University of Stavanger

Faculty of Science and Technology

MASTER'S THESIS

Study program / Specialization: **Master's Thesis in Computer Science**	Spring semester, 2021 **Open**
Author: **Chukwudi Aboaja**	……………………… (Author's signature)

Program coordinator:

Professor Kjersti Engan

Supervisor(s):

Øyvind Meinich-Bache
Professor Kjersti Engan

Thesis title:

Generalized vs Specialized activity recognition system for newborn resuscitation videos using Deep Neural Networks

Credits (ECTS): **30**

Key words:
Deep Learning, Convolutional Neural Network, Object Detection, Activity Recognition, Image Processing, Newborn Resuscitation

Number of pages: 73

Stavanger, June 15th, 2021

Abstract

Birth asphyxia is a global problem which has resulted in a high mortality rate of newborn babies all over the globe, it is a newborn's inability to establish breathing at birth. A notable breakthrough is the marrying of medical technology with information technology in an attempt to tackle this global health problem. An example of this is the Safer Births project which is focused on establishing technological advancement to curb newborn deaths. In the year 2013, the Safer Births project started and has till date gathered a lot of data captured during resuscitation sessions. The Haydom data used for the Safer Births project and additional data from Nepal and SUS will be used with the aim of comparing a specialized and generalized model trained on activity recognition system I3D and RGB stream excluding optical flow. With focus on only the newborn region, the reason for this is to simplify the existing model. The experiment was conducted in view of the possibility of achieving a system that can generalize or specialize with a combination of different hospital data on some specific activities of interest namely Ventilation, Suction, Stimulation. A new simplified pipeline, which is a reduction of the previous work done by the saferbirth group, showed a very poor performance when generalized.

Acknowledgements

I would like to thank the University of Stavanger, The Department of Electrical Engineering and Computer Science, my supervisors Professor Kjersti Engan and Øyvind Meinich-Bache for helping me realize this thesis.

I would also like to express my sincere gratitude to my family and friends; I'm really thankful of all the times spent together.

I would also like to a big thank you to Theodor Ivesdal for helping with remote access to the Unix server at the University of Stavanger and quick response with technical issues.

Contents

List of Figures

List of Tables

List of Abbreviations

SUS	**S**tavanger **U**niversity **S**ykehus
ANN	**A**rtifical **N**eural **N**etwork
CNN	**C**onvolutional **N**eural **N**etwork
HCP	**H**ealth **C**are **P**roviders
WHO	**W**orld **H**ealth **O**rganiztion
BA	**B**irth **A**sphyzia
ADNB	**A**pparent **D**eath **O**f **N**ewborn
NN	**N**eural **N**etwork
ReLU	**R**ectified **L**inear **U**nit
AR	**A**ctivity **R**ecognition
LFI	**L**inear **F**rame Interpolation
RGB	**R**ed **G**reen **B**lue
TNR	**T**rue **N**egative **R**ate
FP	**F**alse **P**ositives
TN	**T**rue **N**egatives
FN	**F**alse **N**egatives
TN	**T**rue **P**ositives
HBB	**H**elping **B**abies **B**reathe
FPS	**F**rame **P**er **S**econds

Chapter 1

Introduction

1.1 Introduction

Birth asphyxia is the inability of newborn babies to initiate and sustain breathing at birth,as a result of this condition approximately four million babies die prematurely every year globally[5]. This health concerns has led to the need for effective resuscitation methods which can prevent a great number of post natal deaths.

More so, the World Health Organization(WHO) considers neonatal birth asphyxia a birth complication which is an health challenge for newly born babies worldwide. Therefore, the implementation of resuscitation guidelines by the World Health Organization was established to help reduce the risk of high mortality rates in newborn globally especially in economically challenged countries where this birth complication is at an increased rate [47].

WHO concluded that the Sub-Saharan Africa had the highest neonatal mortality rate in the year 2019 with 27 deaths per 1,000 live births [61], this was seconded by Central and Southern Asia with an estimate of 24 deaths per 1,000 live births. For instance, most newborn babies in low income countries in Asia or in Sub Saharan Africa are more likely to die in the first month than a children born in a high-income country developed countries. [27]. Most studies on the birth asphyxia shows that 99% of neonatal deaths takes place more in low and middle income countries [27]. Based on this observation, the need for specialized measures and methods for newborn resuscitation increases especially in these aforementioned continents.

Also, various surveys by the World Health Organization reflects the occurrence of birth asphyxia and its mismanagement is more prominent in low resource countries than developed countries, thus the importance of more specialized methods of resuscitation with consideration of both environmental and socio-economic factors [57]. Due to the evaluation of these statistics, it is remarkable that the video data collected for this project represents high-income countries and low-income countries.

This research aims to look into different resuscitation activities conducted by health care providers in the neonatal care unit of three hospitals, with the aim of impacting on medical technology which will help in recognizing the most effective resuscitation activity this will influence how neonatal care experts tackle every birth asphyxia complication and prevent BA related deaths on a global scale. Due to essential new born care which has been implemented in every neonatal care unit worldwide, there has been a decrease in neonatal deaths from five million in 1990 to 2.4 million in 2019 [67].

This reflects much progress perhaps due to the inclusion of artificial intelligence in the field of medicine. Resuscitation activities like Ventilation , Stimulation, suction and others are extracted from these videos and are further analysed using machine learning models and statistical methods. This plays a key role in medical technology, because, it can be used to recognize and detect inefficiency in manual labour and offer information of more effective resuscitation activities which could deliver better results.Furthermore, machine learning and activity recognition systems will be used to recognize whether the midwives and doctors are following the resuscitation guidelines in the resuscitation sessions.

1.2 Earlier projects

Earlier research has been conducted to tackle and manage complications arising from birth asphyxia. A notable project established to ensure that babies born with issues of asphyxia are resuscitated with the most effective procedure with the outcome of healthy and alive newborn babies is the Safer Births project, established in 2013 [32, 39] . The Safer Births project was enabled by the collection of newborn resuscitation video recordings from the Haydom Lutheran Hospital in Tanzania [39, 40]. This existing data from Tanzania mentioned above will be two of the three data to be used for this master project.

1.3 Aims and Objective

This thesis aims to simplify the already existing system and investigate whether analysing only the newborn area specifically the resuscitation table with the available data which is more than the one used in the previous work would be enough. A notable problem found in the previous work was how computational heavy and technically complex analyzing all other object regions was. Hence, to avoid repetition and achieve the possibility of a different outcome, there would be no need for optical flow as only use RGB is utilized.

Also, to achieve this aim, we compare generalized and specialized models using newborn resuscitation data from Haydom Tanzania, Nepal, and SUS Norway.

Adopting the working proposed solution from the previous work by Meinich-Bache et al., generalizing the solution by using more data from different hospitals instead of one hospital would be investigated. The available data which is more than the previous data would enable better activity recognition especially in premise of suction activity using device.

1.4 Related Works

Preceding studies on Safer Births was conducted by Øyvind Meinich-Bache, this was done by analyzing the new born resuscitation videos using an object detection system to discover important objects in the resuscitation videos and activity recognition to trace and identify sections of the video where resuscitation activities are executed [39].

RetinaNet an object detection model was used to locate objects in the resuscitation videos, the model locates object through predictions.Furthermore, Meinich-Bache used Inception I3D to observe which resuscitation activity is being performed. By using these models Meinich-Bache attempts to identify which resuscitation activity is most effective in the Safer Births Project.

1.5 Thesis Outline

- Chapter 2: Medical background

  This chapter consists of the medical background of newborn resuscitation, a step by step description of process resuscitation activities.

- Chapter 3: Technical background

  This chapter describes the technical knowledge on which the methods applied in the project is based on.

- Chapter 4: Data And Materials

  This chapter enumerates and explains the number of data and its composition which was used,the sources and distribution.

- Chapter 5: Methodology

  This Chapter explains the methods proposed to solve thesis objective.

- Chapter 6: Experiment and Results

  This Chapter explains how different experiments were performed to achieve thesis objective, and presented results.

- Chapter 7: Discussion and Conclusion

  This chapter discusses the results from the methods implemented for the purpose of this thesis, a conclusion and future works.

Chapter 2

Medical Background

This chapter delves into the medical background studied for the thesis.

2.1 Birth Asphyxia

Birth asphyxia (BA)is a worldwide burden that requires, as mentioned earlier, immediate measures to minimize the risk of sudden infant deaths. In order to help minimize these deaths, there is a hasty need for well-equipped, trained health care providers and provision of all thermal care facilities, especially in low-resource countries.

Different resuscitation methods have been put in place by the WHO experts and medical specialists in the field of neonatal care for the purpose of safer births. Resuscitation activities are ventilation, suction and stimulation. The accepted resuscitation activities are regulated with guidelines that every Midwife, doctor, or healthcare professional in the field of pediatric and neonatal care must adhere to. Ventilation is a process of neonatal resuscitation whereby the newborn suffering from asphyxia is given breathing assistance. A ventilation procedure entails the use of a bag and mask placed on the face of the newborn and secured with a seal to avoid a lack of oxygenation. It is a necessary procedure that, if implemented efficiently, can increase oxygen levels in the newborn. And its outcome can be a successful resuscitation of an asphyxiated newborn. Suction is implemented for the sole purpose of clearing the airways. The process of opening the airways involves the suctioning of the mouth and the nose. If there is an observation of normal breathing in the newborn, it means the resuscitation method worked. Therefore there is no need for further resuscitation. If this procedure is not successful, ventilation or stimulation can be implemented immediately by health care providers.

A century before the discovery, a newborn's inability to breathe now known as birth asphyxia, was diagnosed by many physicians and midwives as 'apparent death of the newborn. it was in many cases misdiagnosed as cerebral palsy. Until the 19^{th} century, precisely in the year 1992, the name 'Apparent death of the newborn' or stillbirth was replaced with what we now know as birth asphyxia [46].

The apparent death of the new born phenomenon was paradoxical because not only was it associated with cerebral palsy, and lack of oxygen but was believed to be as a result of incompetency and inefficiency on the part of the doctors or midwives at the time.

In addition, there was a level of acceptance of the phenomenon as unpreventable in the sixteenth century and early seventeenth century. However, late seventeenth century the period oxygen was discovered. Oxygen was introduced as one of the earlier neonatal resuscitation measures (artificial oxygenation), until the nineteenth century due to research and observations (BA) not only replaced (ADNB)[1] more preventive measures were carried out on newborn with this condition [46].

Although, BA and the asphyxiated newborn high mortality rates were preventable, some of the resuscitation methods implemented before the nineteenth century were less effective and detrimental to a successful resuscitation of the newborn. The founding of the World Health Organization in the year 1948, introduced

[1] Apparent death of Newborn

regulated guidelines for newborn resuscitation with this initiative, updates guidelines were created and distributed globally. For example, the document "Basic newborn resuscitation;a practical guideline" was developed in 2009 by the WHO. This document contained updated guidelines on how to resuscitate asphyxiated neonates[47] . This helped regulate newborn resuscitation activities implemented in every neonatal care unit globally. In addition, the nineteenth century ushered in inventions of newborn resuscitation devices, which has been upgraded over the years and has helped prevent a great number of neonatal deaths [56]. Merging of medical technology and information technology has contributed hugely to fight the early deaths of asphyxiated newborns. For instance, creating and training models to recognize the most effective resuscitation activity for newborn resuscitation is a technological advancement which has had an impact on the survival of asphyxiated neonates in recent years [63]. Activity recognition systems trained to recognize newborn resuscitation activities and signal processing for newborn resuscitation are good examples of the impact of Information technology on the survival of asphyxiated neonates [65].

FIGURE 2.1: Newborn with asphyxic deficits.*Image collected from [21]*

2.1.1 Risk Factors of Birth Asphyxia

There are risk factor which needs to be considered when tackling birth asphyxia. These risk factors can be intrapartum and fetal. This occurrence is usually before birth. Due to lack of oxygen, vital nutrients needed for normal growth cannot get to the baby. It can also lead to organ damage as it would have a tremendous effect on the nerves. Risk groups are premature babies and babies with low birth weight[5].

