In l22], a detection-based tracking model was established. It was possible to effectively identify potential dangers to workers, but the method that was used only recorded incidents that involved a single worker; multi-labour analysis was not taken into consideration. The spatial-temporal interaction between workers and machinery is rarely evaluated in a holistic manner, which implies that potential effects from a range of viewpoints are not taken into account. In the meantime, HOG features were utilized in order to detect workers and machines in a different study [24], and a particle filter was utilized in order to follow the activities of these individuals. By utilizing the SIFT algorithm to extract visual characteristics from the images. Despite the fact that a variety of tracking algorithms that are dependent on detection have been developed for the purpose of monitoring construction sites, the potential that deep learning holds in this area has not yet been thoroughly investigated. The third category of talents encompasses the capability of identifying activities as one of its components. In order to support autonomous and real-time monitoring of on-site safety, a robust strategy that evaluates the spatial-temporal interaction between employees and equipment is necessary. The data were inputted into a Kalman filter, which was used to make forecasts about the course that the system will take in the future based on the measurements that had already been done. A few examples of these behaviors are climbing ladders while carrying heavy goods in their hands, climbing ladders while facing the wrong direction, and reaching too far. The authors in l23] were able to differentiate between humans and computers in the images. These projections were based on the fact that the system had been measured previously. In [25], the authors to identify potentially dangerous activities in employees. This can have negative health consequences for workers. It is necessary to put into practice this strategy.