Residual Neural Network for the Accurate Recognition of Human Action and Compared with Bayesian Regression

Narmatha.V¹, Ramesh. S²,*

¹Research Scholar, ²Project Guide & Corresponding Author,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu. India. Pincode: 602105

ABSTRACT : Aim: In this research article, the aim is to analyze and compare the performance of Residual Neural Network and Bayesian Regression for accurate recognition of human actions.

Materials and Methods: The proposed machine learning classifier model uses 80% of the UCF101 dataset for training and the remaining 20% for testing. For the SPSS analysis, the results of two classifiers are grouped with 20 samples in each group. The sample size is determined using a pretest with G-power, with a sample size of 80%, a confidence interval of 95%, and a significance level of 0.014 (p<0.05).

Result: The findings suggest that the novel residual neural network classifier and Bayesian regression classifier achieved accuracy rates of 95.63% and 93.97%, respectively, in identifying human activities accurately. The statistical significance value between residual neural networks and Bayesian regression has been calculated to be p=0.014 (independent sample t-test p<0.05), indicating a statistically significant difference between the two classifiers.

Keywords: Bayesian Regression, Classifiers, Human Action, Machine Learning, Novel Residual Neural Network, Recognition, Technology.

INTRODUCTION

Automatic human activity recognition has gained significant attention in video analysis technology in recent years due to its widespread applications in areas such as surveillance, entertainment, and healthcare systems (Salah et al. 2012). For instance, in the sports field, poor video surveillance can lead to challenges in making accurate decisions (Buizza 2021). The Author’s research shows that fixed surveillance cameras can capture daily

* Corresponding Author Mail: rameshsssecse@gmail.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
human actions that can be studied using computational systems. These systems extract essential features from video clips to recognize a person's actions. Some techniques also focus on the shadowing of background objects, which can improve the accuracy of the recognition rate (R et al. 2019). To ease the recognition process, the proposed system is presented. Applications include detecting gastrointestinal disorders and gender prediction, theft detection (de la Loge et al. 2004).

A total of 1650 articles from the years 2001 to 2022 were reviewed; 350 of those came from IEEE Xplore, 200 from ResearchGate, 300 from Elsevier, and 800 from Springer (Ji et al. 2013). Scene contexts have allowed viewers of movies and TV shows to identify and locate human sports, allowing them to better grasp the dynamics of social interactions (Nair et al. 2020). It is possible to boost the status of such complicated videos by integrating context and descriptions into a single framework (R et al. 2019). It is strongly suggested that the classifier be used in practice. The proposed model's efficiency and accuracy are in model and specification dependent. (Jaimes and Sebe 2007).

Recognizing human actions using video-based technology is a rapidly growing research area, but there are still significant gaps that need to be addressed. One major challenge is differentiating between actions across classes, as actions across classes can have similar representations while actions within the same class can have varying representations due to differences in body movements between individuals (Mahrishi et al. 2020). Several challenges have been identified in accurately capturing human actions through visual perception, including the impact of changes in viewing angle, variations in movement speed, and camera motion (Salah et al. 2012). Human Action and face detection remains a challenging task despite numerous research methods to be enhanced in future technology (Karpouzis and Yannakakis 2016). The surveillance system must take many photographs over time and then analyze the data to determine things like the ages and genders of the people in the scene and what they are doing ((Islam et al. 2022). When conducting surveillance, knowing a person's gender is crucial. In gender prediction, humans face the issue of recognizing people in camera motion (Golestani and Moghaddam 2020). From the literature survey, it can be inferred that existing models still need many improvements to achieve higher accuracy in human action detection. While the proposed novel residual neural network and Bayesian regression models improve accuracy through deep learning techniques, there may still be room for further improvement to achieve even higher accuracy rates. Five distinct locations have analyzed the work's results, providing averages of 95.63% and 93.97% accuracy, respectively.

MATERIAL AND METHOD

This research was conducted in the Computational Intelligence Lab of the Department of Computer Science and Engineering at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The widely used UCF101 dataset (Kaggle) was utilized to examine the results, which comprises 13,320 video clips classified into 101 categories, including body motion, human-human interactions, human-object interactions, playing musical instruments, and sports. These video clips have a total length of over 27 hours and are collected from YouTube with a fixed frame rate of 25 FPS and a resolution of 320 × 240. For this research, 20 sample videos were selected from the human-human interactions category to build the novel residual neural network and Bayesian regression models. The analysis in this research is based on a total of 40 samples, divided into two groups of equal size. The novel residual neural network algorithm is implemented first, followed by the Bayesian regression algorithm. The recommended Machine Learning classifier model trains with 80% of the dataset volume and tests with 20%. In the previous
study (Vrskova et al. 2023), the sample size was also set at twenty in each of the two groups, totaling forty, and determined using a G power pretest with a sample size of 80%, a threshold of 0.05, and 95% confidence intervals. This research project utilized the Python-Tensor Flow Package on a Jupyter Notebook platform and the Windows 11 operating system with an i5 processor, 64-bit architecture, and 16 GB of RAM. A minimum of 20 GB of hard disk space was required to store the images of the dataset downloaded from Kaggle. DreamsPlus Consulting Pvt Ltd, Chennai, provided technical consultation for this study.