2.1.2 Medical Procedures

This refers to preventive measures, that is, treatments used to tackle and prevent the loss of asphyxiated neonates. For example, the various resuscitation activities performed and how these activities are performed in accordance with the accepted resuscitation guidelines to ensure safer births. Asphyxiated new born babies worldwide require neonatal care doctors and midwives who adhere to the basic resuscitation guidelines provided by the WHO. These breathing interventions are important for a successful resuscitation [47].

2.1.3 Breathing Intervention

Newborn who does not breathe at birth needs immediate help in other to establish breathing. One minute after birth which is tagged the "Golden Minute" a newborn should be breathing well or has to undergo a ventilation process. There are specific steps which can be seen on figure 2.2 which can be undergone to help babies breathe(HBB) at birth. These procedures where put into use in some hospitals in Tanzania and it showed a good reduction of early neonatal deaths within 24hour period [45].

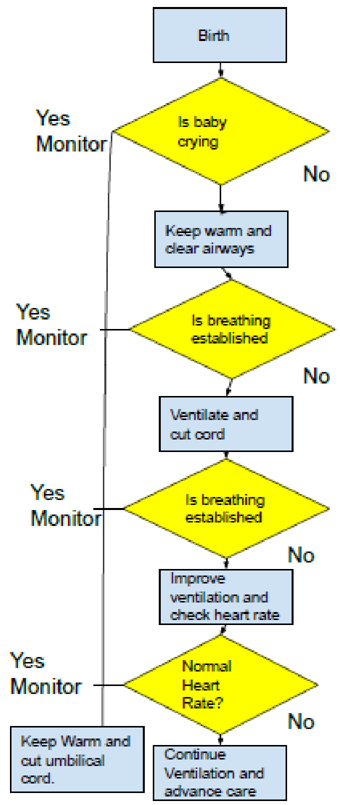

FIGURE 2.2: Flow Chart guidelines on helping babies breathe [32, 31]

Ventilation

Ventilation is a process in Neonatal resuscitation whereby an asphyxiated newborn is given a breathing assistance. Ventilation takes place when the bag and mask is placed on the face of an asphyxiated newborn and secured with a seal to avoid lack of oxygenation [47].

FIGURE 2.3: Ventilation performed on Newborn during resuscitation

Suction

Suction on the other hand, is implemented for the sole purpose of clearing the airways. Clearing of the airways includes suctioning the nose and mouth. If there is an observation of normal breathing in the newborn, it means the resuscitation method worked therefore there is no need for further resuscitation. But if there seems to be no changes, then further resuscitation activities such as ventilation or stimulation can be implemented immediately. The process can be performed with a suction tube or a suction penguin as it differs from different health care providers. See Figure 2.4 and 2.5 for a better illustration.

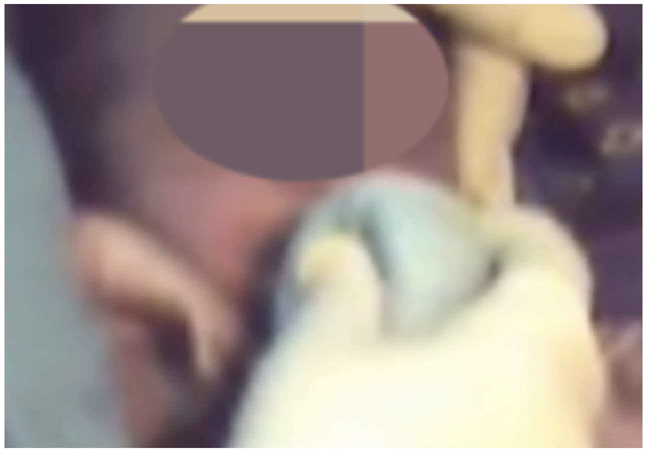

Figure 2.4: Suction performed on Newborn during resuscitation using suction device

Figure 2.5: Suction performed on Newborn during resuscitation using suction tube

Stimulation

Stimulation or gentle tactile stimulation is used in cases of perinatal asphyxia and is categorized as early sensory motor stimulation, this is done for the purpose of resuscitation and neurodevelopment of asphyxiated newborn infants [25] . Early stimulation steps are drying the baby, providing warmth and lastly stimulation, stimulation is effective for non breathing infants with primary apnea but it is less effective in very severe cases of infant asphyxia [44].

Figure 2.6: Stimulation performed on Newborn during resuscitation

Chapter 3

Technical background

3.1 Artificial Neural Networks (ANN)

The discipline Artificial Neural Networks is rapidly-evolving field in Information Technology and in particular, Artificial Intelligence and Machine Learning [22]. Artificial neural network, ANN for short, is a simplified version of the human brain mathematically modeled [71]. It is a make-up of large number of neurons. They are systems that readily adjust to their design to fit into its objective [23]. These networks are composed of nodes and their interconnections. Biologically, the human brain is a configuration of a biological neural network which contains approximately 100 billion neurons [41]. Each neuron executes its functions utilizing extensive connections to each other, called synapses.

3.1.1 How artificial neural networks functions

The Artificial Neural Networks are interpreted as weighted directed graphs. In this case, "the nodes are set up by the artificial neurons and the connection between the neuron outputs, neuron inputs are possibly defined by the directed edges with weights" [52].

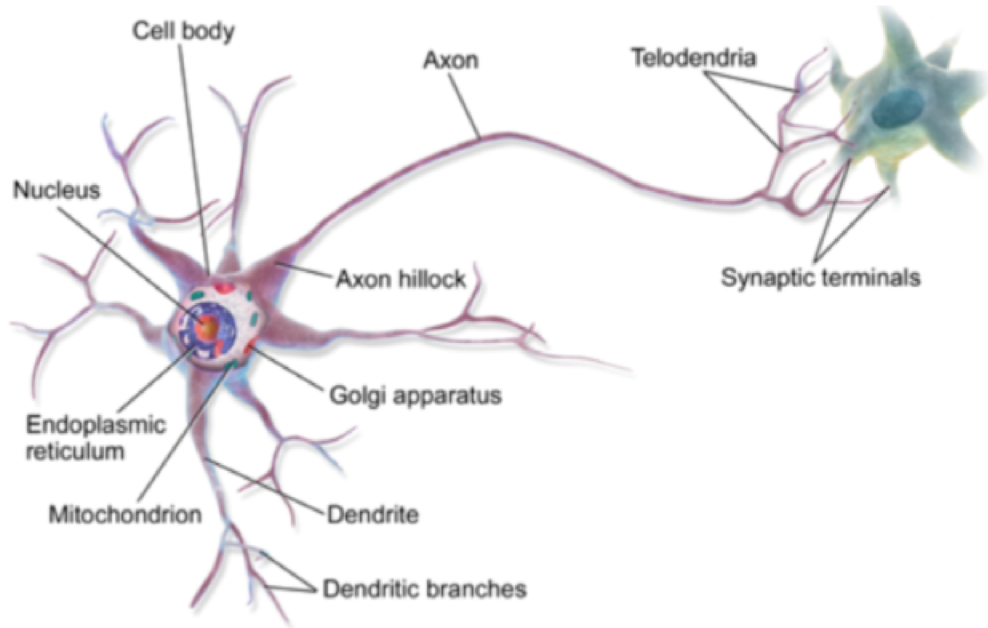

FIGURE 3.1: Structure of A Biological Neuron [37]. *This image is used under the licence of Creative licence Attribution 3.0 Unported CC BY 3.0*

This network composed of an input layer, hidden layer and an output layer. The input layer is multiplied with its weights , these weights are specific based on different applications. An ANN node applies a function to the weighted sum of its inputs [70].

3.1.2 Basic Concepts of Neural Networks

According to Knut Hinkelmann [26] in his lecture, named the basic concepts of artificial networks as follows:

1. **Input Nodes (input layer):** Data is feed through at this layer without any form of computation.

2. **Output Nodes (output layer):** This layer is the final layer of the network that outputs the final outputs by the use of activation functions.

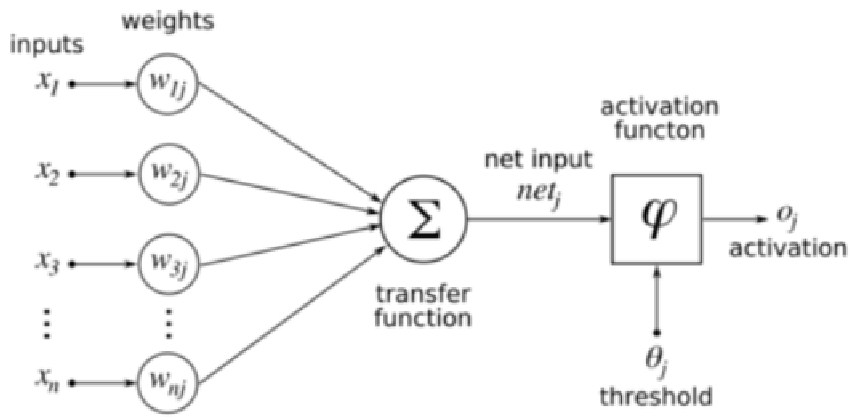

FIGURE 3.2: Structure of An Artificial Neural Network [15]

3. **Hidden nodes (hidden layer):** Hidden layer takes the weighted inputs from the first layer and undergoes a nonlinear map using an activation function.

4. **Connections and weights:** A neural network composed of a network of interconnected layers depicting a flow of information from one network neuron to another [14].

5. **Activation function:** The activation function decides the output of a node when given an input or set of inputs. On the other hand a non linear activation function allow such networks to evaluate nontrivial problems with the use of smaller number of nodes.

6. **Learning rule:** Its work is to adjust the thresholds and weights of a NN with intent to produce a desirable output.

3.1.3 Activation Functions

In artificial neural networks, the activation functions determine the output of a node given the input of a node. There are different types of activation functions depending on whether the node is in a hidden layer or output layer depending on the problem we are trying to solve—the activation functions in the hidden and output layer impact non-linearity to ANN.

3.1.4 Types of Activation Functions

1. Sigmoid: The sigmoid function plays an essential role in the context of logistic regression. It is a technique to predict the outcome of a binary classification problem. It serves as an activation function. It takes the weighted sum of the input features as an input and outputs the probability score of the outcome [42]. The sigmoid function by on plotted graph is an S-shaped curve. For any value of x, the function outputs a value between the rage 0 and 1 .

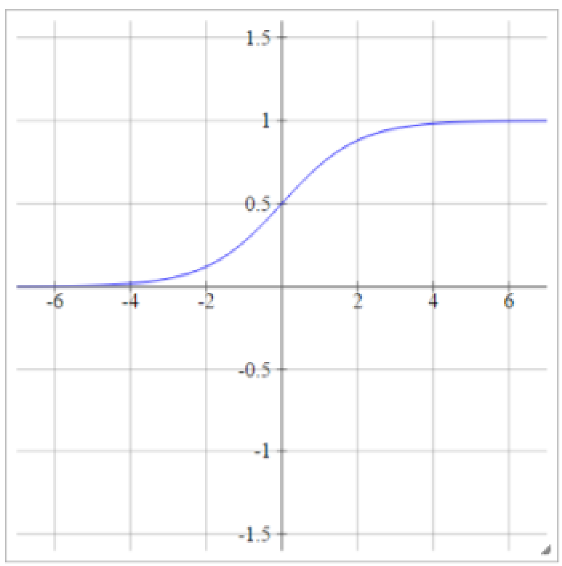

FIGURE 3.3: Graphical representation of Sigmoid Function [13]

Hence, the application of a sigmoid function in an ANN setting leads to what is called "Vanishing Gradient [7]" [49]. This equation can be shown in equation

[7]difficulties encountered while training certain Artificial Neural Networks with backpropagation

3.1 below.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3.1}$$

2. TanH: Hyperbolic functions contains a function of range between [-1,1]. It is very similar to the sigmoid function in looks. The pictorial representation of Figure 3.3 and the 3.4 clearly shows the difference between two function.Tanh is a scaled sigmoid function as it has proven to have a more steeper derivative when applied to ANN.

