REAL TIME HUMAN ACTIVITY DETECTION USING IMPROVED DCNN BASED ON TRANSFER LEARNING.

BY

NAHID HASAN
181-15-1970
AND
NAFISA TABASSUM
181-15-2046

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Md. Mahfujur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By
Zarin Tasnim Shejuti
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project/internship titled **"**Real time human activity detection using Improved DCNN based on Transfer Learning**"**, submitted by Nahid Hasan ID: 181-15-1970 and Nafisa Tabassum ID: 181-15-2046 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 19.0.09.2021

<u>BOARD OF EXAMINERS</u>

____

Dr. Md. Ismail Jabiullah **Internal Examiner**

Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

____

Narayan Ranjan Chakraborty **Internal Examiner**

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Dr. Mohammad Shorif Uddin **External Examiner**

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md. Mahfujur Rahman, Senior Lecturer, Department of CSE Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

15.1.2022

Md. Mahfujur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Zarin Tasnim Shejuti
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Nahid Hasan

Nahid Hasan
ID: 181-15-1970
Department of CSE
Daffodil International University

Nafisa Tabassum

Nafisa Tabassum
ID: 181-15-2046
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we would love to express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes it possible to complete this final year research project successfully.

We are really grateful and wish our profound indebtedness to **Md. Mahfujur Rahman, Senior Lecturer, Department of CSE Daffodil International University, Dhaka and co-supervisor Zarin Tasnim Shejuti, Lecturer, Department of CSE, Daffodil International University, Dhaka**. We would like to thank the deep knowledge & keen interest of our supervisor in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor and Head, Department of CSE and Dr. S. M. Aminul Haque, Associate Professor and Associate Head Department of CSE** for their kind help to finish our project and also to all of the faculty members and the staff of CSE department of Daffodil International University for kind support and help on the technical and the administrative aspect of the study.

We would like to thank all of our course mates in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge the constant support and patience of our parents. They have been with us our entire lives, to support and encourage us in all of our endeavors.

ABSTRACT

Nowadays the world become a digitalization. That's why now security problems are rising every day. We have too much concern about the security system. So human activity detection's significance is also increased argent. Human activity detection (HAD) is a significant series of data in the computer vision community. Nowadays we are utilizing CCTV cameras, Smartphone cameras for reliability plan. Human action acknowledgment is worried about recognizing various kinds of human developments and activities utilizing information accumulated from different ways. In this project report, we are proposing an improved DCNN that can recognize eating, walking, working, playing, fighting, and other types of real-time human activity are being from images. Using the Deep Convolutional Neural Network model (DCNN) images were fed for image classification. Then, by means of joining the DCNN model with a custom human action identification dataset, limits and new attaching is one great step. The benefit of an improved Deep Convolution Neural Network or (DCNN) is its capacity to separate attributes from the data. Transfer Learning was used to feature extract the images and also the methodology we used is transfer learning. Used Keras framework to train the images. In our Project, we used Keras and TensorFlow as a framework. Moreover, we have contrasted the further improved DCNN model and other conventional techniques, and here the improved DCNN model accomplished an accuracy pace of 98.82% and outperforms different models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

The human activity Detection (HAD) task is one of the most significant problems and afterward decides or predicts conditions of activity or behavior. It has different applications, traversing from movement understanding for wise reconnaissance frameworks to further developing human-PC connections. The undertaking of Human Action Detection (HAD) is to arrange body signals or movement, and afterward, decide or anticipate conditions of activity or conduct.

Here to examine the item principally zeroing in on Further developed CNN for action discovery. So by utilizing a CNN strategy distinguish the Human movement and afterward break down the exercises to identify the people. To distinguish people from the exercises, we need a pre-prepared arrangement of pictures for correlation purposes. Particularly, in general, Walking, Eating, working with the pervasion of portable personal digital devices such as smartphones, and multi-media terminals, creating an extraordinary number of various kinds of constant information, for example, video recorders, photographs. There will be a critical need for individual customization utilizing human activity recognition [1]

Applying Improved Deep Convolution neural networks for our experiments, we collected image data from video data of nine human subjects performing nine different activities in sudden places: Eating, Walking, Working, fighting, Skating, Hospital, Playing different types such as Basketball, Baseball, and Football. We execute Forms of the DCNN and compare they are produced with a baseline support Vector Machine (SVM) that carries out the high-quality highlights [2] . Firstly DCNN is the one who trained the customer data. Secondly, DCNN is the Transfer learned Improved DCNN, in particular, we take a pre-prepared DCNN, which is prepared on a different, wide scope RGB picture grouping dataset, ImageNet, and fine-tune the organization boundaries utilizing the custom data. Improved DCNN model combination of Resnet50 and Transfer Learning for better results.

Then we find out outstanding accuracy for our model. In the accompanying segments, we summarize our reproduction study and custom data process, clarify the Improved DCNN preparation in additional detail, and present better results. We develop a neural organization design that empowers us to construct a customized HAR model with negligible human management [3] .

1.2 Motivation

As the world has been going through an information age, there is a corresponding increase in different types of activities committed. This causes concerns about the safety of our society. We will use our object and abnormality recognition system so that all the kinds of human activity especially criminal activity or human behavior we can easily detect. Not criminal activity but also detect all the types of human activity like health, working, playing, eating etc. Our motto is “detect all the kinds of human behavior”. We also believe that the electronic presence of an effective CCTV system will detect all behave and deter humans from the activities that they actually do. Even a simple “Monitored by CCTV” sign will deter human activity.

1.3 Rationale of the Study

This study was conducted to identify persons of interest using images from a database and learn how to classify any object the person is holding. We also studied how to identify situations like walking, working, playing, eating, and lastly some basic situation of the hospital. We also trained our models to identify what kind of playing human play. We have worked on this research to help human beings as well as the community. Scientist will find support from our review that which methodology gives faster output and better accuracy.

1.4 Research Questions

i. How do we identify the human working?
ii. How do we identify human walking?
iii. How do we identify the situation in the hospital?
iv. How do we identify human skating?

v. How do we identify human fighting and eating?
vi. How do we identify human Playing like Baseball, Basketball, and Football?
vii. What dataset do we use?
viii. What methodology do we use?
ix. What framework will we use?

1.5 Main Objective

i. To propose an improved DCNN model which can swiftly identify a human activity.
ii. Identify the situation in an image.
iii. Using this system to reduce the workload and time needed to capture a person that can be used in other tasks instead.

1.6 Report Layout

Our project report is coordinated as follows:

i. Chapter One includes the introduction to our research, motivation for this study, objectives, and main objective of this study.
ii. Chapter Two includes the background of the research, related works, research summary and challenges faced during this study.
iii. Chapter Three includes research methodology, the proposed systems, datasets, the implementation procedure, data preprocessing and the improved model.
iv. Chapter Four includes Experimental Results and Discussion including experimental setup, confusion matrix, performance and comparative analysis.
v. Chapter five includes the Conclusion and Future Scope of this project.
vi. Lastly the References section contains the resources used during the research.

CHAPTER 2

BACKGROUND

2.1 Introduction

The current period of our human activity progress can be known as the time of data, and the vitally innovative drive behind it is the computer and the internet. We have used DCNN and Resnet50 architecture and Transfer Learning in our research program. DCNN is a kind of particular neural network used for preparing picture information. DCNN is generally used for picture detection and classification. Resnet50 is a famous deep learning model used for picture recognition. ResNet-50 is a convolutional neural network that is 50 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database. It can be used to add extra layers in deep neural networks which can improve accuracy and performance. The feature of taking the information or components of one problem and applying that to take care of another problem is called transfer learning [1] . Like taking a product used to recognize unpredictable bread shapes to distinguish malignant growth shapes. Transfer learning is an open-source structure for profound learning [5] .

2.2 Related Works

There has been a large number of research done on Convolutional neural networks and object, face, and scenario recognition. We've reviewed a few of these research papers and have taken some ideas from them.

Jinyong Pang et a1 proposed a Human Activity Recognition Based on Transfer Learning that utilizes CNN to dissect the extricated features [2]. They discovered larger part casting a ballot was the most reliable out of a gathering comprising of ResNet-50, AlexNet, VGG-19, GoogleNet, and VGG-16 [2] . Elnaz Soleimania, Ehsan Nazerfardb et a1 [3] ProposedTransfer Learning-based a Cross-Subject in Human Activity Detection Systems utilizing Generative Adversarial Networks that can recognize human activity from still pictures. Here they have utilized the LFW picture dataset and MOBIO video dataset. Here, normal MRH is utilized for video datasets and it is quicker than the looked at ones. The picture dataset gives 92.59% precision for 7 pictures and takes .76 seconds to handle 41 casings. Its destruction is its exactness is very not exactly the thought about ones [3] .

Li Wei and Shishir K. Shah et a1. [4] Proposed a Human Activity Recognition utilizing Deep Neural Network with Contextual Information that focuses on two particular elements are considered in this acknowledgment organization, the principal in light of human motion and the second based on the context recognizing highlights from restricted biometric content in low-goal pictures and matching it to data rich high-goal pictures. The proposed calculation yields better outcomes for lower goals [4] . Kandethody Ramachandran and Jinyong Pang et a1. [5] Proposed a Transfer Learning Technique for Human Activity Recognition based on Smartphone Data. Fix-based the problem of classifying various patterns. It has applications in health care and also has huge commercial benefits. However, it performs better for human activity detection and adjusting pictures. The model gives 92.3% accuracy. The weighted prediction precision of recognizing six human activities is 94% on the PIE light dataset for different mathematical varieties. For the video dataset, they have utilized MRH and LBP include extraction combined with MSM for grouping pictures.

Renjie Ding et al. [6] analyzed human movement acknowledgment (HAR) in view of sensor information as a critical issue in broad figuring. This task thought about a few generally utilized calculations and tracked down that the Maximum Mean Discrepancy (MMD) technique is generally reasonable for HAR. However it gives a superior consequence of around 85% precision for a solitary individual picture, it gets diminished for a gathering of individuals discovery which has an exactness of around 77%. Josepd Redmon et al. [7] research the new headway way to deal with object identification. Earlier work on object discovery reuses classifiers to perform location of CCTV video progression dealing with. They investigated Pre-Processing, Conversion to Grayscale, Background Subtraction, Image Segmentation, Image Enhancement, Object Detection, Region of Interest, and the method for picture dealing with.

