AUTOMATED CHILD DAYCARE MONITORING SYSTEM-A DEEP LEARNING METHOD TO DETECT UNSAFE BEHAVIOUR UPON CHILDREN

Submitted by

Ratib-Al-Karim

183-35-2578

Department of Software Engineering

Daffodil International University

Supervised by

Md. Shohel Arman

Assistant Professor

Department of Software Engineering

Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering

© All right Reserved by Daffodil International Universit

APPROVAL

This thesis titled on **“Automated Child Daycare Monitoring system- A Deep Learning Method to Detect Unsafe Behavior Upon Children”**, submitted by **Ratib-Al-Karim (ID: 183-35-2578)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. Imran Mahmud
Head and Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

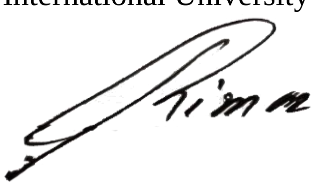
Internal Examiner 1

Md. Khaled sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

External Examiner



Rimaz Khan
Managing Director
Tecognize Solution Limited

 ©Daffodil International University

THESIS DECLARATION

I announce hereby that I am rendering this study document under Md. Shohel Arman, Lecturer, Department of Software Engineering, Daffodil International University. I therefore state that this work or any portion of it was not proposed here therefore for Bachelor's degree or any graduation.

<u>Supervised by</u>

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Daffodil International University

<u>Submitted by</u>

Ratib Al- Karim

Ratib-Al-Karim
183-35-2578
Department of Software Engineering
Daffodil International University

 ©Daffodil International University

Acknowledgment

I would want to send my profound thanks to the All-Powerful God who gave my family and me with loving care during this journey till the accomplishment of this undertaking.

I also want to express my appreciation to Md. Shohel Arman, an assistant professor in the department of software engineering at Daffodil International University, served as my honorary supervisor, for his proper direction, assistance, support, and collaboration. without whom my thesis work would not be completed. His endless patience, intellectual leadership, encouragement, frequent and forceful supervision, constructive criticism, intelligent guidance, and reading of multiple mediocre versions and subsequent improvements made this endeavor possible.

I'd want to express my gratitude to all of the students at Daffodil International University who joined in on this discussion while also focusing on their studies.

Finally, I want to convey my thankfulness to my parents and my brother, both of whom have always encouraged me. They constantly give my input and inspiration priority.

 ©Daffodil International University

Abstract

Abuse in daycare centers is a very serious problem in modern parenting. Our lives are becoming busier every day as we place a greater emphasis on our professional lives. Since most parents choose to send their kids to daycare while they are at work, there have been a lot of recent reports of child abuse at daycare. Every year, there are more abused kids in daycare than the year before, and sometimes this results in significant injuries. Daycare is now seen by parents as a risk to their kids. As a result, the daycare industry is ultimately failing, and many parents' jobs will be hampered by the issue of child care. Therefore, a method that can instantly classify child abuse in daycare using CCTV footage must be developed. I decided to create a mechanism that will aid in the authority's classification of child abuse detection as a result. To find child maltreatment in daycare, I employed image processing and a custom convolutional neural network method. The method produced a promising result with an astonishing 84% accuracy in this case. In order to improve accuracy and incorporate this in CCTV to receive notifications of abuse in real-time, I will work on a large dataset with more Transfer learning techniques in the near future.

 ©Daffodil International University

Table of Contents

©Daffodil International University

LIST OF FIGURES

 ©Daffodil International University

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Abuse of children takes place all too regularly in today's daycares, making it a huge source of concern for the parents of today. Because we place a higher value on our professional lives than on our personal lives, our calendars fill up quite rapidly. In recent times, there have been a number of charges leveled against daycare institutions for child abuse. This should not come as a surprise considering that the vast majority of parents choose to drop off their children at the facility while they are at work. Due to the fact that the majority of parents in today's society are employed, it is impossible for them to constantly be by their child's side. Recent news articles have shed light on a number of instances of child abuse, most of which occurred at the hands of individuals employed in childcare facilities and cleaning services. Abuse of children who are entrusted to the care of daycare workers is on the rise, and some of these incidents result in the child suffering severe physical harm. These days, a lot of parents are concerned that their kids will get hurt while they're at daycare. As a result, the daycare industry is eventually going to implode, and the lack of available child care will make it difficult for many parents to maintain employment. Therefore, it is absolutely necessary to have a remote monitoring system set up.

A rise in the number of reported cases of child abuse has led to an increase in the number of people interested in installing surveillance systems in their own homes or in childcare facilities. The existing surveillance systems require a significant financial investment to install, as well as a large number of cameras, in order to adequately monitor an area. As a result, there is an urgent need to develop a system that can instantly identify instances of child abuse in daycare settings by analyzing CCTV data.

Dealing with the problem of detection can be done using a variety of strategies and approaches. During the course of their investigations, researchers have utilized a great number of distinct algorithms, the most of which have been focused on deep learning. since the work involves the processing of images. Because of this, deep learning may be utilized to discover and organize them in an easy and efficient manner.

Image recognition achieved through the application of deep learning is a common activity. Why is it so crucial that we categorize things the way that we do? Why? Because it is vital to advise on how to spot the indications of abuse and provide a model of who could or could not commit such actions on children, and because it is important to provide a model of who could or could not commit such acts on children. These questions comprise the entirety of the information that you require. The workers at the daycare center where it all began had been divided up into divisions.

 ©Daffodil International University

1.2 MOTIVATION OF THE RESEARCH

Undoubtedly, the majority of parents have the intention of giving their offspring the very greatest opportunities in life. According to research, parents and other caretakers, including those who work in child care, who have access to resources and support are more likely to provide children with surroundings that are both safe and healthy. Indeed, when children grow up in an environment that provides them with stable relationships and interesting experiences, they have the opportunity to draw from that setting to become self-assured and caring individuals.

However, the abuse or neglect of children can and can still take place.

Abuse and neglect of children can take the form of a single incident, such as a caregiver shaking an infant to silence its cries, or it can take the form of a pattern of behavior, such as a parent providing insufficient supervision or sexually abusing a child over the course of several months or years. Both of these scenarios are forms of child maltreatment. It is not always easy to diagnose, and this is especially true in younger children who may not be seen on a consistent basis by anyone other than their parents or those who care for them in a daycare setting. Because of their consistent interaction with the children in their care, child care workers are in an advantageous position to detect any signs of possible child abuse. In addition to this, they are required by law to report it.

Therefore, the goal of the study is to create a system for the parents that will produce the best outcomes based on data from the daycare center’s administration system.

 ©Daffodil International University

1.3 PROBLEM STATEMENT

Children under the age of three have a significant dependence on the adults who care for them, whether those adults are members of the child's family or friends. Babysitters, nannies, friends, neighbors, and members of the general community are not regarded to be members of the same family; nevertheless, siblings, parents, and extended family members are considered to be members of the same family.

According to the findings of the United States Census Bureau, 11.6 million children under the age of five spent a typical week participating in some sort of child care, which represents 63% of the total. Forty percent of all children younger than five years old had a relative watch them at least once throughout the week, while 35 percent of all children in this age range had someone who was not a relative watch them, and 0.2% of all children in this age range took care of themselves. A little less than one in four children were cared for in an establishment such as a daycare or a preschool such as Head Start, while another one in six children were looked after in the homes of their respective families.

Every adult who takes care of children has the responsibility of ensuring that the children in their charge are housed in an environment that is safe and welcoming, as well as one that fosters the children's physical and mental development in a way that is conducive to a full and healthy life. Child care providers, in light of the current emphasis on the significance of early brain development and the desire of communities to send their children to school "ready to learn," are obligated to make the promotion of both positive social and emotional development as well as early learning a priority in the care they provide for young children.

A lot of work has already been done with Child abuse. Different sorts of algorithms, including Deep Convolutional Neural Networks, CNN, R-CNN, embedded systems, and many more, are used by many of the researcher's authors. In other countries, they work with their own data on the daycare centres. But the climate and situation of Bangladesh are completely different from other countries. The study, therefore, has relevance for their own nation. On the dataset obtained from the daycares, we wish to use deep learning models.

