PERFORMANCE ANALYSIS OF DEEP LEARNING ALGORITHMS FOR HUMAN ACTION RECOGNITION USING SPATIO TEMPORAL FEATURES FROM VIDEO IMAGES

Thesis submitted to Bharathiar University
in partial fulfillment of the requirements for the award of the Degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

By

N. SRI LAKSHMI

Under the Guidance and Supervision of

Dr. (Mrs.) N. RADHA, M.Sc., M.Phil., Ph.D.,
Associate Professor and Head
Department of Data Analytics (PG)
PSGR Krishnammal College for Women, Coimbatore-641004

DEPARTMENT OF COMPUTER SCIENCE
PSGR KRISHNAMMAL COLLEGE FOR WOMEN
College of Excellence-Awarded by UGC
(An Autonomous Institution - Affiliated to Bharathiar University)
(Reaccredited with "A++" Grade by NAAC- 4th Rank in NIRF 2023 by MHRD)
(An ISO 9001:2015 Certified Institution)
COIMBATORE - 641004
TAMIL NADU, INDIA

SEPTEMBER 2023

CHAPTER IX

RESULTS AND DISCUSSION

In this chapter the effective model of JTDGPAHBRD-GCN is analysed by comparing with the other existing models like JTDD, JTPADBRD, JTDPAHBRD, JTDGPAHBRD for predicting the human action recognition. The goal of this comparison is to show that the JTDGPAHBRD-GCN can improve recognition and accuracy rates more effectively than previously available models.

Using JTDD extraction effectively raises the HAR to 0.987%. However, predicting the locations of body joints using a huge dataset required a lot of time, and the skeleton estimation algorithm was computationally expensive. The JTPADBRD were utilized in order to address these issues. It improves recognition accuracy on the Penn Action dataset to roughly 99.4% when compared to state-of-the-art methods. The problem with this method is that it smoothes out crucial changes in time and space between categories. The vanishing gradient problem manifested itself in PABRNN because of the larger number of parameters. The JTDPAHBRD framework was used to find solutions to these issues. It outperforms previous methods by a factor of 99.6 in terms of recognition accuracy. The trajectory of a single joint only communicates gesture information, not contour or geometrical details. An alternative HAR framework based on the JTDGPAHBRD is offered to address these problems. The accuracy of its recognition is 99.7%. However, further investigation into the spatial-temporal dynamics of the various skeletal structure geometric aspects was lacking. The JTDGPAHBRD model was modified to include the GCN in order to learn spatio-temporal information from the skeleton graph and so solve these problems. The proposed study predicts a recognition rate of 99.82% when compared to the other HAR models by merging features from the conv5b and conv4b layers with the GCN characteristics.

9.1 SUMMARY OF RESULTS AND DISCUSSION

To evaluate the effectiveness of the presented models, MATLAB 2017b is used. The Penn Action dataset is taken into account for the experimental analysis; it contains 2326 640x480 video sequences, each of which is labeled with one of 15 action classifications. All clips are assembled from several web video libraries and involve

50–100 blocks, each of which has 13 annotated joints. Of these sequences, 1861 are utilized for learning, while 465 video sequences are utilized for testing. Sources include C3D features, coordinates of primitive geometries, trajectory coordinates, and spatiotemporal correlations.

Table 9.1. Performance Analysis of Existing and Proposed HAR Models

Models		Accuracy	Precision	Recall	F-measure
Phase I	JTDD	98.7%	0.975	0.982	0.979
Phase II	JTPADBRD	99.4%	0.983	0.991	0.987
Phase III	JTDPAHBRD	99.6%	0.989	0.994	0.992
Phase IV	JTDGPAHBRD	99.7%	0.992	0.996	0.994
Phase V	JTDGPAHBRD-GCN	**99.82%**	**0.995**	**0.998**	**0.997**

From Table 9.1, it is observed that the performance of the JTDGPAHBRD-GCN model (concatenating features from the conv5b and conv4b layers) is greater than all other models for recognizing human actions in the videos. This model can learn various features from the multiple both long-range and short-range video sequences for generating video descriptors, which is useful to recognize human actions, in contrast with the other models. Hence, according to the observed accuracy, the JTDGPAHBRD-GCN model is useful for effective HAR.

