

Classification of Congenital Hypothyroidism in Newborn Screening using Self-Organizing Maps

Iris Ivy M. Gauran

University of the Philippines Diliman

Maria Sofia Criselda A. Poblador

University of the Philippines Manila

The Newborn Screening Reference Center (NSRC) of the National Institutes of Health in University of the Philippines Manila collects measurements from five attributes to determine whether Congenital Hypothyroidism (CH) is present in a neonate. Detecting the CH cases is a major concern of medical practitioners because it provides richer information than the healthy ones. However, because of the rarity of this metabolic condition, existing classification algorithms oftentimes misclassify a newborn as “normal.” We investigate the efficiency of Self-Organizing Kohonen Maps (SOM)—a type of artificial neural network—as a tool for classification, particularly in detecting outliers.

A SOM with higher learning rate and larger training sample size yield low misclassification rate. A bootstrap estimate of the variability of the misclassification error rate is approximately 5%. The misclassification error rate is lower when the original validation sample is used, compared to the average misclassification error rate computed from the bootstrap validation samples. Particularly, for a learning rate of 0.8 and a ratio of 2:1 training to validation sample, a 2.04% misclassification against 7.93% misclassification with 4.86% standard deviation is observed.

Keywords: Self-Organizing Kohonen Maps (SOM), Classification Algorithm, Outlier Detection, Newborn Screening for Congenital Hypothyroidism

1. Introduction

In the advent of surging interest in machine learning techniques, issues pertaining to difficult “real-world” problems became more pronounced. Of these issues, data sets with unbalanced class distributions are quite common in many

domains. Fawcett and Provost (1996) dealt with fraudulent telephone calls which are outnumbered by legitimate usage of the company services. Dumais et al. (1998) investigated the use of Support Vector Machines (SVMs) for text categorization while Kubat et al. (1998) employed machine learning for the detection of oil spills in satellite radar images. Other applications also include fault diagnosis of helicopter gearboxes and seismic discrimination between earthquakes and nuclear explosions. In all of these applications, there is a disproportionate number of cases that belong to the different classes. It could be noted that the degree of imbalance varies from one application to another. For instance, a manufacturing plant operating under the six sigma principle may discover four defects in a million products shipped to their customers, while the amount of credit card fraud may be of the order of 1 in 100 (Krishna Veni and Sobha Rani, 2011).

In this paper, we aim to develop a methodology for coming up with a correct classification of the newborns with Congenital Hypothyroidism (CH). CH results from an abnormality in the thyroid gland development and is one of the most common preventable causes of mental retardation. The study utilizes a one-week newborn screening data collected by the Newborn Screening Reference Center (NSRC) of the National Institute of Health in University of the Philippines Manila. In one-week's time, a total of 5,844 newborns underwent a screening test for Congenital Hypothyroidism, which entailed measurements from five distinct attributes. A newborn is classified in any of the three mutually exclusive categories namely "Normal," "Retest" or "Abnormal." In both practice and literature, the amount of thyroid-stimulating hormone (TSH) found in the blood sample serves as the deciding factor for the classification of newborns. An elevated TSH value indicates an impaired functioning of the thyroid gland, thus a newborn with a high value of TSH is deemed "abnormal" and is highly likely to develop congenital hypothyroidism. Newborn screening of neonates for congenital hypothyroidism, followed by appropriate treatment among those affected by this condition, results in the disappearance of overt intellectual disability (Grosse and Van Vliet, 2011).

Worldwide, it is known that the incidence of Congenital Hypothyroidism is 1:4,000. Thus, in the given data set, it is expected that there are no more than ten cases of CH observed. This scenario is highly susceptible to misclassification errors since the proportion of CH cases is relatively low compared to that of healthy newborns. However, despite their infrequent occurrences, health professionals are more concerned with CH cases because it provides much more valuable information than the healthy cases.