 ©Daffodil International University

1.4 Research Questions

- Can this Method detect child abuse in Daycare center?
- How to Detect whether the incident of Child abuse is occurring or not in a selected daycare center?
- Can this Method suggest any alarm ringing system for the daycare center?

1.5 RESEARCH OBJECTIVE

Increasing reports of abuse against children have led many parents to view daycare as a potentially dangerous environment for their children. Consequently, the daycare business is doomed, and many parents' ability to work will be impaired as a result of the child care crisis. Thus, it is necessary to create a system that can immediately categorize child abuse in daycare using CCTV footage. Because of this, I've resolved to develop a system that will facilitate the authority's categorization of child abuse detection. My paper's major goal is to detect any physical abuse on children in Daycare center automatically. Additionally, we aimed for a better outcome to increase the applicability of our concept. Our thesis is:

- Establishing an automated system.

- For Detecting whether the incident of Child abuse is occurring or not in a selected daycare center.

- Using Different daycare centers footage to suggest the best one for any child.

- Ensuring Better Environment for a child in a Daycare Centre.

 ©Daffodil International University

1.6 RESEARCH SCOPE

The following are the primary research areas:

- To create a method based on daycare center's CCTV footages in order to collect data on abusers.
- Will benefit the Bangladeshi Working parents by having a system that is quickly detect whether their child is safe or not, and will also identify the child abuser immediately. Therefore, it will aid the authorities in monitoring their own organization and employees well.
- We can easily find out numbers of child abusing incidents.
- We will try to make an alarm system so that the message goes to the parents and the authority after sensing a bit of physical abuse.
- Parents will be greatly benefited from this system. They will be able to detect and file cases very easily and it will reduce their suffering a lot.
- As a result of this system, recognition of Child abusing can be done very easily and the data of the footage can be checked and the case file in abuser's name will be convenient, which will be a boon for the management.

 ©Daffodil International University

1.7 THESIS ORGANIZATION

The first chapter focuses on the Daycare Centre and Child abuse incident identification system and its application, providing context for the study, justifying the research, describing the issue, posing research questions, and outlining the research aim. Our study also includes the following elements:

In Chapter 2, I will discuss the literature review, which provides a look at previous studies addressing questions about child abuse, the methodologies employed, the gaps in the research, and a comparison between my own study and those of other researchers based on our shared understanding of our findings.

The final section will focus on how we conducted our research. My paper's methodology part will discuss my approach to data collecting, data preprocessing, and analysis of results. Methodological details are also presented. When is a certain technique employed? Which model is responsible for which technique?

The findings of the technique as well as a comprehensive discussion of the many different possible scores for the correctness of the results will be presented in Chapter 4. All of the accuracy, helmet detection, and license plate identification discussed in this study are readily apparent.

As the title suggests, the last chapter wraps things off. In this final section, I will present a thorough overview of my entire paper.

©Daffodil International University

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

A literature review is an analysis of previously published materials such as studies, articles, books, and conference proceedings. With this data, you may find out what has been done to address the issue, get a summary of that effort, and spot any holes in the research. They may then zero in on constraints and brainstorm workarounds in order to achieve the desired outcomes.

2.2 PREVIOUS LITERATURE

In India, there has been a rise in the number of reported cases of both domestic violence and child abuse. Recently, in the month of November of 2016, a maid-cum-helper at a daycare in Mumbai violently abused a 10-month-old daughter. When the parents noticed the child's face had marks consistent with an injury, they filed a complaint. The maid was detained after the act was caught on camera. There have been reports of similar instances in other regions as well. A daycare worker in Florida was arrested after she was seen on camera kicking a sleeping child in the head. The prevalence of child maltreatment has increased the demand for monitoring systems in homes and daycares. It is important for parents to keep an eye on their children at all times, not just when they are with a caretaker but even when they are home alone. Existing surveillance systems are expensive to set up and require numerous cameras to cover all possible angles.

Many researchers started working on it. I got very good results from this. Again, many of them could not go to their main goal properly. Many researchers have done detection classification using many models.

 ©Daffodil International University

In Fatima Mahmood, Jehangir Arshad[5] , They introduce a deep learning–based system, specifically a Faster Regional Convolution Neural Network, for in-exam automation (RCNN). While MTCNN (Multi-task Cascaded Convolutional Neural Networks) is used for face detection and recognition, Faster RCNN (Recurrent Convolutional Neural Networks) is utilized to detect suspicious actions of students based on their head movements during tests. The suggested model achieves a remarkable 99.5% accuracy during training and 98.5% during testing. The model is accurate enough to reliably detect and track over a hundred test-takers in a single image.

In LIN, & Jhang.[20]'s paper, they combine you only look once (YOLO) with a convolutional fuzzy neural network (CFNN) to create a smart traffic-monitoring system that can accurately capture data on traffic volumes and vehicle types. Vehicle detection is handled using YOLO, and then traffic flow is determined using a vehicle counting method. Then, a network mapping fusion technique and two efficient models (CFNN and VectorCFNN) are suggested for vehicle classification.

In Ngo, & Lin.[18], In light of the popularity of convolutional neural networks for object detection—and in particular, person detection—we propose a room monitoring system that employs deep learning and perspective correction to identify individuals who are in various stages of motion, as well as high-density groups and individuals who appear small due to their physical distance from the camera. The input images for this system are successive frames captured by the surveillance camera. Perspective correction and person detection are the two main methods used. At the outset, an image is converted from its original 3D viewpoint into a 2D top-down version via perspective correction. Users can get a better look at the system from a number of perspectives thanks to this. For the second part, the suggested person identification system combines the Mask region-based convolutional neural network (R-CNN) scheme and the tile technique for person detection, notably for recognizing small-sized humans.

 ©Daffodil International University

Chandan.[19] shows the effort is to develop a real-time video processing–based facial recognition system for an automatic gate pass. The database contains a number of images taken of the pupil in a range of settings. The proposed system requires students to present their faces in front of a camera as they leave campus, and it then identifies them using the information already saved about them. When a student is identified, their information is retrieved from the system and they are added to the e-register. In the meantime, the computerized system will alert the relevant student. The system's graphical user interface (GUI) makes it possible to keep tabs on everything happening within it with relative ease.

Nadia Zouba, François Bremond, and Monique Thonnat[10] suggest a system of activity monitors to carry out the aforementioned behavior study of the senior population. The suggested system is based on a method that merges data from several types of sensors in order to identify domestic routines. This method integrates footage from security cameras with readings from sensors fastened to common areas of the home. In this study, 9 genuine senior volunteers from the experimental flat are used to verify the accuracy of the proposed activity monitoring system in recognizing a set of daily tasks (such as using kitchen equipment, making a meal). Each of the nine senior volunteers' behavioral profiles is compared to the others. This research demonstrates that the suggested approach is highly valued by both the older population and the medical community.

Muhammad Ramzan , Adnan Abid and Shahid Mahmood Awan's [6] is the first study to create a dataset of anomalous examination behaviors and propose a deep learning algorithm to identify them. The suggested model, which uses motion-based frame extraction techniques to extract key-frames and then applies an advanced deep learning Convolutional Neural Network algorithm with varying configurations, has been dubbed Automatic Unusual Activity Recognition (AUAR). The AUAR model has been shown to perform better than previously proposed methods of abnormal behavior recognition in standardized evaluations. In addition to testing the proposed model on the examination dataset, They also implement it on the Violent and Movies datasets, which are commonly utilized in the aforementioned literature for this purpose. The outcomes show that AUAR outperforms the state-of-the-art models on multiple data sets.

 ©Daffodil International University

In Fuerties, & Carbelleira.[7], They offer a novel omnidirectional camera-based deep learning method for persons detection that just needs point-based annotations, in contrast to the majority of state-of-the-art approaches, which require bounding box annotations. As a result, the time and effort required to manually annotate the necessary training databases is drastically reduced, allowing for a quicker rollout of the system. The approach relies on an innovative deep neural network design that not only implements the idea of a Grid of Spatial-Aware Classifiers, but also permits the kind of end-to-end training that boosts the system's performance as a whole.

