.77	0.77	31

Fig. 8.3 Classification report for testing data of sequential model

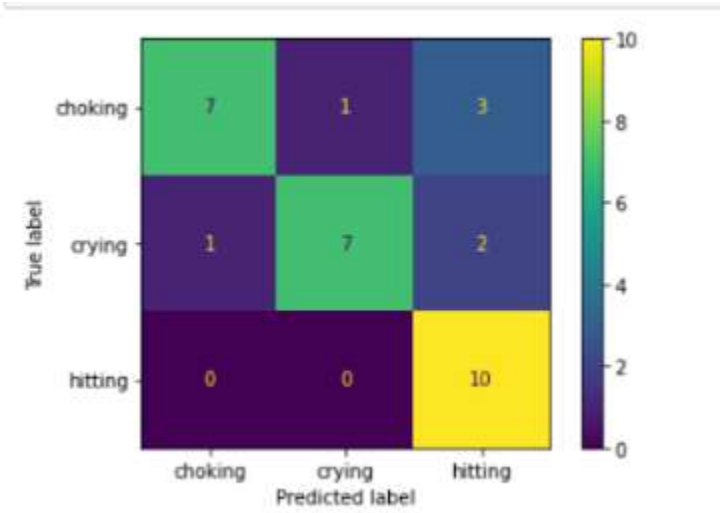


Fig. 8.4 Confusion matrix for testing data of sequential model

Fig No: 8.5 Live streaming on application

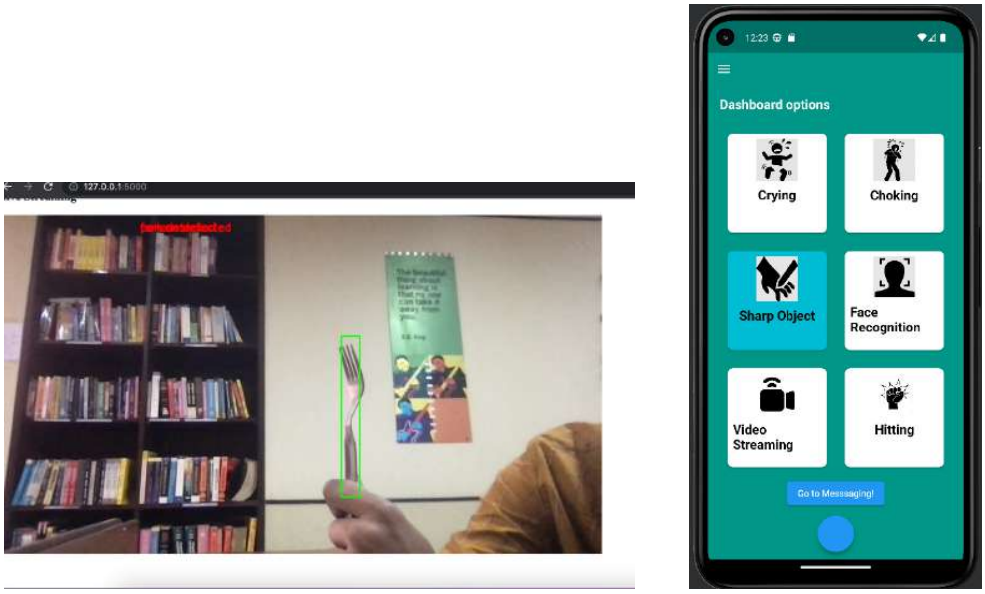


Fig No: 8.6 The alert is displayed on the Application on harmful object detection.