The fact that most of the medical datasets contain rare cases introduces additional challenges to the classification algorithm. To uncover the subtle patterns that may be hidden in these medical datasets and to solve the classification problem, Artificial Neural Networks (ANNs) are utilized. Exploratory data mining using

ANNs offers an alternative dimension to data mining since they have a natural propensity to learn; they learn how to solve problems from data as opposed to solving problems based on explicit problem specification (Craven and Shavlik, 1997). Also, Artificial Neural Networks provide tremendous opportunities pertaining to data classification because the learning characteristics of ANNs enable them to deal efficiently with noisy data — partial, possibly incorrect and potentially conflicting data (Lu et al., 1995). The proposed methodology involves the use of Self-Organizing Kohonen Maps (SOM), a type of artificial neural network, as an unsupervised classification technique.

There are four existing problems which health professionals involved in newborn screening would be most interested in solving: (i) correctly classify a baby as “normal,” “retest” or “abnormal”; (ii) investigate the significance of utilizing age in the classification; (iii) assess the performance of the method by coming up with an indication of the variability of the error rate; and (iv) identify the possible cost of misclassifying a baby as normal or abnormal. Due to the lack of statistical researches in this particular field, health professionals usually rely on their expertise and knowledge in order to determine the answers to the abovementioned problems.

Hence, this paper aims to provide a novel technique in determining cases as either “normal,” “retest” or “abnormal.” In addition, it aims to provide a bootstrap estimate of the variability of the misclassification error rate. It also addresses the imbalance among the classes by oversampling the minority class and downsizing the majority class.

2. Self-Organizing Maps

The self-organizing map (SOM) is a type of artificial neural network which is a widely popular visualization and clustering tool. It projects input data on prototypes of a low-dimensional regular grid that can be effectively utilized to visualize and explore properties of the data. These prototypes are sometimes called nodes or neurons. A typical SOM network consists of two layers of nodes, an input and output layer. Each node in the input layer is fully connected to the nodes in the output layer. The number of input nodes corresponds to the number of input variables while the number of output nodes usually depends on the given problem and is specified by the user. The key concept in training SOM is the neighborhood around a best matching or winning node, which is the collection of all nodes with the same radial distance.

At the start of the training process, weights are initially assigned to each node in the output layer and a neighborhood size large enough to cover half of the neurons in the output layer is also chosen. When an input pattern, randomly

chosen from the training set is presented to the network, each node in the output layer calculates how similar the input is to its initial weights. The similarity is often measured by some distance between the input pattern and the weight of the neuron. The neuron with the minimum distance is the winning neuron and its weight, as well as the weights of its neighboring neurons are strengthened or updated to be closer to the value of the input pattern. The goal of SOM training is to represent all data points from a high-dimensional space into a two-dimensional feature map, without altering the distance and proximity of the points from each other. Therefore, the training of SOM is unsupervised and competitive with the winner-takes-all strategy.

The algorithm maps the input data to a two-dimensional feature map containing the output neurons with the following steps:

1. An input vector \underline{x} randomly chosen from the training set is compared with all the output nodes i with an initial weight vector w_i and the best matching unit (BMU) on the map is identified. The BMU is the node k whose weight vector w_k has the smallest Euclidean distance from the input vector. That is,

$$\|\underline{x} - w_k\| = \min \|\underline{x} - w_i\|$$

2. The weight vectors of the best matching unit k as well as those of the nodes in the neighborhood of k are updated so as to move towards the current input pattern \underline{x} using the Kohonen learning rule:

$$w_i^{new} = w_i^{old} + h_{ik}(\underline{x} - w_i^{new}), \text{ if } i \text{ is in the neighborhood of the BMU } k$$

$$w_i^{new} = w_i^{old}, \text{ if } i \text{ is not in the neighborhood of the BMU } k$$

where h_{ik} is the neighborhood kernel function, which is a decreasing function of the distance between the i^{th} and k^{th} weight vectors on the feature map. A widely used neighborhood function is based on the Gaussian function,

$$h_{ik} = \alpha \exp \left\{ -\frac{\|r_i - r_k\|^2}{2\sigma^2} \right\}$$

where $0 < \alpha < 1$ is the learning rate factor, r_i and r_k are the positions of neurons i and k on the SOM grid and σ is the neighborhood radius, which is also decreasing monotonically.

