IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 6, Ser. 2 (Nov. – Dec. 2023), PP 22-31
www.iosrjournals.org

Recognition Of Human Actions Using Deep Learning

Ujjwal Deep[1], Sushant Kumar[2], Kanika Singla[3]
[1]*(Department Of Computer Science And Engineering/ Sharda University, India)*
[2]*(Department Of Computer Science And Engineering/ Sharda University, India)*
[3]*(Department Of Computer Science And Engineering/ Sharda University, India)*

Abstract –
The subject of activity recognition by humans or anyone has undergone a revolution because of recent advances in deep learning, which enable models to learn intricate representations and hierarchies from unprocessed data, improving recognition accuracy. Because they require a lot of feature engineering and data preparation, the traditional way by many machine learning techniques like the support vector machines models (SVM) and histogram of gradients (HOG) with k-nearest neighbor classifiers have lost some of their appeal. In order to increase the robustness and predictive accuracy of human activity recognition from raw image data, this research focuses on developing an efficient and robust model by utilizing the strengths of convolutional neural networks (CNN), Vgg16, Inception model, and Resnet model. The suggested model shows promising results and does away with the requirement for sophisticated feature engineering. Additionally, a large and varied dataset with over 12,000 labeled photos of different human activities is used in the research. To guarantee the quality and applicability of the dataset, a number of data pretreatment techniques—including exploratory data analysis—are used before model training and the evaluation and analysis of the model. All things considered, this study advances the field of deep learning methods which are based on recognition of human activities and shows promise for increased efficiency and accuracy in practical settings and situations.

Keywords: *HOG, Deep Learning, VGG16, ResNet50, Inception, Xception and HAR.*

Date of Submission: 02-12-2023 Date of Acceptance: 12-12-2023

I. INTRODUCTION

A. Needed importance of the Recognition of the Human Actions

The changing field of Human Activity Recognition (HAR) is a dynamic area and it has many potentials that has many of its roots in the great history of computer vision and pattern recognition algorithms. Its inception dates back to the late 1960s and 1970s, when the very first initial efforts were starting to made to identify the fundamental human actions through pictures or still images or photographs and video clips snapshots. The very development in this field of HAR went and has also gone through the many several phases: the feature engineering phase, which was characterized by the prevalence of hand-crafted features and rule-based systems; the sensor-based HAR phase, which used accelerometers and gyroscopes to record motion data with the groundbreaking period of deep learning marked by the development of the Recurrent model of Neural and Networks in deep learning field (RNNs) and the Convolutional model of Neural and Networks in deep learning field (CNNs), which allowed for the automatic extraction of hierarchical features from raw visual input. Because of their ability to carry out a vast array of cognitive tasks and activities in daily life, humans have evolved into invaluable resources. The ability to understand and, to some extent, anticipate the acts of others is a fundamental human ability that underpins effective communication and interaction in daily life. Parallel to this, it is often essential that robots comprehend human activities prior to their complete execution. Because of this comprehension, robots can react accurately and quickly. The discipline of Human Activity Recognition (HAR) is a dynamic area of study that focuses on strategies and tactics for automatically recognizing Activities or actions of the Daily Living activities (ADLs). For example, it is possible to identify fall behavior in elderly people living alone by observing their movements; this allows family members to provide early assistance [1]. Recently, the focus of sensor-based HAR research has been on using inertial signals obtained from mobile devices such as cellphones. These methods have implications for healthcare rehabilitation, athletic training, technology for smart homes, transgression identification, and elderly detection of falls. The mainly two types of data that are used or can be fed or trained into the HAR systems as input are the: first, time and series data of physical motions or actions during the various approaches and activities recorded by the sensors like acceleration sensors plus gyroscopes; and second, video clips or images of actions by humans, etc. We have utilized a number of images that we have taken ourselves as well as images from the internet and other sources, including an easily available extracted dataset. Only until artificially intelligent systems are able to discern and categorize human activities from visual input will they be able to comprehend and respond to human behaviors [2]. In technologies such as gesture-based interactions, footage evaluation, and video monitoring, precise action recognition is crucial for making decisions [3]. Recent development and advances in deep learning field have revolutionized the field by allowing algorithms to learn complex representations and structures from raw data, by improving the very preciseness of actions and approaches recognition detection systems [4]. Training a model to recognize specific actions from raw sensor data has become significantly easier thanks to deep learning. One machine learning technology that is still in use today, though nearly obsolete, is the support vector machine (SVM) used in [5]. The k-nearest-neighbor classification in the previously suggested histogram of gradients (HOG) feature extraction method [6]. These techniques produced remarkably accurate recognition results. But there was a lot of labor involved in getting the data ready, including pre-processing and domain expertise. The CNN-LSTM model was proposed by Mutegeki et al. [7] and showed promising accuracy across a number of datasets. This method not only lowers the complexity of the model and does away with the requirement for sophisticated feature engineering, However, it also improves the precision of human behavior predictions made from raw data. Numerous approaches have been proposed, including chronologically deep short-term memory (LSTM) systems, spatially advanced convolutional neural networks (CNN) [8], and deep feed-forward neural networks and their adaptations. There are benefits and drawbacks to each of these tactics. In addition, the publication "The iSPL technique of Inception: A Human Activity Recognition Inception-ResNet Deep Learning Architecture " by Mutegeki Ronald et. al. presented the SPLInception model, that performs better than different approaches thanks to its 88% F1 score. This research also looks into novel approaches, such as "The Recognition of human activities Using the Convolutional Neural Network approach and the Depth Sensor Data approach" by Zeeshan Ahmad et al. [11], that turns depth data into visuals and obtains exceptionally accurate classification rates. When taken as a whole, these studies broaden the body of knowledge about human activity recognition and show the promise of neural networks in this field.