The discussion of Detection of suspicious human actions in automated video surveillance applications in Scariaa, E., Abahai, A., and Isaacc, E.[8] is quite relevant. The unpredictable nature of human motion makes it challenging to reliably categorize suspicious human actions. The project's overarching goal is to establish a methodology for the challenge of autonomously following people and identifying suspicious behavior in footage captured by Closed Circuit Television (CCTV).

It has been demonstrated in a study led by Alexandra König, Carlos Fernando Crispim-Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, et al.[9] that IADL functioning can be assessed with the aid of an automatic video monitoring system, and that even on the basis of the extracted data, significant group differences can be obtained. Dementia patients' ability to do instrumental activities of daily living (IADL) was evaluated with the help of an automated event recognition video surveillance system.

©Daffodil International University

The installation of an affordable video surveillance system in a home or daycare center in Okky Permatasari, Siti Ummi Masruroh, or Arini [4] is something that can be beneficial. One of these projects has already been finished: a Raspberry Pi has been hooked up to a stationary camera, and the video feed is being streamed live to a website where the parents can keep an eye on what's going on. Deep learning models are utilized in order to keep track of one facet of the overall objective. In this part of the process, the RCNN and CNN are used to identify the children. The models were able to successfully deploy replicas of themselves in order to keep an eye on the item. Other models are responsible for handling the additional processing aspects. There is a 98.06% chance that the model is correct. Their outputs indicate that they perform exceptionally well on their dataset.

Fernand, Bathrinarayanan, & Fosty. [11] show the results of testing a prototype video monitoring system for recognizing events involving the elderly. Participants in a clinical protocol for Alzheimer's disease study (n=29) had their recognition accuracy for physical tasks (like the Up and Go test) and instrumental activities of daily living (like watching TV and writing checks) assessed using the prototype.

In Sundar, S., Ghosh, R., & Shahil, H. [12], Using Passive Infrared (PIR) sensors and a Servo motor, the suggested system combines a Raspberry Pi microcomputer with a camera to create a dynamic setup. As the kid wanders around the room, the camera will follow him or her with automatic rotation. The camera will be placed in the room's geographic center for maximum coverage. Parents can watch their child's activities in real time by going into a website and viewing the collected video feed. Whenever any sensor is triggered, an SMS alert is sent to the parents through a GSM module. The proposed system is a cheap surveillance setup that may be installed in homes or daycares.

 ©Daffodil International University

Reddy, & Vamsi. [13] To ensure the proper care and wellbeing of their children, busy parents can use their work, which provides a baby watching framework. This framework is able to recognize the child's movement and aid in detecting sounds, notably crying and newborn's present position, which can be forecasted using CNN so the parent can monitor the state of the infant together with the sensor data while away from the kid.Within the framework of the proposed work, a variety of sensors will be polled for data, which will subsequently be continually processed by a Raspberry PI.

The article by Karageorge, & Kendall. [14] explains how daycare workers may help stop, spot, and report cases of child abuse and neglect both inside and outside of formal childcare settings. This guide provides an overview of child maltreatment, including preventative measures, reporting legislation, care for abused children, and resources for parents and professionals. Professionals working with young children in a range of settings, such as daycares, Head Start programs, preschools, nursery schools, and family child care homes, are the major readers of this guide. In addition, this group contains early childhood educators, home visitors, parent educators, program administrators, and trainers.

Pais , & Buluschek.[15] talk about The demands placed on healthcare systems and its associated budgets are rising as a direct result of the aging of the population. The ability to function independently declines with age because of frailty and the chronic diseases that plague the elderly. However, the vast majority of seniors aspire to remain in their own homes as they age. There is hope that the emerging field of in-home monitoring will help the elderly maintain their autonomy as they age while still facilitating their fitness, well-being, and social engagement. This prospective, observational study followed home-dwelling older adults (OA), their family carers (FC), and nurses over the course of 12 months to assess the efficacy of a novel in-home monitoring system designed to supplement OAs' in-home care.

 ©Daffodil International University

Jelwy, & mahdi.[16] indicate in theie article Abnormal behavior in the modern environment often poses a danger to others. A phenomenon that stands out from the ordinary is considered an anomaly. Intelligent video surveillance is required since it is difficult to maintain constant monitoring of public areas. When artificial intelligence, machine learning, and deep learning were integrated into the system, the technology had evolved much too far. All of the aforementioned combinations can be used in a variety of ways to aid in differentiating between normal and suspicious activities in real time surveillance film. Human behavior is the most unpredictable, and assessing whether it is suspicious or normal is pretty tricky. A deep learning technique is used in a university setting to identify normal and aberrant behavior and notify the proper authorities if suspicious behaviour is predicted. A common method of monitoring involves taking still images from a video in order. The structure consists of a top and a bottom half. In the first step, features are extracted from video frames, and in the second, a classifier uses those characteristics to make a prediction about whether a certain class is suspicious or not.
This study provides a practical approach to the design of a system that can be deployed in both indoor and outdoor locations inside a university campus and is capable of automatically detecting any unexpected or abnormal condition and alerting the proper authority. Overall, the proposed approach was 95.3% accurate.

In Dinga, & Fang.[17], they propose a new convolution neural network (CNN)/long short-term memory (LSTM) hybrid deep learning model for detecting dangerous behaviors in the workplace. The suggested hybrid deep learning model can be used to (1) recognize potentially dangerous behaviors, (2) gather motion data and locate movies, (3) extract visual characteristics from videos with a CNN model, and (4) sequence learning features made possible by the use of LSTM models. The model's capacity to identify potentially harmful behaviors is evaluated experimentally. According to the findings, the created hybrid model (CNN + LSTM) can effectively identify safe/unsafe acts performed by personnel on-site. The model's accuracy surpasses that of the best existing descriptor-based approaches for identifying image hotspots.

 ©Daffodil International University

The work of Zejun Huang, Huijuan Lu, and others[1] describes the development of a system that makes use of an android phone to determine whether the child is in a safe zone or a hazardous zone and then utilizes GPS to update the location. Deep learning was integrated with hardware attachments in this strategy so that the child's location could be determined and his safety could be guaranteed. They utilized GPS to monitor their own positions, and they utilized deep learning models to detect and protect themselves from potential threats. They are applying YOLOv3 in order to locate the object as well as the area around it. The CNN model, which is based on the YOLOV3 model, has a rate of accuracy that is equivalent to 90.23 percent. The model had sufficient granularity to identify them and check that their current position was in line with the goals they had set for themselves.

Similarly, in Dhiraj Sunehra, Pottabathini Laxmi Priya, and Ayesha Bano[2,] another system was developed that allows the whereabouts of a kid to be tracked using GPS and Google, and this location can be monitored by the school as well as the child's parents. This article employed the CNN (CONVOLUTIONAL NEURAL NETWORK) as its model, which suggests that approaches from the field of deep learning are being applied in order to discover the location of the child by reacting negatively to its presence. There are some hidden layers that have a high accuracy of 97.45% thanks to the fact that the task of object detection was completed by a Custom CNN. Maintaining the highest possible degree of precision while keeping track of your position was a very accurate google map.

Damodaran Sanipath and Sankar Sundarani[3] discuss the process of developing a communications system that, in the event of a catastrophic medical emergency affecting a kid, is capable of issuing a warning through the use of a remote alert. On the other hand, there is no visual representation of the child in any of them. Deep learning techniques, such as CNN and LSTM models, are utilized to evaluate the youngster's state of health and to keep track of him by evaluating the facial expressions and emotions that the child displays. When it comes to carrying out their monitoring responsibilities, these models are very efficient. The hybrid deep learning model that was suggested is capable of identifying potentially harmful activities, locating movies using motion data, extracting visual characteristics from films using LSTM models, and extracting visual features from videos using CNN models. They achieved a level of accuracy for their dataset of 90%.

 ©Daffodil International University

2.3 CONCLUSION

Algorithms come in a variety of forms. They used a variety of techniques, including feature extraction, augmentation, annotation, and more, to achieve a decent outcome. Leave real-time performance and increased accuracy behind. In our efforts, we also tried to make it simple to identify abusers who abuses child in daycare centre and to work on real-time data.

 ©Daffodil International University

CHAPTER 3
RESEARCH METHODOLOGY

3.1 RESEARCH METHODOLOGY

On a dataset obtained from my own dataset, I have utilized Custom CNN.