3. Repeat steps 1 and 2 until all input patterns from the training set are processed. To achieve a better convergence towards the desired mapping, it is usually required to repeat the previous loop until weights are stabilized, while decreasing the size of the neighborhood.

The overall loop is repeated several times until there are no significant changes

in the weight vectors of the output neurons. Once the SOM training algorithm has converged, the computed feature map displays important statistical characteristics of the input space. Goncalves et al. (2011) summarized these characteristics as follows:

- i. *Vector Quantization*: the basic objective of SOM is to store a large set of input vectors by finding a smaller set of prototypes that provides a good approximation of the input space.
- ii. *Topological Ordering*: the feature map computed by SOM is ordered topologically. Similar input vectors are mapped close to each other, while dissimilar ones are mapped far apart.
- iii. *Density matching*: the SOM reflects the probability distribution of data in the input space. Regions in the input space in which the input patterns are taken with a high probability of occurrence are mapped onto larger domains of the output space, and thus have better resolution than regions in the output space from which input patterns are taken with a low probability of occurrence.

3. Methodology

The classification of a newborn as possessing a potentially fatal metabolic disorder or not is formulated using Self-Organizing Maps (SOM). SOM is used to map the original patterns of the data to a reduced set of prototypes arranged in a two-dimensional rectangular grid, seeking to preserve the probability distribution and topology of the input space. The key point of the proposed methodology is to classify a newborn utilizing a set of SOM prototypes instead of working directly with the original attributes of the data.

3.1 Specifications for the training map

In the processing level, two important factors should be taken into account. First, it is necessary to determine the SOM training parameters. The following parameters and criteria to train SOM will be utilized by the proposed methodology:

1. *Number of output nodes*: Deboeck and Kohonen (1998) recommend using ten times the dimension of the input patterns as the number of neurons, and this will be adopted in the methodology. The two-dimensional feature map will be constructed with a map size of 15x15, with a total of 225 output neurons.
2. *Weights of the input nodes*: Input neurons will have randomly initialized weights. The weights are derived from the Uniform Distribution multiplied by the corresponding range of values of the variables.

3. *Number of iterations:* The accuracy of the map also depends on the number of iterations of the SOM algorithm. A rule of thumb states, for good statistical accuracy, the number of iterations should be at least 500 times the number of neurons.
4. *Neighborhood radius:* Neighborhood radius is initially set to 7.5. A larger neighborhood radius is usually used at the beginning of the training which gradually decreases to a suitable final radius. Past literature suggests that the initial radius size be set to half the size of the feature map.
5. *Learning rate:* There are no guidelines in suggesting good learning rates to any given learning problem. In standard SOM, too large and too small learning rates can lead to poor network performance. However, some researchers posit that the learning rate is associated with the number of iterations.

Secondly, the sampling process of the input data will be considered in order to address the problem posed by outliers. Mining rare cases is an area of challenge in pattern recognition and neural network research (Muñoz and Murazabal, 1998). Thus, there are several proposed methodologies regarding outlier detection in the literature. Among these, Ling (1998) proposed that the rare class must be oversampled while Kubat (1997) posited the downsizing of the majority class. Also, Chawla (2002) formulated the Synthetic Minority Oversampling Technique (SMOTE) which aims to incorporate both downsizing of the majority class and oversampling of the minority class.

Given the data set on 5,844 newborns, 10 are classified as abnormal, 36 cases must undergo a retest and 5,798 newborns were normal. The input data set contains all cases which are classified as retest or abnormal. In addition, 100 normal newborns are sampled to incorporate the idea of downsizing the majority case. These 146 cases will then be divided into the training set and the validation set.

