PREFACE

The field of computer vision is dominated by research on Human Action Recognition (HAR) because of its significant applications in human-computer interaction, multimedia indexing, surveillance, and healthcare, among other areas. The need for sophisticated techniques that can accurately record the complex dynamics of human movements is growing along with the desire for intelligent systems that can comprehend and interpret human actions. The merging of LSTM and CNN is the focus of this project. This hybrid technique has demonstrated significant potential in improving the accuracy and efficiency of HAR systems.

The difficulties and restrictions found in current HAR approaches—particularly those that depend on conventional staged action datasets—served as the impetus for this effort. With its realistic and varied YouTube video samples, the UCF50 Action Recognition Dataset is a perfect platform to experiment with cutting-edge HAR methods. Our goal in concentrating on this dataset is to tackle the intricacies brought forth by real-world scenarios, like different camera movements, different object appearances, and crowded backdrops.

Our approach builds two powerful frameworks: known as Long-term Recurrent Convolutional neural network along with the Convolutional LSTM, by leveraging CNNs' advantage in extracting spatial features as well as LSTMs' power in temporal sequence modelling. This integration opens the door for strong action recognition systems that can function in dynamic and unpredictable contexts, as well as for the thorough analysis of video data.

This project's introduction summarises the reasons for our investigation and emphasises the need of using hybrid deep learning models for HAR. It provides an overview of the thorough procedures, experimental analyses, and conclusions drawn from this study. We are confident that our effort will further the current state of HAR research and offer insightful references for next studies and applications in this area.

We have worked hard to strike a balance between theoretical investigation and real-world application during this project, making sure that our conclusions are both rigorously scientific and applicable to real-world situations. This report is to provide a thorough and readable presentation of our research, including insights into the methods used, the outcomes attained, and the possible next steps for this fascinating flied of study.

In addition to adding to the corpus of knowledge already known about HAR, my goal with this effort is to spur additional advancements and discoveries. The future of human-computer interaction and beyond will surely be greatly influenced by our capacity to precisely identify and interpret human activities as we move closer to more intelligent and responsive systems.