Marc Kurz et al. [8] have explored Continuous Transfer and Evaluation of Activity Recognition. This paper depicts and assesses the difficult element of a crafty movement acknowledgment framework to prepare a newfound sensor with the accessible detecting gadgets to perceive exercises at runtime. It has lower computational time appeared differently in relation to a Google-Net Inception-V3 model. The principal commitment of the paper is the assessment of the methodology by portraying an exploratory arrangement and introducing brings about terms of accuracy rate from the machine learning point of view also as from the structure point view by looking at predicted classes from the teaching sensor set and the recently prepared sensor to acquire QoS parameters. Nils Y. Hammerla et al. [9] have proposed a framework that is utilized as a Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables that can identify human activity from smartphone cameras which are utilized for participation. For human detection, Multi-task CNN is utilized here.

Ejaz Ul Haq et al. [10] have proposed a Human detection framework in CCTV Systems. The proposed Human detection and tracking is a vital viewpoint in observation frameworks because of its significance in conveniently recognizable proof of people, acknowledgment of human action, and scene investigation. Ajai V Babu et al. [11] have proposed a novel human detection morphology method to detect human presence. The creators utilized Features from the Accelerated Segment Test for highlight identification and Multi-Resolution Analysis for characterization. This model gives 93.78% precision. Where HOG gave 59.89% precision. Here, the structure can recognize just four edges each subsequent which can be worked on further.

Martin Milenkoski et al. [12] have proposed Real-Time Human Activity Recognition on Smartphones using LSTM Networks. Here, classifier and max-margin human activity recognition with CNN-based features are utilized for human motion identification from the picture. Develop another lightweight algorithm for activity detection based on Long Short Term Memory networks, which can take in features from crude accelerometer information, totally bypassing the process of producing hand-created features. Shaohua Wan et al. [13] have proposed a Deep Learning Model for Real-time Human Activity Recognition that is ready with a progressing dataset made by photograph shooting. This paper plans a cell phone inertial accelerometer-based engineering for HAR.HAR research has hung out considering the advantages of the limitless application in canny perception structures, clinical benefits frameworks, computer-generated reality interchanges, shrewd homes, strange conduct locations, and different fields. CNN, LSTM, BLSTM, and MLP were prepared at a learning pace of 0.001 with 200 emphases. The dataset was then ordered by SVM. For classification, CNN based model provided 96.36% accuracy where LSTM based accuracy gave only 92.73% accuracy.

N.Albukhary and Y.M. Mustafah et al. [14] have deep learning-based model which can Real-time Human Activity Recognition. The creators have utilized an adjusted human activity dataset that contains 9040 colored pictures. However, they just took 1600 pictures from the dataset. In the proposed technique, they have utilized Multitask Cascade CNN for human activity identification and implanted utilizing the feature. This proposed model gives 92.56% accuracy in the event that we give 19 pictures of that individual and 85.72% precision in the event that we give 30 pictures of that individual.

2.3 Limitations of existing work

The limitations of these papers are that some of them work only with text data, some of them work with image data and some of them deal with video data. But they did not work to convert video data into image data. Most existing methodologies for human activity recognition principally use CNN or transfer learning and some are using Deep learning method but they have not worked on improved CNN anywhere. Level of accuracy of these paper is less than our paper. Because of we used improved CNN. We have discovered that the human detection and activity recognition part of our intended process can be done by many different methods of machine learning [4]. In any case, the most practical choice for us is the open-source ImageNet like Resnet50, transfer learning, and Convolutional Neural Network that have freely accessible documentation and codes. We likewise expected to make our own custom dataset for this task.

2.4 Scope of the Problem

The objective of this research-based project is to make a system that can separate features from an image to recognize from the record. The system likewise needs to extract features from images that recognize any fighting people or ill people in the photos, the target that we can notify the authorities. The system additionally needs to remove a few qualities to identify if there is nothing in the picture. All the pictures were also labeled in our project.

2.5 Challenges

Data collection: There may be some valid requirements in the use of human recognition innovation in a public spot. We need to talk with the legal authorities before we can carry out this framework in a public setting.

Data storage: There are restrictions to how much data can be stored from multiple cameras for a lengthy period of time. So we might have to carry out an information scouring method each 3-5 months depending on the capacity size accessible.

Calculation limitations: There are restrictions to how quickly information can be prepared and items and situations recognized. This would then be able to make a human activity be detected well after they have left the area. We can correct this by utilizing cloud computing for our framework.

Hardware limitations: There can be restrictions in our processing hardware where they will linger behind the captured record. This can undoubtedly be amended by adding more powerful hardware like CPU, HDD, or GPU to our architecture. Truth be told, we have dealt with this issue in our preparation of the model where we needed to leave our computer on for a couple of days.

Model limitations: From our research, we have seen that a few models will underperform to meet expectations in either human recognition or activity detection. We have attempted to redress this by using CNN and Resnet50 and Transfer learning models.

Weather and Light: The performance of the framework can be seriously affected by unfavorable climate conditions like tropical storms, weighty downpours, or dense fog. There may be limitations on identifying objects in a low-light climate like during the evening or night. This can be solved by using Infrared Cameras during the evening or having some amazing lights in the field of vision of the cameras. Also, gathering area is difficult to detect human activity.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Our research goal is to detect human scenes and identify activity. We have used DCNN or Deep Convolution neural network algorithms which we have implemented using transfer learning. For training our dataset we have used a custom dataset of images. The research of human action detection (HAD) is to portray body movements or development, and thereafter choose or foresee conditions of activity or conduct. Its spacious applications, showing up in medical services, actual recuperation from incapacity or injury, and clinical distortion remedy, or any fighting place, are causing increasingly more to notice the further turn of events and abuse from industry and academe. Particularly, in general, medical services, with the invasion of convenient individual computerized gadgets, for example, smartphones, smartwatches, and multi-media terminals, creating an amazing number of various sorts of long term information, for instance, video recorders, photos moves, and spatial-common logs, there will be a critical prerequisite for individual customization utilizing human action Detection [22] .

3.2 Proposed System

As our research base project objective is to detect human activity, we have used Deep CNN, transfer learning, and Resnet50. Transfer learning and Resnet50 based on improved DCNN model to recognize the people activity. Finally, we analyzed these three models and improved DCNN outflanks different models. From the beginning, we have used a Deep CNN model with four different filters and used ReLU and Softmax activation layer. Then, we have used a pre-trained model as Resnet50 to extricate the features from pictures, train this pre-trained model and get accuracy. At long last, we construct our transfer learning-based improved DCNN model where pre-trained model Resnet50 worked as a feature extractor and Deep CNN worked as a trainable model.

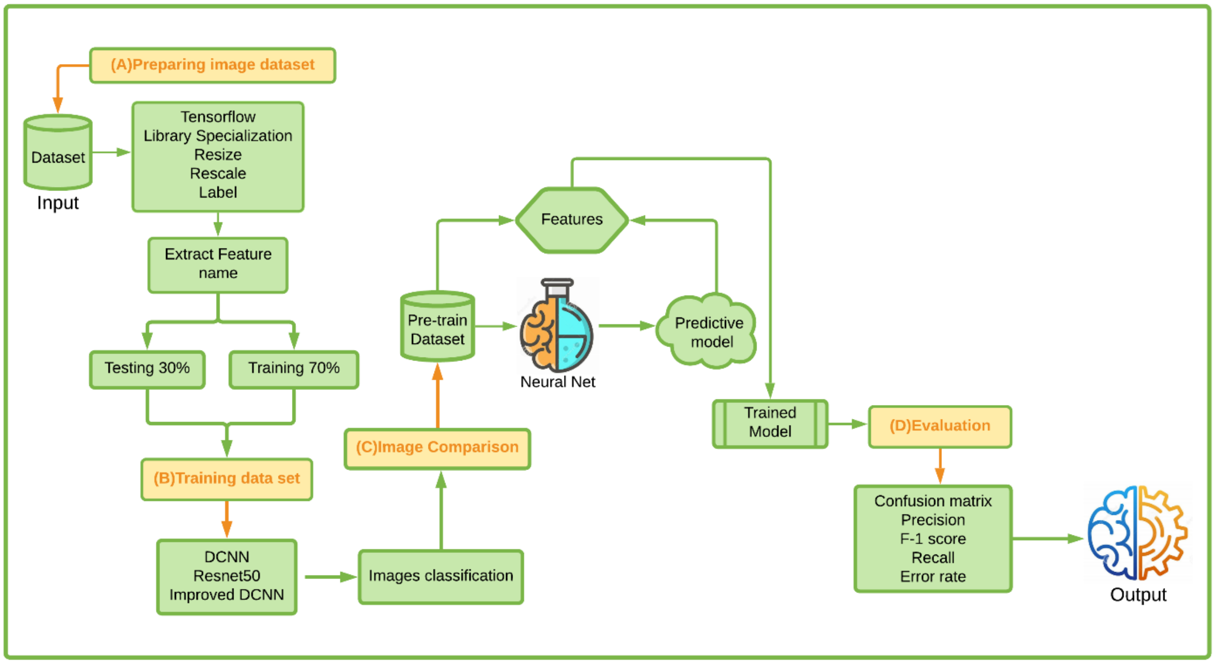

Figure 1: Working method of proposed methodology.

3.3 Implementation Procedure

3.3.1 Dataset

The initial step of carrying out our model is collecting data. As our model requires a picture dataset, we have taken pictures from the video data as custom datasets. Firstly take videos data set then convert video dataset to the image dataset. Every one-second sequent chapter saving is an image data from video data. There are 9 classes taking all things together. They are mainly pictures of eating, walking, working, some kind of playing such as football, baseball, basketball, the situation of hospital, and fighting types. Our dataset comprises 3850 pictures. Our all dataset is a custom dataset and some images in the training and testing datasets. Our raw images were high-quality images and for that reason, it set aside somewhat long effort for stacking information.