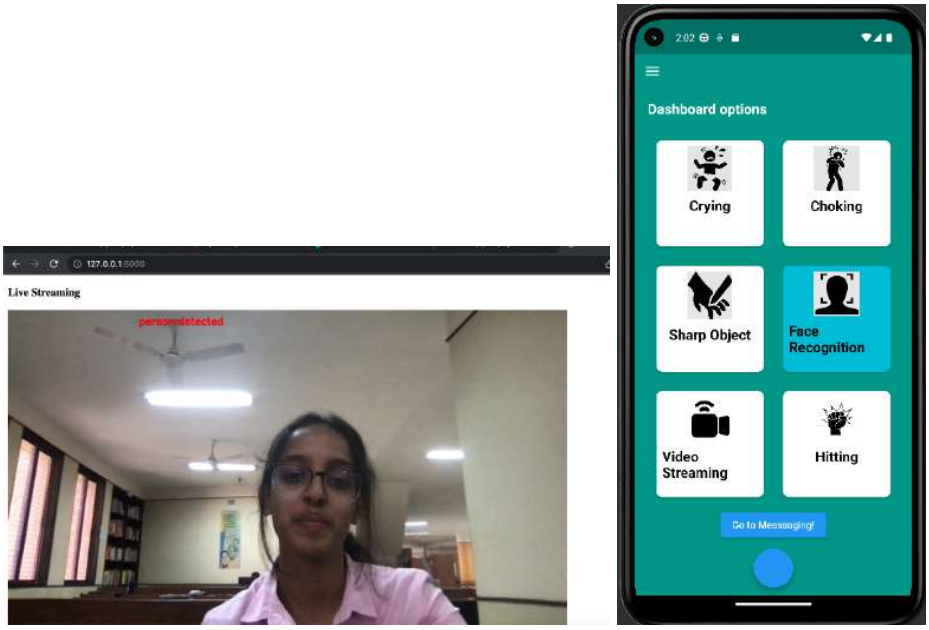


Fig No: 8.7 Alert on Stranger face detection

CHAPTER 9

CONCLUSION AND FUTURE WORK

In this project, a novel approach for infant monitoring has been presented. The framework of the infant monitoring system is novel and revolutionary.The five cases that the project handles are infant crying, infant choking, infant being hit, harmful object detection and stranger detection. Live streamed video is provided as input and alerts are displayed in the application in real time. The alerts generated would be extremely useful for parents and caretakers as it would help them to constantly monitor the infant even while being away from them. Hence this project presents an approach for an infant monitoring system that is the need of the hour.We have attained a training accuracy of 94% and a testing accuracy of 77% for the sequential model.

As a part of the future work, we seek to enhance the accuracy of the models as well as reduce the response time with respect to receiving the alerts on the application.We will also publish a conference paper post completion of the project.

REFERENCES

[1] P. Sivakumar, J. V, R. R and K. S, "Real Time Crime Detection Using Deep Learning Algorithm," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021, pp. 1-5, doi: 10.1109/ICSCAN53069.2021.9526393.

[2] C. -Y. Chang and F. R. Chen, "Application of Deep Learning for Infant Vomitin and Crying Detection," 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2018, pp. 633-635, doi:10.1109/WAINA.2018.00158.

[3] Anirudha B Shetty, Bhoomika, Deeksha, Jeevan Rebeiro, Ramyashree,Facial recognition using Haar cascade and LBP classifiers,Global Transitions Proceedings, Volume 2, Issue 2,2021,Pages 330-335, ISSN 2666-285X

[4] Guodong Guo, S. Z. Li and Kapluk Chan, "Face recognition by support vector machines," Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), 2000, pp. 196-201, doi: 10.1109/AFGR.2000.840634.

[5] N. Bordoloi, A. K. Talukdar and K. K. Sarma, "Suspicious Activity Detection from Videos using YOLOv3," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342230.

[6] S. T. Ratnaparkhi, A. Tandasi and S. Saraswat, "Face Detection and Recognition for Criminal Identification System," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 773-777, doi: 0.1109/Confluence 51648.2021.9377205.

[7] R. Pathak and Y. Singh, "Real Time Baby Facial Expression Recognition Using Deep Learning and IoT Edge Computing," 2020 5th International Conference on Computing, Communication and Security(ICCCS), 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277428.

[8] Manju, D. and Radha, V., 2020. A novel approach for pose invariant face recognition in surveillance videos. *Procedia Computer Science*, *167*, pp.890-899.

[9] M. N. Mansor, S. H. -F. S. M. Jamil, M. N. Rejab and A. H. -F. S. M. Jamil, "Fuzzy k-NN for choke infant detection," 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), 2012, pp. 349-351, doi: 10.1109/MSNA.2012.6324590.

[10] M. Sein, K. S. Htet, K. T. Murata and S. Phon-Amnuaisuk, Object Detection, Classification and Counting for Analysis of Visual Events," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), 2020, pp. 274-275, doi: 10.1109/GCCE50665.2020.9292058.

[11] Ali, Md. Forhad & Khatun, Mehenag & Turzo, Nakib. (2020). Facial Emotio Detection Using Neural Network. International Journal of Scientific and Engineering Research. 11. 1318-1325.

[12] A. Lebedev, V. Khryashchev, A. Priorov and O. Stepanova, "Face verification based on convolutional neural network and deep learning," 2017 IEEE East-West Design & Test Symposium (EWDTS), 2017, pp. 1-5, doi: 10.1109/EWDTS.2017.8110157.

[13] V. Kulkarni and K. Talele, "Video Analytics for Face Detection and Tracking," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),2020, pp. 962-965.