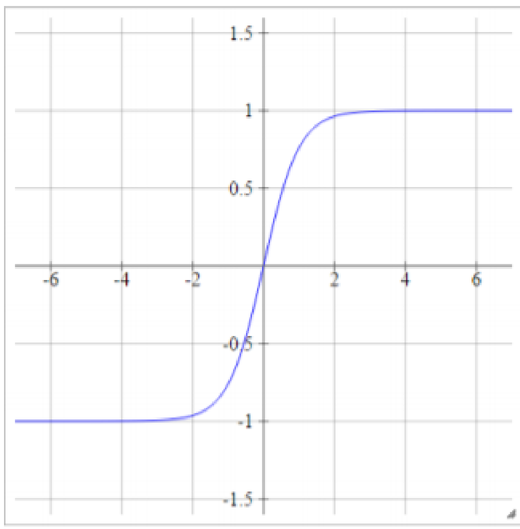

FIGURE 3.4: Graphical representation of TanH Function [13]

3. ReLu: In Artificial Neural Network, the rectified linear activation function, equation 3.2 denoted as ReLU has gained popularity over the years in the field of CNN [6, 8]. The function is represented with the equation below:

$$f(x){=}max(0,x) \tag{3.2}$$

Compared to other traditional existing activation function, the ReLu function speeds up computation during training. It also solves the problem of

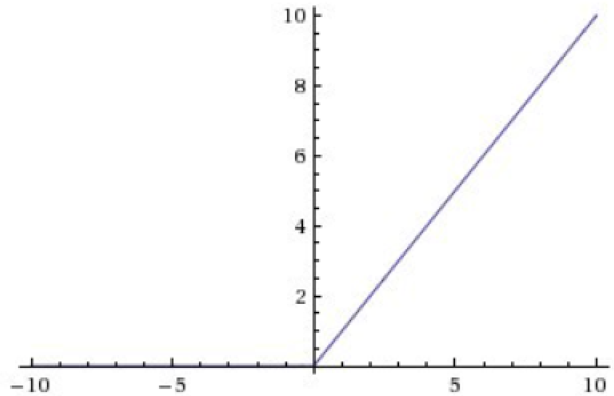

Figure 3.5: ReLU Activation Function [54]

vanishing gradient as layers of a network increases. This is because ReLu does not possess an asymptotic upper and lower bound [9].

4. Softmax: In the field of Machine Learning, the softmax function is often used to transform the output of the last layer of the neural network into probabilities. The number of softmax unit in output is always equal to the number of classes in such a way that each unit holds the probability of a class. This can also be called a probability distribution where the sum of possibilities for each class is equal to 1. In ANN, the softmax function is oftentimes combined as a set with cross-entropy. The softmax equation [20] is written as shown in figure 3.3 below.

$$softmax(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j))} \tag{3.3}$$

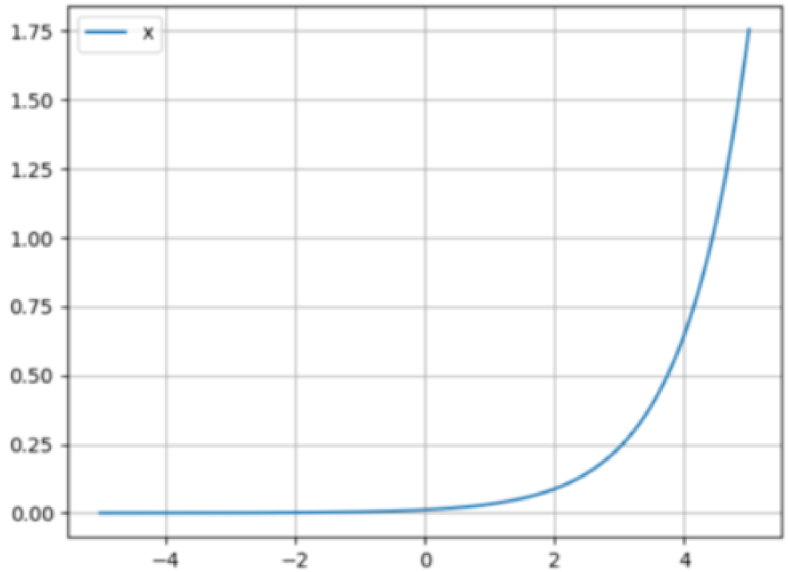

FIGURE 3.6: Softmax Function of a logit scenario [64]

3.2 Transfer Learning

Transfer Learning is a vast advancement in the field of deep learning whereby a machine learning problem trained on a particular task is applied as a starting point to solve a related task [66]. Transfer learning allows the utilization of pretrained models for the retraining of new datasets and tasks. In transfer learning, weights are loaded and reused to train new data.

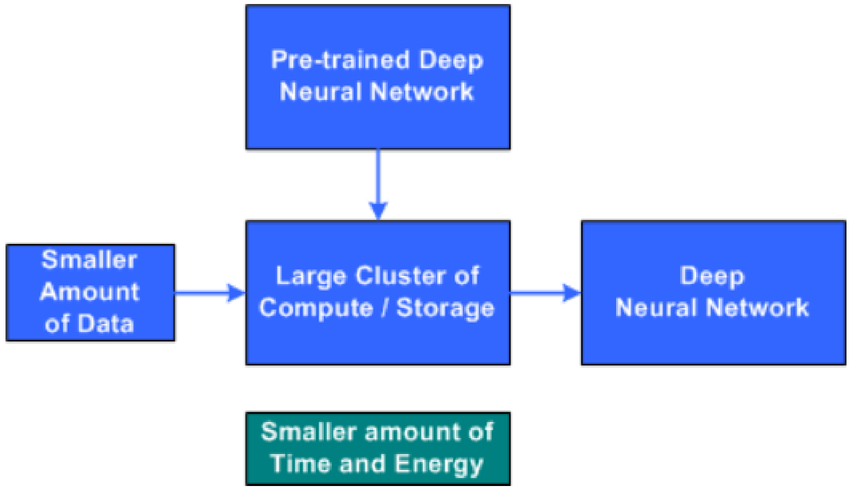

FIGURE 3.7: Transfer Learning with a pretrained Network [30]

3.3 Convolutional Neural Network

"Convolutional neural network (CNN) is a unique kind of multilayer neural network or deep learning architecture inspired by the visual system of living creatures" [18]. It takes an image as input, learnable weights and biases to different objects inside the image, and distinguish one from the other [28]. The basic idea that comes with CNN is that it makes use of filters [1] to filter an image before training the deep neural network. These filters assists in capturing the Spatial and Temporal dependencies in an image through the application. After these images are filtered, features within the images then come to the forefront, and then these features being spotted out would then be used to identify something.

[1] A filter, in this case, is simply a set of multipliers

3.3.1 Convolutional Layer

The convolution layer is the foundation of the CNN which is quite different from the traditional neural networks in the sense that it does not make use weighted sum and connected weights. This layer consists of filters or kernels which perform an operation called "convolution [2]". The feature map is retrieved by passing an image through the convolution filters.. After the feature map is generated, it is then processed by the activation function and results as an output produced by the layer also feature maps are simply a matrix representation of the of the original image[58].

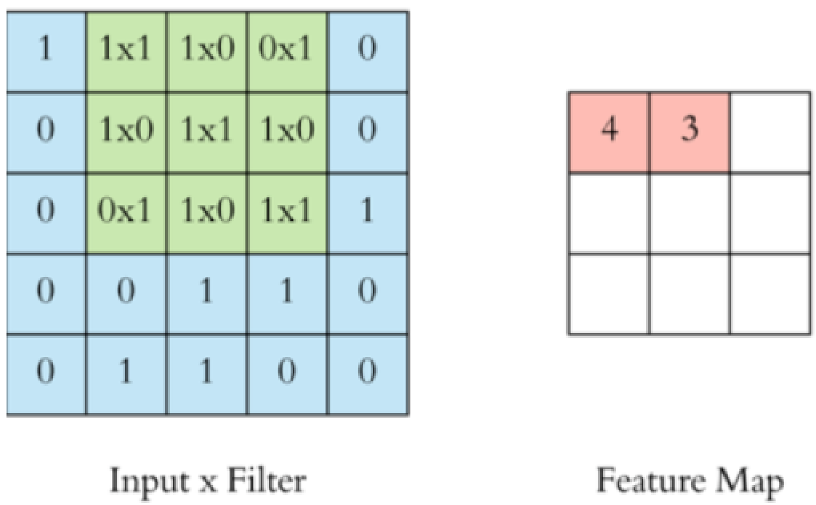

FIGURE 3.8: Convolution Operation. From the left side, filter slides over the input image.To the right: the obtained value is totalled and added to the feature map [11].

The convolution operation is done by taking a reverse kernel and moving it row by row, column by column matching the part of an image multiplying all of its elements one at a time by each other, and adding up the results. One crucial problem with the feature map is, they tend to be receptive to the area of the

[2]Convolution is a linear process of summing each element of the image to its local neighbors, weighted by the kernel [17]

features in the input. In other to solve this problem, there would be a need to downsample the feature maps. The downsampling operation is called pooling.

3.3.2 Max Pooling

Max pooling is an operation that is frequently added to a CNN following individual convolution layers [10]. When added to a model, max-pooling decrease the individual dimensionalities of images by simply decreasing the number of pixels. This can be clearly seen as a transformation of the original size of an image to its smallest. Max pooling works by defining an end to end region as a corresponding filter for the pooling operation followed by a defined stride [6]

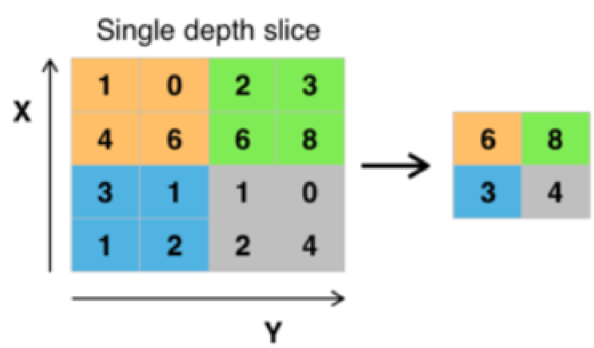

FIGURE 3.9: Max Pooling Operation on a 4x4 input feature map through a spatial window which occupies a 2x2 grid sliding across the image with a stride of 2 [4].*Figure is reprinted in an unaltered form from Wikimedia commons, File:Max pooling.png, by Aphex34, licensed under CCBY-SA 4.0*

3.4 Fully Connected layer

In an artificial neural network, fully Connected layers in a neural network are a set of layers where all the inputs from one layer are connected to all activation unit of

[6]Adjusts how many pixels we want our filter to move as it slides across the image

the following layer [33]. Most known machine learning models have adopted this pattern because the last few layers are fully connected layers, which compile the data extracted by previous layers to form the final output [69].