To be able to compare the performance of the classification algorithm, the learning rates are varied. Learning rate could be set to 0.4 or 0.8. To maintain a balance between the learning rate and the number of iterations, if the chosen value of the learning rate is 0.4 then 1,500 iterations were considered. Meanwhile, if the learning rate is initially set to 0.8, then the number of iterations specified is 5,000. The drawback of having a large number of iterations is the amount of time it will take in processing the results. Even in the presence of modern computers, the running time of the algorithm is almost twice as long if there are 5,000 iterations compared to 1,500 iterations.

In addition, the number of cases assigned to the training set and the validation set are varied. First, the 146 cases will be equally divided into the training set and the validation set. The results of this 1:1 scenario will be compared to the results when the division considered has 2:1 allotment.

3.2 SOM as a tool for classification

To perform the classification of newborns using Self-Organizing Maps, we first get an undersample of the majority case (nearly 2%) and obtain a census of the rare cases. Divide the sample into training and validation data sets. After which, 200 replicates of the original training set and the original validation set are formed to constitute the samples for nonparametric bootstrap estimation.

The following steps are performed on the 200 pairs of bootstrap resamples:

1. Construct a self-organizing map using the parameters specified earlier and train the self-organizing map using the training set. The Self-Organizing Map was trained using complete information on the six attributes of the newborns. The original map contains 225 nodes wherein each node contains the weight vector \underline{w}_i . Every weight vector \underline{w}_i contains six values corresponding to the six attributes.
2. Construct a second map by removing the last element of the weight vector on each node of the original SOM. Note that this second map is not retrained. It will still consist of the values derived except for the last element of the original weight vector which corresponds to the variable containing the information on the classification of the newborns. This is done for purposes of validation.
3. Likewise, disregard the last variable pertaining to the classification of the newborn on each validation sample. This is done to determine if the training set can come up with the correct classification had we not known the value of the classification variable.
4. Use validation samples as input vector on the second map. As each observation in the validation set is inputted into the second map, the best matching unit is identified. This is the unit which gives the minimum Euclidean distance.
5. Perform classification based on the classification of the nodes of the original SOM. Since the second map does not have the classification variable, it can give us information about the best matching node only. This node will then be identified and matched in the original map to come with the value of the classification variable.

Compute for the misclassification error rate. This is done by comparing the classification value assigned by the map and the “real” classification of the newborns.

4. Results and Discussions

Given the sample of 146 observations with five distinct attributes from the newborn screening data for Congenital Hypothyroidism, classification is

performed following the methodology discussed earlier. The self-organizing map as a tool for classification is assessed on two different levels of learning rate and training sample size, and by calculating the misclassification rate on the original validation sample.

Table 1 provides a comparison of the percentage of misclassification on the original validation sample when learning rate and input data allocation are varied. It can be seen that a higher learning rate (0.8) yields lower misclassification rate regardless of the allocation of the input data. Comparing between equal and unequal allocation of input data, it can be noticed that a higher training sample size, that is, a ratio of 2:1 training sample to validation sample, results to lower percentage of misclassification irrespective of learning rate.

A closer look on the misclassified cases under 0.4 learning rate and equal input data allocation (1:1) is given on Table 2. Likewise, Table 3 gives the characteristics of the misclassified cases under 0.4 learning rate and unequal allocation of training and validation samples.

Under both situations, the most common misclassification occurs as assigning a newborn as “retest” when TSH level is actually “abnormal” and assigning a newborn as “normal” when TSH level requires a “retest.” The error of assignment of classes relegates the adversity of the impact of the true classification of the newborns.