II. LITERATURE SURVEY

With a focus on the different use and approaches of deep learning techniques, this section provides an overview or the brief of the various studies and methodology applied in the field of human activity recognition. It illustrates the noteworthy developments in this field and the efficacy of various models and datasets. Notably, studies like as "The CNN and the LSTM method to the Human Activity Recognition model" by Ronald Mutegeki et. al. demonstrated the impressive performance of both of these systems (the LSTM and the CNN model) in tandem, yielding phenomenal outcomes with a success rate of 99.06% according to the UCI HAR Database. Furthermore, the SPLInception model—which outperformed alternative approaches with an 88% F1 score—was introduced in the publication " The iSPL technique of Inception: A Human Activity Recognition Inception-ResNet Deep Learning Architecture " by Mutegeki Ronald et. Al. Innovative methods are also investigated in the study, as in Zeeshan Ahmad et al.'s " The Recognition of Human actions or activities Using the Convolutional Neural Network model and the Depth Sensor Data approach" [11], which achieves high classification accuracy rates by converting depth data into images. Together, these works add to the expanding corpus of research on human activity recognition, demonstrating the potential of deep learning in this area.

Table 1: Description of different Technology

Paper About	Methodology	Accuracy and Results
Convolutional neural networks as tools for human activity recognition: A conducted investigation of the very state art and the data sets, problems, and the future directions [12] 2022	investigated the four types of equipment categories that are most frequently utilized for HAR: visual and audio signals, signals from radars, sensor data gathered from mobile phones, and multidisciplinary sensing equipment.	The maximum batch size, dropout, optimizer, loss, and hyper-parameters are 64, 0.5, Adam, and cross-entropy, respectively.
A CNN-LSTM Method for Identifying Human Activity [13] 2021	LSTM classification and The CNN network to detect human activities. the advantages of combining both of these networks, especially with regard to recognizing human activities.	The accuracy achieved by this approach was 99.06%
The iSPL technique of Inception: A Human Activity Recognition Inception-ResNet Deep Learning Architecture [14] 2021	Here we employed SPLInception, which is based on Inception-ResNet, and we compared it with other previous approaches.	The accuracy achieved by this approach was 81%
The Real in time Humans Activity Detection with 3D Convolutional Neural Networks and ResNet [15] 2021	Using 3D CNN and Resnet without using the LSTM-attention model.	The accuracy achieved by this approach was 94.75%