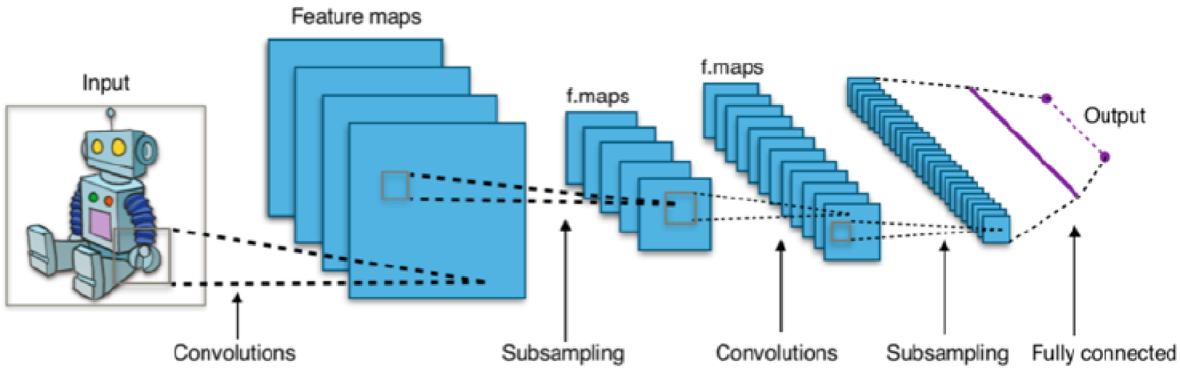

FIGURE 3.10: *Figure created by Aphex34 under Creative Commons Attribution-Share Alike 4.0 license International* [8]

A fully connected layer often learns features from all the combinations of the features of the previous layer, where a convolutional layer relies on local spatial coherence with a small receptive field. Consequently, training a fully connected network to "convergence" is not a relevant measure. The network continues to train and learn as long as the user is patient to stand by [48].

$$y_{i'} = \sum_{i} w_{ii'} x_i + b_{i'} \tag{3.4}$$

3.5 Supervised Learning

In machine learning, Supervised learning has been widely used. This type of learning whereby machines learn under guidance by feeding label data and explicitly telling them what the input is and what is expected as an output. Supervised learning is basically of two types: Classification [3] or Regression [4]. The main aim of supervised learning is to build an AI system that can largely learn the mapping relationship between the input and the output and predict the system's output given a set of new inputs. Fundamentally, a supervised learning algorithm can be mathematically represented as:

$$Y = fx \tag{3.5}$$

Y is the output gotten after prediction, which is resolved by a mapping function that allocates a class to an input value x.

3.5.1 Backpropagration

Backpropagation is a supervised learning algorithm used for training Neural Networks using a gradient. Backpropagation has its central idea to optimize parameters while training a neural network. The yield from a backpropagation neural network is calculated by a process identified as the forward pass [50, 2] [51]. Conceptually, backpropagation starts with the last parameter and works its way backward to estimate all other parameters with an artificial neural network and an error function. The gradient of the error function is calculated with reference to the neural network's weights.

Before backpropagation takes effect, feedforward is a vital process before backpropagation.In feedforward, an input is fed to the input layer, and its output is fed through each layer till a final output is retrieved. This final output is therefore compared with the network's expected outcome. With that comparison, an error is calculated through each output node. The process of the outputs being moved

[3] a classification algorithm gets data points with an assigned category [68]

[4] is a predictive statistical process where the connection between dependent and independent variables are determined by a model.

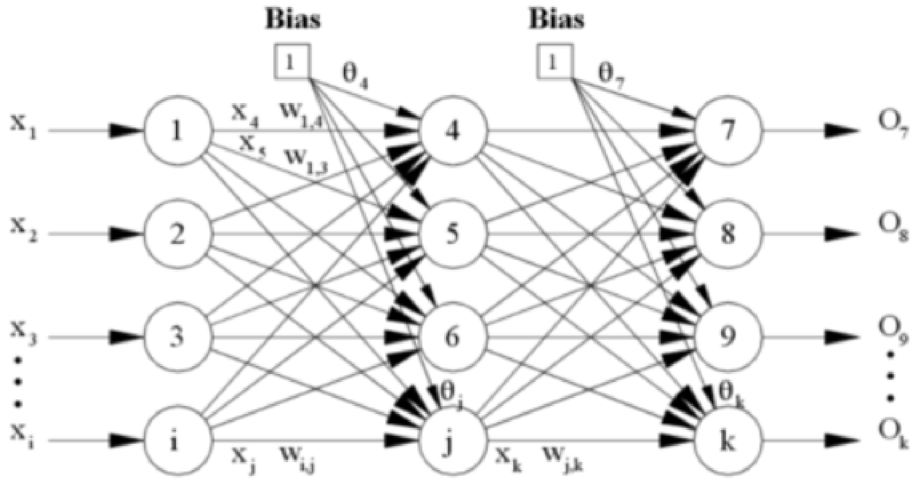

FIGURE 3.11: Backpropagation Neural Network with one hidden layer [36]

backward from each node's hidden layer is called backpropagation. Furthermore, this action is computed repeatedly until each node in the network has returned an error signal.

3.6 Loss Functions

The loss function is one of the useful components of the Neural network. It is simply the prediction error of a neural network. The calculation methodology of this loss is known as the Loss function. Loss is used in calculating gradients. With the obtained gradients the neural network weights are improved. [29].

3.6.1 L2 Loss

L2 Loss function stands for Least Square Errors [55]. They can also be denoted as LS. The L2 Loss is also commonly utilized for error reduction. Generally, the L2 Loss Function is preferred in most cases. But when the outliers [6] are present in

[6]an object that deviates significantly from the rest of the objects [24]

the dataset, it does not work properly. In CNN networks, L2-loss is generally used because it converges faster. The L2 loss can be mathematically shown in equation 3.6 below:

$$L_2 = \sum_{j=1}^{n} (target_j + predicted_j)^2 \tag{3.6}$$

3.6.2 Smooth L1 Loss

Smooth L1-loss behaves like an L1 loss when the absolute of the argument is high and specifically performs like L2 when the total value of the argument is close to null[16].

This function was adopted in the Fast R-CNN[5] [19] which is more potent than L1 loss function.

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \tag{3.7}$$

3.6.3 Focal Loss

Focal loss also denoted as (FL) is a loss function created by the Facebook Artificial Intelligence Research Group in the year 2017 [35]. In their paper [35] it is an improved version of Cross-Entropy loss (CE) that attempts to deal with class imbalance by assigning weights to slightly misclassified samples.

$$CE(p, y) = \begin{cases} -log(p), & \text{if } y = 1 \\ -log(1 - p), & \text{otherwise} \end{cases} \tag{3.8}$$

[5] stands for Fast Region-based Convolutional Network method for object detection

The above equation (3.8) is the cross-entropy loss for a given binary classification. Considering $y \in \pm 1$ which represents the ground-truth class and $p \in [0, 1]$ the model's estimated probability [62].

3.6.4 Cross Entropy Loss

This loss function, in comparison, compares each of the predicted probabilities to actual class output either 0,1. Afterward, the penalizing probability score are calculated based on distance from expected value. It is also used during training to adjust the weights of a model [1].

$$L_{CE} = \sum_{i=1}^{n}(t_i + log(p_i)), \text{for n classes} \tag{3.9}$$

A model with a cross entropy of 0 is said to be an ideal model. The log in this equation (3.9)is calculated in base 2. The truth label in this case is denoted as t_i while p_i is the softmax probability of the i^{th} class [62].

3.7 Activity Recognition System

3.7.1 Inception 3D

The Inception 3D model also denoted as I3D model is was introduced by researchers from DeepMind and the University of Oxford in a paper called “Quo Vadis, Action Recognition?". A New Model and the Kinetics Dataset”[7] [53]. Inception I3D is a "inflated" 2D called "Inception" which has been pretrained on ImageNet info. In the VU 2017 Charades Challenges, it came in first place [32, 59]. The word inflated refers to the network's initial state as a 2D neural network. The image classification network was converted to a 3D network, and the square filters and pooling kernels were replaced with cubic kernels. The figure 3.13 shows a pictorial representation of a two stream 3D convolution network.

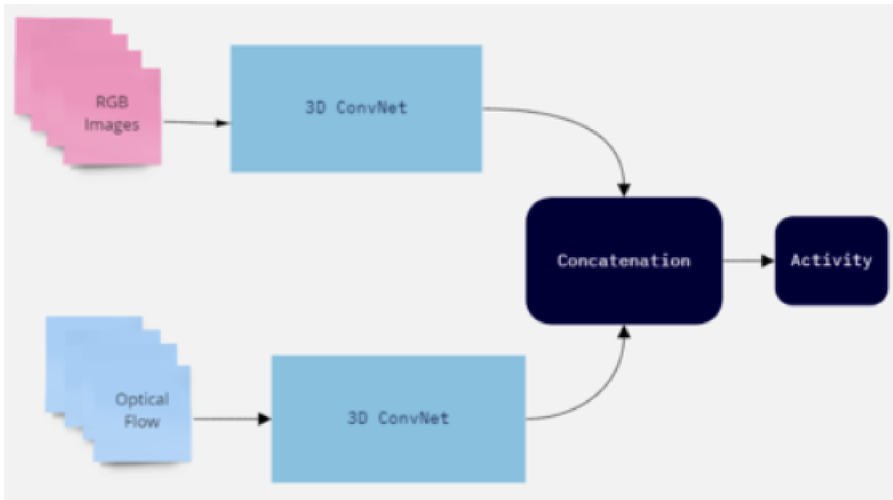

FIGURE 3.12: A model for two-stream 3D convolution networks

The paper [7], used parameters from a pretrained 2D model trained on ImageNet while executing their 3D model. This experiment was achieved by weight repetition of a 2D filters N number of times through the time dimension and adjusting them by dividing through N times[53]. More so, in the paper [60] GoogLeNet also became the winner of the ILSVRC (ImageNet Large Scale Visual Recognition Competition) in 2014, an image classification competition. To minimize computational complexity, GoogLeNet used an 1 x 1 convolution as a dimension reduction module. It can increase depth and width by reducing the computing constraint.

3.8 RetinaNet

RetinaNet is defined as a one-stage object detection network which was created by two improvement on already existing one stage object detection models [12, 35]. It was named due to its dense sampling operations on object regions of an input image [35]. According to [35],it was discovered that one stage networks run into huge object-background class imbalance observed during training in contrast to a two-stage networks, of whom the Region Proposal network gave a decrease in the number of candidate objects areas.

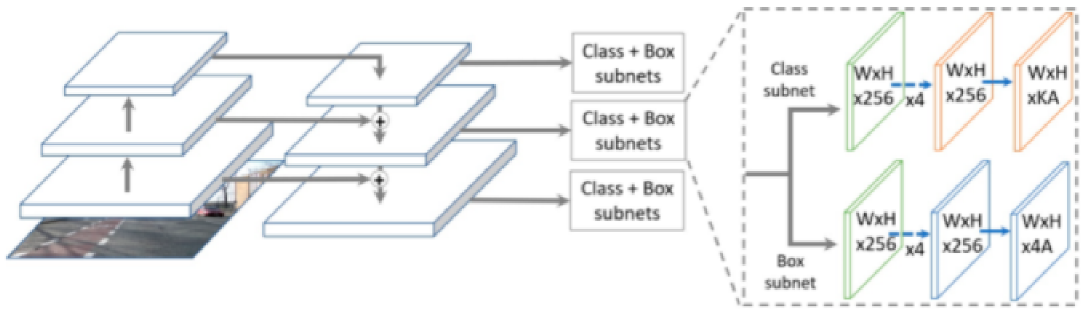

FIGURE 3.13: An illustration of RetinaNet's architecture, proposed in the article "Focal loss for dense object detection" [35].Moving from left to right: CNN base, feature pyramid network and class and bounding box sub-networks.

In computer vision, feature pyramids networks (FPN) have been mostly used in detecting objects with different image scales [34, 12]. The FPN has its architecture by combining semantically strong features with semantically weak features using a top-down pathway and a lateral connection. RetinaNet utilizes focal loss which was presented in section 3.6.3 and also smooth L1-loss in section 3.6.2.

3.9 Performance Metrics

Evaluation techniques is an important factor in every machine learning project. Performance metrics differs based on different Machine learning models and application. Based on differences in data and its distribution, there are various ways to evaluate a classification model which would be discussed below.

		Predicted Class		
		YES	NO	Total
Actual Class	YES	TP	FN	$TP + FN$
	NO	FP	TN	$FP + TN$
	Total	$TP + FP$	$FN + TN$	N

Table 3.1: Confusion Matrix Table

3.9.1 Accuracy

In machine learning, Accuracy is an evaluation metric for classification models. In this case classification accuracy is often know as the accuracy. Naturally , accuracy is known as the overall proportion of correct classification. Indeed, accuracy might not even be the best statistic to use at all especially in the case of class imbalance. The formula for classification accuracy is shown in the equation 3.10 below

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.10}$$

3.9.2 Threshold Types of Discriminator Metrics

The discrimination estimation of the best (ideal) method during a training phase can be determined in any given binary classification using a confusion matrix, as shown in Table 1.