3.3.2 Data preprocessing

Our dataset contains pictures of the different sizes. Which will cause an issue while training it. So to stay away from that we have transformed and re-sized the pictures so every one of the pictures is of a similar size where we used picture size 120. Additionally, as we utilized the DCNN model for preparing our dataset, it requires a lot of pictures. We had sufficient data to train our dataset. We have also labeled our images and 3 channel color images (RGB) used in the pre-processing data part. That's why to get improved performance and quality pictures to train the dataset. It is a process of expanding the measure of knowledge by adding somewhat adjusted information into the dataset. It helps with limiting data waste while keeping up with great execution.

3.3.3 Deep Convolutional Neural Networks (DCNN)

Deep Convolutional Neural network or DCNN is a kind of artificial neural network, which is generally utilized for picture/object classification and detection as a result of their high accuracy. Deep learning is a subfield of transfer machine realizing which is generally utilized for big data and picture processing. Here, we will consider DCNN or Deep Convolution Neural network which we have utilized in our proposed model. One of the main benefits is that it can naturally identify important elements without any human oversight. Moreover, DCNN stays away from the requirement for complex picture preprocessing because they can utilize the real picture directly as input. DCNN is mainly composed of a convolutional layer, a pooled layer, and a completely associated layer. Convolutional layers are the layers where filters are applied to the original image, or to other feature maps in a deep CNN. Convolutional layers use different types of parameters like the number of kernels and the size of the kernels, Activation function, stride, and padding. The most significant parameters are the size of the kernels and the number of kernels. By DCNNs to execute operations on items used a ReLU layer. The result is a corrected feature map. ReLU layers work obviously better in light of the fact that the network can prepare much quicker without having a critical effect on the exactness. The ReLU layer applies the capacity $f(x) = max(0, x)$ to each of the qualities in the information volume. In fundamental terms, this layer simply changes every one of the negative enactments to 0. This layer expands the nonlinear properties of the model and the general organization without influencing the open fields of the Convolutional layer. Pooling layers are like convolutional layers, yet they fill a particular role, for example, max pooling, which takes the maximum value in a certain filter region, or average pooling, which takes the average value in a filter region. Pooling layers use two parameters stride and size of the window. The main function of the fully connected layer is to incorporate the different picture maps got later the picture is gone through various convolution layers and pooling layers to acquire the high-layer semantic elements of the picture for ensuing picture characterization. Fully connected layers are used when the pooling layer's flatten matrix is input and placed before the classification output of a DCNN.

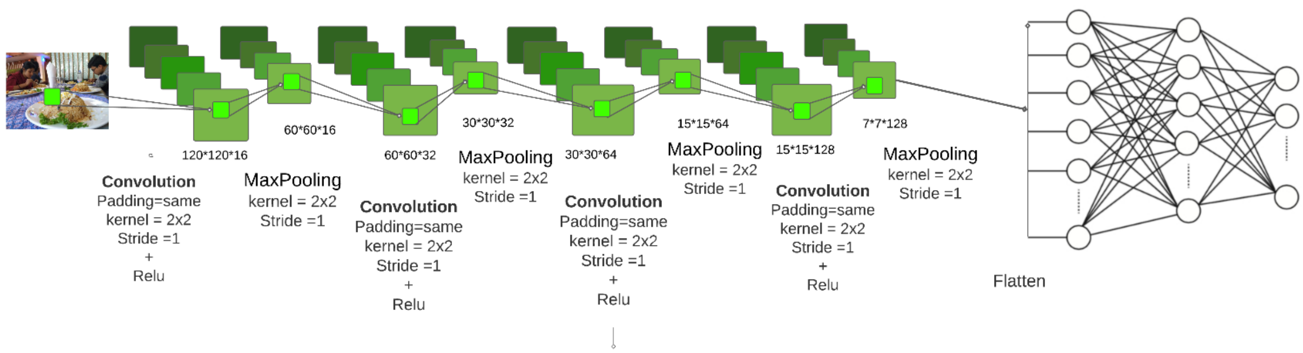

Figure 2: Working prototype of DCNN.

3.3.3.1 Model Training

Firstly create our own images dataset from video data. This dataset contains 4000 images extension (png). There are approximately 3700 images grouped by 9 different folders. The classes name Eating, Walking, Working, Fighting, Skating, some situation in the hospital Playing 3 types like football, Basketball, Baseball. Then prepare our dataset for training will involve creating labels and resizing our images 120 X 120 by using a convolution layer and ReLU. Then assigning labels and features. Split X and Y for use in CNN. Using the pooling layer compiling and training the CNN model. Finally, fully connected layers are used to find out the accuracy and score of the model.

3.3.3.1 Model Evaluation

According to the calculation of a confusion matrix, we have to know about TP, FP, TN and FN.

True Positive (TP): The number of instances that a model predicts accurately with the end goal that the Actual Labels are Positive and Anticipated Labels are Positive.TP (true positive) = accurately predicated classes.

A false positive is an error in a double grouping where an experimental outcome inaccurately demonstrates the presence of a condition. FP (false positive) = Actuating negative classes as positive.

False-negative is the contrary mistake, where the experimental outcome mistakenly demonstrates the absence of a condition when it is really present. FN (false negative) = Actuating negative classes as negative.

A true negative is a result where the model accurately predicts the negative class. TN (true negative) = Actuating negative classes as negative.

Using the above variables, we have calculated the values of accuracy, F1-score (micro), precision (micro) [14] .

Accuracy: It's the proportion of right forecasts to add up to predictions made. Accuracy for a multiclass confusion matrix is the average number of right expectations.

Here, accuracy = (total TP + total TN) / (total TP + total TN +total FP + total FN).

F1 Score: It's the weighted average of precision and Recall. Hence, this score considers both false positives and false negatives. On the off chance that the expense of false positives and false negatives are totally different, it's better to check out both precision and Recall. For multiclass confusion we utilize a miniature f1-score.

Here, f1-score = (2*total TP / (2*total TP +total FP + total FN))

The precision: It's determined as the proportion between the numbers of Positive examples effectively characterized to the absolute number of tests classified Positive. The accuracy estimates the model's exactness in characterizing a simple positive.

Here, Precision = Total TP / (total TP + total FN)

Recall: It's the number of pertinent archives recovered by a search divided by the absolute number of existing significant records.

The error rate: It's communicated as a ratio and is determined by dividing the total number of words read by the complete number of errors made. Error rate indicates the fraction of incorrect predictions.

Here, E = 1- accuracy.

We trained the model to find accuracy, f1-score, error-rate and precision of the model for our dataset. It took 15/26 minutes for each epoch.

Table 1: Accuracy after training with DCNN model

epoch	train loss	value loss	val_accuracy	val_recall	val_precision	Time(s)
0	0.6834	0.0511	0.2412	0.2412	0.2412	574
1	0.2647	0.0375	0.2147	0.2147	0.2147	564
2	0.8155	0.0187	0.1176	0.1176	0.1176	561
3	1.2797	0.0344	0.0824	0.0824	0.0824	549
4	0.2281	0.0285	0.1441	0.1441	0.1441	550
5	0.0477	0.0382	0.2412	0.2412	0.2412	556
6	0.0164	0.0281	0.0824	0.0824	0.0824	556
7	0.0063	0.0453	0.0824	0.0824	0.0824	550
8	0.0021	0.0656	0.0824	0.0824	0.0824	551
9	0.0087	0.0588	0.0824	0.0824	0.0824	566

In Table 1, however, the accuracy was acceptable, the model had massive training and validation loss. So our model has space for development. After checking Figure 2, we can say it didn't function as well as it should be.

3.3.4 RESNET 50

ResNet-50 is a pre-trained Deep Learning model of the Convolutional Neural Network (CNN) for picture arrangement, which is most generally applied to investigating optical symbolism and it is a class of deep neural networks. ResNet-50 is trained on 1,000,000 pictures of 1000 classes from the ImageNet database and it will be 50 layers deep. Also, the model has more than 23 million workable boundaries, which indicates a deep design that further develops it for picture acknowledgment. Utilizing a pre-trained model is an exceptionally viable methodology, compared to on the off chance that you want to construct it without any preparation, where you want to gather incredible measures of data and train it yourself. Obviously, there are other pre-trained deep models to utilize like AlexNet, GoogleNet, or VGG19, yet the ResNet-50 is noted for brilliant speculation execution with fewer blunder rates on acknowledgment assignments and is, subsequently, a valuable device to know. As the pre-trained Resnet50 model's features transfer into the Deep CNN model it provided better performance. We didn't need to train the layers of Resnet50 to create transfer learning but extracted the features from it and we also add 3 more custom layers with this model. These three custom layers here we add that are ReLU and softmax activation function with a dropout function of 0.4. Our model has advanced the parameters and changed the pre-trained model to Resnet50 (Without tweaking) this is our give further improvement model. We will find better outcomes when we apply more than 50 epochs. We get actually improve the outcome than the previous table after decreasing the epochs.

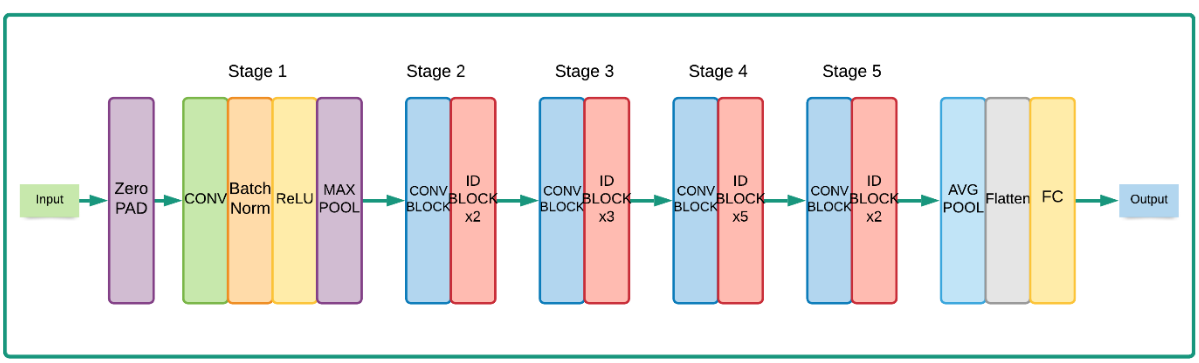

Figure 3: Resnet50 model's architecture.

