

## ABSTRAK

*Human Activity Recognition* (HAR) merupakan kunci untuk lebih memahami berbagai semantik yang terkandung dalam video. Dalam penelitian ini, menerapkan kombinasi metode CNN dan LSTM dengan menggunakan beberapa variasi nilai parameter model pada *dropout layer* dan *batch size*. Pada prosesnya, lapisan konvolusi yang digunakan untuk mengekstraksi fitur spasial dari frame, akan diumpangkan ke lapisan LSTM pada setiap jaringan untuk pemodelan urutan temporal. Dengan cara ini jaringan pada model akan mempelajari fitur *spatiotemporal* secara langsung dalam pelatihan *end-to-end*, untuk menghasilkan model yang kuat. Data yang digunakan yaitu 10 aktivitas olahraga yang diperoleh dari topik penelitian terkait milik *University of Central Florida* (UCF). Hasil penelitian menunjukkan kinerja yang cukup baik, meskipun masih terdapat kekeliruan klasifikasi aktivitas olahraga karena memiliki kemiripan pergerakan aktivitas yang dilakukan. Hasil klasifikasi menunjukkan nilai akurasi tertinggi diperoleh pada variasi yang tidak menerapkan *dropout layer* dengan *batch size* bernilai 4, menghasilkan nilai akurasi 0,94 dan nilai loss 0,4, sedangkan nilai akurasi terendah diperoleh pada variasi yang menerapkan *dropout layer* bernilai 0,4 dengan *batch size* bernilai 8, menghasilkan nilai akurasi 0,84 dan nilai loss 0,57.

**Kata Kunci:** *Human Activity Recognition (HAR)*, Klasifikasi, *Convolutional Neural Network*, *Long Short-Term Memory*.

## **ABSTRACT**

*Human Activity Recognition (HAR) is the key to better understanding the various semantics contained in videos. In this study, the application of the CNN and LSTM combination method used several variations of model parameter values in the dropout layer and batch size. In the process, the convolution layer which is used to extract the spatial features from the frame, will be fed to the LSTM layer in each network for temporal sequence modeling. In this way the network on the model will learn spatiotemporal features directly in end-to-end training, to produce a robust model. The data used are 10 sports activities obtained from related research topics belonging to the University of Central Florida (UCF). The results showed that the performance was quite good, although there were still errors in the classification of sports activities because they had similarities in the movements of the activities carried out. The classification results show that the highest accuracy value is obtained for variations that do not apply a dropout layer with a batch size of 4, resulting in an accuracy value of 0,94 and a loss value of 0,4, while the lowest accuracy value is obtained for a variation that applies a dropout layer of 0,4 with a batch size of 8, resulting in accuracy value of 0,84 and loss value of 0,57.*

**Keywords:** *Human Activity Recognition (HAR), Classification, Convolutional Neural Network, Long Short-Term Memory.*