Precision (Positive predictive value)

Precision determines what fraction of actual positives within observations were predicted as positive [43]. The formula for precision is calculated with equation 3.11 below.

$$Precision = \frac{TP}{TP + FP} \tag{3.11}$$

Specificity (True negative rate)

Specificity also called TNR refers to the proportion of negative data points that are falsely considered as positive, apropos to all negative data points [43]. Specificity (SP) is calculated as the number of correct negative predictions divided by the total number of negatives. The best specificity is 1.0, whereas the worst is 0.0 [43]. specificity is calculated with the equation 3.12 below.

$$SN = \frac{TN}{TN + FP} \tag{3.12}$$

Sensitivity (Recall or True positive rate)

Recall or Sensitivity estimates the fraction of actual positives which has been identified as positives. This measurement ranges from 0 to 1. A larger value indicates better predictive accuracy. Therefore, Recall or Sensitivity corresponds to the proportion of positive data points that are correctly considered as positive, in relation to all positive data points. [43]. The equation 3.13 is the formula for sensitivity.

$$SP = \frac{TP}{TP + FN} \tag{3.13}$$

False positive rate

The false positive rate (FPR) computes the false alarm rate or the fraction of actual negatives predicted as positives. It ranges from 0 to 1. FPR is calculated as the aggregate of incorrect positive predictions divided by the total number of negatives [3]. Also, it can be calculated as 1 – specificity [43].

$$FPR = \frac{FP}{FP + TN} \tag{3.14}$$

Chapter 4

Data and Materials

This chapter contains an overview of the data sources and its application on different experiments

4.1 Data Sources

Data has been gathered from Safer Births project(Haydom Data) and from hospitals in Stavanger and Nepal. The information are collected from video recordings which were captured during resuscitation sessions of babies. The data utilized for this work were acquired from medical facilities from three countries namely: Tanzania(Haydom), Nepal(Unlabelled) and Norway(SUS). The quality of videos differs from each hospital which is expected due to the effects from environmental factors and camera calibration. One of these factors are is camera position. Knowing the factors that can impact the quality of screen recording can be useful as it will add improvements where necessary. The variations observed from this videos lead to a more tedious job while performing activity recognition.

The data collected as part of the Safer Birth project includes includes a total of 400 videos containing resuscitation episodes during Newborn resuscitation activities which are used to model training and 20 videos set aside for evaluation.The newborn region was split into 3sec video(45frames) overlapping video clips. Images with labels were augmented through histogram matching. These datasets where adopted from the previous work by [39] with an addition from SUS and Nepal with were provided by Øyvind Meinich-Bache. These video data only depicts a time interval when a newborn was kept on a resuscitation table.

FIGURE 4.1: A blurry view of Resuscitation Episodes captured by an overhead camera

4.2 Activity set

On this set , each of these hospital videos are Linear Frame Interpolated (LFI) with a fixed frame rate of 15 fps, which is much closer to the pre-trained I3D weights that were initially trained on 25 fps videos. Initially these input channel were split into RGB and optical flow on x and y directions. But in the course of this new research optical flow was not utilized. Each of these clips has a fixed resolution of 256x256 pixels. The training and validation sets contains class labels as true labels labelled 1 or 0 where 1 means an activity is being performed and 0 means no activity.

4.3 Data Distribution

The datasets are divided into a Training list, Validation list, and test lists generated using an existing Matlab script originally created by Øyvind Meinich-Bache with minor adjustments to fit activities of interest. For a specific activity, this script automatically assigns a value of 1 for activity, and class 0 for not activity.

For example, when a specific activity such as ventilation is being trained it is classed at a value of 1 while activities not trained is given a value of 0, thus, falls under the not classed label. In this case (activity) labelled as classed is ventilation, while (not activity) describes other therapeutic activities apart from ventilation.

These therapeutic activities can be suction, oxygen and others. When training the model for a specific activity, 75percent of the data is assigned to the training lists and 25 percent to the validation list.

4.4 Nepal

The dataset gathered from Nepal was of two batches. It contained videos of distinct qualities. The low-quality videos It has a resolution of 480 x 240. It also includes a stable frame rate per second of 24.

On the other hand, the dataset from Nepal has a resolution of 1920x1080 and a stable frame rate of 30. The differences in quality come from the qualities affecting video data, such as camera positions.

Table 4.1: Information on amount of training data used to train and test different activities of interest for Nepal.

Class	**Training**	**Validation**	**Test**
Stimulation	3403	1147	856
Ventilation	31014	10365	7339
Suction	274	95	80
Suction with tube	1267	424	212
Drying	1334	800	600

4.5 Haydom

The dataset from Haydom had differences in quality due to bad camera calibrations and quality. The dataset had a varying frame rate between 2-30 frames per second. These videos were interpolated using Matlab by Meinich-Bache et al. [39].

Table 4.2: Information on amount of training data used to train and test different activities of interest for Haydom.

Class	**Training**	**Validation**	**Test**
Stimulation	4099	1389	2237
Suction	6896	2325	3400
Ventilation	17671	5916	8536

4.6 SUS

The data set from SUS had a better quality and was manually annotated by Øyvind Meinich-Bache et al. Data also had a more a stable frame rate of 15 FPS. Its resolution is 1280 x 720.

TABLE 4.3: Information on amount of training data used to train and test different activities of interest for SUS.

Class	Training	Validation	Test
Stimulation	1409	446	600
Suction	2298	1165	1100
Drying	1314	744	150
Suction with tube	1015	500	200
Ventilation	5400	1874	4000

Table 4.1 ,4.3 and 4.2 shows the amount of data used as training and validation lists which were extracted from the resuscitation data described above in section 4.5, 4.4 and 4.6 for experimentation purposes. Below is the distrubition on how various data lists were generated from different hospital data. These datasets were generated using a matlab script inherited from [39] by manually specifying each activity of interest.

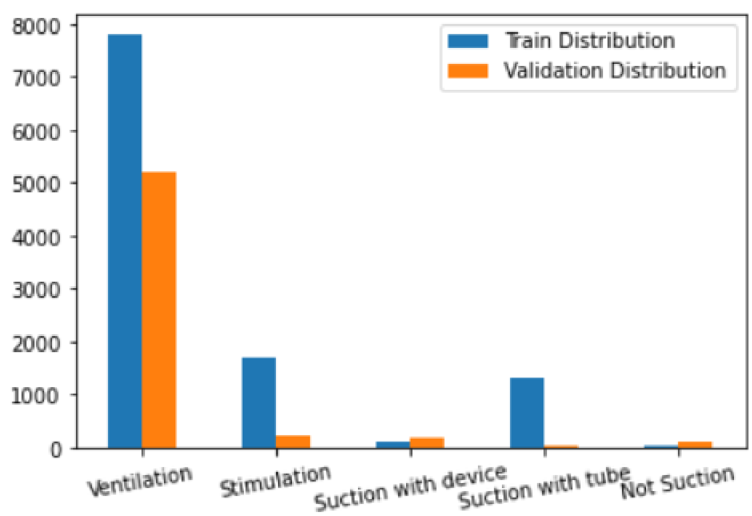

FIGURE 4.2: This bar plot shows how different therapeutic activities were used to generate a training list for Nepal Ventilation

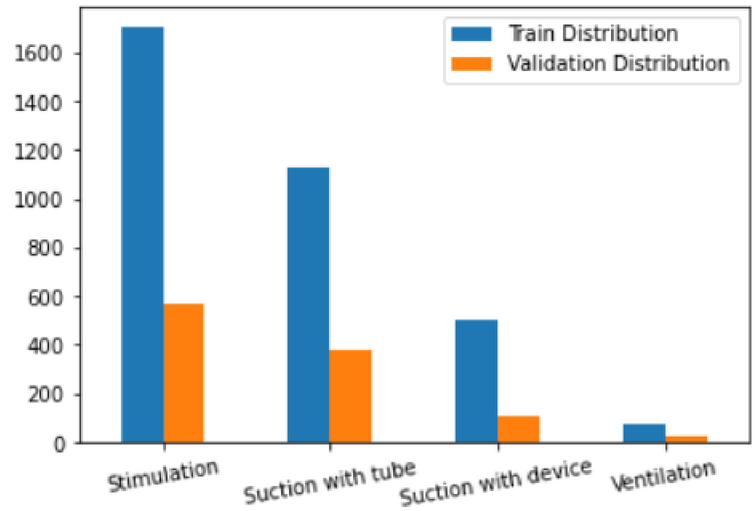

FIGURE 4.3: This bar plot shows how different therapeutic activities where used to generate a training and validation list for Nepal Stimulation

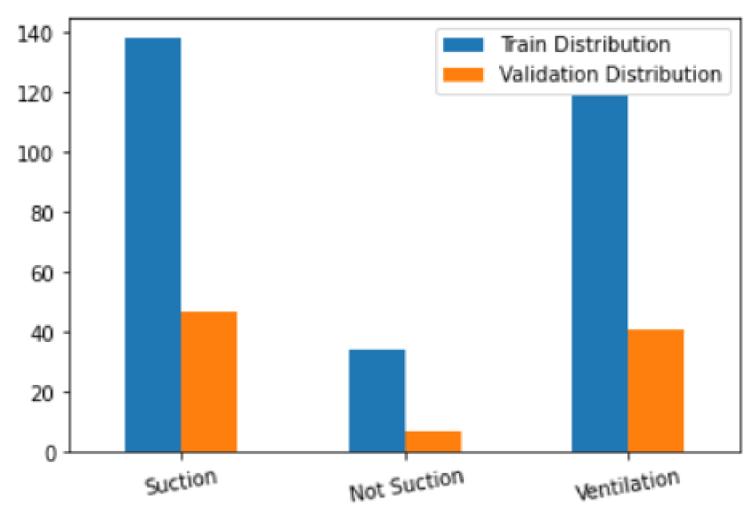

FIGURE 4.4: This bar plot shows how different therapeutic activities were used to generate a training list for Nepal Suction

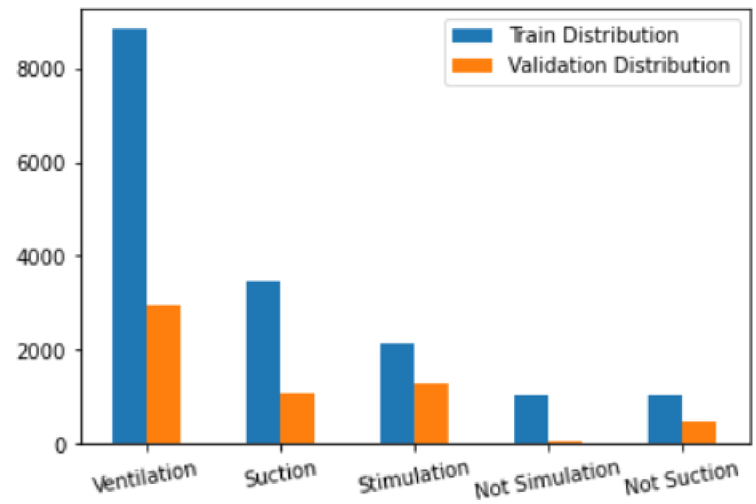

FIGURE 4.5: This bar plot shows how different therapeutic activities were used to generate a training list for Haydom Ventilation

FIGURE 4.6: This figure shows different theraputic activities performed by different hospitals

Chapter 5

Methodology

This chapter gives an overview of the methods applied and its implementation in this study.

5.1 Methods Overview

The first paper *Activity Recognition From Newborn Resuscitation Videos* by Meinich-Bache et al. [39], a one-stage object detection model (RetinaNet) was used for object detection and Region proposal. In his work, Meinich-Bache et al. used I3D, which utilized RGB stream and Optical flow.

The datasets gathered from Haydom hospital in Tanzania [38]. Tracking of resuscitation devices, HCP hands, and the newborn region were further analyzed to recognize different activities of interest and RetinaNet performed best.