3.3.4.1 Model Training

ResNet-50 is a convolutional neural network that is 50 layers deep. We can load a pre-trained version of the network trained on more than 1,000,000 images from the ImageNet database. The pre-trained network can classify pictures into 1000 item classes. For ResNet 50 firstly we have to import resnet50 from TensorFlow. Then declare the image shape and declare weight is none. After that connect the dataset and split the train set into training and validation sets. Finally fit the created resnet50 model.

3.3.4.2 Model Evaluation

According to the calculation of a confusion matrix, we have to know about TP, FP, TN and FN.

True Positive (TP): The number of instances that a model predicts accurately with the end goal that the Actual Labels are Positive and Anticipated Labels are Positive.TP (true positive) = accurately predicated classes.

A false positive is an error in a double grouping where an experimental outcome inaccurately demonstrates the presence of a condition. FP (false positive) = Actuating negative classes as positive.

False-negative is the contrary mistake, where the experimental outcome mistakenly demonstrates the absence of a condition when it is really present. FN (false negative) = Actuating negative classes as negative.

A true negative is a result where the model accurately predicts the negative class. TN (true negative) = Actuating negative classes as negative.

Using the above variables, we have calculated the values of accuracy, F1-score (micro), precision (micro) [14] .

Accuracy: It’s the proportion of right forecasts to add up to predictions made. Accuracy for a multiclass confusion matrix is the average number of right expectations.

Here, accuracy = (total TP + total TN) / (total TP + total TN +total FP + total FN).

F1 Score: It's the weighted average of precision and Recall. Hence, this score considers both false positives and false negatives. On the off chance that the expense of false positives and false negatives are totally different, it's better to check out both precision and Recall. For multiclass confusion we utilize a miniature f1-score.

Here, f1-score = (2*total TP / (2*total TP +total FP + total FN))

The precision: It's determined as the proportion between the numbers of Positive examples effectively characterized to the absolute number of tests classified Positive. The accuracy estimates the model's exactness in characterizing a simple positive.

Here, Precision = Total TP / (total TP + total FN)

Recall: It's the number of pertinent archives recovered by a search divided by the absolute number of existing significant records.

The error rate: It's communicated as a ratio and is determined by dividing the total number of words read by the complete number of errors made. Error rate indicates the fraction of incorrect predictions.

Here, E = 1- accuracy.

We trained the model to find accuracy, f1-score, error-rate and precision of the model for our dataset. It took 15/26 minutes for each epoch.

Table 2: Performance after training with ResNet-50

epoch	train loss	Val accuracy	Val recall	Val precision	time
0	0.6834	0.2412	0.2412	0.2412	00:21
1	0.2647	0.2147	0.2147	0.2147	00:18
2	0.8155	0.1176	0.1176	0.1176	00:17
3	1.2797	0.0824	0.0824	0.0824	00:15
4	0.2281	0.1441	0.1441	0.1441	00:18
5	0.0477	0.2412	0.2412	0.2412	00:15
6	0.0164	0.0824	0.0824	0.0824	00:15
7	0.0063	0.0824	0.0824	0.0824	00:15
8	0.0021	0.0824	0.0824	0.0824	00:15
9	0.0087	0.0824	0.0824	0.0824	00:15

The accuracy was acceptable, the model had massive training and validation loss. So our model has space for development. After checking Figure 3, we can say it didn't function as well as it should be.

3.3.5 Improved DCNN

In our proposed model for better performance, we implemented Improved DCNN the combination with resnet50 and transfer learning. Transfer learning is an exceptionally supportive instrument in this review, improving/ flourishing the performance of Human Activity Detection (HAD) system. In the field of machine learning, introducing transfer learning algorithm transfer learning calculation would make a leap forward on the normal assumption that a preparation dataset should be of a similar source as a future testing dataset, display two datasets are interchangeable dispersed. For those last datasets gathered from various distributions with different components in comparative mission, Transfer learning could get ready conventional machine-learning algorithms to have an extraordinary handle of new information from future datasets from one more distribution by reusing past preprocessed data. This is a fundamental capacity for machine learning dependent on transfer learning getting the hang of, decreasing the expense of naming new information, retraining new models and computational assets. Transfer Learning started from the investigation of various assignment learning models, zeroing in on learning typical or lethargic quantifiable parts from both source and target undertakings in playing out different tasks.. A more clear meaning of transfer learning was from the Guard Progressed. Research Activities Organization's Information Handling Innovation Office that holds the information from single or different information source tasks and uses the legitimate information a pointed objective undertaking is called transfer learning. Unique in relation to previous ideas, the new definition focused more on track assignments without the limit of finding normal elements from various tasks. The goal of transfer learning is to transfer information between related sources and target areas. As such, transfer learning can utilize the information learned from unique source material to another comparable material in another climate, expanding models transformation and application in different associated information just as assignments that are similar.

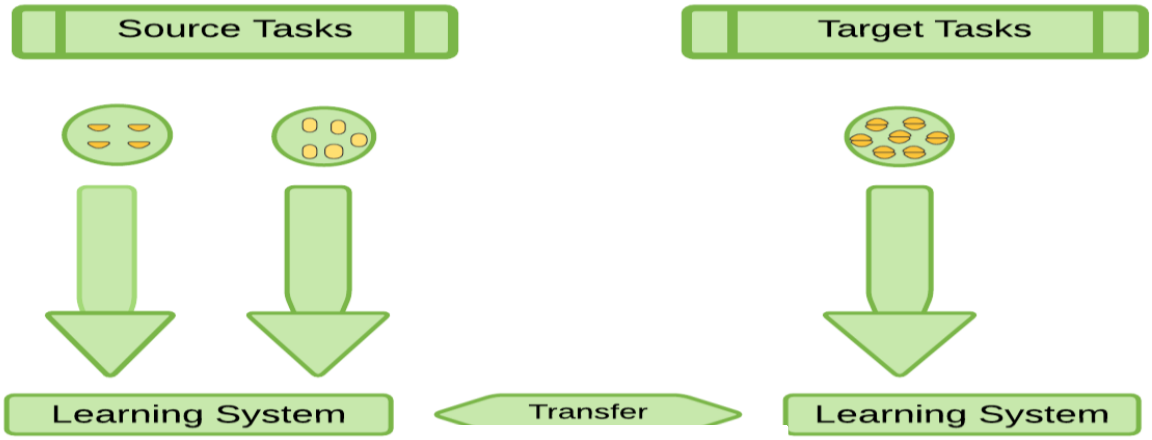

Figure 4: Working prototype of Transfer Learning.

In the Human activity recognition (HAR) task, transfer learning could customize the task of perceiving of perceiving behaviors in a brief time frame with a higher expectation accuracy. Based on transfer learning, another Human activity recognition (HAR) framework would be created without a lot of cost of time, computational assess, and extra data processing. Transfer learning provides strong accuracy in our model. In our proposed model we have utilized transfer learning to get the features from it. From the pre-trained model, we have taken the outcome and accumulated it with our Deep Convolution Neural Network (DCNN). For additional improvement, we have advanced the parameters and changed the pre-trained model to ResNet-50 (Without tweaking). We will find better outcomes when we apply more than 50 epochs [17] . Even after decreasing the ages to 20, we actually improve results than the past table. Another speed record for training ResNet-50 in just 224 seconds (three minutes and 44 seconds) with 75% accuracy utilizing 2,100 NVIDIA Tesla V100 Tensor Center GPUs. This achievement addresses the fastest revealed training time at any point distributed on ResNet-50. We actually improve results than Table 2.

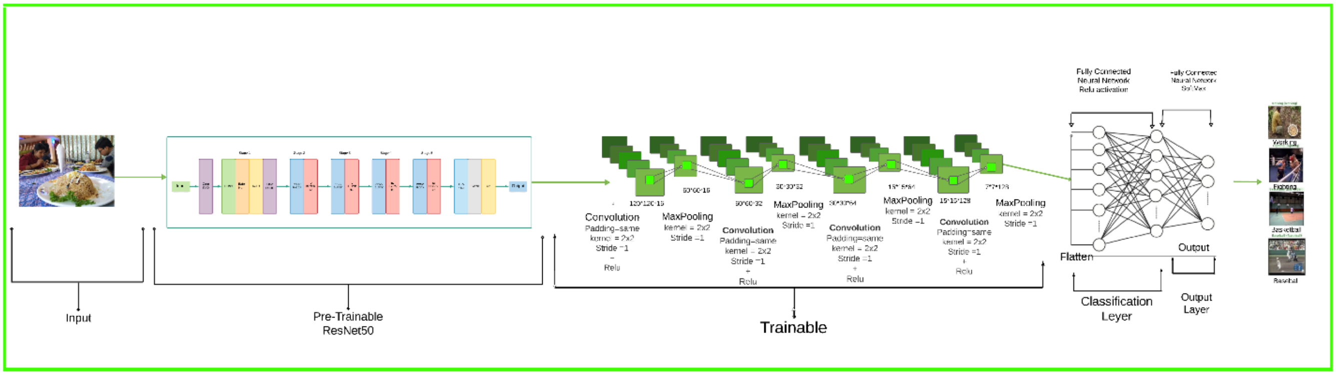

Figure 5: Working prototype of improved DCNN model.

3.3.5.1 Model Training

Loaded data utilizing python libraries. This means we are imported all libraries that are required. After that we are downloaded resnet50 with image shape and weights is "imagenet".Preprocess of data incorporates reshaping, one-hot encoding, and parting. Developing the model layers of CNN followed by model assembling, model preparing. Afterward, we create an Improved CNN model with the combination of resnet50 and transfer Learning. Assessing the model on test data. At last, predicting the right and inaccurate labels.

3.3.5.2 Model Evaluation

According to the calculation of a confusion matrix, we have to know about TP, FP, TN and FN.

True Positive (TP): The number of instances that a model predicts accurately with the end goal that the Actual Labels are Positive and Anticipated Labels are Positive.TP (true positive) = accurately predicated classes.

A false positive is an error in a double grouping where an experimental outcome inaccurately demonstrates the presence of a condition. FP (false positive) = Actuating negative classes as positive.

False-negative is the contrary mistake, where the experimental outcome mistakenly demonstrates the absence of a condition when it is really present. FN (false negative) = Actuating negative classes as negative.

A true negative is a result where the model accurately predicts the negative class. TN (true negative) = Actuating negative classes as negative.

Using the above variables, we have calculated the values of accuracy, F1-score (micro), precision (micro) [14] .

