

# A Review on Machine Learning Approach on Data Imbalance for wireless Sensor Network

Reshma Lakra

rashmi.rohitas0001@gmail.com

Dept. of Computer Science & Engineering

Sagar Institute of Research and Technology

Bhopal, M.P., India

Prof. Ruchi Dronawat

dron.ruchi@gmail.com

**Abstract-** Machine learning is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the pattern or extract information from the data. In that case, we apply machine learning with the abundance of datasets available. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves. Data imbalance problem become greatest issue in for machine learning algorithm. Imbalance problem occur where one of the two classes having more sample than other classes. The most of algorithm are more focusing on classification of major sample while ignoring or misclassifying minority sample. The most of algorithm are more focusing on classification of major sample while ignoring or misclassifying minority sample. The minority samples are those that rarely occur but very important. There are different methods available for classification of imbalance data set which is divided into three main categories, the algorithmic approach, data-preprocessing approach and feature selection approach. We will apply on wireless imbalance data to identify correct information. In this paper systematic study for define which gives the right direction for research in class imbalance problem.

**Keywords-** Unsupervised learning, Supervised learning, Wireless Sensor, Machine Learning algorithm.

## I. INTRODUCTION

Wireless sensor networks (WSNs) have been applied in monitoring systems that are capable of controlling and monitoring various indoor premises. WSN are collections of stand-alone devices which, typically, have one or more sensors (e.g. temperature, light level), some limited processing capability and a wireless interface allowing communication with a base station. As they are usually battery powered, the biggest challenge is to achieve the necessary monitoring whilst using the least amount of power.

A Wireless sensor network (WSN) is composed typically of multiple autonomous, tiny, low cost and low power sensor nodes. These nodes gather data about their environment and collaborate to forward sensed data to centralized backend units called base stations or sinks for further processing. The sensor nodes could be equipped with various types of sensors, such as thermal, acoustic, chemical, pressure, weather, and optical sensors. In particular, WSN designers have to address common issues related to data aggregation, data reliability, localization, node clustering, energy aware routing, events scheduling, fault detection and security [1]. Machine learning (ML) was introduced in the late 1950's as a technique for artificial intelligence (AI).

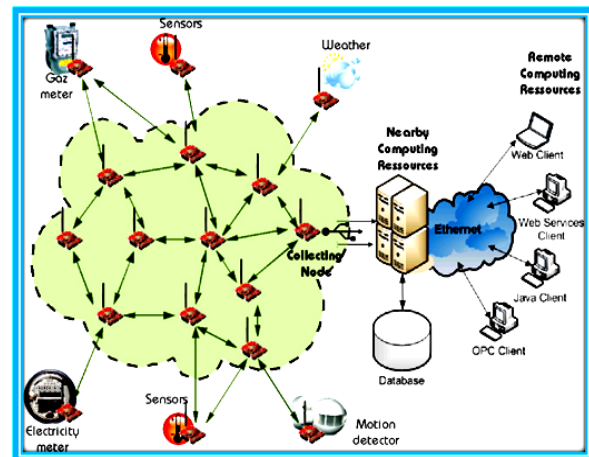


Fig.1 Architecture of Wireless Sensor.

Over time, its focus evolved and shifted more to algorithms which are computationally viable and robust. In the last decade, machine learning techniques have been used extensively for a wide range of tasks including classification, regression and density estimation in a variety of application areas such as bioinformatics, speech recognition, spam detection, computer vision, fraud detection and advertising networks. The algorithms and techniques used come from many diverse fields including statistics, mathematics, neuroscience, and computer science. [1]

## II. LITERATURE REVIEW

In this literature a systematic study for define which gives the right direction for research in class imbalance problem.

### 1. Author: - Remah Alshinina and Khaled Elleithy

**Problem:-** Author indentified Generative Adversarial Networks (GANs) algorithm. GANs contain two networks: a generator (G) network and a detector (D) network. The G creates fake data similar to the real samples and combines it with real data from the sensors to confuse the attacker. The D contains multi-layers that have the ability to differentiate between real and fake data.