The Personalized used Models for Deep Learning-Based Recognition of Human Activities [16] 2020	Progressive learning is combined with algorithmic methods of deep learning to generate specific models which perform better than alternatives.	Research demonstrated that neural networks adjust to a new user more quickly than the baseline.
Identification of Human Action Behavior in Still Pictures via Suggested Frames Selection and Transfer Learning [17] 2023	The models used in this study were Inception V3, xception, and VGG16	The accuracy achieved by this approach was 58.125%
Convolutional Neural Network and Depth Sensor Data for Human Action Recognition [18] 2019	creates images from series of depth data. The specially for this purpose of recognition, created CNN model would be trained on this dataset (with the photos), and its output results would be used to categorize the action or the behaviors.	Using a Kinect two-dimensional, a 97.23% confidence in classification was attained. On the traditional MSR 3D database, the degree of precision was 87.1%.
Using LSTM/BiLSTM for Human Activity Recognition [19] 2022	Divided according to the method of extracting significant frames from the video dataset	80.12% LSTM Inception v3 v3 80.75% BiLSTM Inception − 77.02% LSTM Xception
Classification of Pulmonary Images Using the Inception-v3 Transfer Learning Model [20] 2019	Softmax function, the Logistic function, and the SVM classifiers were used for extracting features from the pulmonary imaging data employing the revised Inception-v3 approach, which depends on learning through transfer.	The trial's and used pulmonary image in the model testing categorization efficiency was superior to the rest of the studies, with 95.41% and 80.09%, accordingly, the highest sensitivity and specificity.
Use of the Hybrid Deep Evolving and the Neural Networks used for the identification of human activities [21] 2022	For the task of classifying video actionsThe trial's pulmonary image categorization efficiency was superior to the rest of the studies, with 95.41% and 80.09%, accordingly, the highest sensitivity and specificity.The Convolutional and the Long Short-Term Memory method (ConvLSTM) Networks and the Long-term Recurrent approach Convolutional Networks model (LRCN) are utilized.	In accordance and study with the real results, the LRCN model which was used here using Mobile-net as the encoder and a new network model, BiLSTM network which here acted as the decoder achieves the highest possible accuracy rate (i.e., 87% regarding the categorization of 50 activity categories in UCF50).

III Methodology

A. DATA SET

The recognition of human activities or actions is a crucial task in the field of computer vision, with the diverse and huge number of implications such as surveillance and human-computer interaction. The paper focuses on utilizing a comprehensive dataset for human action recognition, which includes 15 distinct and different classes representing various human activities. The purpose of this used dataset is to support research and to analyze on the creation and study and assessment of the different machine learning approaches and the technical models used for the recognition of human actions. The dataset we are utilizing was assembled using information obtained from both our own research and other resources [9]. For the purpose of training and assessing the algorithms, we set out to compile a broad and varied dataset encompassing a variety of human behaviors. The collection consists of more than 12,000 labeled photos, including validation images, that represent distinct categories of human activity. Each of the 15 identified courses has its own folder containing these pictures. "Calls," "doing cycle," "clap," "dance," "eat," "drinking anything," "hugs," "fight," "listen to audio," "laughs," "sit," "run," "text," "sleep," and "laptop" are the target classes. The names and labels assigned to each image in each class of our dataset are shown in Figure 1. We used the exploratory data analysis which is most stable in this case (EDA) and other different data and information pre-processing approaches before utilizing the dataset for model training and evaluation. Our research's success is largely due to the extensive range of data preparation techniques we have used. First things first, the foundation of our data preparation activities is data cleaning. We create a strong base with dependable and accurate data by locating and removing the duplicity or damaged photographs from our dataset. This thorough data pruning ensures that our subsequent research and training of models are built upon an impartial and trustworthy dataset, reducing the likelihood of skewed results or biased model behavior. Approaches to data augmentation will play a critical role in the years to come in enhancing the variety and accurate representation and analysis of the data of training set which we used here. We here, also have utilized the different techniques developed and implemented by Hamza Amrani et al. [10]. By including and combining the transformations like flipping, rotations, and scaling of the data, our models with these approaches may be able to easily learn from a greater variety of perspectives and disparities in the information we provide. As a result, our simulations may function admirably in real-life situations where objects may be encountered at different scales, introductions, or perspectives. This promotes excellent generalization skills even more. The efficiency and accuracy and stability and optimization of our tested model training process are then tested and then impacted by the normalization (so that uniformity be there) step of the data preparation (which we are using) pipeline. By uniformizing the values of pixels to a scale, typically ranging from 0 to 1 or -1 to 1, we ensure that gradients during optimization remain within acceptable bounds. This promotes more gradual convergence during training, which lowers the likelihood of gradient-related issues that can impede learning. As a result, our models show better overall performance and train more quickly. To make certain that the information we provide is consistent with the machine learning algorithms we utilize, the encoding of labels is crucial when used in conjunction with these techniques. Once the category class labels are converted to quantitative representation, our models are able to understand and derive information from them. This translation is important because it allows the mathematical framework to make decisions, comparisons, and predictions based on the encoded numerical visualizations, so bridging the gap across raw data and insightful knowledge. Finally, the last layer of our data pretreatment method is the wise practice of train-validation splitting. Our dataset is split up into separate training and validation sets, which makes it possible for us to keep an eye on the performance of our model as it is being trained. We ensure that our algorithms do not overfit the training set by regularly evaluating them, which could lead to poor applicability on fresh data. Through the use of validation measures to constantly track our machine learning model's efficiency, we are able to assure the robustness and reliability of our study findings, ultimately yielding noteworthy and reliable results.

