**INTRODUCTION**

In HAR, an action is an observable thing that may be seen by the human eye or by some sensor device. In fact, engaging in an activity like walking requires constant attention to a person located inside the field of view. Actions may be separated into four categories depending on the body components needed to do them. [1]. x Gesture: It's based on facial expression. Don't need any action or verbal way of communication. x Action: Walking, playing, punching human action comprised. x Interaction: It Consist human object interaction as well human interaction like handshake, hugging is example of interaction. x Group activity: when more than two action is happening is known as group activity like combination of gesture and interaction. Two or more than two actor is involved for performing action. Over the course of the last two decades, HAR has emerged as an essential component of research into computer vision. Based on a collection of observations, HAR is designed to detect and identify activities carried out by one or more persons. This may be done for any number of people. This had developed into a need for the further development of human–computer interaction. Many researchers from different parts of the globe are drawn to this field of study due to the wide range of domains in which it may be used. Surveillance video, the labeling and retrieving of visuals, monitoring of health, automation, and environmental modelling are but a few of its most prominent applications[1]. Human activities have an intrinsic hierarchical structure that denotes their many levels, which may be categorized at three levels. First, there is an atomic element at the lowest level, and these action primitives represent increasingly complicated human actions. The actions/activities level follows the action primitive level as the second level. The greatest level of categorization for human activities is represented by complex interactions. Each of these categories is sufficiently extensive to need its own area of study. This is mostly due to the unpredictability and ambiguity of human acts in a real-world situation. HAR faces several barriers and obstacles. Gender discrimination, multi-subject interactions, and disparities in inter-class activity are some examples. Human action recognition form videos have four stage process. In first step we extract feature from given sequences of images. In the feature extraction process, handcrafted methods just like SIFT (scale invariant feature transform), SURF (Speed up Robust Feature), Shape-based, Pose-based, and optical flow, to mention a few, may be used [1]. Feature extraction can be done with deep learning method automatically in this approach model learn itself all feature from sequences of images. It involves the extraction of poses and gesture patterns from frames and sequences of visuals depicting human activities. Therefore, it is a challenging task due to obstacles such as size fluctuations, poor illumination, incorrect viewpoints, and backdrop clutter.