Accuracy: It's the proportion of right forecasts to add up to predictions made. Accuracy for a multiclass confusion matrix is the average number of right expectations.

Here, accuracy = (total TP + total TN) / (total TP + total TN +total FP + total FN).

F1 Score: It's the weighted average of precision and Recall. Hence, this score considers both false positives and false negatives. On the off chance that the expense of false positives and false negatives are totally different, it's better to check out both precision and Recall. For multiclass confusion we utilize a miniature f1-score.

Here, f1-score = (2*total TP / (2*total TP +total FP + total FN))

The precision: It's determined as the proportion between the numbers of Positive examples effectively characterized to the absolute number of tests classified Positive. The accuracy estimates the model's exactness in characterizing a simple positive.

Here, Precision = Total TP / (total TP + total FN)

Recall: It's the number of pertinent archives recovered by a search divided by the absolute number of existing significant records.

The error rate: It's communicated as a ratio and is determined by dividing the total number of words read by the complete number of errors made. Error rate indicates the fraction of incorrect predictions. Here, E = 1- accuracy. We trained the model to find accuracy, f1-score, error-rate and precision of the model for our dataset. It took 15/26 minutes for each epoch.

Table 3: Performance transfer learning with Resnet50

epoch	train loss	Val_ loss	val_accuracy	val_recall	val_precision	time
0	0.2885	0.1673	0.9647	0.9294	0.9783	00:21
1	0.1330	0.0982	0.9794	0.9735	0.9822	00:18
2	0.0657	0.0622	0.9853	0.9853	0.9853	00:17
3	0.0486	0.0656	0.9794	0.9794	0.9794	00:15
4	0.0386	0.0463	0.9853	0.9824	0.9882	00:18
5	0.0210	0.0594	0.9794	0.9794	0.9828	00:15
6	0.0331	0.0541	0.9824	0.9794	0.9828	00:15
7	0.0225	0.0541	0.9853	0.9853	0.9882	00:15
8	0.0163	0.0586	09853	0.9853	0.9853	00:15
9	0.0176	0.0482	0.9882	0.9853	0.9882	00:15

As Resnet50 requires twice as much time as DCNN for preparing, so it will require more GPU and smash. So we may run out of memory. Right now, pretty much every NVidia GPU upholds tensor centers which can accelerate neural organization preparing by 2x-3x. They likewise require less GPU memory. In Table [3], it takes significantly additional time than Table 2 to pursue every epoch utilizing the tensor center component. Transfer learning takes more time to epoch 10 where the DCNN model takes l.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1.1 Experimental environment

An experiment is a controlled observational test of a hypothesis. The term environment refers to the state of a PC, controlled by a combination of software, basic equipment, and which programs are running. The experiment is directed on a personal PC where the microprocessor is 3.60 GHz Intel Core(TM) i3-8130U CPU 2.20GHz 2.21GHz, 8 GB RAM, and Windows 10 64 bits operating system working in the system. The model is implemented on the Tensorflow and Keras framework. The model is executed on an open-source stage Google Colab Notebook.

4.1.2 Parameter Optimization

During the reproduction of our proposed model, we have focused on certain boundaries to work on the presentation of our model. Here, we have expanded the batch size to 128. It expanded our model exhibition. However we utilized a learning rate between 1e-5 and 1e-4, it didn't work on our model exhibition. Indeed, even in the wake of attempting up to 20 epochs, it didn't improve develop our presentation that much. So we utilized 15 epochs where we focused on 10 epochs. On account of the performing work or activation function, we have utilized soft-max and ReLU as a classification algorithm. Which decreased our training loss.

4.1.3 Confusion Matrix

Human activity recognition (HAR) is turning into an increasingly more appealing examination theme with a few applications, like video observation, computer-generated reality, intelligent human-PC communications, and so forth Be that as it may, exact recognition of activities is an exceptionally moving errand because of jumbled foundations, impediments, and perspective varieties. Human Movement recognition (HAR) refers to the programmed identification of different proactive tasks performed by individuals in their regular routines. A HAR framework perceives the exercises performed by an individual and gives educational input to mediation. To get the best out of our model, we have prepared our model with various classes of pictures and utilized it as the last layer of our DCNN model. In order to get the exhibition of our model, we have utilized confusion matrices. After preparing with Resnet50 in transfer learning for better accuracy. So our further developed model is improved.

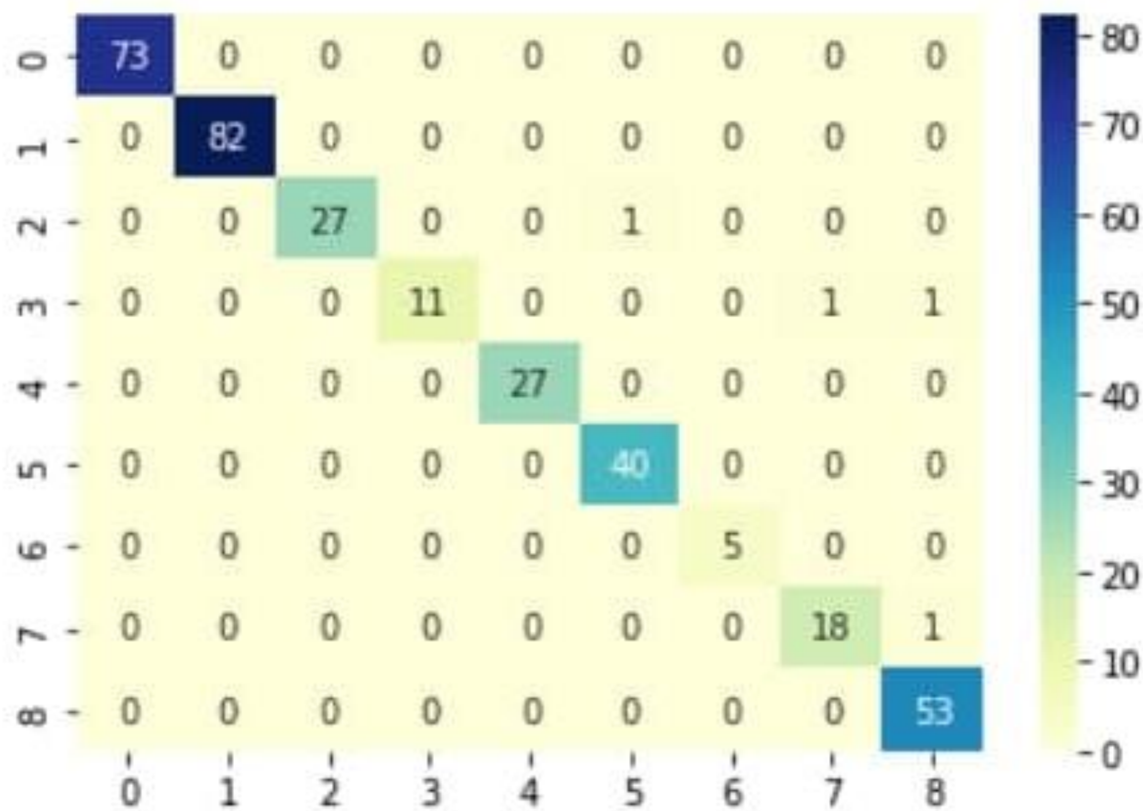

Figure 6: Confusion matrix of DCNN

We have shown the confusion matrix of the Resnet50 model after preparing the model with our given pictures. The inclining part demonstrates the number of pictures it has been detected accurately.

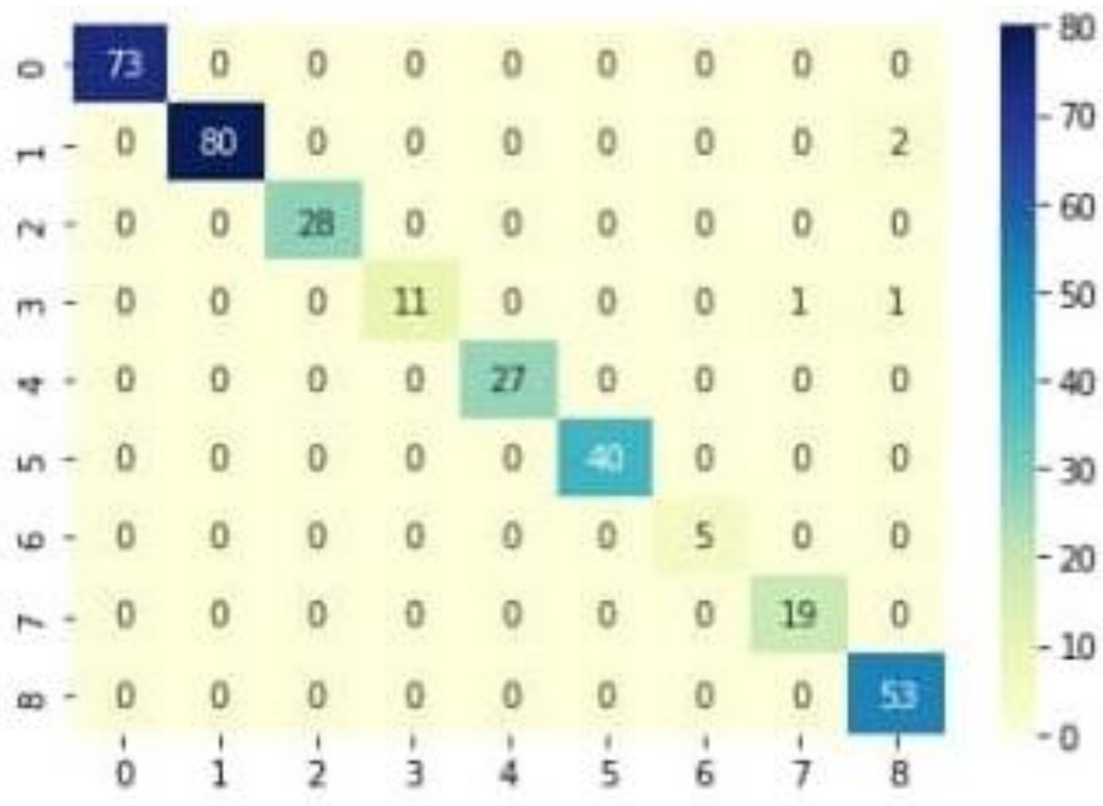

Figure 7: Confusion matrix of Resnet50

We have shown the confusion matrix of the Resnet50 model after preparing the model with our given pictures. The inclining part demonstrates the number of pictures it has been detected accurately.