[14] S. Mane and S. Mangale, "Moving Object Detection and Tracking Using Convolutional Neural Networks," 2018 Second International Conference on Intelligent Computing and Control Systems(ICICCS),2018, pp. 1809-1813, doi: 10.1109/ICCONS.2018.8662921.

[15] Jaiswal, Akriti, A. Krishnama Raju and Suman Deb. “Facial Emotion Detection Using Deep Learning.” *2020 International Conference for Emerging Technology(INCET)* (2020): 1-5.

APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Definitions and Acronyms

- CNN- Convolutional Neural Network.
- NN - Neural Network.
- DL - Deep Learning.
- Algorithm - A process or set of rules to be followed in calculations or other problem solving operations, especially by a computer.
- Novelty - Novelty is the quality of being new, or following from that, of being striking, original or unusual.
- Reliability - The probability that a product, system, or service will perform its intended function adequately for a specified period of time or will operate in a defined environment without failure.

Annexure I

Infantza: Computer Vision and Deep Learning Enabled Infant Surveillance System

Anagha Suresh
Department of Computer Science
PES University
Bengaluru,India
anaghas2001@gmail.com

Jitta Amit Sai
Department of Computer Science
PES University
Bengaluru,India
jsamit27@gmail.com

Jeevan Anil
Department of Computer Science
PES University
Bengaluru,India
jeevananil03@gmail.com

Immadisetty Sai Jayanth
Department of Computer Science
PES University
Bengaluru,India
saijayanth092001@gmail.com

Dr. R Bharathi
Department of Computer Science
PES University
Bengaluru,India
rbharathi@pes.edu

***Abstract*—Through this capstone project we seek to provide alerts to parents with respect to few cases of prospectively dangerous situations that an infant might be exposed to. This would avoid the need for constant monitoring of the infant from the parent's side as the monitoring would mainly be required only when an alert has been issued indicating the occurrence of an abnormal situation.**

Keywords—CNN, Flutter, Computer Vision, Deep Learning, Ngrok, Haar Cascade, FaceNet, YOLOV3.

I. Introduction

With the rise in complexities in the job roles of today's parents and their hectic schedules, the need for infants to be observed frequently when left in the care of a caretaker to avoid any kind of injuries and to constantly have an eye upon them all day becomes a tedious task. In today's world infants and children being subjected to abuse from caretakers and others has become a serious issue due to which parents are unable to entrust them with the safety of their child.

A huge percentage of women end up leaving their professional dreams behind to take care of their infants because of concerns with respect to their safety Parents shall not be able to spend all their time in their workspaces devoted to check if their child is fine, hence an efficient alert system is required in order to detect any kind of dangerous situations.

II. PROBLEM STATEMENT

In the scenario when an infant is left with a caretaker, we seek to develop a model using computer vision and deep learning for infant monitoring in order to alert the parents when the infant is subjected to any harm or is in any unusual situation, thereby avoiding the need for 24/7 constant monitoring of the child. Some of the cases we intend to address include the infant crying, choking, being hit, any harmful objects in its vicinity and the presence of any stranger near the infant.

Environment:

The project environment is strictly constrained to that of a home environment within which the infant's room would be monitored.

Constraints:

- This infant monitoring system will be developed keeping in mind an infant below the age of 1 year.
- The system shall be able to monitor a single infant at a time.

The infant monitoring system can be used generically by anyone with the prerequisite of having to feed the photos of themselves prior to using the Application.

III. LITERATURE SURVEY

In [1] a deep learning algorithm has been proposed to identify abnormal situations with respect to crimes and to send alerts to police stations.A real time CCTV camera has been used in order to capture the image of the criminal along with the weapons.The Ybat annotation tool has been used in order to prepare the dataset and the boxes for the Ybat annotation tool are manually annotated to ensure precision and regularity.The object detection algorithm that has been used is YOLO.The Darknet framework has been used for training neural networks for YOLO. The live CCTV image that has been captured is compared with the criminal and weapon database and an alert is sent to the nearby police station on matching.The architecture that has been used is extremely fast and detects faces at the rate of 45 fps.It can be used to detect criminals even in crowded areas.

To determine whether the infant's mouth is covered with vomit or a quilt, Chang and Chen in [2] used a deep learning neural network to detect the baby's face and the SSD+ Mobilenet network architecture in Tensorflow for infant vomit detection. The dataset that was used was the public face dataset-WIDERFACE. The method includes the following steps: infant face detection, vomit detection, and finally the classification results. The infant's mouth is detected, and noise is removed using gaussian filtering. The average pixel value of the mouth region is calculated, along with the difference between the previous and following frames..If the value of r(mouth area) is less than 0.5 then it is detected as vomit or the mouth being covered.This approach can successfully recognise the baby's face in a variety of lighting conditions or complicated backgrounds

In [3] there is a comparison between two techniques Haar Cascade and LBP Classifier for the purpose of facial recognition.A real time camera is used in order to capture the image.The input images are then converted to grayscale .The two techniques are then applied to the image using which the face and eyes are detected.There is a comparison between the two based on the factors of accuracy and time.The results indicate that Haar Cascade has a much higher level of accuracy and has the ability to detect a higher number of faces than the LBP classifier but LBP classifier is much faster than Haar Cascade.