3.2 DATA COLLECTION

The research project began with a collection of data. The data in my research methodology includes second-hand data collection, as it was quite hard to get such sensitive data firsthand. I systematically collected and evaluated the sensitive data related to child abuse in daycare from empirical evidence collected through YouTube. Although there is a lot of data related to child abuse online only a few are related to daycare and tried to get the data and that was restricted. So I go through different case videos and CCTV footage available on different websites, YouTube, and social media to get the data and then took screenshots from the videos. Many videos are blurred with low quality so I have to ignore those videos and took only clean images so that machine can identify and can recognize violence here. I tried to get all class data equally so that distribution can be identical for taking training and test dataset. Maximum data I took for the violence class which is 512 and then the non-violence class where I got 440 images. As data is very low otherwise, I will try to at least get 1000 images for each class. The dataset capacity is very low for the perspective of the research work, if this set of data will be more than 5000 then the research work will be nice and great. The maximum videos of CCTV footages are very poor in quality and the availability of those videos are very rare. The amount of violence dataset 512 is collected from very few videos because of the unavailability of these types of videos and time is one more element to play here for collecting the data. The Non-Violence dataset images of 440 are very easy to get because of the occurrence of timing, so this type of data sets is easily to get.

©Daffodil International University

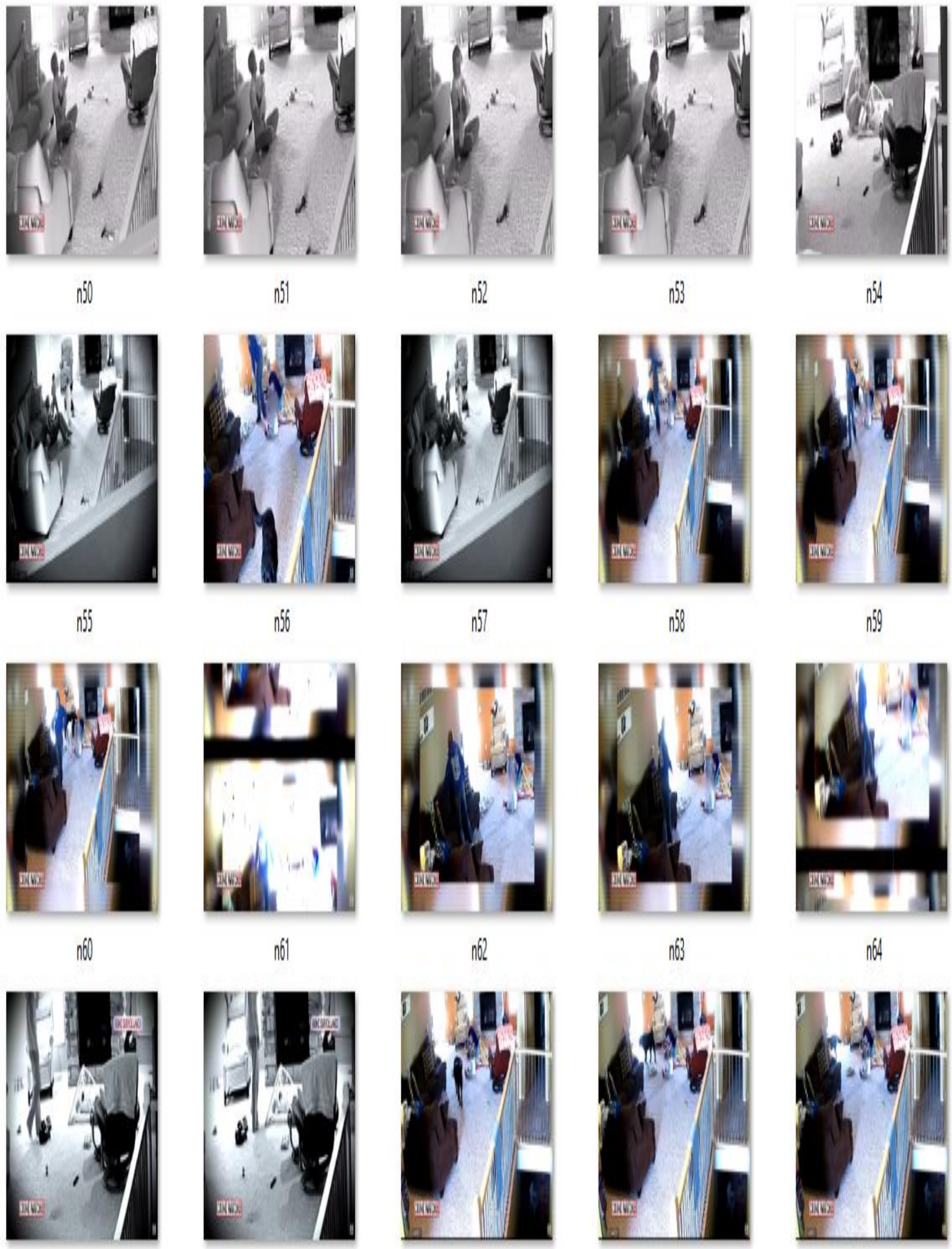


Figure 1: Non-violence Dataset

 ©Daffodil International University

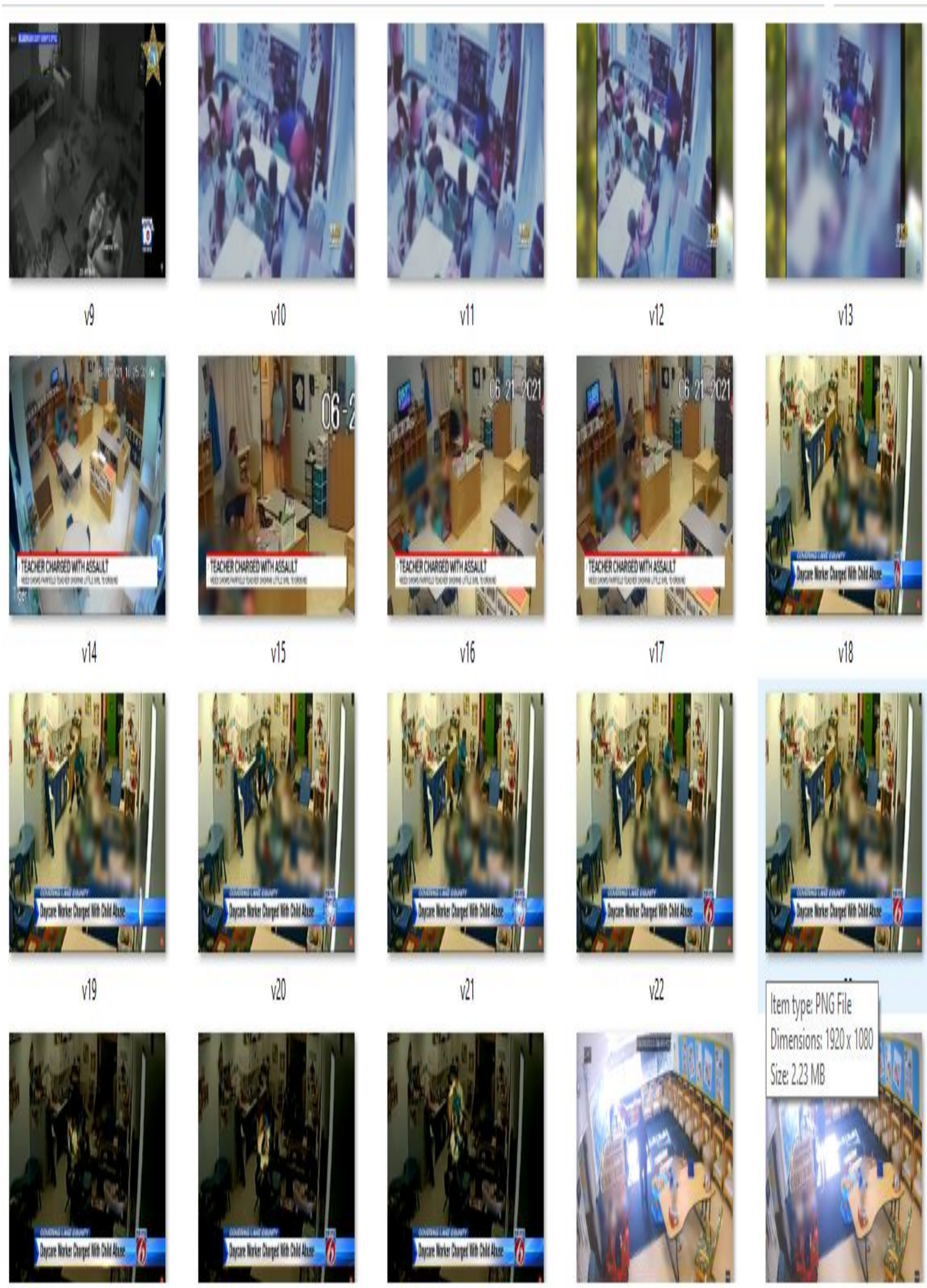


Figure 2: Violence Dataset

©Daffodil International University

3.3 DATA PREPROCESSING

Data preprocessing is an important stage in putting the data into a usable format. Data mining includes the stage of data preprocessing. I collected data from authentic video frames from the internet. The dataset has 512 images of violence and 439 images of non-violence. The images are stored in separate labeled folders. Prior to training, all the images were converted to greyscale images. Consequently, images were resized to a predefined width size of 128 and height size of 128.