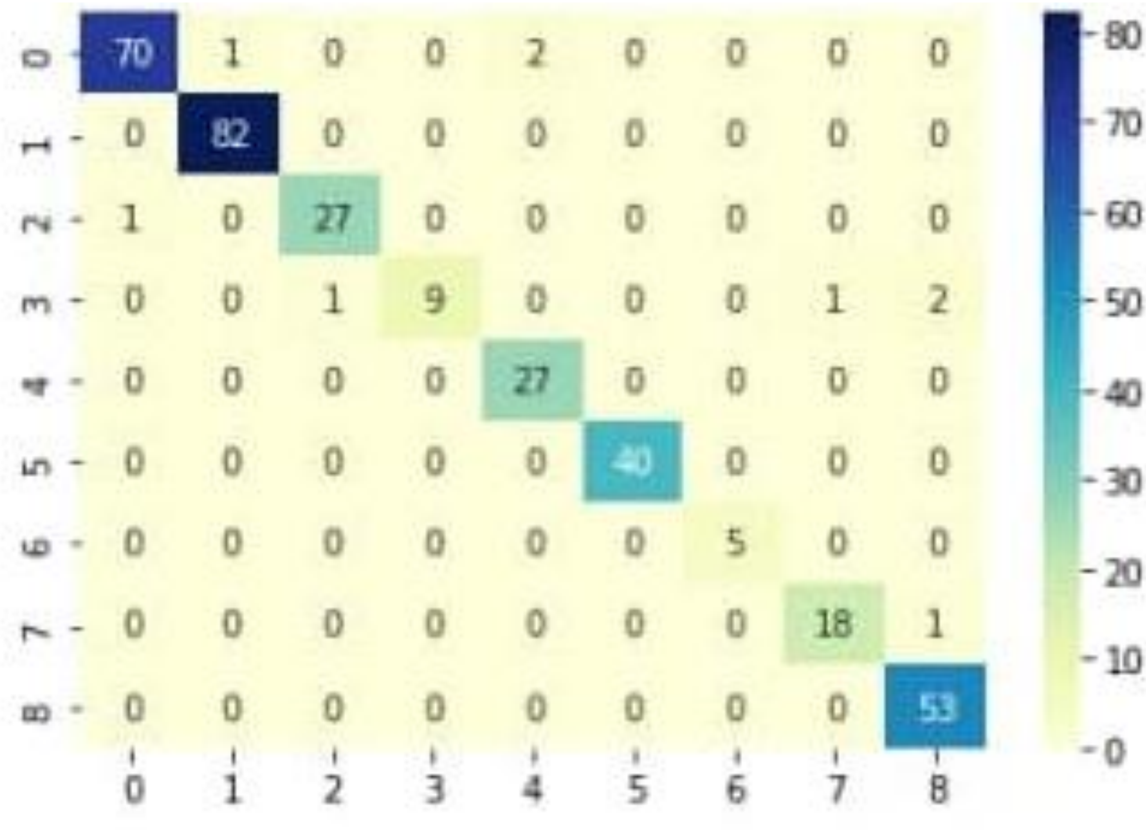

Figure 8: Confusion matrix of Improved DCNN

4.2 Performance Analysis

In our model, we proposed DCNN model based on transfer learning outperformed the pre-trained ResNet50 model and the general DCNN model approaches for human activity detection. Simple to implement in Keras and Tensorflow framework in DCNN model. It additionally can identify features from pictures consequently which makes it so natural to utilize. Assuming we analyzed the 3 unique models we utilized in our research-based project we can see that the performance of the transfer learning-based DCNN model is very strong. For multiclass picture classification, DCNN is the best model. It additionally requires some investment to run the system which is vital for a model like this. Similar analysis implies contrasting our model and other conventional ones. After that comparing accuracy with the ResNet50 model and an overall DCNN model, we get Table 4. Here, Table 4 gives detail of the result analysis. For multiclass or different picture Grouping Improved DCNN model is better than others models.

Table 4: comparison of Accuracy Precision, recall, f1-score for Deep CNN, ResNet-50, improved DCNN.

Class	Algorithms	Accuracy	Precision	Recall	F1 Score	Error Rate
Baseball	**DCNN**	98.82%	0.96	0.99	0.97	0.0118
	ResNet-50	100%	1.0	1.0	1.0	0.0
	Improved DCNN	100%	1.0	1.0	1.0	0.0
Basketball	**DCNN**	99.71%	1.0	0.99	0.99	0.0029
	ResNet-50	99.41%	0.98	1.0	0.99	0.59
	Improved DCNN	100%	1.0	1.0	1.0	0.0
Fighting	**DCNN**	99.41%	0.96	0.96	0.96	0.0059
	ResNet-50	100%	1.0	1.0	1.0	0.0
	Improved DCNN	99.71%	0.96	1.0	0.98	0.0029
Football playing	**DCNN**	98.82%	0.69	1.0	0.82	0.0118
	ResNet-50	99.41%	0.85	1.0	0.92	0.0059
	Improved DCNN	99.41%	0.85	1.0	0.92	0.0059
Hospital	**DCNN**	99.41%	1.0	0.93	0.96	0.0059
	ResNet-50	100%	1.0	1.0	1.0	0.0
	Improved DCNN	100%	1.0	1.0	1.0	0.0

Dataset	Algorithms	Accuracy	Precision	Recall	F1 Score	Error rate
Eating	**DCNN**	100%	1.0	1.0	1.0	0.0
	ResNet-50	100%	1.0	1.0	1.0	0.0
	Improved DCNN	99.71%	1.0	0.98	0.99	0.0029
Skating	**DCNN**	100%	1.0	1.0	1.0	0.0
	ResNet-50	100%	1.0	1.0	1.0	0.0
	Improved DCNN	100%	1.0	1.0	1.0	0.0
walking	**DCNN**	99.41%	0.95	0.95	0.95	0.0059
	ResNet-50	99.71%	1.0	0.95	0.97	0.0029
	Improved DCNN	99.41%	0.85	0.95	0.95	0.0059
working	**DCNN**	99.12%	1.0	0.95	0.97	0.0088
	ResNet-50	99.12%	1.0	0.95	0.97	0.0088
	Improved DCNN	99.41%	1.0	0.96	0.98	0.0059

Here we have shown Activity Detection which represents playing (Baseball, Basketball, Fighting, Football), Hospital, Eating, Skating, walking, working, and it is correctly predicted and defined.

Figure 9: Human Activity representation

On test data, 98.82% accuracy was achieved utilizing the proposed model. The recommended model has a validation loss of 2.22%. The model has been made and prepared on the arranged dataset so that it gives practically exact outcomes. The performance charts give a clearer image of the results. The vertical and horizontal axes have been rescaled by 1/100 and 1/10, separately. The horizontal axis shows the number of the epoch, while the vertical axis is the percent accuracy or loss. The recommended model's performance accuracy curve is displayed in Figure 14. The proposed model's most validation accuracy has been determined to be 99.50%. For expanding epoch, no extra interesting improvement was found for the proposed model [18] .

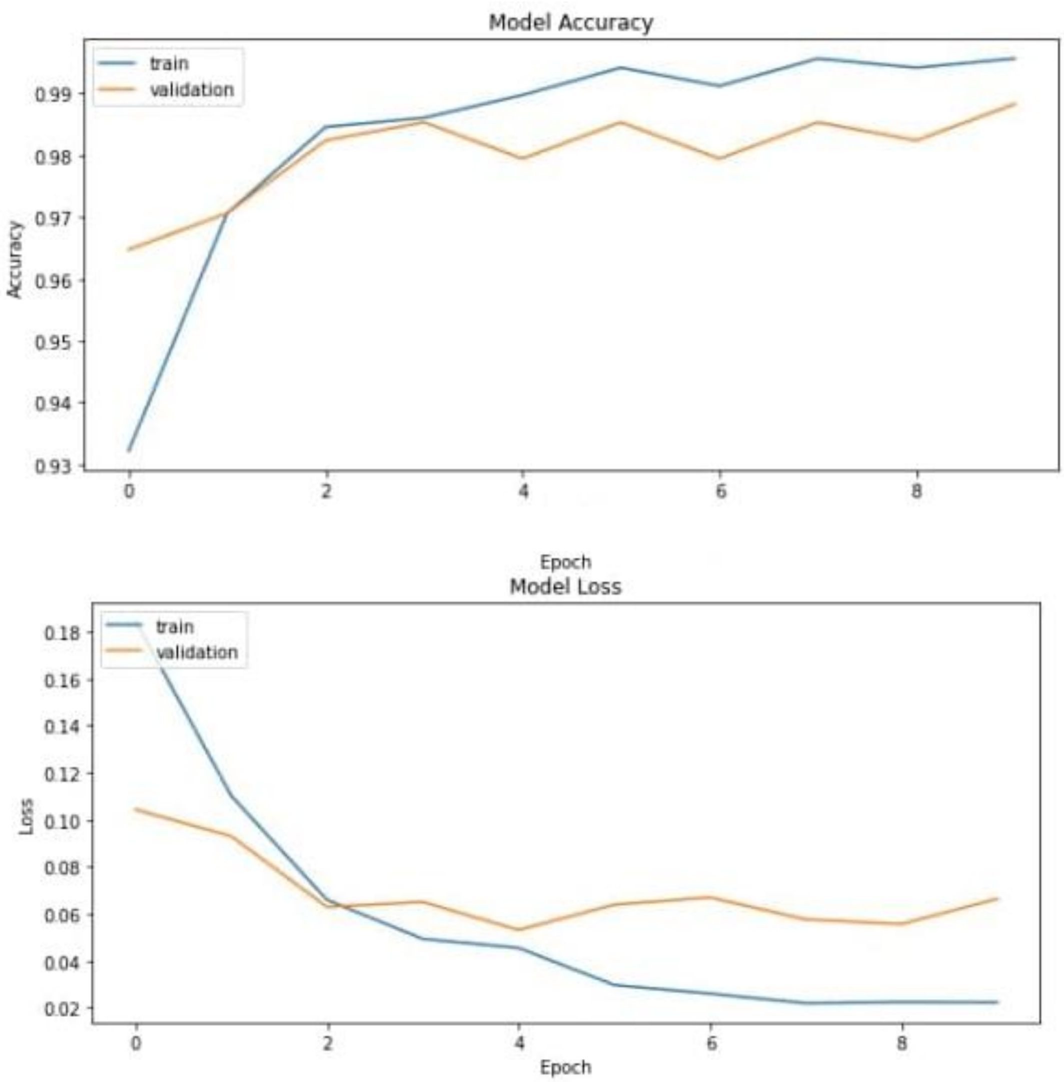

Figure 10: Learning curve of Deep CNN model.