Table 1. Misclassification Error Rate based on the learning rate and the allocation of the input data set

Learning Rate	Allocation of the Training and Validation data sets	
	1:1	2:1
0.4	10.96%	6.12%
0.8	5.48%	2.04%

Table 2. Misclassified Cases under learning rate of 0.4 and equal allocation of the input data set

OBS	Class	Assignment	Age	TSH1	TSH2	TSH3	TSH99
31	RETEST	NORMAL	1	11.18	4.49	5.22	4.86
37	RETEST	NORMAL	1	10.83	9.86	9.59	9.72
43	RETEST	NORMAL	2	10.03	9.19	9.01	9.1
48	RETEST	NORMAL	2	10.02	10.2	9.95	10.08
67	ABNORMAL	RETEST	1	12.46	11.13	13.02	15
70	ABNORMAL	RETEST	2	14.96	13.71	13.79	15
71	ABNORMAL	RETEST	2	12.74	14.08	12.8	15
72	ABNORMAL	RETEST	9	14.04	14.2	12.56	15

Table 3. Misclassified Cases under learning rate of 0.4 and unequal allocation of the input data set

OBS	Class	Assignment	Age	TSH1	TSH2	TSH3	TSH99
31	RETEST	NORMAL	1	11.18	4.49	5.22	4.86
45	RETEST	ABNORMAL	2	13.37	12.41	11.3	11.86
63	RETEST	NORMAL	25	10.08	9.53	9.88	9.7
72	ABNORMAL	RETEST	9	14.04	14.2	12.56	15

For instance, the cost of screening for CH is much lower than the cost of diagnosing CH at an older age. This estimate does not include the loss of tax income resulting from impaired intellectual capacity in the untreated but non-institutionalized person (Rose et al., 2006). As a consequence, misclassifying “retest” as “normal” is always detrimental for patients. Some infants with normal scan findings at birth may have a transient form of hypothyroidism. It is critical that rescreening test be obtained in a timely manner and that there be no loss of follow-up. Transient CH is not usually evident at birth and, therefore, may not be detected if newborn screening is performed in the first few days postnatally. In some cases, hypothyroidism can manifest after the newborn screening and if not detected, may result to failure of normal development in infants.

In general, when an abnormal screening result is found, the responsible physician is notified immediately so that he or she can arrange for subsequent testing. Such infants should be examined immediately and have confirmatory serum testing performed to verify the diagnosis. Treatment with replacement levothyroxine ($L-T_4$) should be initiated as soon as confirmatory tests have been drawn and before the results of the confirmatory tests are available. Delayed treatment may result in decreased intellectual capacity. This implies that misclassifying “abnormal” as “retest” may put the life of a newborn in huge amount of peril owing to the delay of medical mitigation. On the contrary, misclassifying “retest” as “abnormal” may entail higher medication cost if unfortunately the slight elevation in the TSH values is in fact a false positive. However, ten percent of infants with confirmed CH have slightly elevated TSH values (Rose et al., 2006).

Another closer examination on the misclassification pattern observed under a learning rate of 0.8 with equal allocation of training and validation samples is given in Table 4. Similarly, Table 5 shows the characteristic of the misclassified case under the same learning rate of 0.8 but with unequal allocation of input data.

Regardless of the allocation of training and validation samples, a higher learning rate seem to more commonly misclassify “normal” as “retest.” A second screening poses the problem of increased cost as well as alarm on the infant’s parents.

Table 4. Misclassified Cases under learning rate of 0.8 and equal allocation of the input data set

OBS	Class	Assignment	Age	TSH1	TSH2	TSH3	TSH99
11	NORMAL	RETEST	2	3.34	3.34	3.34	3.34
15	NORMAL	RETEST	3	9.98	9.98	9.98	9.98
46	ABNORMAL	RETEST	1	12.46	11.13	13.02	15

Table 5. Misclassified Cases under learning rate of 0.8 and unequal allocation of the input data set

OBS	Class	Assignment	Age	TSH1	TSH2	TSH3	TSH99
15	NORMAL	RETEST	3	9.98	9.98	9.98	9.98