Guodong Guo, Stan Z. Li. and Kapluk Chan in [4] have used SVM's with a binary tree recognition strategy in order to perform facial recognition. The dataset that has been used is the Cambridge ORL face database ,which is composed of 40 distinct persons with 10 different images corresponding to each person.In SVM, there are two strategies for multiclass recognition: one against one and one against all methods.The one against all method gives inaccurate results ,which is why the one against one method has been used.In the paper,the authors have tried to compare the error rates between the standard NCC algorithm and the SVM algorithm.Among the two,SVM has the lowest error rate ,which is 8.79%, which is comparatively better than NCC, which is 15.14%

In paper [5] ,any form of suspicious activities have been detected from the input video using YOLOv3.The dataset has been generated using video data as input by taking into account the three instances of suspicious activity—wallet theft, bag snatching, and lock breaking.The videos are converted to frames at 30 fps.In order to capture the area of interest, frames are annotated. The model has been trained and tested using the YOLOv3 algorithm.s been used for training and testing the model. YOLOv3 helps in detecting the presence of any suspicious actions and has a better performance than FASTER R-CNN.The model has the ability to detect each image at an extremely fast rate and has an accuracy of about 95%.The method of feature extraction that has been used gives accurate results only in a controlled environment.Due to the incredibly less amount of training data, there were differences between the test results and the ground truth.

In [6], the authors have used both detection as well as face recognition to identify criminals.The input image is used to identify the face. The normalising technique allows for the identification of the face landmarks. Feature vectors are made by extracting the features from the face.The face is identified and verified using the process of facial recognition.Multi Task Cascaded Convolutional Networks is used for facial detection and alignment in pictures.It is a 3 part CNN which can recognize landmarks on the face like nose,forehead and eyes.The images are loaded in the form of numpy arrays.Facenet has been used for verification and recognition of images.The implementation has been carried out using python in jupyter notebook.The model has achieved a high accuracy in facial classification.The dataset

has been input in the form of 200 images and the model does not support a dynamic dataset.An accuracy of 92% has been obtained for training and 90% for testing.

In paper [7],the authors have carried out realtime facial expression recognition using deep learning and IoT edge computing. The face detection model is applied to detect the presence and the location of the face.The face segment is cropped and the 128d face embedding is computed model classification is done for the three emotions happy,sad and sleeping.The deep learning face detector is trained using Caffe deep learning framework that is based on the Single Shot Detector framework with a ResNet base network.
Another DNN based model is used to merge the face into a 128-D unit hypersphere that quantifies the face. The DNN model is based on a Deep Convolutional Neural Network (CNN).After training, the fine-tuned deep learning model is optimised as per the hardware and deployed for production on a low-cost Jetson Nano embedded device.The deep learning model is deployed on the edge device. The deployed deep learning model is working as a web service where the image is sent through REST API to the webserver and in return, the model predicts the category of the image.The average precision, recall, and f1-score of the proposed approach for happy,crying, and sleeping categories outperform the machine learning model and all the processing can be performed on the edge device without using the internet.The disadvantages that have been highlighted in the paper are that the edge device has memory and computational constraints, so the size of the deep learning models should meet all constraints for functioning and the insights are sent to the cloud only if the internet is available.
In [8] Manju D and Radha V have proposed a novel approach for pose invariant face recognition in Surveillance videos using viola jones algorithm.The input consists of videos on which frame extraction is done.The concepts of integral image, Adaboost and cascade structure of Viola jones have been used.The features of the image are extracted using HOG and LBP.The method is highly accurate and robust for facial recognition.However it is a little slower in execution than the existing methods.

In paper [9], the authors have used fuzzy-kNN for the purpose of detection of choking in infants.The input data into the system is in the form of video sequences which are preprocessed to ensure that noise as well as any effect of shadow or lighting can be eliminated.After the detection of the face, there is extraction of 8 features from the facial region of the infant.Further, the facial expressions of hunger as well as normal has also been classified in this paper.According to the findings, each module and the fused judgement for each type of cry ,functions effectively during the detection procedure.The overall findings can be significantly improved with algorithms that identify infants' less prominent eyebrow positions..To make the sound processing module more robust, additional metrics could be combined.