Figure 3: Grayscale images after converting.

 ©Daffodil International University

Figure 4: Grayscale images after converting.

©Daffodil International University

Subsequently, the image data were converted to a NumPy array for further processing. For the train-test purpose, I split the dataset into 70 percent training data and 30 percent testing data and the validation set during training was 30 percent of the data. All the image data NumPy array was divided by 255 for data normalization. Furthermore, Label Encoder was applied to the target column where 0 represents Non-Violence and 1 represents Violence.

3.4 CONVOLUTIONAL NEURAL NETWORK(CNN)

Convolutional neural networks are a kind of deep learning, which is important to understand. The most common techniques to identify images and videos are Convolutional neural networks (CNN or DCNN). In 1980, it was first applied. So, here are the explanations of the fundamental concept behind CNNs (Convolutional Neural Networks) (figure 5). Although there are different kinds of neural networks in deep learning, CNNs are the preferred network architecture for identifying and recognizing objects. They are therefore ideally suited for computer vision (CV) activities and for applications where accurate object recognition is crucial, such facial and self-driving automobile systems.

An essential component of deep learning techniques are artificial neural networks (ANNs). Recurrent neural networks (RNNs), one type of ANN, take input from time series or sequential data. It is appropriate for applications involving speech recognition, language translation, natural language processing (NLP), and image captioning.

A different kind of neural network called a CNN may find important information in both time series and picture data. This makes it very beneficial for applications involving images, such as pattern recognition, object classification, and picture identification. A CNN makes use of linear algebraic concepts, including matrix multiplication, to find patterns in an image. CNNs may categorize audio and signal data as well.

 ©Daffodil International University

Multiple layers of a CNN are possible, and each layer trains the CNN to recognize the many aspects of an input image. Each image is given a filter or kernel to create an output that gets better and more detailed with each layer. The filters may begin as basic characteristics in the lower layers.

In order to check and identify features that specifically reflect the input item, the complexity of the filters increases with each additional layer. As a result, the partially recognized image from each layer's output, or convolved image, serves as the input for the subsequent layer.

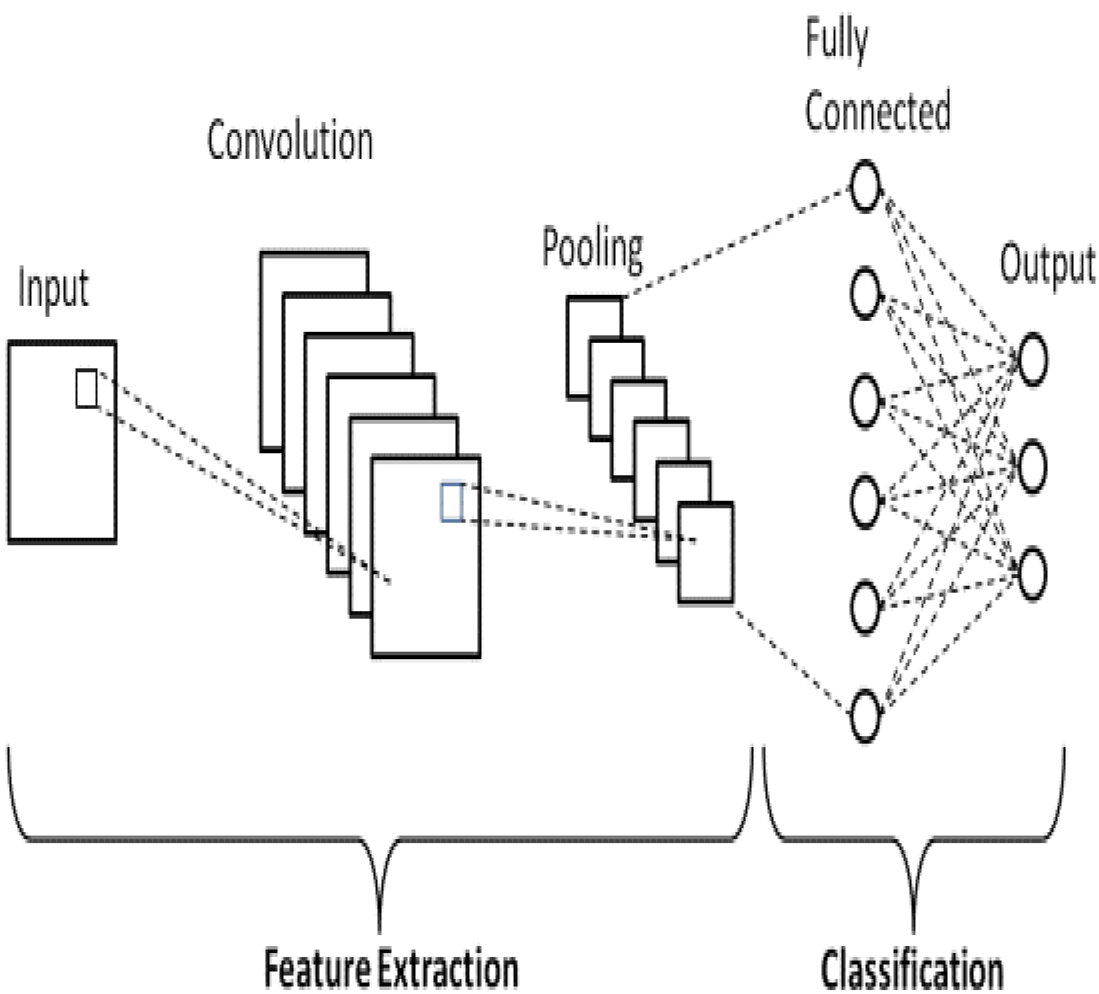


Figure 5: Convolutional Neural Networks Architecture

 ©Daffodil International University

Traditional artificial neural networks gave way to CNNs, which use a three-dimensional neural pattern modeled after the animal visual cortex. Although its major uses are in object identification, image classification, and recommendation systems, convolutional neural networks are commonly employed for natural language processing. Convolutional neural networks employ images as input and use those images to train a classifier. The network employs a unique mathematical technique termed "convolution" in place of matrix multiplication. Convolution, pooling, activation, and fully connected are the four layers that typically make up a convolutional network.

3.5 CUSTOM CONVOLUTIONAL NEURAL NETWORK(CNN)

The detection aim was completed using Custom CNN, which means that in this instance I used Custom CNN to identify physical abuse occurring in a daycare facility. So, what makes CNN and Custom CNN different from one another? The Custom CNN model differs in that it can modify its layers in accordance with its requirements. Custom CNN will contain more layers than CNN, and there will be a hidden layer difference between the two. This change benefits my entire procedure because the pre-processing phase of the process was handled and tailored using Custom CNN at the outset. The following portions are entirely built in a customized manner, which means the conventional CNN layers are adjusted in accordance with the detection requirements.

The nicest thing about Custom CNN's architecture is that feature extraction is not necessary, which is another major benefit. The fundamental idea of Custom CNN is that it generates invariant features by convolution images and filters, which are then passed on to the next layer as the system learns to do feature extraction. The features in the following layer are combined with various filters to produce more invariant and abstract features, and the process is repeated until the final feature is produced that is invariant to occlusions.