**Novel Residual Neural Network**

Novel residual neural network is a deep residual neural network architecture that consists of 50 layers. It was first introduced by Microsoft in 2015 and is widely used for various computer vision tasks, including image classification, object detection, and semantic segmentation (Liu et al. 2021). ResNet-50 introduces a new residual block design that enables the training of very deep neural networks by addressing the vanishing gradient problem. The architecture includes skip connections that bypass one or more layers to propagate the input directly to deeper layers, which helps preserve information during training data. This allows ResNet-50 to achieve state-of-the-art performance on several benchmark datasets and the Residual Neural Network approach is shown in Table 1.

**Procedure**

1. Data Collection: Collect video data of human activities from existing datasets.

2. Data Preprocessing: Preprocess the video data by resizing, normalizing, and dividing it into frames. Extract the frames from the video and use them as input for the ResNet model.

3. Train/Test Split: Split the dataset into training and testing sets. Use 80% of the dataset for training and 20% for testing.

4. Feature Extraction: Extract features from each frame of the video using the pre-trained ResNet model. The ResNet model will extract the most relevant features from each frame.

5. Activity Recognition Model: The novel Resnet model uses a dataset to detect mortal behavior, which is displayed as textual material in the resulting video.

6. Train the Model: Train the activity recognition model on the training set.

7. Evaluate the Model: Evaluate the performance of the model on the testing set. Calculate metrics such as accuracy, precision, and recall.

8. Hyperparameter Tuning: Tune the hyperparameters of the model to achieve better performance.

9. Prediction: Use the trained model to predict the activity in new video data.

**Bayesian Regression**

The parameters of the model are inferred using the Bayesian logistic regression approach, and this inference follows the standard pattern for all other methods of analysis (Sihombing, Rohimah, and Kurnia 2021). This approach treats the model's parameters like random variables while treating the data as constant. The parameters also have a prior distribution. This is because Bayesian estimation permits probabilistic interpretations of the model coefficients. Bayesian estimation addresses the limitations of maximum likelihood
estimation in small samples and, due to its flexibility, overcomes the problem of the rigid assumptions of the classical approach. The main benefit is that you obtain a better range of inferential solutions through this Bayesian processing as opposed to a factor estimate and a self-belief as in classical regression. This is used here because of recognition of human activities compared to Resnet and the algorithm of Bayesian Regression was represented in Table1.

Procedure
1. Data Collection: Collect video data of human activities from existing datasets.

2. Data Preprocessing: Preprocess the video data by resizing, normalizing, and dividing it into frames. Extract the frames from the video and use them as input for the ResNet model.

3. Train/Test Split: Split the dataset into training and testing sets. Use 80% of the dataset for training and 20% for testing.

4. Feature Extraction: Extract features from each frame of the video using the pre-trained ResNet model. The ResNet model will extract the most relevant features from each frame.

5. Activity Recognition Model: The Bayesian Regression model uses a dataset to detect mortal behavior, which is displayed as textual material in the resulting video.

6. Train the Model: Train the activity recognition model on the training set.

7. Evaluate the Model: Evaluate the performance of the model on the testing set. Calculate metrics such as accuracy, precision, and recall.

8. Hyperparameter Tuning: Tune the hyperparameters of the model to achieve better performance.

9. Prediction: Use the trained model to predict the activity in new video data.

Statistical analysis

The statistical analysis of the proposed model was performed using SPSS IBM 2021 to determine the mean, standard deviation, significance, and to plot the graph, among other things, and the Python Collaboration Tool to assess the effectiveness of the algorithm (George and Mallery 2021). The independent variables are collections of dataset which possess human activities, whereas the main goal is to identify the actions of humans (input parameters). The dependent variables are the accuracy variables (output parameters).

RESULTS

The detection and recognition of human activities from selected video dataset is performed using classifiers namely Novel Residual Neural Network and Bayesian Regression from python compiler. Accuracy gain of Novel Residual Neural Network and Bayesian Regression recorded as 95.63% and 93.97%. The proposed Novel Residual Neural Network classifier recognised human activities at a huge rate and its accuracy gained visualizing the standard of such classifiers.