The validation loss and accuracy curve of the proposed model is displayed in Figure 15. The validation loss has been displayed to immerse at 1.18%. An examination graph showing the accuracies of similar works gives a sensible image of execution enhancements. The proposed model in this review perceives every one of the 9 human exercises with high precision and identifies them in real-time. The proposed model was moreover approved utilizing test information got from Human activities photographs that were included in the primary dataset. On the test information given by outsiders, the proposed model gets a 98.82% accuracy. For the test data in this situation, the approval misfortune for the recommended model is 0.12%.

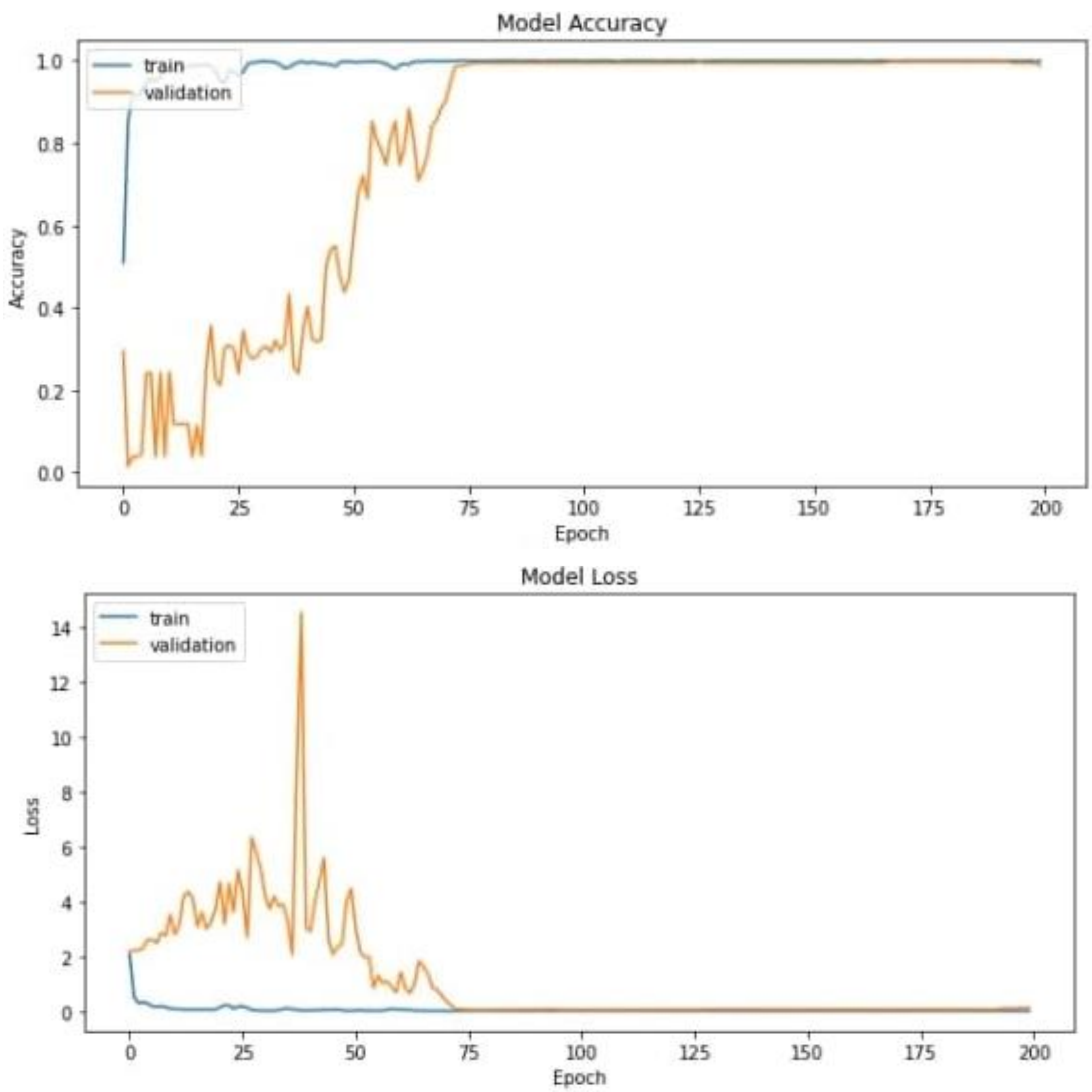

Figure 11: Learning curve of ResNet-50.

The improved Deep CNN model's exhibition accuracy and loss curves in this situation, the accuracy and loss estimations outline the vigor of the proposed model paying little mind to the human activities. The proposed model was utilized to recognize human activity from different kinds of photographs. The acknowledgment model was utilized to predict different kinds of human activity in different lighting conditions. Also, the model was utilized to detect activity between numerous photographs. And the connected probability detection of the human activity was recovered from the model. To recognize the human activity in a picture and to draw bounding boxes around them is an extreme issue solved.

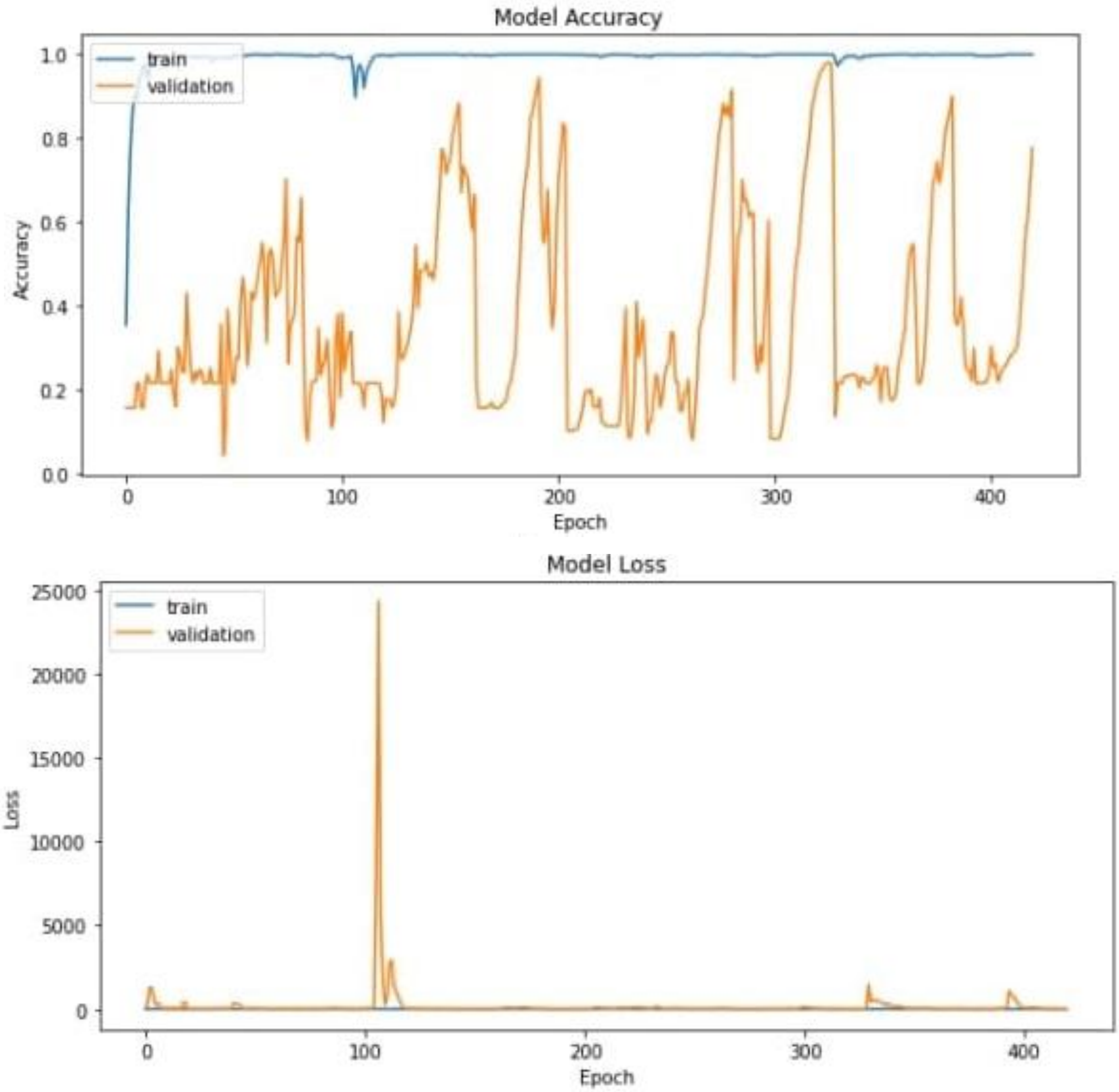

Figure 12: Learning curve of Improved Deep CNN model.

Comparative examination implies comparing our model with other customary ones. Later comparing accuracy with the resnet50 model and an overall general Deep CNN mode, we get Table 5. Here, Table 5 gives detail of result analysis

Table 5: Performance comparison image detection with some traditional methods.

Method Name	Accuracy
DCNN	97.35%
ResNet-50	97.85%
Improved DCNN	**98.82%**

4.3 Comparative Analysis

Justify for our proposed model performance is perfect for all data sets, that's why we collected two more human activity detection-related datasets from Kaggle. Then applied that Kaggle dataset to our model (Improved DCNN). Tests show that our model gives good accuracy for that dataset. So we can mention our model is perfect for all kinds of datasets.

Table 6: Performance comparison of the proposed model using different Dataset.