 ©Daffodil International University

Another crucial aspect of my dataset, which consists of two distinct datasets, one with images of violence and the other with images of non-violence, is how adaptable and effective deep convolutional networks are at processing visual data. So, with a custom CNN, my datasets are incredibly versatile. Since regions are made up of contiguous blocks of pixels, this type of convolutional layer takes advantage of the fact that an intriguing pattern can appear anywhere in the image.

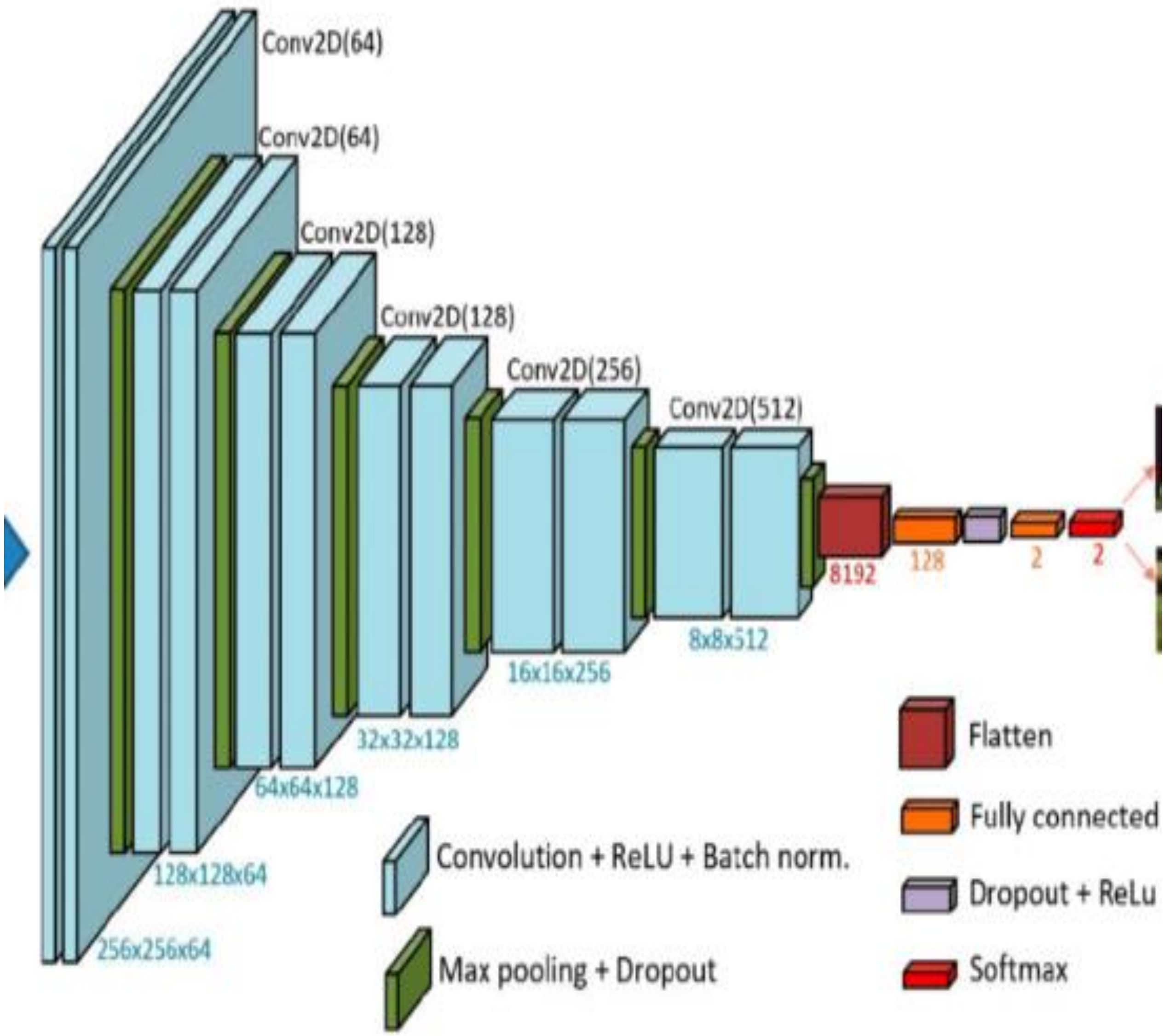


Figure 6: Custom Convolutional Neural Networks Architecture

©Daffodil International University

However, the capability of the model to learn meaningful features from raw data, and the fact that my dataset is recorded with complete raw sets of data, is one of the main arguments for selecting this Custom CNN model because of this deep learning. Both sorts of images can readily be recovered, and the need for conventional human image processing techniques can be avoided thanks to the ability of custom convolutional neural networks to extract relevant features from photos.

As a result, I can employ a variety of hidden layers for my detection purposes and modify the layers as needed for my project to produce acceptable detection accuracy. To identify the violence, I employed a six-layer Custom CNN system. The layers were completely customized to meet the requirements of my project. In order to achieve the goal of detection, the additional hidden layers were intended to create good accuracy. This modified CNN model generates some weights that represent the results of the test data set that I have used. In order to provide good detection with high accuracy, this customization was chosen because it contained more hidden layers.

3.5 Evaluation Methods

I plotted the confusion matrix to evaluate the results. Evaluation requires both true positive and negative results as well as false positive and negative values. This is a real plus because the actual amount was as predicted. rightfully rejecting a true negative Although a positive number is predicted, the value is actually a false positive. False-negative is incorrectly rejected. This methods help to understand the overall perspective of the datasets and it easily describes the actual numbers of images prediction. Without these methods the real prediction can not be measure in a health way where the prediction result is very important to understand the model. The true positive values and true negative values can not describe the whole datasets prediction , the false positive and false negative describes the whole assessments of the dataset prediction. The actual real prediction is measured by the whole 4 aspects of this negative and positive

©Daffodil International University

3.5.1 Accuracy:

The accuracy of the model determines how well a computer can predict outcomes. Something is substantial when each class is equally important. For me, every class is equally important to my work. In order to assess the model's accuracy, finding an excellent accuracy is essential. For binary classification, accuracy can also be assessed in terms of positives and negatives, as shown below:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Where TP, TN, FP, and FN stand for "True Positives," "True Negatives," "False Positives," and "True Negatives," respectively. This section also represents the total accuracy of any models and basically for Custom CNN this type of accuracy matters a lot. The outcome totally depends on these types of accuracy.

3.5.2 Precision:

Precision is the amount of information that a number can express through its digits; it indicates how closely two or more measures are spaced from one another. Basically precision is a measure of how much detailed information is given. It does not depend on accuracy. Precision is used to determine how many predictions from the positive class really fall within it. To calculate accuracy, the true positive value is removed from the overall positive value. The definition of precision is as follows:

©Daffodil International University

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Where TP and FP stand for "True Positives," and "False Positives," respectively. This equation basically shows that the true positives are divided by the sum of true positives and false positives.

3.5.3 Recall:

Recall serves as a measurement tool for determining precise genuine positive identification. Basically, the recall is computed as a percentage of the total number of Positive samples to the number of Positive samples that were accurately identified as Positive. Recall evaluates the model's capability to identify Positive samples. The recall is calculated by dividing the true positive value by the total number of pertinent documents that are currently in existence. The definition of recall is as follows:

$$Recall = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)}$$

Where TP and FN stand for "True Positives," and "False Negatives," respectively. This equation basically shows that the true positives are divided by the sum of true positives and false positives negatives.