Figure: 1 Dataset Overview

In this section, we outline the methodology employed for human action recognition, where we classify images into 15 distinct classes representing various human activities. Our dataset consists of over 12,000 labeled images, including validation images, each depicting a single human activity category. The targeted classes here are "Calls," "doing cycle," "clap," "dance," "eat," "drinking anything," "hugs," "fight," "listen to audio," "laughs," "sit," "run," "text," "sleep," and "laptop".

Before training the models, we perform the following data preprocessing steps:
Image Resizing: We resize all the images to 160 X 160 resolution,

Data Augmentation: The different implication of Data augmentation and processing techniques, including randomly common rotations, flips, and brightness adjustments, which are then employed and are used to enhance the existing diversity of our training data and to develop and improve the generalization and complexity and accuracy of the model used.

B. ALGORITHMS USED
B.1. CONVOLUTIONAL NEURAL NETWORK (CNN)
B.1.1 Model Architecture Figure-2*:*

Our research starts by training a Convolutional model of Neural Network in deep learning (CNN) with the specified architecture.

Input Layer: Takes the re-sized images as input.
Convolutional Layers: ReLU activation functions are used in multiple convolutional layers.

Max-pooling Layers: Pooling the layers are generally used and utilized to reduce and make smaller the size of featuring maps through downsampling.
Fully Connected Layers: The architecture includes Dense layers for classification purposes.
Output layer: utilizes the Softmax activation function for multi-class classification.

B.1.2. Formulas:

In the scenario where we have a shaped like asquare neuron layer of defined size A×A which is then followed by a new convolutional layer which is then used further, the resulting output size of the convolutional layer, when using an f×f filter, will be (A−f+1) × (A−f+1).

Convolution Operation: Y = m*x + c

f(n) = max of (0, n) is the used Rectified Linear Unit function (Relu).

The Softmax function of Activation: $\frac{eZj}{\sum_{a=1}^{N} \quad eZa}$

The a-th entry used in the softmax function output vector softmax(z) can be described as the projected and represented probability which is of the test input belonging to class a. Z here is the used vector of raw outputs which is taken from the neural network and here, e = 2.718.

B.1.3 Optimization algorithm:

To reduce the loss due to categorical cross-entropy, we employ the Adam optimizer.

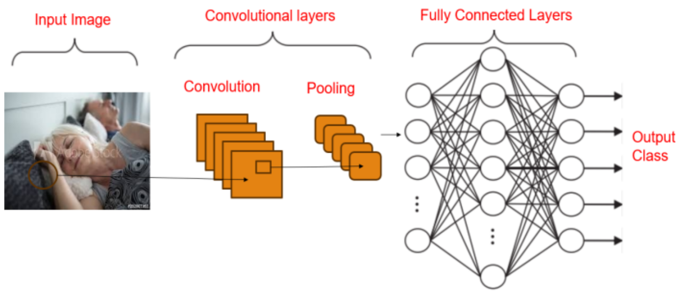

Figure: 2 CNN Model Architecture

B.2 RES-NET MODEL (THE RESIDUAL NETWORK MODEL)

B.2.1 The Architecture of the Model Figure: 3:

Next, in order to improve or to increase model depth and alleviate the vanishing or lost gradient problem or issue, we utilize a Residual Network model (Res-Net). There are leftover blocks with skip connections in the architecture.

Input Layer: Takes the re-sized images as input.

Residual Blocks: Comprising multiple convolutional developed and used layers.

Spatial dimensions or the dimensions are then reduced by the different techniques for example, via the global avg. pooling layer.

Like CNN, but with fully connected layers and an output layer.

B.2.2 Formulas:

Skip Connection: K(x) = L(x) + x, where L(x) here is the used residual function.

Global Average Pooling: Averages the values in each feature map.

B2.3 Optimization Algorithm:

Adam optimizer is utilized, similar to the CNN model.