In paper [10], the objective of the paper is to detect,classify and count objects by analysing visual events. The techniques that have been used are fast region proposal, feature extraction, and segmentation. For the purpose of the fast region proposal the CNN model has been used with hyperparameter optimisation. The Yolo algorithm has been used for object detection. The benefits are that it is robust enough to handle large datasets and has high performance. However due to illumination and shadow of the object, detection and classification errors are likely to occur.

In paper[11], the authors have used neural networks to detect facial emotions. The dataset has been self prepared and feature extraction has been done for the input. The models used were CNN along with Keras,Tensorflow and pretraining concepts. The Viola jones algorithm has been used to detect the eye and lips region, ML, DL, and NN algorithms can be used for emotion recognition. The advantages were the accuracy is high and has been determined using decision trees, 7 emotions have been detected and classified using the method.In order to achieve higher accuracy a large quantity of test data and keywords are required.

In paper[12], the authors have used deep learning and convolutional neural networks for face verification.
The algorithm used produces face feature vectors. The distance between these vectors allows one to determine whether images are from the same class or not. Deep Convolutional Neural Network has been used as the model for this approach. The advantages were that a modern face recognition algorithm was used and testing was carried out under unsupervised learning.However ,preprocessing has not been carried out. Enhancement of the AUC value can be done by carrying out preprocessing.

In [13] Kulkarni and Talele have performed tracking and detection of faces and objects from videos using Viola Jones.Videos have been provided as input to the model.This is followed by preprocessing. The preprocessed image is then subjected to image segmentation.Detection and recognition of objects is then carried out.The detected objects will then be tracked followed by the process of data fusion. The Viola Jones algorithm has been used for the purpose of detection and cropping.Tracking of continuous feature points is done using the KLT algorithm. The advantages were that the algorithm is robust even when noise and clutter is present. Selecting the facial frame from the real-time surveillance videos and analysing at the edge reduces human effort and also eliminates human errors. However the Viola Jones algorithm has a very slow training time and is mainly effective only when the face is in frontal view.

In paper [14], the authors have used convolutional neural networks to detect and track moving objects.In the proposed system, the input will be videos.This is followed by the extraction of frames.The Tensorflow library has been used for the purpose of object detection. Once the object has been located, the tracking of the object is done using CNN. Understanding the location of the object is trivial to the process of tracking.The object detection module in this system robustly detects the objects.Object tracking requires a huge number of features,using CNN for image classification improves the performance significantly as it is trained in millions of classes.The model tracks objects at the speed of 150 frames per second.This is also able to remove the barrier as a result of occlusion. The approach achieves an accuracy of 90.88%, 92.14% sensitivity and 91.24% specificity.

In paper [15], deep learning has been used for facial emotion detection. The technique that has been used consists of face detection. This is followed by feature extraction and emotion classification. The model that has been used is a Convolutional neural network based deep learning model. The advantages of this paper is that it presents the design of an artificial intelligence (AI) system capable of emotion detection through facial expressions. The results of the experiment indicate that the model proposed is better in terms of the results of emotion detection. However, higher accuracy can be obtained in terms of the FERC dataset.