 ©Daffodil International University

3.5.4 F1 Score:

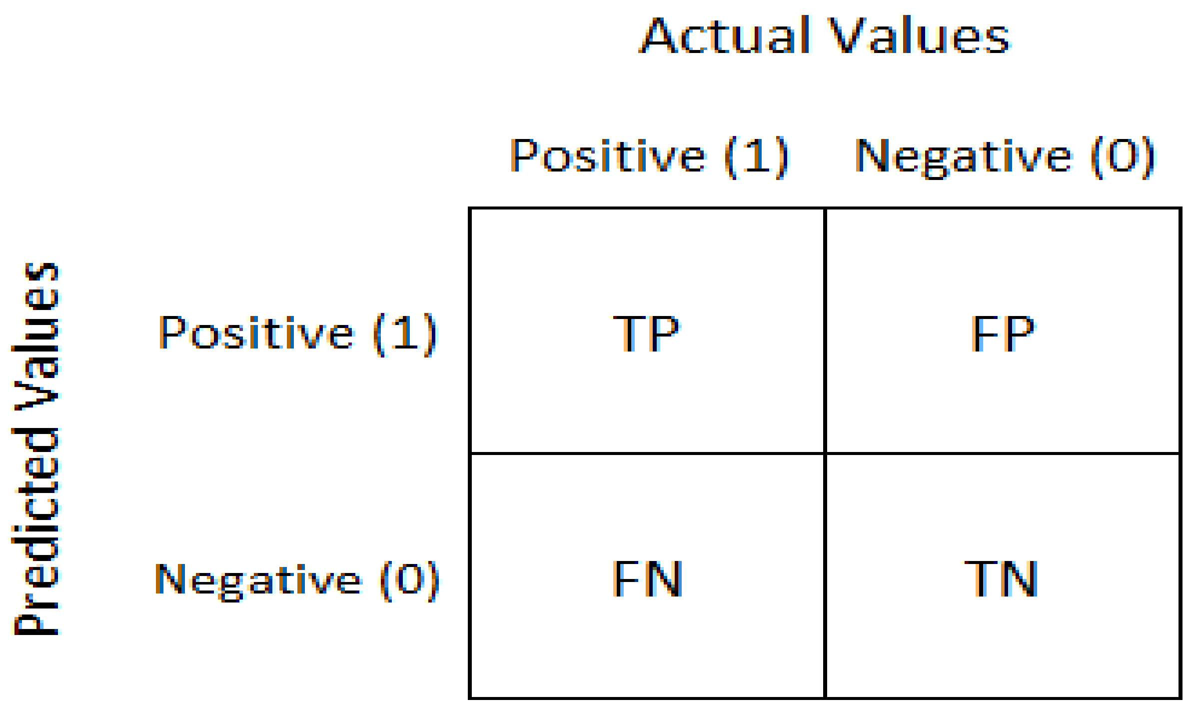
The F1 Score is calculated by combining precision and recall. Therefore, this score takes into account both false positives and false negatives. Despite being less intuitively straightforward to understand, F1 is frequently more advantageous than accuracy, especially if you have an uneven class distribution. For accuracy to work best, false positives and false negatives should approximately cost the same. If the costs of false positives and false negatives differ significantly, it is desirable to take into consideration both Precision and Recall. The definition of an F1 score is as follows:

$$F1\ score = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

3.5.5 Confusion Matrix:

A method for summarizing a classification algorithm's performance is the confusion matrix. If your dataset has more than two classes or if each class has an unequal number of observations, classification accuracy alone may be deceiving. You can acquire a better understanding of the categorization model's successes and failures by calculating a confusion matrix . Looking at the confusion matrix is a much better technique to assess a classifier's performance. The main goal is to determine how frequently examples of class A are put in the class B category. I need the following four properties to make a confusion matrix

©Daffodil International University



- True Positives (TP): matches when the model accurately anticipated the label and the ground truth.
- True Negatives (TN): The label is not predicted by the model, and it is not based on the truth.
- False Positives (FP): When a label is predicted by a model but is not really present in the data (Type I Error).
- False Negatives (FN) are labels that the model does not anticipate, but which are nevertheless true in reality. Error type II.

©Daffodil International University

CHAPTER 4

RESULTS AND DISCUSSION

4.1 INTRODUCTION

The models and procedures for detecting violence using CCTV cameras from Daycare so that no child got abused. This portion basically targeting the physical violence that happened in the day care center. After the data gathering and preparation phase, I detailed the model I used for the detection approach. The outcomes of the model I used will be discussed here.

4.2 RESULT DISCUSSION:

After applying the model, we obtain a large number of findings from our dataset. For the purpose of identifying physical assaults in daycare facilities, we implemented Custom CNN. We determine the model's weights for our dataset, F1 score, precision, and recall value using Custom CNN. We also discover the loss plotting graph for the model. The model's various performance and accuracy characteristics are defined by these weights. This further clarifies the model's entire methodology for identifying violence from beginning to end. This also indicates the possibilities of correct prediction and wrong prediction that this model has generated. Various characteristics of the dataset for prediction are also shown in this model. Any model's true state may be determined by looking at these facts, and for my model, they are crucial. The accuracy and performance of any model are based on these discoveries.

 ©Daffodil International University

4.3 CUSTOM CNN:

The custom CNN I created for my project yielded promising results. The model was run for 20 epochs with the Adam optimizer and loss function binary-cross entropy and accuracy as metrics. The training accuracy was 0.9614 and the training loss was 0.0080. Here, it is demonstrated that the training loss was 0.6254 when the val and train started the epochs, and the training accuracy reached 0.0080 when the entire epoch was completed, with a completely decreasing trajectory. However, in the beginning of the accuracy phase, training accuracy was at 0.269 and the curve was completely downward. The training graph was at the top, indicating upward movement, when the epoch was complete.

Figure 7: Loss and Accuracy graphs

©Daffodil International University

The trained model was applied to test data and there are results are generated. There are 3types of result output we can see here (Figure 8)and the outputs are for precision, recall and f1-score.This results are for both types or classes that means violence and non-violence. Here for non-violence, we can see that it is denoted by 0 and for violence it is denoted by 1. The results are for the precision for non-violence was 0.81 and for violence 0.87. The recall for non-violence was 0.86 and for violence 0.83. The f1-score showed for non-violence was 0.83 and for violence 0.85. Overall, the accuracy for the model was 0.84. This describes that the whole data set prediction can be shown in a way that called Custom CNN results where each and every output for each methods are shown. The accuracy indicates that this model is quite good prediction rate for this raw set off data.

```
              precision    recall  f1-score   support

           0       0.81      0.86      0.83       132
           1       0.87      0.83      0.85       155

    accuracy                           0.84       287
   macro avg       0.84      0.84      0.84       287
weighted avg       0.84      0.84      0.84       287

[[113  19]
 [ 27 128]]
```

Figure8: Custom CNN Results

 ©Daffodil International University

From the Confusion Matrix, it is observed that the model could detect several non-violence and violence data correctly out of the test dataset. A predicted value label and an actual value label are both addressed here. The confusion matrix reveals that the model accurately detected 128 instances of violence in addition to 113 instances of non-violence. However, this matrix also made clear to us that the model is unable to accurately detect some inputs. This matrix showed us that there are 19 non-violence data that this model is unable to detect and that this model incorrectly predicted as a violence data. It also showed us that there are 27 violence data that this model is unable to detect and that this model incorrectly predicted as a non-violence data. The whole dataset detection is shown in this matrix.