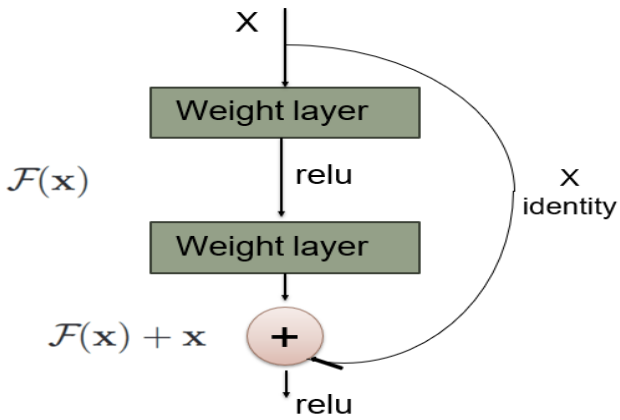

Figure: 3 ResNet Skip Connection Architecture

B.3 INCEPTION MODEL

B.3.1 Model Architecture Figure: 4:

We test the Inception model which we trained, which records characteristics at numerous scales by using the different and unique filter sizes inside the same layer.

Input Layer: Takes the re-sized images as input.

Inception Blocks: Consisting of parallel convolutional operations.

Global Avg. Pooling Layers, with the Fully Connected Layers which are used in the model, and Output Layer: Similar to previous architectures.

B.3.2 Formulas:

Inception Module: Utilizes various filter sizes (e.g., 1x1, 3x3, 5x5) in parallel and concatenates their outputs.

B.3.3 Optimization Algorithms:

Adam optimizer is used for training the Inception model.

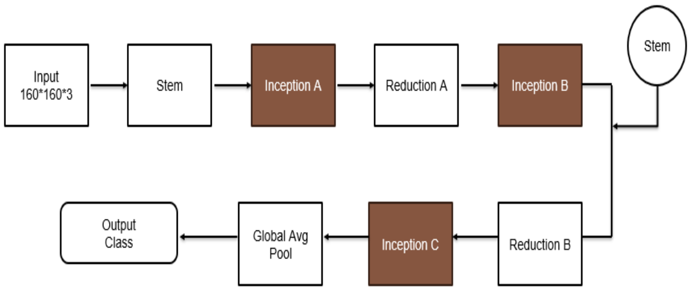

Fig. 4. The full Inception V3 (not single inception) Main Architecture

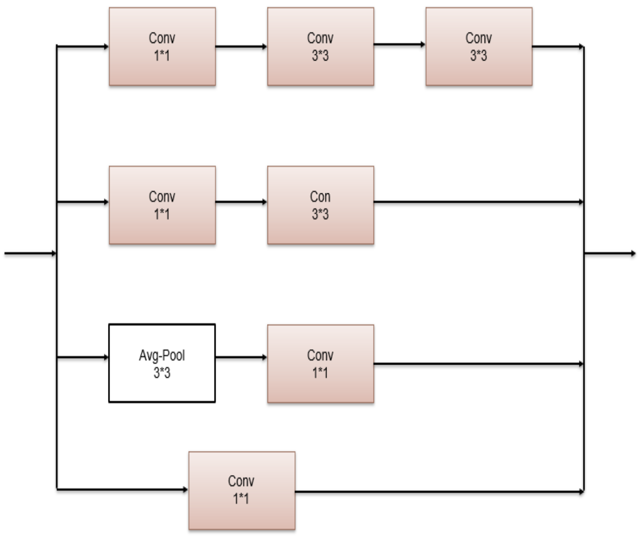

Figure: 4.1. Single Inception Architecture

B.4. THE ENSEMBLE LEARNING MODEL

B.4.1 The Ensemble Learning model Architecture Figure: 5:

To harness the collective power of multiple models, we create an ensemble model by combining predictions from CNN, ResNet, and Inception.

The input layer takes the same data as separate models.

Fusion Layer: Merges CNN, ResNet, and Inception predictions.

Final Fully Connected Layer: Makes the ultimate prediction.

B.4.2 The Ensemble Learning Method:

The work is that, it had been employed a weighted average method or ensemble approach, assigning different weights to the predictions of CNN, ResNet, and Inception. These weights are determined through experimentation.