IV. FRAMEWORK

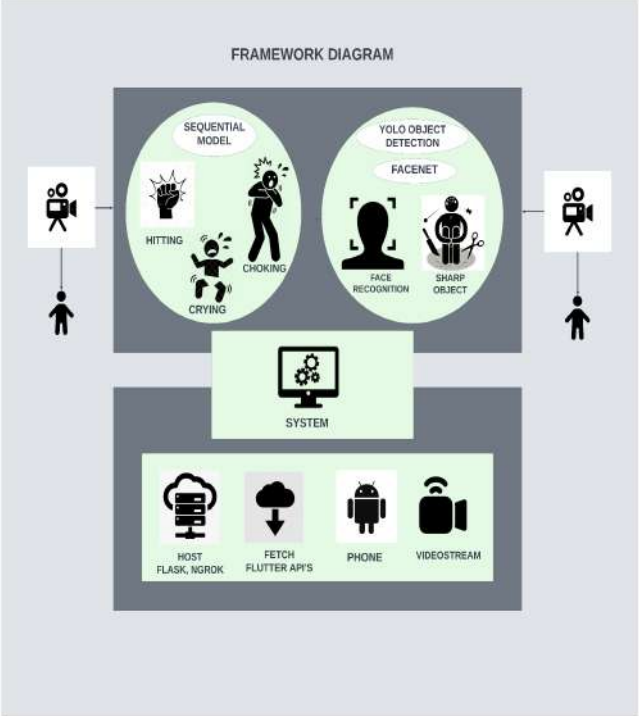


Fig. 1. Framework Diagram

The project framework consists of two modules. An infant activity detection module will be used to handle the cases of infant hitting, infant choking, and infant crying. The second module is composed of two sub parts. The first part is concerned with harmful object detection. It is used to detect sharp and harmful objects like knives, scissors, and forks. It can also detect the presence of a person. When a person is detected, the second part concerned with stranger detection is launched to identify faces by comparing them to the database of photographs initially fed into the application by the user. The primary purpose of this part is to identify the presence of any strangers near the infant. The two modules will work in parallel.An alert will be displayed on the application when any of these five cases are encountered. Live-streamed video via webcam will be used as the input to the framework.

V. IMPLEMENTATION

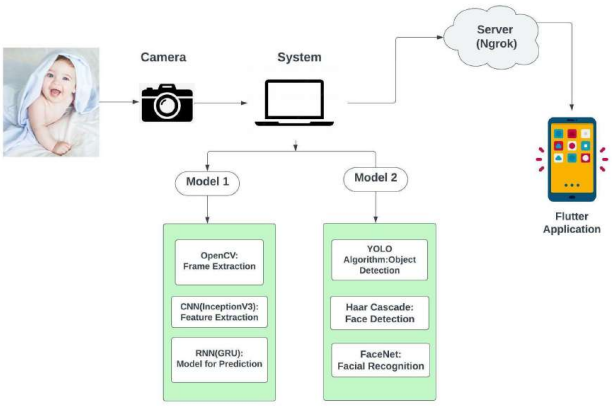


Fig. 2. System Architecture Diagram

The implementation of the project has been done in the form of two models. For the purpose of detecting the infant's activity a sequential model has been used. The model is used to identify instances of infant crying, infant hitting and infant choking. InceptionV3 has been used to extract features followed by GRU that has been used for prediction. The second model is a YOLO framework-based object detection model. It is used to detect harmful objects near the infant like knives, scissors and forks. It is also used to detect the presence of any stranger in the infant's room. When a person is detected, the input is sent to the Haar Cascade classifier. On detection of a face by the classifier, the FaceNet model is launched to identify faces by comparing them to the database of photographs initially fed into the application by the user.The two models will work parallelly. The results computed by the model are hosted using Ngrok as an api. Flask has been used as the web framework for the api. The api is used by the application to fetch the results which are displayed as alerts on the application. The input to the models will be the live streamed video via webcam.

VI RESULTS

The proposed sequential model has been trained on several videos. The model hyperparameters consist of sparse categorical cross entropy as the loss function and Adam as the optimizer.The model has been trained for 100 epochs with a batch size of 64.We have achieved an accuracy of 87.1% on this model.

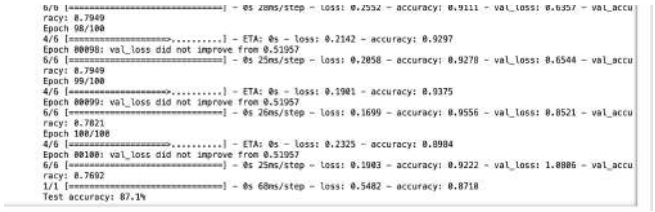


Fig. 3. Accuracy for Sequential Model

Live streamed video via webcam serves as input to the object detection model.

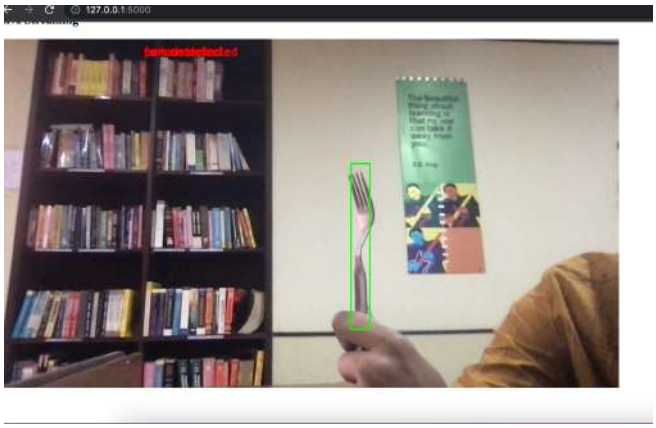


Fig. 4. Sharp object detected

The alert is then displayed on the Application on detection.

