

---

## **Self-Organizing Map with Real-time Updating for Big Data Analysis that Uses Bit Value Addition of the RGB Values of the Overlapped Data Points**

K.K.L.B. Adikaram<sup>a</sup> and P.A. Jayantha<sup>b</sup>

<sup>a</sup>*Computer Unit, Faculty of Agriculture, University of Ruhuna, Sri Lanka*

<sup>b</sup>*Department of Mathematics, Faculty of Science, University of Ruhuna, Sri Lanka*

Corresponding author: lasantha@agricc.ruh.ac.lk

Usually, standard Self-Organizing Maps demand the user to define the number of expected clusters. Most importantly, when there is an update of the data, the data set has to be analyzed using a pre-decided algorithm. Thus, it is required to have a high processing capacity to produce real-time analysis of big data. This paper presents a Self-Organizing Maps with Real-time Updating (SOMRU) which eliminates the above-mentioned drawbacks. The proposed SOMRU uses a bitmap as the plotting area. A suitable marker with pre-determined filled colour is used in SOMRU to represent data points graphically in a manner that they can overlap when adding adjacent points. Data will be added sequentially. If there is an overlap, the existing colour value of each pixel of such overlapped area will be updated using binary addition with the colour value of corresponding pixel of the newly added data point. This process updates only the pixels in the overlapped area with very small processing power and creates regions with different colours. The colour value of a certain point is proportional to the number of overlaps (data density) in any selected point of the plotting area. Because the plotting area is bitmap, the final output is a bitmap which can be considered as a matrix containing colour values that reflects the data density. The proposed method was applied on a dataset of over 35000 data points and the results showed that the method is capable of creating regions separated by automatically generated colour lines which can be considered as isoclines. Thus, the proposed SOMRU can be considered as an efficient algorithm, which is similar to the ones used in the field of artificial neural networks to produce trained data, with low computational cost to analyse big datasets while representing the different clusters visually by creating a matrix of pixels that could be converted to values that reflect the data density.

*Keywords: Big data, Cluster identification, Continuous learning, Kohenin's map, Self-organizing feature map*