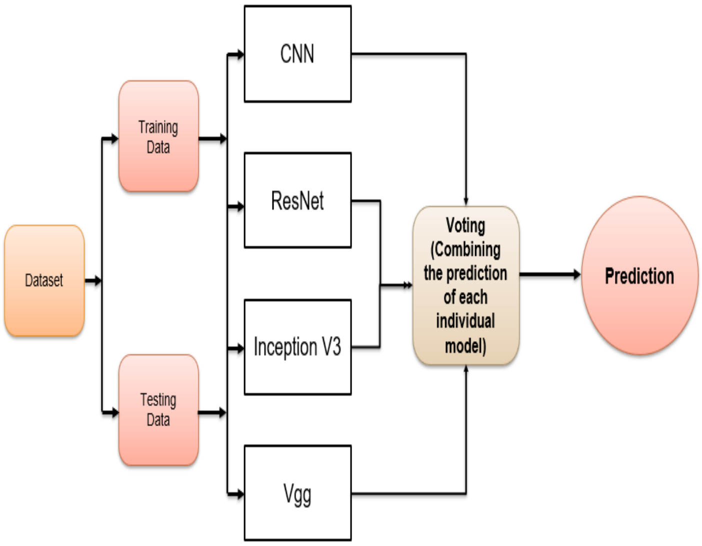

Figure: 5. Ensemble Model Architecture

III Training and Result Analysis

During the training phase, the CNN model employed a forward calculation with the training set to generate the network output. The categorical cross-entropy errors were then computed by comparing the predicted outputs to the actual outputs. To optimize the model's performance, the Adam optimizer was utilized to backpropagate these errors through the layers, updating the hyperparameters accordingly. The model was initially set to train for 100 epochs; however, training ceased at the 60th epoch. Figure 6 illustrates the relationship and dependency between the proposed number of epochs during the training and the accuracy achieved. As the number of epochs increased, the accuracy also improved. At the 60th epoch, the accuracy surpassed the 90% threshold and stabilized, indicating a satisfactory level of performance as shown in table 2.

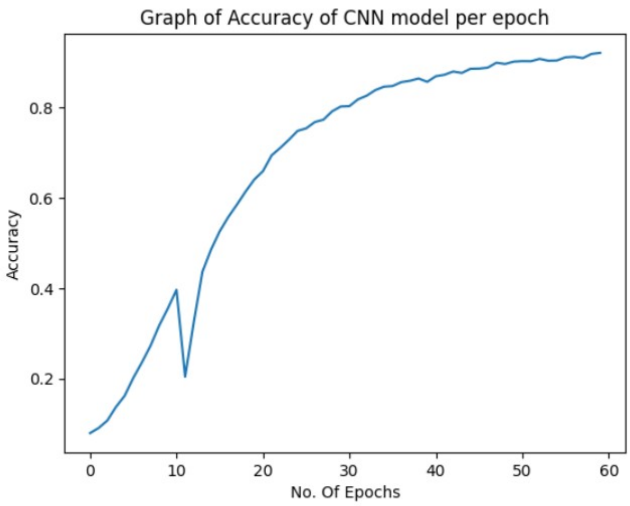

Figure- 6 Graph of Accuracy of CNN model per epoch

A Sequential model was initialized in order to develop the ResNet deep learning model. With some exception like the exception of the top layer and the pre-trained weights from the ImageNet dataset, the basic model that was employed was ResNet50. Now as the output layer, and the different Dense layers with like 15 units and a SoftMax function of activation function was implemented. Accuracy was used as the evaluation metric when the model was assembled using the generally used Adam optimizer which then is used further and the categorical and the cross-entropy used loss function of the model. And as seen in figure 7, it was found that the accuracy of the model peaked after the eighth epoch.