Although much has been learned, some questions remain. These issues include the optimal screening approach and the follow-up of infants with normal TSH concentrations. Delayed TSH elevation is particularly common in infants with low birth weight (<2500g) and very low birth weight (<1500g). Moreover, age seem to have a significant effect on the result of the newborn screening. Specimens collected in the first 24 to 48 hours of life may lead to false-positive TSH elevations. Recent data using a sensitive and specific immunofluorometric assay indicate that normal TSH values before 24 hours of age are not as high as those using previous assays and usually less than the cutoff value of 20 to 25 mU/L (Heyerdahl and Oerbeck, 2003). Allen et al. (1990) mentioned that a 50% reduction in abnormal values occurred when age-adjusted TSH cutoffs were used. It is important that age-appropriate normative cut-off values be used.

To obtain precision estimate of the misclassification rate of the self-organizing map as a tool for classification, a bootstrap estimate of the standard deviation is computed. Table 6 provides the average misclassification rate of the 200 self-organizing maps constructed using the bootstrap training samples and evaluated by means of the bootstrap validation samples. It can be perceived that a higher learning rate causes lower misclassification rate. Likewise, a higher allocation of training samples leads to lower percentage of incorrect classification.

Table 6. Average Misclassification Error Rate based on the learning rate and the allocation of the input data set

Learning Rate	Allocation of the Training and Validation data sets	
	1:1	2:1
0.4	11.25%	8.87%
0.8	9.59%	7.93%

Moreover, it can be seen from Table 7 that the misclassification rate of SOM under different specifications of learning rate and input data allocation has an estimated variability close to 5%.

Table 7. Nonparametric Bootstrap Estimate of the variability of the Misclassification Error Rate based on the learning rate and the allocation of the input data set

Learning Rate	Allocation of the Training and Validation data sets	
	1:1	2:1
0.4	5.48	5.29
0.8	5.46	4.86

A higher precision of classification can be observed for SOM with higher learning rate (0.8) and higher allocation of training samples (2:1). This means that the self-organizing map can be an effective tool for classification of data with rare cases.

5. Summary and Conclusions

In this paper, the Kohonen’s self-organizing map (SOM) is utilized in classifying newborns with congenital hypothyroidism (CH).

5.1 Summary

The data obtained from the National Screening Reference Center of the National Institute of Health consists of a one-week screening test results on 5,844 newborns. The screening approach includes obtaining a blood sample from infants to determine their thyroid-stimulating hormone (TSH) level. Seriously elevated TSH levels (>13) are considered “abnormal,” requiring prompt medical attention. Moderately elevated TSH levels (between 10 and 13) indicate the need for “retest,” while low TSH levels (<10) suggests the newborns to be “normal.”

The main problem of this research is performing classification under the presence of outliers. This problem is mitigated by oversampling from the rare cases and under-sampling from the majority cases. Consequently, nearly 2% of the majority case and a census of the rare case are taken as a sample.

The construction of a self-organizing map is done under different levels of learning rate and training sample size. In particular, 0.4 and 0.8 are considered as learning rates. While SOM training is accomplished by using two different training sample sizes, that is, $1/2$ and $2/3$ of the original sample, which leaves the remaining $1/2$ and $1/3$ as validation samples, respectively.

Evaluation of the performance of SOM as a classification tool is then made by calculating misclassification error rate based on the validation samples. Moreover, bootstrap is implemented to obtain an estimate of the variability of the misclassification error rate.

5.2 Conclusions

Results show that a higher learning rate and training sample allocation yield lower percentage of misclassification. In particular, the self-organizing map trained using the original sample under a learning rate of 0.4 tend to more commonly misclassify “retest” as “normal” and “abnormal” as “retest.” This finding reveals the problem of downsizing the adverse impact of the true classification of the newborns. Costs entailing such misclassification include failure to normal development of infants, even death if proper clinical attention failed to be administered without delay. On the other hand, SOM trained using the original sample under a learning rate of 0.8 is more inclined to misclassify “normal” as “retest.” This would cause distraught on the parents as well as increased medication cost.