Also, with the use of RetinaNet, predictions were faster, and the accuracy of the activity timelines was acquired per video analysis. Meinich-Bache et al. had datasets in three divisions: Training data, Validation data, and test data. The initial objective proposed by [39] was to perform object detection and then analyze the capabilities of activity recognition systems through newborn resuscitation videos of diverse quality.

This proposed system utilized the convolutional Neural network (CNN), which made use of two streams consisting of RGB and Optical flow presented by [7]. From observation, the main idea of Meinich-Bache et al. [39] was to first detect relevant object regions.

The method of this work is still based on this system, and the datasets gathered was sourced from the main system used by Meinich-Bache et al. [39] with additional data from SUS and Nepal, but the focus will be the newborn area excluding other object regions. Also, without the use of optical flow. The RGB stream will be used. There will not be further work on object detection except for the detection of the newborn area which has already been done in the previous work, which is part of the dataset provided by Meinich-Bache et al for this project to facilitate and achieve an efficient activity recognition analysis of the newborn resuscitation videos. additional datasets from Nepal and SUS is used for a more detailed activity recognition analysis within the newborn area.

In the previous works in order to recognize the activities Meinich-Bache et al. [39] inputs short video sequences from activity relevant areas to models trained on a specific activity [38]. Therefore, binary classification was applied for each model, the classification was performed even if there is no evidence of activity. Object dependent activities within the object region and the newborn region are used during predictions. [39] [38].

In his method, Meinich-Bache et al. [40][38] utilized two steps; the first step was object detection by using RetinaNet, an object detection network to detect the location of an object. In the second step, Meinich-Bache et al. progressed into activity recognition by using Inception I3D to determine if an activity was being performed; RGB and optical flow in the Inception I3D model contain two streams perform activity recognition and object detection.

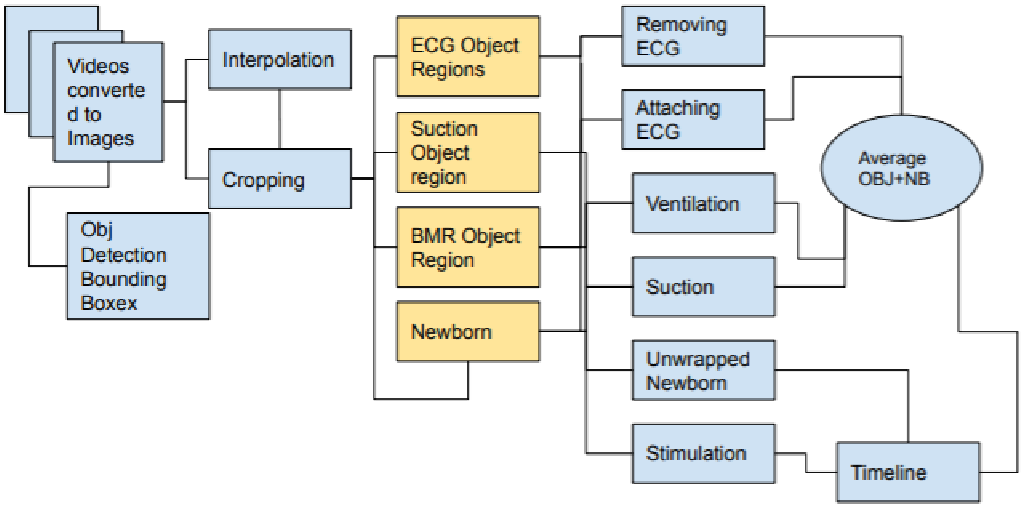

FIGURE 5.1: Old pipeline redesigned

5.2 New Simplified Pipeline

In this project, the region of interest to analyze is the newborn area. The focus would be to investigate whether studying the resuscitation table would be enough. Also, with more data, perhaps this is achievable. The idea is to simplify the initial system by limiting the number of regions to be analyzed, especially by doing less or no analysis on object regions as it can be computationally demanding.

However, in this project, the Generalized Vs. A specialized activity recognition system focused on activity recognition by using RGB sequence inputs to determine if an activity is performed considering the newborn area. RetinaNet is used in this project in generating the datasets containing newborn regions. The activity recognition system in this research is the same as the system used by Meinich-Bache et al. Using the same activity recognition method as Meinich-Bache but excluding optical flow data.

The use of RGB is based on an observation by [39] which shows optical flow was more computationally expensive and also showed little or no performance; therefore

RGB stream is adopted for rapid results.

5.2.1 Activity Recognition (AR)

Convolutional Neural network Inception I3D, which was used in the previous work [39] is the same method inherited for activity recognition. The first step of this project which is detecting the newborn area, was performed while generating the dataset used in this project. The second step of this project is to see how the activity recognition system(I3D) performs while analyzing the resuscitation tables(Newborn area) with the use of RGB stream .

The I3D architecture(a first class activity recognition network) contains two streams the RGB and optical flow only the RGB would be used for the activity recognition process and predictions.This is done with the available pretrained model publicly accessible through transfer learning. With the adopted RGB-imageNet pretrained, further training on this model with focus on the newborn resuscitation activities data, was executed for analysis in the newborn region shown in step1 figure 5.2

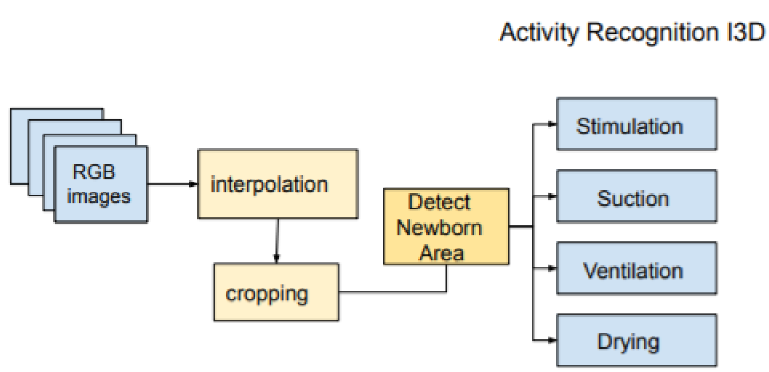

FIGURE 5.2: The new simplified pipeline. This method uses a binary class model for activities of interest

5.3 Newborn Region

In this project, activity recognition was done with the newborn region as the main focus. The newborn region is the area where the resuscitation table is found and where the resuscitation activities are performed. The newborn region was first detected using RetinaNet. The process involved creating a heat-map around health care providers hands, thus deciding the area of the newborn. This first step mentioned above came with the pre-processed data provided by Meinich-Bache et al. for the purpose of this research.

5.3.1 Predictions

For different activities of interest, predictions are generated by using the RGB data and model to recognize the type of activity that is being performed in the newborn region.

It is important to note that the video quality of the data makes AR strenuous to perform. Therefore the RGB model is used to train on individual activities mentioned in Chapter 2. Binary classification is implemented for each activity analysis.

The RGB stream is used to detect whether there is an activity or not. It predicts the type of activity through the data projections. Therefore, the model recognizes, for example, stimulation and makes its predictions based on particular motions associated with stimulation activity. Also, this applied to other activities.

The RGB was used to train and analyze different newborn area activities, and specialization and generalization resulted in 6 I3D models learned with RGB and no optical flow. It begins with the video going through activity recognition tracking and region proposal. Also, the videos from the newborn region are Linear frames interpolated [39] [7].

Furthermore, sub-signals of the RGB stream were generated by a sliding window approach, and these sub-signals are offered as feeds to their corresponding model. Softmax is applied to perform predictions by averaging the logits gotten from the final I3D layer of the RGB model.

Although ventilation and suction are both object dependent and activity recognizable, stimulation is not object dependent. Because all three activities are activity-dependent, predictions are generated and classified based on recognized movements associated with each specific activity.

5.4 Implementation

Most pre-processing techniques applied were performed on the images. These pre-processing techniques included: Histogram Equalization, Homomorphic filtering, and color pre-processing.

Images were color pre-processed by splitting into different channels in the RGB stream, which in turn denotes Red, Green, and Blue color channels. Adopting the I3D trained, which had minor modifications and was trained on Nvidia Tesla V100 GPUs, would help improve training and testing computational capacity.

5.4.1 Training setup

The generalized Vs. The specialized activity recognition system was contingent on a considerable amount of parameters. These parameters were strategically set up by making available a base that significantly differentiated the model performance.

Hyper-parameters which is commonly adapted for all models can be seen in table 5.1 and 5.3. Based on the network's similarity, the hyper-parameters connect to an absolute minima bringing together the smallest or overall value of a set and how it functions(speed) in its entity.

In other to optimize performance, further training was done using the setups. Using the heuristic evaluation method, the values of these hyper-parameters were initiated. Other forms of the search were not efficient enough.

Table 5.1: List of common hyper-parameters for the Activity Recognition networks.

Hyper-parameters	Description
Batch Size	The amount of training examples the network uses for weights to be updated
Learning rate	Controls models adaptability
Epochs	Defines the number of iterations a network has been trained on the entire dataset
Weight Decay	A form of regularization penalizing large values of weights in the model.Reduces over fitting
Learning rate decay	Helps in learning rate reduction

Table 5.2: List of parameters in training activity recognition recognition model.

Parameters Name	Parameter Value
Learning rate	0.0001
Number of frames per clip	45
Crop Size	224
Batch Size	6
Classics	224
Max Steps	18000
Decay When	6000
Decay Rate	0.1

TABLE 5.3: List of parameters in relation to the network architecture. These parameters are the original set of the adapted network model

Additional parameters	Description
Backbone	The neural network(CNN) which was utilized for the sole purpose of feature extraction
Loss function	Different types of functions
Dropouts	Ignores neurons prior to training
Conv Layers	The number of convolutional layers
Kernel size	defines the height and width of filter kernels
Padding	Padding alternative
Activation functions	Different activation functions
Stride	Striding options
Down-sampling	Reducing spatial dimensions of an image

Additional adjustments on adapted architecture source code also included:

- Models being saved during training.
- GPU enabling.
- Retrieval of performance evaluation metrics.
- Layer selection.
- Selection and loading specific pretrained models.

Chapter 6

Experiments and Results

In this section, different experiments were carried out for the activity recognition analysis carried out on the newborn area with the purpose of simplifying the system as stated in Chapter 5. These experiments are carried out in view of two outcomes which is comparing how a model would perform if generalized or specialized with each hospital data.

6.1 Specialization

The first step is a presentation of an activity recognition experiment performed based on a specialized model on three different activities from three available datasets mentioned in Chapter 4. The main experiment is performed to observe the performance of a model when specialized for each hospital by training and testing on specific activities on each hospital data. Due to the amount of False Positives(FP), Thresholds were set to penalize wrongly classified recall metrics from predictions. Figure 6.1 shows different activity thresholds.

TABLE 6.1: Thresholds utilized for different specialized models predictions

Activity	**RGB**
Stimulation	0.5
Suction	0.8
Ventilation	0.7

The activity Suction is trained, validated, and tested on each data from different hospitals. In the same manner, ventilation and stimulation were trained and evaluated. Furthermore, evaluation of the I3D models used to train each specific activity is initiated.

This evaluation was done using accuracy, precision, and recall. Table 6.2 the results obtained from training a model for the therapeutic activity stimulation from the three hospitals dataset denoted with the short form Haydom Hospital(D_H) , Nepal(D_N) and SUS(D_S) for simplicity purposes. The letter "M" with a subscript denotes the model trained on a specific data activity, while "M" with a superscript represents model trained on a combined data. "T" denotes a combined testset.

6.1.1 Experiment 1: Evaluation of model accuracy on Haydom , Nepal and SUS on activity the Stimulation

In experiment 1, the model was trained on the activity stimulation using RGB and I3D. The model utilized 75% of training data for each hospital. As shown in the table 6.2 model $\mathrm{M}_H ST$ had precision 0.6280 and a recall of 0.7076, $\mathrm{M}_N ST$ 0.6650 and 0.7258, $\mathrm{M}_S ST$ 0.8182 and 0.6136.