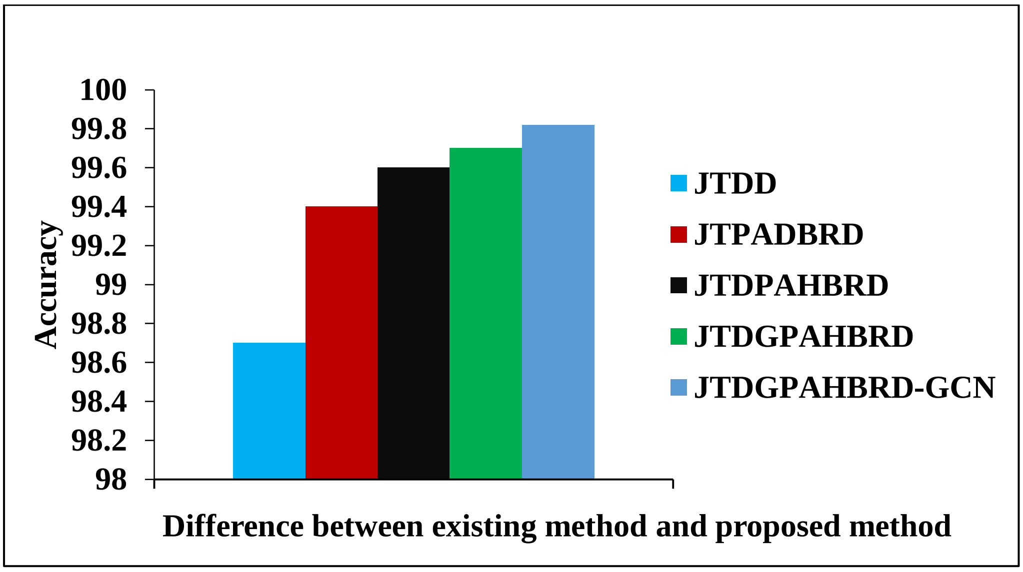

Fig 9.1 Overall Comparison of accuracy of proposed method

Overall accuracy comparison of the JTDD, JTPADBRD, JTDPAHBRD, JTDGPAHBRD, and JTDGPAHBRD-GCN is shown in Figure 9.1. When it comes to forecasting human action recognition, the proposed method of JTDGPAHBRD-GCN has a good accuracy range. When compared to JTDD, JTPADBRD, JTDPAHBRD, and JTDGPAHBRD, the JTDGPAHBRD-GCN are 1.13%, 0.42%, 0.22% and 0.12% higher, respectively. This study demonstrates that the proposed JTDGPAHBRD-GCN achieves higher accuracy in human action recognition than current approaches.

