The detection and recognition of words has grown to be a serious problem in recent years. Deep learning and computer vision developments, as well as an increase in text detection and recognition applications, have all contributed to this trend. A difficult problem for many applications is text information extraction from photos. The wide range of differences in backdrops, textures, typefaces, and lighting conditions make it more difficult to recognize words in natural photos than it is to recognize characters or numbers in scanned texts. This research examines the difficulty of word recognition, one of the most crucial and challenging tasks in image-based sequence recognition. An architecture for a hybrid neural network that combines CNNs (Convolutional Neural Networks), which are good at extracting features, RNNs (Recurrent Neural Networks), which predict output sequentially each time step, and CTC Loss functions, which predict output for each individual time step. The main objective of the proposed method is to predict the word from the images. The parameters of the neural network are tuned in such a way that optimized results are obtained with minimum training loss. The dataset is used here is taken from Visual Geometry Group, the images in the dataset are converted to grey scale after this padding is applied to alter the image size then each image dimension is extended to fit the architecture input shape. After this, the model architecture is compiled through which the dataset is trained using Adam Optimizer. The approximate final model accuracy is 76.59 % with validation accuracy of up to 65.70 %.