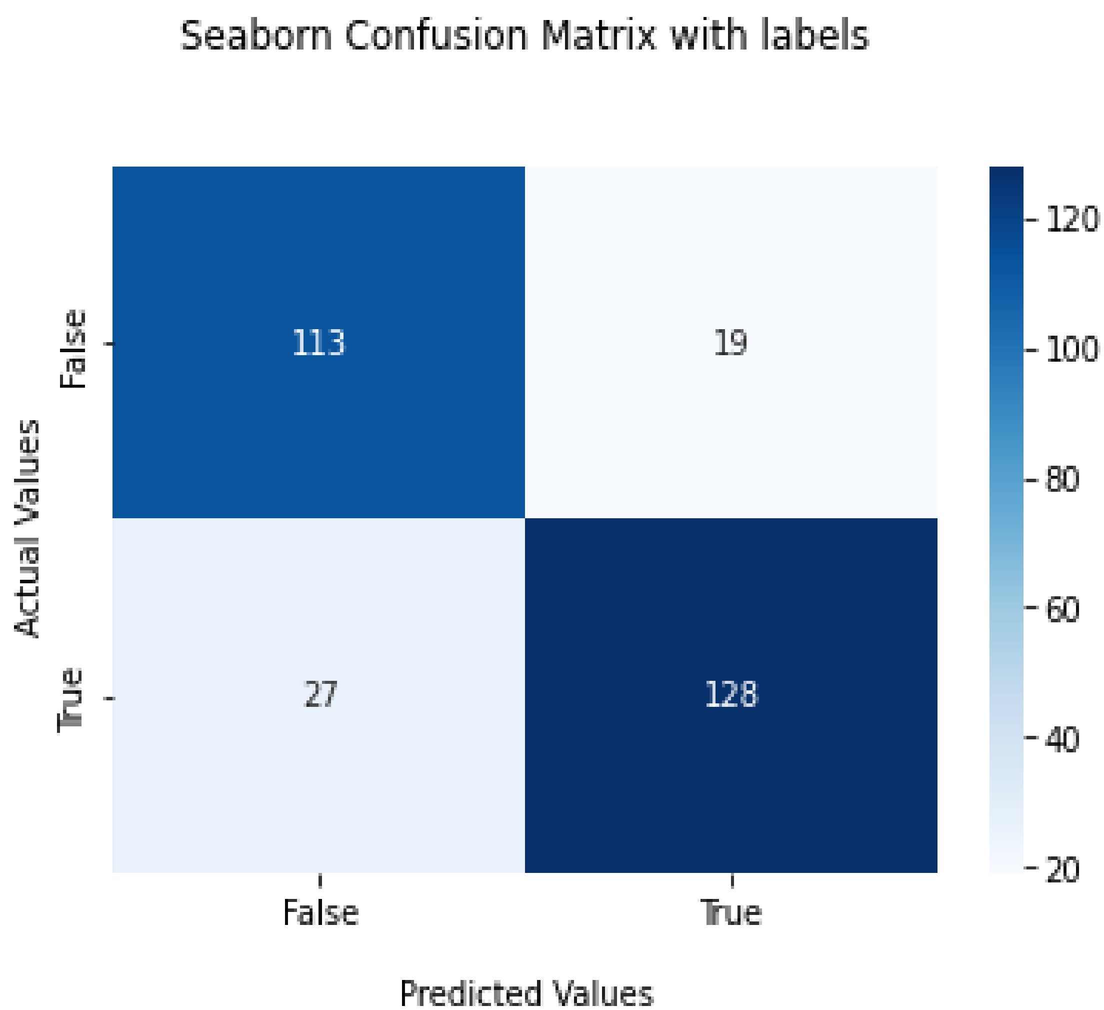


Figure9: Confusion Matrix of Custom CNN

©Daffodil International University

We can see this kind of output after training our data. The discovered images are shown in this output. The violence and non-violence actions are both easily detected by this model, although occasionally the predicted and real values do not appear to be equal.

Figure 10: Predicted label vs True label

 ©Daffodil International University

There are some events that it incorrectly predicts, but the quantity of these errors is quite small, and the model has been successfully implemented as seen by the outcomes. If these output does not show in a big picture, then the whole prediction cannot be seen. There are some events where we can see that the predictable output and actual or true output are not same that indicates the lack of perfect model implementation. Where this implementation shows also that the raw dataset works quite good in this section. So, this is the whole scenario or output of this model where above 80% percent data are predicted correctly and some predictions are shown the lack of model implementation.

 ©Daffodil International University

CHAPTER 5

CONCLUSION AND LIMITATIONS

5.1 CONCLUSION

The data presented above clearly demonstrate that our custom CNN is well suited for real-time processing for identifying child abuse and was effective in correctly categorizing and localizing all object types. The proposed end-to-end architecture has all the elements required to be automated and deployed for child abuse monitoring, and it has been successfully built. The libraries and software used in our project are highly flexible and affordable because they are all open-source. The main goal of the program was to address the problem of inadequate child safety management. We may thus draw the conclusion that, if adopted by any childcare organization, it would facilitate and enhance the efficiency of their job and guarantee kid safety at daycare.

5.2 LIMITATIONS AND FUTURE WORKS

In the work, the output is promising for future works. At first, I didn’t get enough data or videos to get the images and train the machine. And I communicated with so many daycare centers and they denied sharing the footage and there is no online dataset available to train my Custom CNN for massive data to get better accuracy. I don't have that many resources to run more epochs to get a better output. I had to use very bad quality images as I took maximum images from different low-quality CCTV footage available on the internet. Basically, the scarcity of datasets mainly limits our outcomes. We hope in near future we will get better-quality images to train the model and we will capable to use other pre-trained models to obtain very good outcomes. Then we will be able to use these techniques in CCTV to find real-time violence and we will be able to ensure safety in Day Care.

 ©Daffodil International University

Chapter 6

REFERENCES

1. Fatima Mahmood, Jehangir Arshad. “Implementation of an Intelligent Exam Supervision System Using Deep Learning Algorithms”

2. LIN, & Jhang. (n.d.). *Intelligent Traffic-Monitoring System Based on YOLO and Convolutional Fuzzy Neural Networks*.

3. Ngo, & Lin. (n.d.). *A Room Monitoring System Using Deep Learning and Perspective Correction Techniques*.

4. Chandan. (n.d.). *An Automated University Gate Pass Monitoring System Using Deep Learning*.

5. Nadia Zouba, François Bremond, Monique Thonnat. An Activity Monitoring System for Real Elderly at Home: Validation Study. 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance AVSS10, Aug 2010, Boston, United States. ffinria-00503864f

6. Muhammad Ramzan, Adnan Abid, Shahid Mahmood Awan. “Automatic Unusual Activities Recognition Using Deep Learning in Academia”

7. Fuerties, & Carbelleira. (2022, February 18). “*People detection with omnidirectional cameras using a spatial grid of deep learning foveatic classifiers*.” People Detection With Omnidirectional Cameras Using a Spatial Grid of Deep Learning Foveatic Classifiers - ScienceDirect.

 ©Daffodil International University

8. Scariaa, E., Abahai, A., & Isaacc, E. (2020, January 1). *Suspicious Activity Detection in Surveillance Video using Discriminative Deep Belief Network | Semantic Scholar*. Suspicious Activity Detection in Surveillance Video Using Discriminative Deep Belief Network | Semantic Scholar.

9. Alexandra König, Carlos Fernando Crispim-Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, et al.. Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients. Journal of Alzheimer's Disease, IOS Press, 2015, ff10.3233/JAD-141767ff. ffhal-01094093f

10. Okky Permatasari, Siti Ummi Masruroh, Arini, "A Prototype of Child Monitoring System using Motion and Authentication with Raspberry Pi", in 4 th International Conference on Cyber and IT Service Management, 2016, pp. 1-6

11. Fernand, Bathrinarayanan, & Fosty. (n.d.). *Evaluation of a Monitoring System for Event Recognition of Older People*.

12. Sundar, S., Ghosh, R., & Shahil, H. (2020, January 1). *A prototype of automated child monitoring system | Semantic Scholar*. [PDF] a Prototype of Automated Child Monitoring System | Semantic Scholar.

13. Reddy, & Vamsi. (n.d.). *An Automated Baby Monitoring System*.

14. Karageorge, & Kendall. (n.d.). *The Role of Professional Child Care Providers in Preventing and Responding to Child Abuse and Neglect*.

15. Pais , & Buluschek. (n.d.). *Evaluation of 1-Year in-Home Monitoring Technology by Home-Dwelling Older Adults, Family Caregivers, and Nurses*.

16. Jelwy, & mahdi. (n.d.). *Detection of Unusual Activity in Surveillance Video Scenes Based on Deep Learning Strategies*.

17. Dinga, & Fang. (n.d.). *A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory*.

18. Zejun Huang, Huijuan Lu et al., "An Mobile Safety Monitoring System for Children", in 10th International Conference on Mobile Ad-hoc and Sensor Networks, 2014, pp. 323-328.

19. Dhiraj Sunehra, Pottabathini Laxmi Priya, Ayesha Bano, "Children Location Monitoring on Google Maps using GPS and GSM Technologies", in IEEE 6th International Conference on Advanced Computing, 2016, pp. 711-715.

20. Damodaran Sanipath, Sankar Sundarani. "Telecommunication system for remote monitoring and access in child care" IET International Conference on Wireless, Mobile and Multimedia Networks, pp. 81-84, 2008.

 ©Daffodil International University