Table 1 displays the accuracy gain of Novel Residual Neural Network and Bayesian Regression classifier obtained from python compiler at different instants. The SPSS
estimation in small samples and, due to its flexibility, overcomes the problem of the rigid assumptions of the classical approach. The main benefit is that you obtain a better range of inferential solutions through this Bayesian processing as opposed to a factor estimate and a self-belief as in classical regression. This is used here because of recognition of human activities compared to Resnet and the algorithm of Bayesian Regression was represented in Table1.

**Procedure**

1. Data Collection: Collect video data of human activities from existing datasets.
2. Data Preprocessing: Preprocess the video data by resizing, normalizing, and dividing it into frames. Extract the frames from the video and use them as input for the ResNet model.
3. Train/Test Split: Split the dataset into training and testing sets. Use 80% of the dataset for training and 20% for testing.
4. Feature Extraction: Extract features from each frame of the video using the pre-trained ResNet model. The ResNet model will extract the most relevant features from each frame.
5. Activity Recognition Model: The Bayesian Regression model uses a dataset to detect mortal behavior, which is displayed as textual material in the resulting video.
6. Train the Model: Train the activity recognition model on the training set.
7. Evaluate the Model: Evaluate the performance of the model on the testing set. Calculate metrics such as accuracy, precision, and recall.
8. Hyperparameter Tuning: Tune the hyperparameters of the model to achieve better performance.
9. Prediction: Use the trained model to predict the activity in new video data.

**Statistical analysis**
The statistical analysis of the proposed model was performed using SPSS IBM 2021 to determine the mean, standard deviation, significance, and to plot the graph, among other things, and the Python Collaboration Tool to assess the effectiveness of the algorithm (George and Mallery 2021). The independent variables are collections of dataset which possess human activities, whereas the main goal is to identify the actions of humans (input parameters). The dependent variables are the accuracy variables (output parameters).

**RESULTS**
The detection and recognition of human activities from selected video dataset is performed using classifiers namely Novel Residual Neural Network and Bayesian Regression from python compiler. Accuracy gain of Novel Residual Neural Network and Bayesian Regression recorded as 95.63% and 93.97%. The proposed Novel Residual Neural Network classifier recognised human activities at a huge rate and its accuracy gained visualizing the standard of such classifiers.

Table 1 displays the accuracy gain of Novel Residual Neural Network and Bayesian Regression classifier obtained from python compiler at different instants. The SPSS statistical analysis took data available in table 1 as input and performed a comparative mean test.

<table>
<thead>
<tr>
<th>Accuracy (%)</th>
<th>Novel Residual Neural Network</th>
<th>Bayesian Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>92.01</td>
<td>90.62</td>
</tr>
<tr>
<td></td>
<td>92.32</td>
<td>91.00</td>
</tr>
<tr>
<td></td>
<td>92.53</td>
<td>91.37</td>
</tr>
<tr>
<td></td>
<td>93.08</td>
<td>92.13</td>
</tr>
<tr>
<td></td>
<td>93.34</td>
<td>92.88</td>
</tr>
<tr>
<td></td>
<td>94.06</td>
<td>93.02</td>
</tr>
<tr>
<td></td>
<td>94.47</td>
<td>93.17</td>
</tr>
<tr>
<td></td>
<td>94.86</td>
<td>93.28</td>
</tr>
<tr>
<td></td>
<td>95.16</td>
<td>93.43</td>
</tr>
<tr>
<td></td>
<td>95.64</td>
<td>93.67</td>
</tr>
<tr>
<td></td>
<td>96.28</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>96.49</td>
<td>94.52</td>
</tr>
<tr>
<td></td>
<td>96.84</td>
<td>94.91</td>
</tr>
<tr>
<td></td>
<td>97.05</td>
<td>95.23</td>
</tr>
<tr>
<td></td>
<td>97.35</td>
<td>95.53</td>
</tr>
<tr>
<td></td>
<td>97.73</td>
<td>95.76</td>
</tr>
<tr>
<td></td>
<td>98</td>
<td>95.99</td>
</tr>
<tr>
<td></td>
<td>98.26</td>
<td>96.05</td>
</tr>
<tr>
<td></td>
<td>98.52</td>
<td>96.15</td>
</tr>
<tr>
<td></td>
<td>98.63</td>
<td>96.82</td>
</tr>
</tbody>
</table>
Table 2: The mean and standard deviation of the group and accuracy of the Novel Residual Neural Network and BR algorithms were 95.63% and 2.19974, 93.97% and 1.82772, respectively. In comparison to the Novel Residual Neural Network, Bayesian regression had a lower standard error of 0.40869.