**Solution:** - Machine learning Approach for Data transmission from the WSN to the end user then becomes much more secure and accurate compared to conventional techniques. [1]

### 2. Author: -Mohammad Abu Alsheikh, Shaowei Lin

**Problem:** - Wireless sensor networks monitor dynamic environments that change rapidly over time. This dynamic behavior is either caused by external factors or initiated by the system designers themselves.

**Solution:** - By A Visualization of the Q-learning method for solutions for their specific application challenges. [2]

### 3. Author: -Shengguo Hu and Yanfeng Liang

**Problem:** - Not Accurate SMOTE (Synthetic Minority Over-sampling Technique) is specifically designed for learning from imbalanced data sets external factors or initiated by the system designers themselves.

**Solution:** - By modified approach (MSMOTE) for learning from imbalanced data sets, based on the SMOTE algorithm. [3]

### 4. Author: -Shengguo Hu and Yanfeng Liang

**Problem:** - Not Accurate SMOTE (Synthetic Minority Over-sampling Technique) is specifically designed for learning from imbalanced data sets external factors or initiated by the system designers themselves.

**Solution:** - By modified approach (MSMOTE) for learning from imbalanced data sets, based on the SMOTE algorithm. [4]

### 5. Author: -L.Dhanabal1, Dr.S.P. Shantharajah

**Study:-** In this paper the NSL-KDD data set is analyzed and used to study the effectiveness of the various classification algorithms in detecting the anomalies in the network traffic patterns.

**Solution:** - The study has exposed many facts about the bonding between the protocols and network attacks.[4]

### 6. Author:- Huaping Guo, Jun Zhou

**Problem:** - Classification of data with imbalanced class distribution has encountered a significant drawback by

most conventional classification learning methods which assume a relatively balanced class distribution.

**Solution:** - For the learning stage, the proposed method uses the following three steps to learn a class-imbalance oriented model: (1) partitioning the majority class into several clusters using data partition methods such as K-Means, (2) constructing a novel training set using SMOTE on each data set obtained by merging each cluster with the minority class, and (3) learning a classification model on each training set using convention classification learning methods including decision tree, SVM and neural network[3]

## III. PROBLEM STATEMENTS

By Above literature we analyzed wireless sensor network generated a huge amount of imbalance data. Wireless sensor data like weather, temperature, motion detector, mobile data extra. Also analyzed different machine learning approaches for improve for imbalanced data, like SMOTE and MSMOTE.

- SMOTE disadvantage with under sampling is that it discards potentially useful data.
- A second reason for using sampling is that many highly skewed data sets are enormous and the size of the training set must be reduced in order for learning to be feasible.

### 1. Proposed system

In this proposed solution we will reduced data imbalanced problem for wireless sensor network and achieve a accuracy in results steps of proposed algorithm are given blows.

- Start Sensing data from different wireless sensor.
- Collect Data from Wireless sensors and check for balanced or imbalanced.
- If data is balanced we will use directly for Machine Learning classification Approach.
- If data is imbalanced will use a Low Cast High Accurate Synthetic Minority Over-sampling Technique LCHA-SMOTE.
- Finally we will compare with different algorithm for imbalanced data. And also check their accuracy.