TABLE 6.2: Exp1 Performance from model trained on activity Stimulation from each Hospital Data

Activity	**Architecture**	**Model/Evaluated**	**Prec**	**Rec**
Stimulation	RGB+I3D	$\mathrm{M}_H ST$ / $\mathrm{D}_H ST$	0.6280	0.7076
Stimulation	RGB+I3D	$\mathrm{M}_N ST$ / $\mathrm{D}_N ST$	0.6650	0.7258
Stimulation	RGB+I3D	$\mathrm{M}_S ST$ / $\mathrm{D}_S ST$	0.8182	0.6136

This means that the model was able to correctly classify the activity stimulation on all hospital. Though it has a high recall with shows some level of misclassification. The model performance shows an overall good accuracy. In evaluation, the results from this experiment shown in table 6.2.

6.1.2 Experiment 2: Evaluation of model accuracy on Haydom , Nepal and SUS on activity the Ventilation

Experiment 2 observes how the model performs on the activity ventilation by specializing the model on all three hospitals' data.The model used 75% of training data Precision is slightly lower than recall on D_H at 0.8059, D_S at 0.7709, D_N at 0.7709.

TABLE 6.3: Exp2 Performance from model trained on activity Ventilation from each Hospital Data

Activity	**Architecture**	**Model/Evaluated**	**Prec**	**Rec**
Ventilation	RGB+I3D	M_HV / D_HV	0.8059	0.8252
Ventilation	RGB+I3D	M_SV / D_SV	0.7709	0.8070
Ventilation	RGB+I3D	M_NV / D_NV	0.7709	0.7883

The model predictions shows a good performance on all data because the percentage of correctly classified results is high.

6.1.3 Experiment 3: Evaluation of model accuracy on Haydom , Nepal and SUS on activity the Suction

In experiment 3 as shown in figure 6.4 model were trained and evaluated on the activity suction using the three hospital data. 75% of each hospital data was applied to train the model. Precision and recall was generally low with M_HSuc at a precision of 0.5910 and a recall of 0.5512, M_NSuc at precision of 0.4345 and recall 0.6980, M_SSuc at precision of 0.5505 and recall 0.6524.

The models performance was low in making correct predictions.

Table 6.4: Exp3 Performance from model trained on activity Suction from each Hospital Data

Activity	Architecture	Model/Evaluated	Prec	Rec
Suction	RGB+I3D	$\mathrm{M}_H Suc$ / $\mathrm{D}_H Suc$	0.5910	0.5512
Suction	RGB+I3D	$\mathrm{M}_N Suc$ / $\mathrm{D}_N Suc$	0.4345	0.6980
Suction	RGB+I3D	$\mathrm{M}_S Suc$ / $\mathrm{D}_S Suc$	0.5505	0.6524

6.2 Generalization

Generalization, in this case, combines therapeutic activities from different hospital data and training them as a single model provided these hospitals perform the resuscitation activity in the same manner(same technique, same device). For a generalized model, the figure 6.5 shows the different thresholds set on model predictions.

Table 6.5: Thresholds utilized for different generalized models predictions

Activity	RGB
Stimulation	0.5
Suction	0.75
Drying	0.5
Suction with tube	0.5
Ventilation	0.4

6.2.1 Experiment 4: Accuracy of evaluation on a combined dataset on the activity Ventilation

In experiment 4, D_H and D_N were combined and trained as a single model. This is because both hospitals utilize the same technique and same device(bag-mask resuscitator) for the activity ventilation. The dataset is divided on a 50/50 ratio.

TABLE 6.6: Exp4 Performance from model trained on activity Ventilation combining different Hospital Data

Activity	Architecture	Model/Evaluated	Prec	Rec
Ventilation	RGB+I3D	$\mathrm{M}^{D_HV+D_NV}$ / D_HV	0.3555	0.8843
Ventilation	RGB+I3D	$\mathrm{M}^{D_HV+D_NV}$ / D_NV	0.4220	0.8432
Ventilation	RGB+I3D	M_HV / D_NV	0.2050	0.7350
Ventilation	RGB+I3D	M_NV / D_HV	0.4821	0.6713

These results show poor accuracy. The model did not perform well as it struggled to classify activity even though both data have similar technique.

6.2.2 Experiment 5: Accuracy of evaluation on a combined dataset on the activity Stimulation

Experiment 5 was done applying the same method as experiment 4. 25% of data from D_H , D_N and D_S were combined when training the model on stimulation. The combination was done because the three hospital applied the same technique.The performance of the model was evaluated using one hospital data at a time. Models were also trained separately on D_HST and D_NST combined excluding D_SST, furthermore, evaluation was done on D_NST and D_HST test set.

TABLE 6.7: Exp4 Performance from model trained on activity Stimulation combining different Hospital Data

Activity	Architecture	Model/Evaluated	Prec	Rec
Stimulation	RGB+I3D	$\mathrm{M}^{D_HST+D_NST+D_SST}$/ D_HST	0.2062	0.6770
Stimulation	RGB+I3D	$\mathrm{M}^{D_HST+D_NST+D_SST}$/ D_SST	0.3538	0.8201
Stimulation	RGB+I3D	$\mathrm{M}^{D_HST+D_NST+D_SST}$/ D_NST	0.2690	0.4445
Stimulation	RGB+I3D	$\mathrm{M}^{D_HST+D_NST}$ / D_HST	0.4500	0.7420
Stimulation	RGB+I3D	$\mathrm{M}^{D_HST+D_NST}$ / D_NST	0.5045	0.8025
Stimulation	RGB+I3D	D_NV / D_HV	0.2590	0.7910

As a generalized model it did not perform well.

6.2.3 Experiment 6 : Accuracy of evaluation on a combined dataset on the activity Suction(tube)

Experiment 6.8 was conducted by training a model on the activity suction using tubing. 50% of D_S and D_N were combined since both hospitals made use of same suction device(tube). The result seen in 6.8 shows a low accuracy.

TABLE 6.8: Exp6 Performance from model trained on activity Suction using tubing combining different Hospital Data

Activity	Architecture	Model/Evaluated	Prec	Rec
Suction tube	RGB+I3D	$M^{D_N Suc + D_H Suc}$ / $D_H Suc$	0.4655	0.4850
Suction tube	RGB+I3D	$M^{D_N Suc + D_H Suc}$ / $D_N Suc$	0.4350	0.5601

Model performance reflects a failed generalization in this case, because for a good model performance it is either a high precision and low recall, a low precision and a high recall. The model hardly made a correct classification.

6.2.4 Experiment 7: Accuracy of evaluation on a combined dataset on the activity Stimulation and Drying

In Experiment 7, as seen in the table 6.9 Model was trained four times. The model on drying for D_N data and stimulation for D_S data divided on a 50% ratio. Results were evaluated on both data as a combined test set. In addition, the results portray a low performance on correct predictions. The model trained on D_N was evaluated on D_S dry and vice versa.

TABLE 6.9: Exp7 Performance from model trained on activity Drying, Stimulation and drying combining different Hospital Data

Activity	Architecture	Model/Evaluated	Prec	Rec
Stimulation and Drying	RGB+I3D	$\mathrm{M}^{D_N Dry+D_S ST}$ / $\mathrm{T}^{D_N Dry+D_S ST}$	0.3124	0.6575
Drying	RGB+I3D	$\mathrm{M}_N Dry$,/ $\mathrm{D}_S Dry$	0.4033	0.4211
Drying	RGB+I3D	$\mathrm{M}_S Dry$,/ $\mathrm{D}_N Dry$	0.4719	0.4467
Drying	RGB+I3D	$\mathrm{M}^{D_N Dry+D_S Dry}$ / $\mathrm{T}^{D_N Dry+D_S Dry}$	0.5291	0.6610

Model trained on D_N and D_S on the activity drying had a better performance.

Chapter 7

Discussion and Conclusion

7.1 Experiment 1

Based on the results shown in table 6.2, training a model on the stimulation activity using D_N, D_N and D_H dataset the results from the evaluation reflects an improvement in accuracy. This signifies that the model performed well with enough data.

TABLE 7.1: True positives(TF) and false positives(FT) from experiment 1.

Model	**TP**	**FP**
$M_H ST$	726	430
$M_N ST$	3143	666
$M_S ST$	243	150

7.2 Experiment 2 and Experiment 3

In experiment 2 shown in figure 6.3, we can see that the specialized model trained on the activity ventilation had a higher number of correctly classified results and an average number of actual relevant results. In view of this estimation, we can proceed to say the specialized model reflects a good performance on all three hospital ventilation activity. In experiment 3 above on figure 6.4 model struggles to correctly predict class activity. These result also shows the existence of False positives. Models trained on activity suction has some misclassifications. Misclassification can happen due to a number of reasons, perhaps, one of the reasons could be quality of the image and variance in implementation of the suction procedures in D_N and D_S data.

FIGURE 7.1: SUS Suction activity video that looks ventilation activity.

FIGURE 7.2: Haydom Suction episode.

Above in figure 7.1, we can see the activity in D_S varying in suction activity implementation from D_N and D_H. In D_S activity recognition might be a bit difficult because from the image we can see ventilation and suction being carried out almost simultaneously as opposed to D_N and D_H where the suction activity is performed exclusively. This might have an effect on the results as proper performance of the activity(suction) could aid a better system performance in terms of the model being able to distinguish and recognise what activity is being performed.

FIGURE 7.3: Nepal Suction Episode

7.3 Experiment 4, 5 and 6

Generalizing the model in experiment 4,5 and 6 shows that the system did not perform well on the combined data therefore the proposed solution on generalization might need further work. Experiment 5 had worse result when compared to experiment 4 with the result is shown in table 6.6 perhaps with further training and data the model might perform better with the combined stimulation activity.

In Experiment 6 we can see that there is separation of the activity suction on tubing for both D_S and D_N. The outcome indicates precision and recall with poor accuracy. Therefore, as shown in the result the system did not perform well as a generalized model. One of the problems could be different quality of data, another problem could be availability of more data on suction in the Nepal data is not equal to the amount of data from SUS.

7.4 Experiment 7

Experiment 7 was performed by training and a model on a combined activity Stimulation and drying, and also on a single activity "Drying".The experiment was done to see if the model will perform as a generalized model trained on a combined data from D_S and D_N. Both hospitals apply the same technique in the stimulation activity in which drying is a part of the procedure. The intent is to see whether separating drying and stimulation will produce better model performance.The results shows a huge misclassification which means combining the data and generalizing it as one model did not improve the performance of the model, rather, there was difficulty in differentiating between drying and stimulation. More so, the model was also trained on the activity drying separately on both hospital data and the outcome of the model's performance on D_N data was better compared to the model's performance on D_S data which had a higher misclassification. Variation in data also added to the performance of the model as seen in the figures below. Another factor to consider for this outcome is model over-fitting.

FIGURE 7.4: Nepal stimulation Episode

FIGURE 7.5: A clear view of Nepal stimulation Episode

FIGURE 7.6: SUS drying Episode

7.5 Conclusion

Using the main system(I3D and RGB stream) excluding optical flow and object regions as a backbone for this research contributed hugely on the model performance especially in the area of activity recognition using a specialized model. More data from Nepal and SUS allowed for a successful training of the model on a combined data set which was needed to achieve the aim of this research.

Inheriting the weights on the pretrained ImageNet dataset through transfer learning and using just the RGB stream fulfilled the purpose of this study in terms of simplifying the system. This was achieved with the use I3D architecture for activity recognition with focus on newborn area(Resuscitation table).This aided a better analysis and evaluation and allowed for a much transparent distinction on the model performance.

In view of the outcome of the trained models we can say that the RGB model was more effective in accuracy in the area of specialization and did not fair well in the aspect of generalization. Thus, the proposed Haydom solution of generalizing the model for a better system performance was not achieved in this study due to observed misclassifications,over fitting and data variability.

However, for future work we can suggest more time to be spent on prepossessing, data labeling and augmentation. Also, consider using both RGB and Optical flow for generalizing the model as this could produce better system performance.