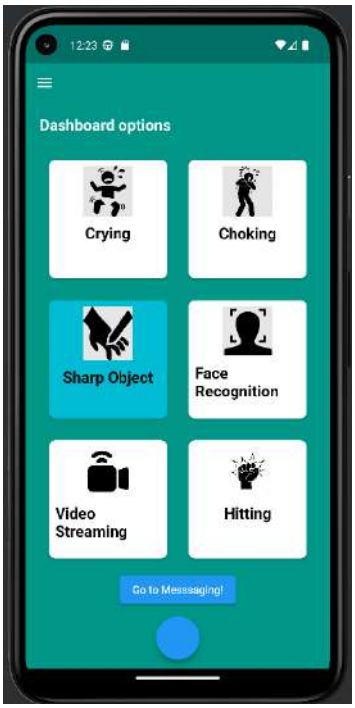


Fig. 5. Alert indicating detection of sharp objects on the Application

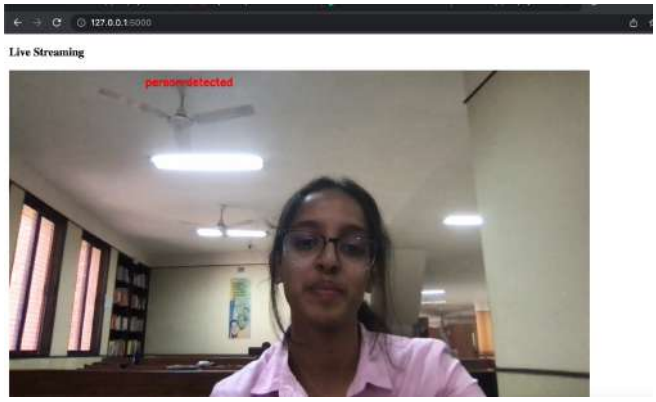


Fig. 6. Stranger detected

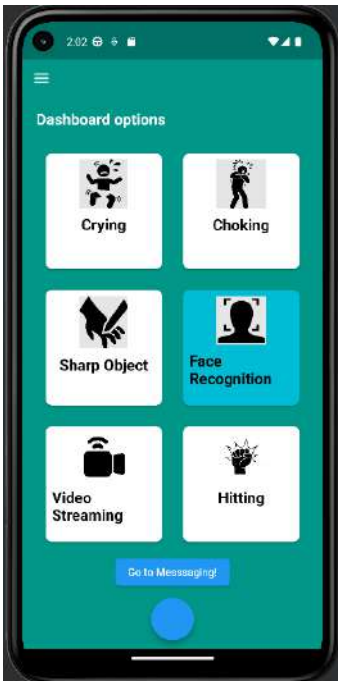


Fig. 7. Alert indicating detection of stranger on the Application

Fig. 8. Livestreaming on the Application

VII CONCLUSION

In this paper, a novel approach for infant monitoring has been presented. Live streamed video is provided as input and alerts are displayed in the application in real time. The framework of the infant monitoring system is novel and

revolutionary. The alerts generated would be extremely useful for parents and caretakers as it would help them to constantly monitor the infant even while being away from them. Hence this paper presents an approach for an infant monitoring system that is the need of the hour.An accuracy of 87.1% has been achieved by the sequential model.

REFERENCES

[1] P. Sivakumar, J. V, R. R and K. S, "Real Time Crime Detection Using Deep Learning Algorithm," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021, pp. 1-5, doi: 10.1109/ICSCAN53069.2021.9526393.

[2] C. -Y. Chang and F. R. Chen, "Application of Deep Learning for Infant Vomiting and Crying Detection," 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2018, pp. 633-635, doi: 10.1109/WAINA.2018.00158.

[3] Anirudha B Shetty, Bhoomika, Deeksha, Jeevan Rebeiro, Ramyashree, Facial recognition using Haar cascade and LBP classifiers, Global Transitions Proceedings, Volume 2, Issue 2, 2021, Pages 330-335, ISSN 2666-285X,https://doi.org/10.1016/j.gltp.2021.08.044.

[4] Guodong Guo, S. Z. Li and Kapluk Chan, "Face recognition by support vector machines," Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), 2000, pp. 196-201, doi: 10.1109/AFGR.2000.840634.

[5] N. Bordoloi, A. K. Talukdar and K. K. Sarma, "Suspicious Activity Detection from Videos using YOLOv3," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342230.

[6] S. T. Ratnaparkhi, A. Tandasi and S. Saraswat, "Face Detection and Recognition for Criminal Identification System," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 773-777, doi: 0.1109/Confluence 51648.2021.9377205.

[7] R. Pathak and Y. Singh, "Real Time Baby Facial Expression Recognition Using Deep Learning and IoT Edge Computing," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277428.