<table>
<thead>
<tr>
<th>GROUP NAME</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel Residual Neural Network</td>
<td>20</td>
<td>95.6310</td>
<td>2.19974</td>
<td>.49188</td>
</tr>
<tr>
<td>Bayesian Regression</td>
<td>20</td>
<td>93.9765</td>
<td>1.82772</td>
<td>.40869</td>
</tr>
</tbody>
</table>

Table 3: The independent sample test revealed a substantial variation in accuracy among the suggested Novel Residual Neural Network and BR (Bayesian Regression) classifiers. Since, significance value of $p=0.014$ that is $p<0.05$ which is statistical significance. There is a substantial variation among two methods.

<table>
<thead>
<tr>
<th>Two- Tailed Test</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene’s Test for Equality of Variances</td>
<td></td>
<td></td>
<td></td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal Variances assumed</td>
<td>1.329</td>
<td>.256</td>
<td>2.587</td>
<td>38</td>
<td>.014</td>
<td>1.6545</td>
</tr>
<tr>
<td>Equal Variances not assumed</td>
<td>2.587</td>
<td>36.766</td>
<td>.014</td>
<td>1.6545</td>
<td>.6395</td>
<td>.3585</td>
</tr>
</tbody>
</table>
**Table 2:**
The mean and standard deviation of the group and accuracy of the Novel Residual Neural Network and BR algorithms were 95.63% and 2.19974, 93.97% and 1.82772, respectively. In comparison to the Novel Residual Neural Network, Bayesian regression had a lower standard error of 0.40869.

**Table 3:**
The independent sample test revealed a substantial variation in accuracy among the suggested Novel Residual Neural Network and BR (Bayesian Regression) classifiers. Since, significance value of $p=0.014$ that is $p<0.05$ which is statistical significance. There is a substantial variation among two methods.

**Table 2** displays the comparative mean test is categorized as group statistical analysis and independent sample test. By taking 20 samples per group, the mean accuracy, standard deviation and standard error mean is obtained. The values obtained from group 1 are 95.63%, 2.19974, 0.49188 and group 2 having 93.97%, 1.82772, 0.40869 respectively. Table 3 implies assumption and non assumption of equal variance in accuracy for selected classifiers. For this analysis, significance value of $p=0.014$ (Independent Sample t-test $p<0.05$) which is statistical significance.

**Figure 1**
The graph obtained from statistical analysis is visualized in figure 1. From table 2, the mean accuracy value is chosen and the mean accuracy comparison graph is prepared. The X-axis denotes suggested classifiers and Y-axis denotes accuracy value. The mean accuracy of proposed and conventional classifiers is 95.63% and 93.97%

**DISCUSSION**

The mean accuracy of Bayesian Regression is 93.97%, whereas the Novel Residual Neural Network Algorithm has higher accuracy of 95.63% which shows that Novel Residual Neural Network Algorithm performs better when compared with Bayesian Regression. The SPSS carried out has a significance value $p= 0.014$ (Independent Sample t-test $p<0.05$), which show that it is statistically significant.

Novel ResNet are one of the maximum efficient Neural Network Architectures, as they assist in retaining a low blunders rate a great deal deeper withinside the network.(Weiss, Yoneda, and Hayajneh 2019). To understand short-duration sports, a novel RNN technique is proposed. There is a growing diversity of virtual camera deployments technology, which is having a negative impact on worker output.(R et al. 2019). Despite the fact that numerous
researchers have developed a variety of prediction models, many of them fail to correctly predict higher values. (Jaimes and Sebe). Novel ResNets has drawbacks in the fact that error detection is difficult for deeper networks. Also, training can be very inefficient if the network is too small (Etemad). Novel ResNets led to deeper networks and led to wider networks (Wang, Yue, and Faraway 2018). By creating Novel Residual Neural networks in the future technology, this quality of detection can be further enhanced.

The drawback of this examination is that it takes a completely long term to teach a Novel Residual Neural Network particularly with huge datasets. The destiny scope of this examination is that machine has to be accelerated to consist of a massive quantity of photo documents with much less time intake in future. Further may be completed within the area of photo processing to enhance accuracy in residual neural network. A Novel Residual Neural Network may be created to enhance the accuracy of the model.

CONCLUSION

A video clip is selected as input for this analysis and it contains the following actions namely a man walking with a dog, high jump, Playing tabla. These actions are recognised by the recommended classifiers and the suggested classifier achieved overall accuracy of 95.63% which is larger than the overall gain of 93.97% attained by Bayesian Regression. The selected classifier can detect multiple actions of humans in a menial time period.

DECLARATION

Conflicts of Interest

No conflicts of interest in this manuscript.

Author Contributions

Author NV was responsible for all data collection, data analysis, and manuscript preparation. Author RS was responsible for all conceptualization, data validation, and critical evaluation of papers..

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

2. Saveetha School of Engineering
3. Saveetha University
4. Saveetha Institute of Medical and Technical Science

REFERENCE


