Preface

Recent advances in Internet of Things (IoT) and sensor technology has influenced various aspects of human-life. It has highly improved the monitoring and control applications. IoT describes the network of physical objects (or things) that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet. IoT finds its one of the applications in water pollution assessment and monitoring which requires real-time monitoring of prime water parameters such as, potential of Hydrogen (pH), Dissolved Oxygen (DO), turbidity, Nitrates (NO2), Biochemical Oxygen Demand (BOD), Fecal Coliforms (FC) *etc.* to take intelligent decisions. With the availability of low-cost sensors such as pH, Turbidity, for monitoring water parameters, it becomes easier and convenient to assess the quality of river water or to identify pollution level. Information about the pollution level helps to take preventive actions to control the contamination of the river water. The estimation of water quality is also referred as Water Quality Index (WQI). It helps to identify water suitability for different purposes like drinking, irrigation, bathing, *etc.*. In water quality assessment, the sensors measure various water parameters and the generated data can be used to analyze the pollution level in the water bodies (e.g., river, pond, lake, *etc.*). In other words, WQI is a parameter which tells about the purity of water for a given purpose of its use. It ranks the water quality data into terms that can be easily understood by public.

In this thesis, we investigate the different challenges encountered while monitoring the real-time river water pollution using various smart sensing-based approaches. We consider two tasks in this work: a) sensor-based river water pollution assessment using deep neural network and b) river water pollution monitoring using knowledge distillation and game theory. While considering the task of developing a deep neural network for river water pollution assessment, we identify challenges, *i.e.*, unlabeled limited lab dataset, unlabeled huge sensory dataset and noisy labels in the dataset. An annotation of unlabeled lab dataset is a time consuming process. All the water parameters are analysed in the lab and then annotated the lab data. It can take several days and sometimes one month also. Next, there does not exist any direct technique to annotate the massive sensory data. The automatic annotation technique inserts the noisy labels in the dataset. These challenges deteriorate the performance and increase the training time. We develop approaches to annotate unlabeled limited lab data and huge sensory data automatically and manage noisy labels in the dataset to enhance the performance of the classifier. Further, the real-time river water pollution monitoring also incurs challenges of resources inadequacy and colossal energy consumption. We also present the approaches to compress the cumbersome deep neural network and reduce the energy consumption required for transmission of pollution data from a river to the base station.

First, we propose a mathematical model to estimate the water quality index giving information about river water pollution level using various water parameters incorporated in limited unlabeled lab data. The mathematical model to estimate WQI comprises four steps: selection of distinct parameters, establishing the weights for selected parameters, calculating the sub-indices for selected parameters and finally, aggregation of weights and subindices to estimate the final WQI. Further, we automatically labeled the limited lab data. Next, we present an automatic label transfer mechanism to annotate the massive sensory data automatically. The automatic annotation technique inserts the noisy lables into the training instances which deteriorates the classifier performance. Therefore, we present a noise handling loss function. The loss function comprises two terms. The first term assigns weights to the true labels and the second term assigns weights to the predicted labels. The noise handling loss function mitigate the damage caused by noisy labels. Next, we present a deep learning-based approach to detect a river water pollution level using LSTM model in the presence of noisy labels. The model builds a classifier by learning a mapping between the extracted features and the class labels of the massive sensory data. Further, we propose a river water pollution monitoring system that incorporates knowledge distillation approach to train the compressed model under the supervision of cumbersome model. Further, the trained compressed model is deployed on limited resources Internet of things devices and achieves the acceptable accuracy in identifying the water pollution level. Finally, we design a game theory-based approach to estimate the suitable time duration for using the spreading factor of the long range network to transmit the pollution data from a river to the base station with minimum energy consumption and ensures the successful transmission of data.