Dataset Name	**Model Name**	**Accuracy**
Custom Dataset	**Improved DCNN**	98.82%
Huang Huan 1et a1 Kaggle dataset		92.50%
Jithin Nambiar J et a1 Kaggle dataset		95.59%

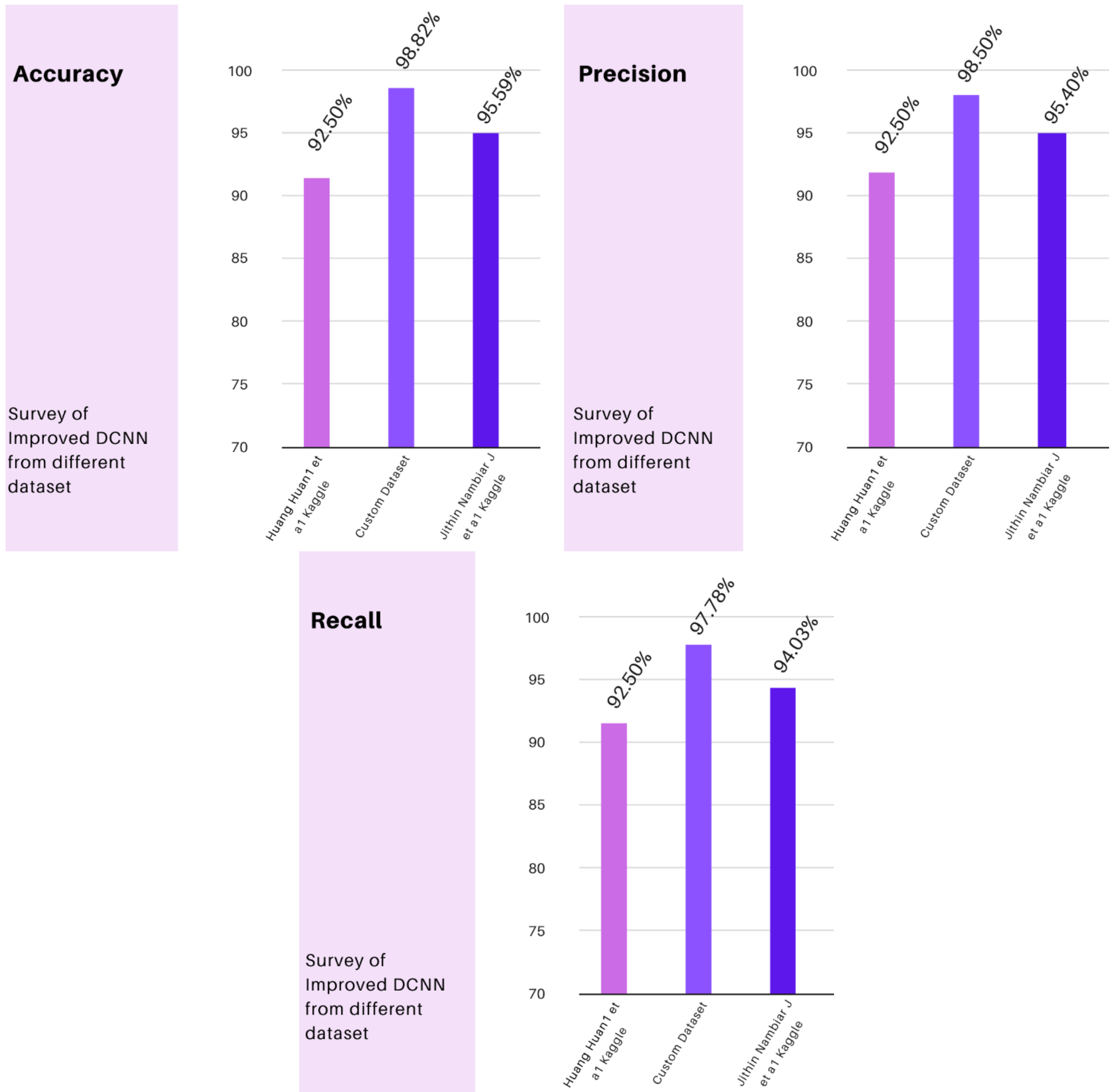

Figure 13: Survey of improved DCNN from different dataset.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In Our research, we proposed human activities and a human movement scene observation system utilizing DCNN with Resnet50 of ImageNet. As far as observing working, walking, eating, playing, etc. so forth it gave great outcomes. Additionally, it gives 99.50% accuracy over every one of the classes. Here we utilized custom data of the various types of human movement that they are doing. I trust that assists us with identifying a wide range of human activity and we utilize one sort of proof. At long last, it assists us with observing all the activity. In the future, we will build a model which can all human activity by CCTV camera footage. Our model will identify human activities and tell us what is going on there. If any suspicious activity is started, it will be informed through alarm. As a result, no observation will be required for 24 hours.

References

[1] S. S. ,. H. C. &. J. F. N. J. Hongyi Zhu, "Human Identification for Activities of Daily Living:A Deep Transfer Learning Approach," *Management Information Systems*, pp. 457-483, 2020.

[2] S. H. T. P. Nils Y. Hammerla1, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables," *arxiv*, 2016.

[3] K. R. a. J. Pang, "Transfer Learning Technique for Human Activity Recognition based on Smartphone Data," *International Journal of Civil Engineering Research,* vol. 11, no. 1, pp. 1-17, (2020).

[4] J. Pang, "Human Activity Recognition Based on Transfer Learning," *Graduate Theses and Dissertations.*, vol. 3, no. 2, pp. 1-58, (2018).

[5] J. L. 1. Y. P. 2. ,. M. S. A. 3. Z. S. 1. a. M. A. Saqib Ali 1, "An Effective and Improved CNN-ELM Classifier for Handwritten Digits Recognition and Classification," *Symmetry*, pp. 965-8580, 17 October 2020.

[6] E. N. Elnaz Soleimania, "Cross-Subject Transfer Learning in Human Activity Recognition Systems using Generative Adversarial Networks," *IEEE,* 2019.

[7] K. R. a. J. Pang, "Transfer Learning Technique for Human Activity Recognition based on Smartphone Data," *International Journal of Civil Engineering Research.,* vol. 11, no. 1, pp. 1-17, (2020.

[8] X. L. L. N. J. L. X. S. D. C. G. L. D. Z. Renjie Ding, "Empirical Study and Improvement on Deep Transfer Learning for Human Activity Recognition," *sensors,* (2018).

[9] S. D. R. G. A. F. Joseph Redmon, "Real-Time Object Detection," pp. 50-62, (2019).

[10] G. H. A. F. Marc Kurz, "Real-Time Transfer and Evaluation of Activity Recognition Capabilities in an Opportunistic System," *The Third International Conference on Adaptive and Self-Adaptive Systems and Applications,* pp. 22-268, (2018).

[11] E. U. H. &. H. J. &. K. L. &. H. U. Haq2, "Human detection and tracking with deep convolutional neural networks under the constrained of noise and occluded scenes," *Multimedia Tools and Applications*, 2020.

[12] D. X. Ajai V Babu, "Morphology Based Real Time Human Detection," *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*, (2018).

[13] K. T. S. K. M. J. B. R. S. Martin Milenkoski, "Real Time Human Activity Recognition on Smartphones using LSTM Networks," *Opatija Croatia,* pp. 1126-1134, (2018).

[14] 2. ·. L. Q. ·. X. X. ·. C. T. ·. Z. G. Shaohua Wan1, "Deep Learning Models for Real-time Human Activity Recognition," *Mobile Networks and Applications,* vol. 5, no. 6, pp. 1120-2110, 2019.

[15] N. a. Y. Mustafah, "Real-time Human Activity Recognition," *6th International Conference on Mechatronics - ICOM'17,* vol. 45, no. 12, pp. 521-929, 2017.

[16] T. P. D. B. H. V. S. Duong, "Efficient duration and hierarchical modeling for human activity recognition.," vol. 173, p. 830–856, October 2006.

[17] N. H. G. S. L. Z. A. Z. C. PENG LUO, "DCNN With Explicable Training Guide and Its Application to Fault Diagnosis of the Planetary Gearboxes," *IEEE,* vol. 8, p. 858–862, July 3, 2020.

[18] F. F. I. S. M. S. K. M. S. R. M. A. S. M. S. H. a. G. H. C. Nahida Islam1, "Towards Machine Learning Based Intrusion Detection in IoT Networks," *Tech Science Press,* vol. 69, 09 April 2021.

[19] K. Y. Y. Kawano, "FOOD IMAGE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK WITH PRE-TRAINING AND FINE-TUNING," *Proc. of European Conference on Computer Vision,* 2021.

[20] K. Z. J. Z. Q. S. a. X. Z. Yuwen Chen, "LSTM Networks for Mobile Human Activity Recognition," *Chongqing institute of green and intelligent technology Chinese Academy of Sciences Corresponding author,* p. 2264–2271, 2016.

[21] N. a. Y. Mustafah, "Real-time Human Activity Recognition," *6th International Conference on Mechatronics - ICOM'17 ,* vol. III, no. 12, pp. 50-62, 2017.

[22] Y. W. T. M. A. A.-g. R. A. Henry Friday Nwekea, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *ELSEVIER,* pp. 233-261, 2018.

[23] R. C. D. M. d. G. Ó. b. A. Godfrey, "Direct measurement of human movement by accelerometry," *ELSEVIER*, p. 1364–1386, 25 September 2008.

[24] Y. C. S. H. X. P. L. H. Jindong Wang, "Deep Learning for Sensor-based Activity Recognition: A Survey," *Manuscript,* 2018.

[25] J. G. Z. Y. Y. Zhang, *MIT Press, Cambridge,* vol. 18, p. 1585–1592, 2006.

[26] G. E. a. B. K. T.L.M. van Kasteren, "Transferring Knowledge of Activity Recognition across Sensor Networks," *Science Park,* p. 1, 2014.

[27] L. Q. X. T. G. Shaohua Wan, "Deep Learning Models for Real-time Human Activity Recognition with Smartphones," *Mobile Networks and Applications,* vol. 25, no. 75, p. (743–755), 2020.

[28] T. A. M. A. A. J. L. Luay Alawneh, "Enhancing human activity recognition using deep learning and time series augmented data," *Ambient Intelligence and Humanized Computing,* vol. 587, no. 16, p. 10565–10580, 23 April 2020.