Bibliography

[1] Angus Addlesee and Pierre Albert. "Ethically Collecting Multi-Modal Spontaneous Conversations with People that have Cognitive Impairments". In: *arXiv preprint arXiv:2009.14361* (2020).

[2] Sylvain K Akpo et al. "A Kohonen self organizing-map (KSOM) approach to analyze pesticides biodisponibility into the Marahoué watershed, Côte d'Ivoire". In: *Eur. J. Sci. Res* 61.3 (2011), pp. 355–366.

[3] Syed Sumera Ershad Ali and Sayyad Ajij Dildar. "An Efficient Quality Inspection of Food Products Using Neural Network Classification". In: *Journal of Intelligent Systems* 29.1 (2019), pp. 1425–1440.

[4] Aphex34. $max_poolingwith2x2filterandstride = 2$. 2015. URL: https://en.wikipedia.org/wiki/Convolutional_neural_network#/media/File:Max_pooling.png (visited on 05/22/2021).

[5] Hafiz Muhammad Aslam et al. "Risk factors of birth asphyxia". In: *Italian journal of pediatrics* 40.1 (2014), pp. 1–9.

[6] Jason Brownlee. *What Is Artificial Neural Network And How It Works*. 2020. URL: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ (visited on 02/20/2021).

[7] Joao Carreira and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset". In: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 6299–6308.

[8] Zhi Chen and Pin-Han Ho. "Global-connected network with generalized ReLU activation". In: *Pattern Recognition* 96 (2019), p. 106961.

[9] Francesco Croce, Maksym Andriushchenko, and Matthias Hein. "Provable robustness of relu networks via maximization of linear regions". In: *the 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2057–2066.

[10] DEEPLIZARD. *Machine Learning Deep Learning Fundamentals*. 2018. URL: https://deeplizard.com/learn/video/ZjM_XQa5s6s (visited on 06/09/2021).

[11] Arden Dertat. *Applied Deep Learning - Part 4: Convolutional Neural Network*. 2017. URL: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2 (visited on 04/04/2021).

[12] ArcGIS Developers. *How RetinaNet works?* 2021. URL: https://developers.arcgis.com/python/guide/how-retinanet-works/#:~:text=RetinaNet%20is%20one%20of%20the,dense%20and%20small%20scale%20objects.&text=RetinaNet%20has%20been%20formed%20by,and%20Focal%20Loss%20%5B2%5D (visited on 06/05/2021).

[13] Anshul Dhingra. "Activation Functions in Neural Networks". In: (2017).

[14] Ben Dickson. *What are artificial neural networks (ANN)?* 2019. URL: https://bdtechtalks.com/2019/08/05/what-is-artificial-neural-network-ann (visited on 05/22/2021).

[15] Ben Dickson. *What Is Artificial Neural Networks(ANN)?* 2019. URL: https://bdtechtalks.com/2019/08/05/what-is-artificial-neural-network-ann/ (visited on 04/05/2021).

[16] Stack Exchange. *How to interpret smooth l1 loss?* 2018. URL: https://stats.stackexchange.com/questions/351874/how-to-interpret-smooth-l1-loss (visited on 06/15/2021).

[17] Jay's Father. *DEEP LEARNING WITH PYTHON*. 2018. URL: http://junyelee.blogspot.com/2018/01/deep-learning-with-python.html (visited on 06/11/2021).

[18] Anirudha Ghosh et al. "Fundamental concepts of convolutional neural network". In: *Recent Trends and Advances in Artificial Intelligence and Internet of Things*. Springer, 2020, pp. 519–567.

[19] Ross Girshick. "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.

[20] Peter Goldsborough. "A tour of tensorflow". In: *arXiv preprint arXiv:1610.01178* (2016).

[21] Olga Golubnitschaja. *Predictive diagnostics and personalized treatment: dream or reality*. Nova Biomedical Books New York, 2009.

[22] Daniel Graupe. *Principles of artificial neural networks: basic designs to deep learning*. Vol. 8. World Scientific, 2019.

[23] Enzo Grossi and Massimo Buscema. "Introduction to artificial neural networks". In: *European journal of gastroenterology & hepatology* 19.12 (2007), pp. 1046–1054.

[24] Jiawei Han, Micheline Kamber, and Jian Pei. "Data mining concepts and techniques third edition". In: *The Morgan Kaufmann Series in Data Management Systems* 5.4 (2011), pp. 83–124.

[25] RG Harper et al. “249 FAILURE OF EARLY INFANT STIMULATION IN IMPROVING THE PROGNOSIS OF SEVERELY ASPHYXIATED NEONATES”. In: *Pediatric Research* 15.4 (1981), pp. 481–481.

[26] Knut Hinkelmann. *Neural Networks*. University Lecture. 2000.

[27] Lucia Hug et al. “National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis”. In: *The Lancet Global Health* 7.6 (2019), e710–e720.

[28] i2tutorials. *What is Convolutional Neural Network? What are all the layers used in it?* 2019. URL: https://www.i2tutorials.com/what-is-convolutional-neural-network-what-are-all-the-layers-used-in-it (visited on 06/01/2021).

[29] Santhoopa Jayawardhana. *What Are L1 and L2 Loss Functions?* 2020. URL: https://towardsdatascience.com/sequence-models-and-recurrent-neural-networks-rnns-62cadeb4f1e1 (visited on 05/01/2021).

[30] M Tim Jones. *Transfer learning for deep learning*. 2020. URL: https://developer.ibm.com/technologies/artificial-intelligence/articles/transfer-learning-for-deep-learning/ (visited on 05/11/2021).

[31] William J Keenan et al. “Delivery and Immediate Neonatal Care”. In: *Pediatric Education in Disasters Manual. Buenos Aires: American Academy of Pediatrics* (2009), pp. 217–37.

[32] Håvard Lapin. “Adapt and Generalize Deep Learning Methods for Activity Recognition on Newborn Resuscitation Videos”. MA thesis. University of Stavanger, Norway, 2020.

[33] Woo Seok Lee and Sergej Flach. “Deep learning of chaos classification”. In: *Machine Learning: Science and Technology* 1.4 (2020), p. 045019.

[34] Tsung-Yi Lin et al. "Feature pyramid networks for object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2117–2125.

[35] Tsung-Yi Lin et al. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[36] Wei Lu. *Neural network model for distortional buckling behaviour of cold-formed steel compression members*. Helsinki University of Technology, 2000.

[37] Blausen Medical. "Medical gallery of blausen medical 2014". In: *WikiJournal of Medicine* 1.2 (2014), pp. 1–79.

[38] Øyvind Meinich-Bache. "Automatic Video Analysis in Resuscitation". In: (2020).

[39] Øyvind Meinich-Bache et al. "Activity Recognition From Newborn Resuscitation Videos". In: *IEEE journal of biomedical and health informatics* 24.11 (2020), pp. 3258–3267.

[40] Øyvind Meinich-Bache et al. "Object detection during newborn resuscitation activities". In: *IEEE journal of biomedical and health informatics* 24.3 (2019), pp. 796–803.

[41] Paul A Merolla et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface". In: *Science* 345.6197 (2014), pp. 668–673.

[42] John William Mooney. "A Biomagnetic Field Mapping System for Detection of Heart Disease in a Clinical Environment". PhD thesis. University of Leeds, 2018.

[43] Vikas D More. *What is true positive and true negative – confusion matrix*. 2017. URL: https://moredvikas.wordpress.com/2017/09/12/what-is-true-positive-and-true-negative-confusion-matrix/ (visited on 06/11/2021).

[44] Robert Moshiro, Paschal Mdoe, and Jeffrey M Perlman. "A global view of neonatal asphyxia and resuscitation". In: *Frontiers in pediatrics* 7 (2019), p. 489.

[45] Georgina Msemo et al. "Newborn mortality and fresh stillbirth rates in Tanzania after helping babies breathe training". In: *Pediatrics* 131.2 (2013), e353–e360.

[46] Michael Obladen. "From "apparent death" to "birth asphyxia": a history of blame". In: *Pediatric research* 83.2 (2018), pp. 403–411.

[47] World Health Organization et al. "Guidelines on basic newborn resuscitation". In: (2012).

[48] Bharath Ramsundar and Reza Bosagh Zadeh. *TensorFlow for deep learning: from linear regression to reinforcement learning*. " O'Reilly Media, Inc.", 2018.

[49] Matías Roodschild, Jorge Gotay Sardiñas, and Adrián Will. "A new approach for the vanishing gradient problem on sigmoid activation". In: *Progress in Artificial Intelligence* 9.4 (2020), pp. 351–360.

[50] Stuart J Russell and Peter Norvig. *Instructor's solution manual for artificial intelligence: a modern approach*. 2003.

[51] Magdy Saeb. "The Stone Cipher-192 (SC-192): A Metamorphic Cipher". In: *The International Journal on Computers and Network Security (IJCNS)* 1.2 (2009), pp. 1–7.

[52] Ravindra Savaram. *What Is Artificial Neural Network And How It Works*. 2021. URL: https://mindmajix.com/artificial-neural-network-and-how-it-works (visited on 02/02/2021).

[53] Madeline Schiappa. *Understanding the Backbone of Video Classification: The I3D Architecture*. 2020. URL: https://maddyschiappa.medium.com/ (visited on 05/01/2021).

[54] Avinash Sharma. *Understanding Activation Functions in Neural Networks*. 2017. URL: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ (visited on 05/20/2021).

[55] Amit Shekhar. *What Are L1 and L2 Loss Functions?* 2019. URL: https://afteracademy.com/blog/what-are-l1-and-l2-loss-functions (visited on 05/11/2021).

[56] George Franklin Smith and Dharmapuri Vidyasagar. "Historical review and recent advances in neonatal and perinatal medicine". In: (1983).

[57] Jonathan M Spector and Subhash Daga. "Preventing those so-called stillbirths". In: *Bulletin of the World Health Organization* 86 (2008), pp. 315–316.

[58] Abdulhamit Subasi. *Practical Machine Learning for Data Analysis Using Python*. Academic Press, 2020.

[59] Waqas Sultani, Qazi Ammar Arshad, and Chen Chen. "Action recognition in real-world videos". In: *arXiv preprint arXiv:2004.10774* (2020).

[60] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[61] Taha Hasson Taha et al. "The Official Journal of The Kuwait Medical Association". In: ().

[62] Sik-Ho Tsang. *Review: RetinaNet — Focal Loss (Object Detection)*. 2019. URL: https://towardsdatascience.com/review-retinanet-focal-loss-object-detection-38fba6afabe4 (visited on 06/10/2021).

[63] Jarle Urdal. "Signal Processing for Newborn Survival: from labour to resuscitation". In: (2020).

[64] Christian Versloot. *How does the softmax function work*. 2021. URL: https://www.machinecurve.com/index.php/2020/01/08/how-does-the-softmax-activation-function-work/ (visited on 01/08/2020).

[65] Huyen Vu et al. "Automatic classification of resuscitation activities on birth-asphyxiated newborns using acceleration and ECG signals". In: *Biomedical Signal Processing and Control* 36 (2017), pp. 20–26.

[66] Jeremy West, Dan Ventura, and Sean Warnick. "Spring research presentation: A theoretical foundation for inductive transfer". In: *Brigham Young University, College of Physical and Mathematical Sciences* 1.08 (2007).

[67] WHO. *Newborns: improving survival and well-being*. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality (visited on 05/25/2021).

[68] Aidan Wilson. *A Brief Introduction to Supervised Learning*. 2019. URL: https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590 (visited on 06/11/2021).

[69] Rikiya Yamashita et al. "Convolutional neural networks: an overview and application in radiology". In: *Insights into imaging* 9.4 (2018), pp. 611–629.

[70] Yu Zhang and Qiang Yang. "An overview of multi-task learning". In: *National Science Review* 5.1 (2018), pp. 30–43.

[71] Jacek M Zurada. *Introduction to Artificial Neural Systems, vol. 4*. 1992.