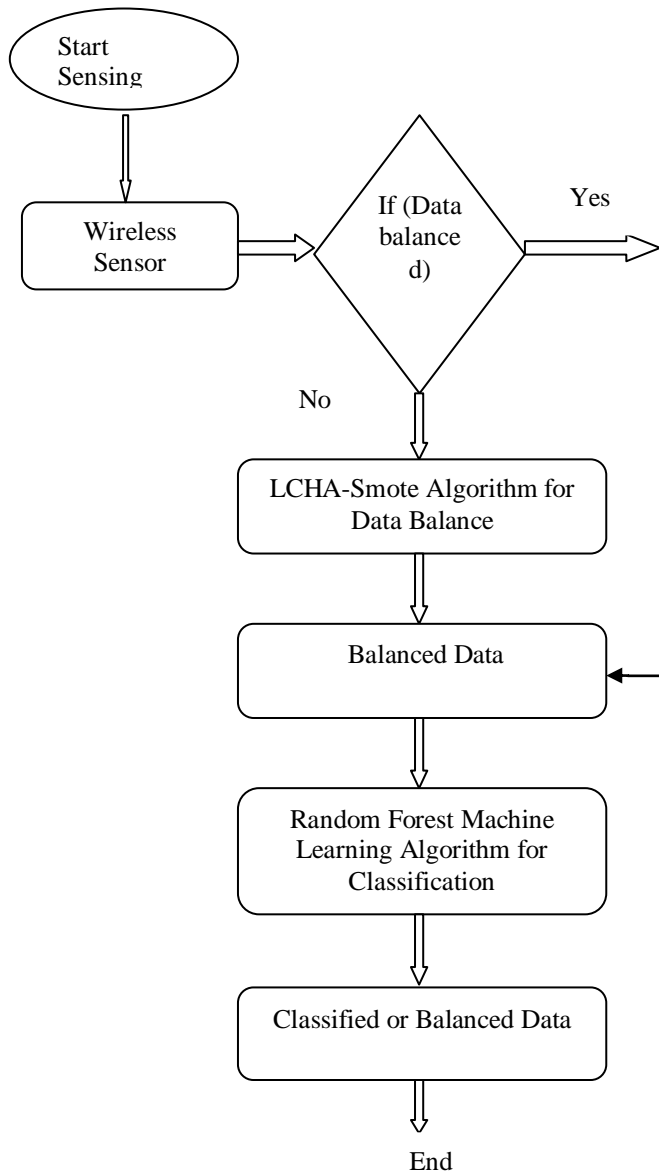


Fig. 2 Block Diagram of Propose system.

## V.CONCLUSION

In this work we present LCHA-SMOTE, a modified technique for learning from different types of imbalanced datasets. LCHA is a variant of the SMOTE algorithm, for improving the performances of model for the minority class. We will use the NSL-KDD for Experiments and improving the problem on imbalanced data and achieve a accuracy in results.

## REFERENCES

[1]. Remah Alshinina and Khaled Elleithy "A Highly Accurate Machine Learning Approach for Developing Wireless Sensor Network Middleware". WTS 2018 1570433702 IEEE.

[2]. Mohammad Abu Alsheikh, Shaowei Lin" Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications" 1553-877X (c) 2013 IEEE.

[3]. Shengguo Hu and Yanfeng Liang "MSMOTE: Improving Classification Performance when Training Data is imbalanced" 978-0-7695-3881-5/09 \$26.00 © 2009 IEEE DOI 10.1109/WCSE.2009.137

[4]. L.Dhanabal1, Dr. S.P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification.Algorithms".DOI10.17148/IJARCCCE. 2015.4696

[5]. K. Bispo, N. Rosa, and P. Cunha, "SITRUS: Semantic Infrastructure for Wireless Sensor Networks," Sensors, vol. 15, no. 11, p. 27436, 2015

[6]. R. Alshinina and K. Elleithy, "Performance and Challenges of Service-Oriented Architecture for Wireless Sensor Networks," Sensors, vol. 17,no. 3, p. 536, 2017

[7]. J. Al-Jaroodi and A. Al-Dhaheri, "Security issues of service-oriented middleware," International Journal of Computer Science and Network Security, vol. 11, no. 1, pp. PP.153-160, 2011.

[8]. X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security:a survey," IEEE Communications Surveys & Tutorials, vol. 11, no. 2,02 June 2009 2009.

[9]. L. M. Ibrahim, D. T. Basheer, and M. S. Mahmud, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," Journal ofEngineering Science and Technology, vol. 8, no. 1, pp. 107-119, 2013.