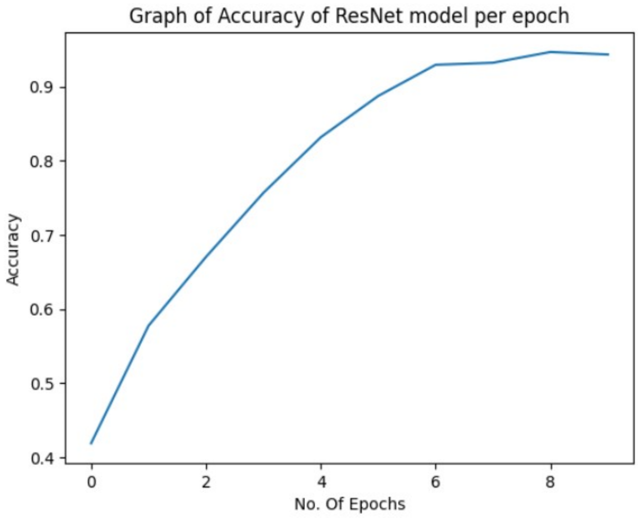

Figure-7 Graph of Accuracy of ResNet model per 50 epochs

An initial Sequential model was used to train the Inception model. With the exception of the top layer, the basic model was the InceptionV3 architecture, and pre-trained weights from the ImageNet dataset were employed. The output layer, which is a Dense layer with like 15 units and a SoftMax function activation function, was then linked to the basic model of deep learning. Accuracy was used as the evaluation metric when the model was assembled using the most commonly used Adam optimizer and the categorically used cross-entropy loss function. Figure 2 illustrates the model's training accuracy, which is higher than 90%.

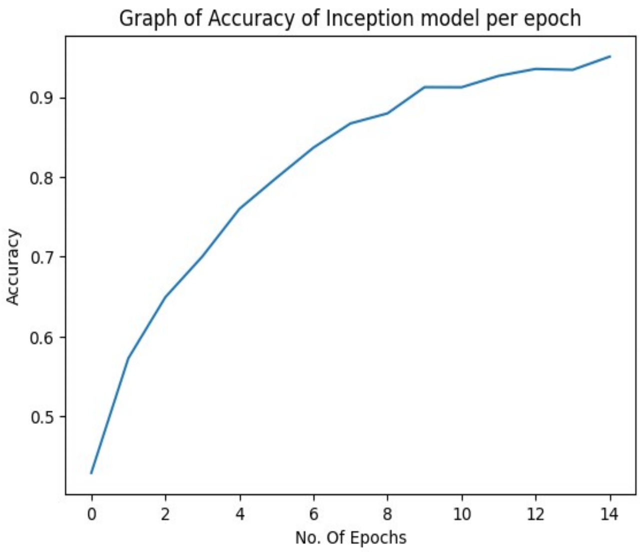

Figure-8 Graph of Accuracy of Inception model per epoch

We began by building the VGG16 model by combining a Sequential model with the VGG16 basic model, omitting the top layer, in order to train the Ensemble mode. Next, the categorical cross-entropy loss function and Adam optimizer are used to assemble the model. Next, we used a combination of MaxPooling and Convolutional layers to generate a CNN model, which was then followed by Dense layers with dropout regularization. The categorical cross-entropy loss function and Adam optimizer are used to compile the model. A Sequential model is combined with the ResNet50 basic model, omitting the top layer, to form the ResNet50 model. After adding a layer which is the Dense layer with SoftMax function of activation, the model is constructed using the categorical cross-entropy loss function and the generally used Adam optimizer. Similar to this, the InceptionV3 model is made by combining a Sequential model with the InceptionV3 base model—that is, without the top layer. After adding a different Dense layer with SoftMax function of activation, the model is constructed using the categorical cross-entropy loss function and the Adam optimizer. Prior to training and assessment, the model is loaded with all of the models' saved weights. Next, we assembled our Ensemble model using accuracy as the evaluation metric and a categorical cross entropy used here as the loss function. After 14 epochs, the accuracy of the model stopped rising since it had reached a reasonable level, as seen in figure 9 and table 2.

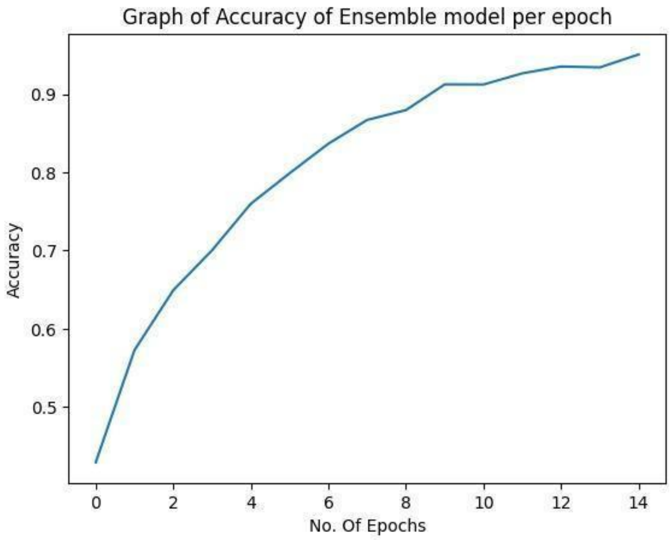

Figure:9. Ensemble Model Architecture

TABLE 2: TRAINING LOSS, ACCURACY AND PREDICTED ACCURACY

S.NO	MODEL	TRAINING LOSS	TRAINING ACCURACY	PREDICTED ACCURACY
1.	CNN	0.3	0.914	0.921
2.	RESNET	0.15	0.921	0.943
3.	INCEPTION	0.15	0.901	0.95
4.	ENSEMBLE	0.22	0.967	0.999