[8] Manju, D. and Radha, V., 2020. A novel approach for pose invariant face recognition in surveillance videos. *Procedia Computer Science*, *167*, pp.890-899.

[9] M. N. Mansor, S. H. -F. S. M. Jamil, M. N. Rejab and A. H. -F. S. M. Jamil, "Fuzzy k-NN for choke infant detection," 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), 2012, pp. 349-351, doi: 10.1109/MSNA.2012.6324590.

[10]M. M. Sein, K. S. Htet, K. T. Murata and S. Phon-Amnuaisuk, "Object Detection, Classification and Counting for Analysis of Visual Events," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), 2020, pp. 274-275, doi: 10.1109/GCCE50665.2020.9292058.

[11] Ali, Md. Forhad & Khatun, Mehenag & Turzo, Nakib. (2020). Facial Emotion Detection Using Neural Network. International Journal of Scientific and Engineering Research. 11. 1318-1325.

[12] A. Lebedev, V. Khryashchev, A. Priorov and O. Stepanova, "Face verification based on convolutional neural network and deep learning," 2017 IEEE East-West Design & Test Symposium (EWDTS), 2017, pp. 1-5, doi: 10.1109/EWDTS.2017.8110157.

[13] V. Kulkarni and K. Talele, "Video Analytics for Face Detection and Tracking," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 962-965, doi: 10.1109/ICACCCN51052.2020.9362900.

[14] S. Mane and S. Mangale, "Moving Object Detection and Tracking Using Convolutional Neural Networks," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1809-1813, doi: 10.1109/ICCONS.2018.8662921.

[15] Jaiswal, Akriti, A. Krishnama Raju and Suman Deb. “Facial Emotion Detection Using Deep Learning.” *2020 International Conference for Emerging Technology (INCET)* (2020): 1-5.

Appendix B

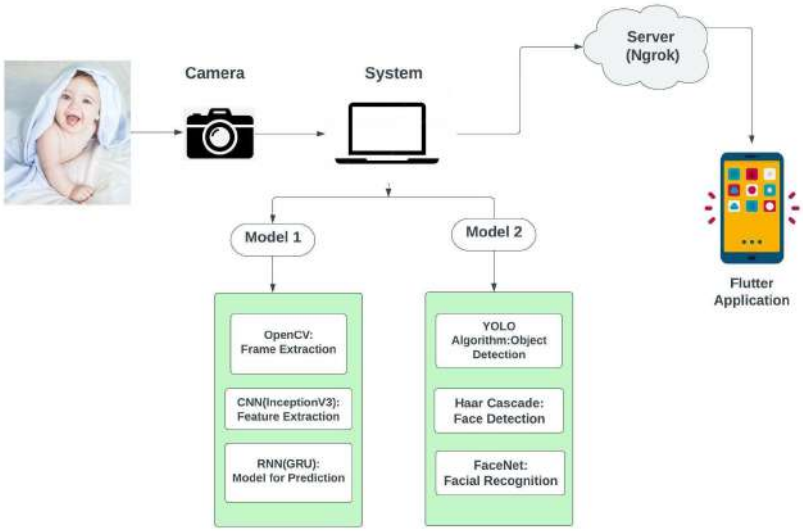
<table>
<tr><td>Group No:48</td><td colspan="2">Title: Infantza:Computer Vision and Deep Learning Enabled Infant Surveillance System
Domain:Computer Vision and Deep Learning</td></tr>
<tr><td colspan="3">Abstract: With the rise in complexity of the job roles of today's parents and their hectic schedules, the need for infants to be observed frequently when left in the care of a caretaker to avoid any kind of injury and to constantly have an eye upon them all day becomes a tedious task. In today's world, small infants and children being subjected to abuse from caretakers and others has become a serious issue, due to which parents are unable to entrust them with the safety of their child. We seek to provide a novel approach for infant monitoring that sends alerts to parents with respect to few cases of prospectively dangerous situations that an infant might be exposed to.To implement this a framework has been built.It consists of two phases. The first phase consists of the models for infant activity detection, harmful object detection and stranger detection. The second phase consists of the mobile application through which parents receive alerts if the infant is exposed to danger.This would avoid the need for constant monitoring of the infant from the parent's side as the monitoring would mainly be required only when an alert has been issued indicating the occurrence of an abnormal situation.</td></tr>
<tr><td>Team:

Supervisor:</td><td>Anagha Suresh
PES2UG19CS037

Immadisetty Sai Jayanth
PES2UG19CS152

Jeevan Anil
PES2UG19CS166

Jitta Amit Sai
PES2UG19CS169

Dr. Bharathi R</td><td>Architecture/flow diagram
 
</td></tr>
</table>