A bootstrap estimate of the standard errors of roughly around 5% is also obtained. The misclassification error rate is lower when the original validation sample is used, compared to the average misclassification error rate computed from the bootstrap validation samples.

References

- ALLEN, D., J.E. SIEGER, T. LITSHEIM, and S.C. DUCK, 1990, Age-adjusted thyrotropin criteria for neonatal screening for hypothyroidism, *Journal of Pediatrics*, 117:309–312.
- CHAWLA, N., K. BOWYER, L. HALL, and W.P. KEGELMEYER, 2002, SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Research*, 16: 341-378.
- CRAVEN, M. and J. SHAVLIK, 1997, Using Neural Networks for Data Mining, *Future Generation Computer Systems*, 13: 211-229.
- DEBOECK, G. and T. KOHONEN, 1998, *Visual Explorations in Finance with Self-organizing Maps*, Springer-Verlag London.
- DUMAIS, S., J. PLATT, D. HECKERMAN and M. SAHAMI, 1998, Inductive Learning Algorithms and Representations for Text Categorization, In Proceedings of the 7th International Conference on Information and Knowledge Management., pp. 148-155.
- FAWCETT, T., and F. PROVOST, 1996, Combining Data Mining and Machine Learning for Effective User Profile, In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp. 8-13.
- GONCALVES, M.L., J.A.F. COSTA, and M.L.A. NETTO, 2011, Land-Cover Classification Using Self-Organizing Maps Clustered with Spectral and Spatial Information, in Mwasiagi, J.I. (ed.), *Self-Organizing Maps-Applications and Novel Algorithm Design InTech*, pp 299-322.
- GROSSE, S.D. and G. VAN VLIET, 2011, Prevention of intellectual disability through screening for congenital hypothyroidism: how much and at what level?, *Arch Dis Child*, 96:374–379.
- HEYERDAHL, S. and B. OERBECK, 2003, Congenital hypothyroidism: developmental outcome in relation to levothyroxine treatment variables, *Thyroid*, 13:1029-1038.
- KEMPER, A.R., L. OUYANG, and S.D. GROSSE, 2010, Discontinuation of thyroid hormone treatment among children in the United States with congenital hypothyroidism: findings from health insurance claims data, *BMC Pediatr*, 10:9.
- KRISHNA VENI, C.V. and T. SOBHA RANI, 2011, On the classification of imbalanced datasets, *International Journal of Computer Science and Technology*, 2: 145-148.
- KUBAT, M. and S. MATWIN, 1997, Addressing the curse of imbalanced data sets: One-sided Sampling, In Proceedings of the 14th International Conference on Machine Learning, pp 179-186.
- KUBAT, M., R. HOLTE, and S. MATWIN, 1998, Machine learning for the detection of oil spills in satellite radar images, *Machine Learning*, 30(2-3):195-215.
- LING, C. and C. LI, 1998, Data mining for direct marketing: Problems and solutions, In Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, pp 73-79.
- LU, H., R. SETIONO, and H. LIU, 1996, Effective data mining using neural networks, *IEEE Transactions on Knowledge Discovery and Data Engineering*, 8(6):957-961.
- MUNOZ, A., and J. MURUZABAL, 1998, Self-organizing maps for outlier detection, *Neuroscience and Computing*, 18(1-3):33-60.
- ROSE, S., R. BROWN, T. FOLEY, P. KAPLOWITZ, C. KAYE, S. SUNDARARAJAN and S. VARMA, 2006, Update of Newborn Screening and Therapy for Congenital Hypothyroidism, *Pediatrics*, 117(6): 2290 -2303.

Acknowledgment

The authors are thankful to Mark Lester Ghany for generously sharing his time and expertise in programming. Also, they are grateful for the encouragement, motivation, and helpful insights from Dr. Erniel Barrios and Dr. Joselito Magadia, which greatly contributed to the improvement of this paper.