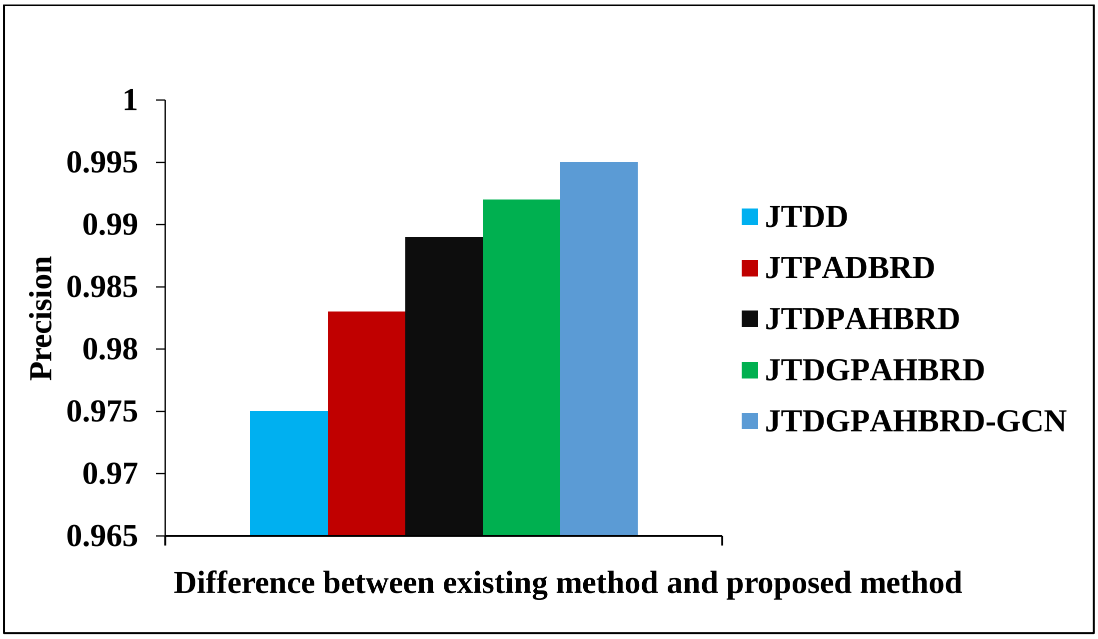

Fig 9.2 Overall Comparison of precision of proposed method

Figure 9.2 depict the overall comparison of precision of the JTDD, JTPADBRD, JTDPAHBRD, JTDGPAHBRD, JTDGPAHBRD-GCN. When it comes to forecasting human action recognition, the proposed method of JTDGPAHBRD-GCN has a high Precision range. The JTDGPAHBRD-GCN are 2.05% higher than the JTDD, 1.22% higher than the JTPADBRD, 0.60% higher than the JTDPAHBRD, and 0.302% higher than the JTDGPAHBRD. The results of this study show that the proposed JTDGPAHBRD-GCN is more precision at recognizing human actions than current approaches.