IV. CONCLUSION AND FUTURE WORK

Deep learning techniques, including Convolutional Neural Networks (CNN), Residual Networks (ResNets), and Ensemble methods, have demonstrated significant promise in the field of human activity recognition. In our research project, we have employed an ensemble approach that combines predictions from different models, including CNN, ResNet, VGG16, and Inception. The ensemble model architecture employs a weighted average approach to amalgamate individual model predictions. Empirical results, as shown in Table 2, highlight the impressive accuracy rate achieved by our ensemble model. Future research directions can be Exploring various deep learning architectures, such as ResNets and Inception models, to further enhance human activity recognition. Investigating the use of transfer learning techniques to improve model performance in human activity recognition. Exploring the application of personalized models in human activity recognition using deep learning techniques. Conducting experiments with diverse datasets and evaluating model performance across a wider range of activities and scenarios. Investigating the use of incremental learning techniques to enable faster adaptation of models to new users or activities. Our research underscores the potential of deep learning in human activity recognition. investigations in the areas mentioned above have the potential to advance the field and contribute to the development of more effective and user-centric human activity recognition systems, which hold significance in improving healthcare, quality of life, and human-computer interaction.

REFERENCES

[1]. Al-Qaness MA, Helmi AM, Dahou A, Elaziz MA. The Applications Of Metaheuristics For Human Activity Recognition And Fall Detection Using Wearable Sensors: A Comprehensive Analysis. Biosensors. 2022 Oct 3;12(10):821.

[2]. Lara OD, Labrador MA. A Survey On Human Activity Recognition Using Wearable Sensors. IEEE Communications Surveys & Tutorials. 2012 Nov 29;15(3):1192-209.

[3]. Ranieri, Caetano Mazzoni, Scott Macleod, Mauro Dragone, Patricia Amancio Vargas, And Roseli Aparecida Francelin Romero. "Activity Recognition For Ambient Assisted Living With Videos, Inertial Units And Ambient Sensors." Sensors 21, No. 3 (2021): 768.
[4]. Moshiri PF, Shahbazian R, Nabati M, Ghorashi SA. A CSI-Based Human Activity Recognition Using Deep Learning. Sensors. 2021 Oct 30;21(21):7225.
[5]. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A Public Domain Dataset For Human Activity Recognition Using Smartphones. Inesann 2013 Apr 24 (Vol. 3, P. 3).
[6]. Jain A, Kanhangad V. Human Activity Classification In Smartphones Using Accelerometer And Gyroscope Sensors. IEEE Sensors Journal. 2017 Dec 11;18(3):1169-77.
[7]. Mutegeki R, Han DS. A CNN-LSTM Approach To Human Activity Recognition. In2020 International Conference On Artificial Intelligence In Information And Communication (ICAIIC) 2020 Feb 19 (Pp. 362-366). IEEE.
[8]. Moya Rueda F, Grzeszick R, Fink GA, Feldhorst S, Ten Hompel M. Convolutional Neural Networks For Human Activity Recognition Using Body-Worn Sensors. Ininformatics 2018 May 25 (Vol. 5, No. 2, P. 26). MDPI.
[9]. Https://Www.Kaggle.Com/Datasets
[10]. Amrani H, Micucci D, Napoletano P. Personalized Models In Human Activity Recognition Using Deep Learning. In2020 25th International Conference On Pattern Recognition (ICPR) 2021 Jan 10 (Pp. 9682-9688). IEEE.
[11]. Ahmad Z, Illanko K, Khan N, Androutsos D. Human Action Recognition Using Convolutional Neural Network And Depth Sensor Data. Inproceedings Of The 2019 International Conference On Information Technology And Computer Communications 2019 Aug 16 (Pp. 1-5).
[12]. Milon Islam M, Nooruddin S, Karray F, Muhammad G. Human Activity Recognition Using Tools Of Convolutional Neural Networks: A State Of The Art Review, Data Sets, Challenges And Future Prospects. Arxiv E-Prints. 2022 Feb:Arxiv-2202.