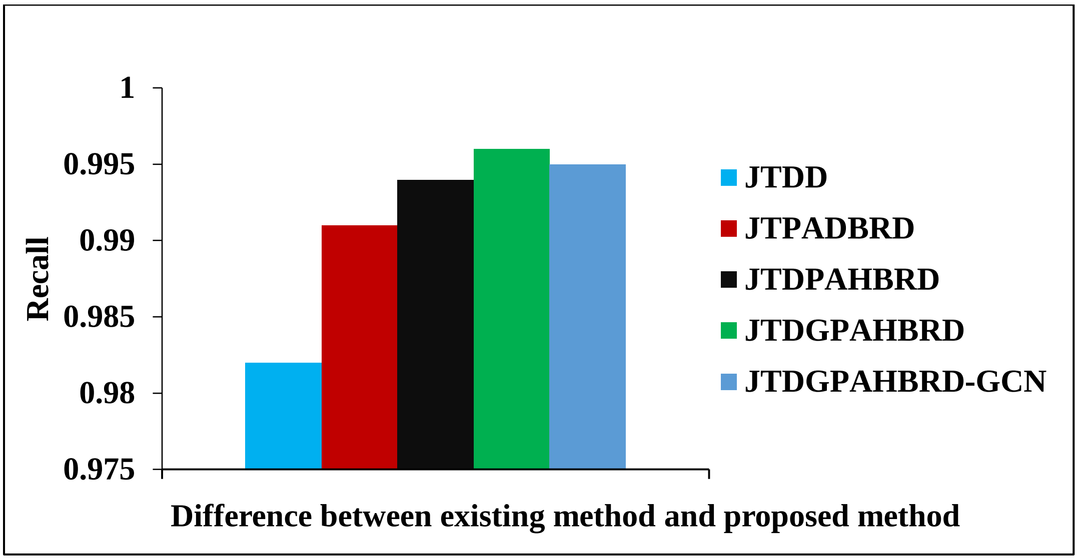

Fig 9.3 Overall Comparison of recall of proposed method

Figure 9.3 depict the overall comparison of recall of the JTDD, JTPADBRD, JTDPAHBRD, JTDGPAHBRD, JTDGPAHBRD-GCN. JTDGPAHBRD-GCN is a proposed method with a high recall range for predicting human action recognition. It is observed that the JTDGPAHBRD-GCN are 1.62% higher than the JTDD, 0.706% higher than JTPADBRD, 0.402% higher than JTDPAHBRD, 0.200% higher than JTDGPAHBRD. Compared to state-of-the-art approaches for human action recognition, the suggested JTDGPAHBRD-GCN is shown to have superior recall in this study.

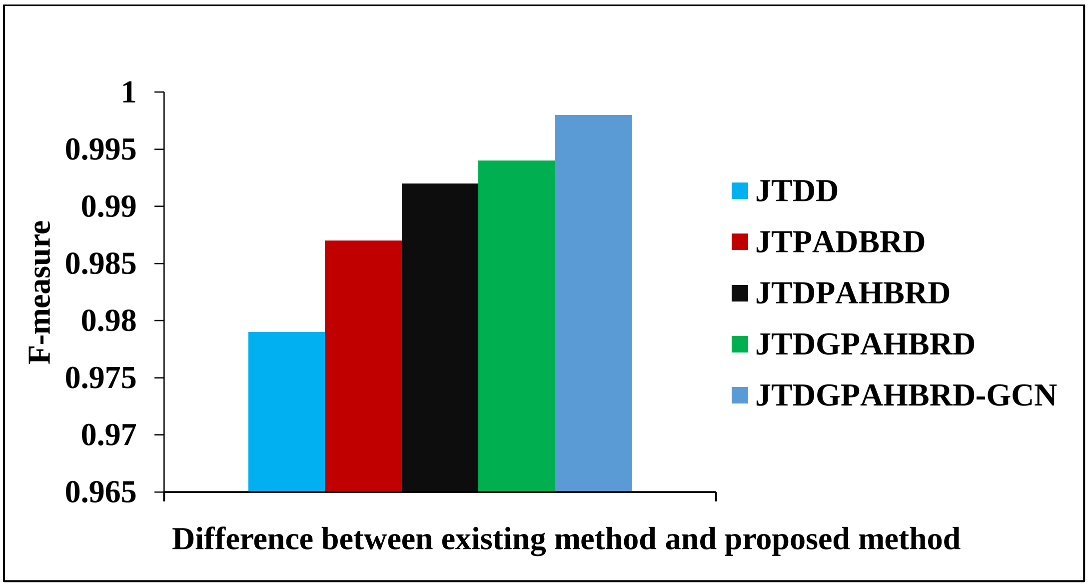

Fig 9.4 Overall Comparison of F-measure of proposed method

Figure 9.4 depict the overall comparison of F-measure of the JTDD, JTPADBRD, JTDPAHBRD, JTDGPAHBRD, JTDGPAHBRD-GCN. When it comes to forecasting human action recognition, the F-measure range of JTDGPAHBRD-GCN is very high. There is a 1.83% gap between JTDD and JTDGPAHBRD-GCN, 1.13% between JTPADBRD and JTDD, 0.5% between JTDPAHBRD and JTDD, and 0.3% between JTDD and JTDGPAHBRD-GCN. This study demonstrates that the proposed JTDGPAHBRD-GCN outperforms state-of-the-art approaches in human action recognition on the basis of the F-measure.

9.2 CHAPTER SUMMARY

For the purpose of anticipating human action recognition in the video sequence, this chapter proposes an overall comparison with JTDGPAHBRD-GCN. In terms of accuracy, precision, recall, and F-measure for predicting human action recognition in video sequences, experimental results suggest that the proposed JTDGPAHBRD-GCN technique performs better than state-of-the-art methods.

CHAPTER X

RESEARCH FINDINGS

The primary goal of this study is to improve the accuracy of HAR from video sequences by creating a strong deep-learning model. The primary goal of this work is to generate video descriptors for human activity recognition by learning features such as body joints, trajectory points, geometric elements, and spatiotemporal data. In order to improve the accuracy of human activity recognition, five models are proposed here to learn various aspects associated with human activities. Experiments are conducted with the Penn Action dataset. The following conclusions can be drawn from the experimental data:

- By learning both body joints and trajectories in video frames from the Penn Action dataset, the JTDD model achieved 98.7% recognition accuracy, compared to the JDD model.
- Based on the learning of relevant spatial dissimilarities among various video frames belonging to different classes from the Penn Action dataset, the JTPADBRD model attained 99.4% recognition accuracy, compared to the JTDD model.
- By learning long-range body joint correlations across various classes of video sequences in the Penn Action dataset, the JTDPAHBRD model achieves 99.6% recognition accuracy, which is an improvement over the JTPADBRD.
- The JTDGPAHBRD model, by recording different geometries together with the body joints and trajectories in several video frames, outperformed the JTDPAHBRD model on the Penn Action dataset, with a 99.7% recognition accuracy.
- Finally, the JTDGPAHBRD-GCN reaches the maximum recognition accuracy of 99.82% compared to the other models on the Penn Action dataset, because of learning body joints, trajectory points, geometrics, as well as spatiotemporal dependencies from long-range video sequences for video descriptor generation.
- When comparing models, the JTDGPAHBRD-GCN outperforms the JTDD, JTDPABRD, JTDPAHBRD, and JTDGPAHBRD models on the Penn Action

dataset by a margin of 1.13%, 0.42%, 0.22%, and 0.12, respectively, when concatenating features from the conv5b and conv4b layers.

- Thus, it is realized that the JTDGPAHBRD-GCN model obtains the maximum recognition accuracy, whereas the JTDD produces the minimum results.

10.1 CONCLUSION

This research work address issues and problems related to inconsistencies in the literature survey. In several contexts, Human Activity Recognition (HAR) has risen to prominence. The model developed must satisfy the objectives specified in the research work proposed and function well in sports applications.

- The existing deep learning models are analyzed, and the JTDD model is proposed for learning the trajectory points of human activities along with the body joints efficiently.
- In order to learn the important spatiotemporal connections between body joints and trajectory points from the various classes of human activity video sequences, the JTDPABRD model was developed.
- The JTDPAHBRD model is developed for enhancing the feature pooling strategy in the C3D network.
- Different geometries, body joints, and trajectory points can be extracted from various video sequence types using the JTDGPAHBRD model.
- The JTDGPAHBRD-GCN model is developed for learning the spatiotemporal dependencies among various geometrics from the human activities and improving the feature learning ability for video descriptor generation in the HAR.
- Using the Penn Action dataset, the accuracy, precision, recall, and f-measure of five different recognition models were evaluate.
- The JTDGPAHBRD-GCN model showed an accuracy of 99.82%, precision of 0.995, recall of 0.998 and f-measure of 0.997, compared with the existing models.

10.2 FUTURE WORK

This proposed research can involve the following future enhancements:

1. The work be extended to integrate metaheuristic optimization algorithms to optimize the proposed models according to various datasets.
2. Data fusion algorithms can be explored to effectively integrated information from diverse modalities like face, palmprint, eye movements, etc.
3. Third, the scope can be broadened to incorporate the creation of transfer learning mechanisms that can transport expertise from a domain with an abundance of labeled data to one with fewer such labels. This can make models more robust to different environments and user variations.
4. HAR scenarios often involve continuous data streams. In the future, online learning can be introduced that can adapt the model to changing activity patterns over time without the need to retrain from scratch.