This full-text paper was peer-reviewed and accepted to be presented at the IEEE ICCSP 2015 conference

Spectral Analysis of Photonic crystal based Bio-Sensor using AdaBoost Algorithm

Sriram G, Vivek M, Sandeep Kumar Roy, Preeta Sharan

Abstract: **Photonic is the study that includes the use of light, where the fundamental element is photon. The optics is the branch of the physics which deals with the properties of the light. In this paper the sensor is designed to detect the glucose content of human body and the spectrum from the output of the sensor is analyzed by using AdaBoost algorithm. The Adaptive Boosting in shortconsideredto be AdaBoost algorithm. The classification of the multiple discrete weak classifiers allowing the formation of strong classifier by integrating, provides betterimprovedperformance using the concept of AdaBoost. The AdaBoost is used widely in almost all fields like physical sensing, chemical sensing, bio sensing etc. This paper primarily based on analyzing the spectrum from the bio sensors. The major role of the machine language is that it gives the accurate outputs which is easily understood by all the humans without having the knowledge on that particular field. Using the AdaBoost along with the photonic crystals based designs, the sensing technology can be improved in accuracy, sensitivity and specificity. Thus this can be successfully implemented in detection of Diabetic patient. The simulation tool that is used in the paper is MEEP, Mat Lab.**

Keywords: *Optical sensor, Spectrum analysis, Saliva, Machine language, photonic crystal, AdaBoost.*

I. INTRODUCTION

The earlier techniques used to detect the glucose in fluids like blood and urine has challenged the clinical chemist. The chemistry test to detect glucose content in human body is the common test because of high incidence of diabetes mellitus. Currently glucose measurement is done by pricking human body which is harmful can be overcome by taking saliva as analyte and the AdaBoost algorithm used to detect whether patient is diabetic by comparing the dataset with database.

Sriram G is with the Department of electronics and communication, The Oxford College of Engineering, Bangalore-, India. (Phone: +918105656242; e-mail: sriramnus13@gmail.com).

Dr. Vivek M is with the Electronics and communication Department, The Oxford College of Engineering, Bangalore-, India.

Sandeep Kumar Roy is the research scholar from AMET University, (e-mail: sandipr@hotmail.com)

Dr. Preeta sharan is with the Electronics and communication Department, The Oxford College of Engineering, Bangalore-, India.

The Photonic crystal are periodic systems that possess high and low dielectric region which are symmetrically placed.Photonic crystal is in three forms 1-dimensional, 2-dimensional, 3-D.Here the 2-dimensional photonic crystal is used and it is the periodic array of parallel holes where no direction show total invariance under translation. Photonic crystal band gap structure can be utilized for glucose detection. The concept behind the photonic sensing technology is that each material has distinct permittivity (ε) that is greater than air as result the propagation of electromagnetic waves that pass through them is altered in response to change in refractive index. The analyte used here is the saliva thus making the process a non-invasive method. This sensor can be developed as lab-on-a-chip sensor for detection of glucose concentration in saliva. The light is passed through the one end and the transmission spectrum is observed at the other end. The transmission spectrum is observed is unique for the specific value assigned. The photonic crystal (PC) consist of light source, sensors and the detectors which are integrated on one single chip. Since the topology and the formulations for three dimensional photonic crystal is complex so we have considered the two dimensional photonic crystal in this paper for the input part.

Machine learning is the study of computer algorithm for learning to do the stuff. Based on the observation of the data, direct experience or instruction the learning has to be done. The core subarea of artificial intelligence is the machine learning. The advantage of this machine learning is that we can build any kind of the system which is capable for our convenient algorithm that we design. It is an automatic methods which can be read as “programming by example”. The primary goal of machine learning research is to develop general purpose algorithms of practical value. These kind of the algorithm should be efficient, as the time and the space matters a lot for the users.

AdaBoost, stands for “Adaptive Boosting” which is a machine learning meta-algorithm. AdaBoost is sensitive to noisy data. AdaBoost is referred to as the best out-of-the-box classifier. The problem of the

978-1-4799-8081-9/15/$31.00 © 2015 IEEE

machine learning is mainly based on the dimensions and evaluating features also reduces the speed of the classifier training and execution and also reduces the predictive power. The advantage of machine learning over the direct programming is that the results obtained by the machine learning is more accurate compared to the other one. This is because the machine learning are data driven are we can detect large amount of data.

AdaBoost is an algorithm for constructing a "strong" classifier as linear combination of "simple" "weak" classifiers $h_t(x)$.

$$H(x) = \sum_{t=1}^{T} \propto_t h_t(x) \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where,

- $h_t(x)$……… "Weak" classifier or basis classifier.
- $H(x)$ ……. "Strong" or final classifier.
- T ……….. Number of Weak classifier.
- α_t ……….. Weight of each weak classifier.

Where each f_t is a weak learner that takes an object x as input and returns a real valued result indicating the class of the object.

II. STRUCTURE DESIGN

In this paper we have designed a photonic crystal sensor for detecting the presence of sugar in human body. Silicon rods in air configuration are used. The rods are contacted by saliva sample when the sensor device is dipped in saliva sample. The light passed through the photonic crystal from one end will interact with the saliva components and get detected from another end. Depending on the refractive indices of the saliva analytes the propagation of the light will vary in the photonic crystal. Design of the sensor device is shown in Figure 1. Designing and simulation is done with the help of MEEP simulation tool. Design1 Specifications are given below:

- Rods in air configuration
- Hexagonal lattice structure
- Lattice constant 'a'=1µm
- Radius of rods 'r'=0.19µm
- Dielectric constant of rods 11.74
- Background dielectric constant is changed with respect to sample taken
- Height of rods infinity
- Light source: Gaussian Pulse with center frequency at 0.295 and width of the pulse is 0.1
- Lattice size 26 * 32
- Wavelength of light 1570nm.

The structure is designed by using hexagonal lattice where this lattice provide the confined passing of light compared to other lattice and the holes in air configuration can be easily done. Here the width of the line defect is chosen by w=$\sqrt{3}a$ so that there won't be any losses in light, passing between in and out of the structure.

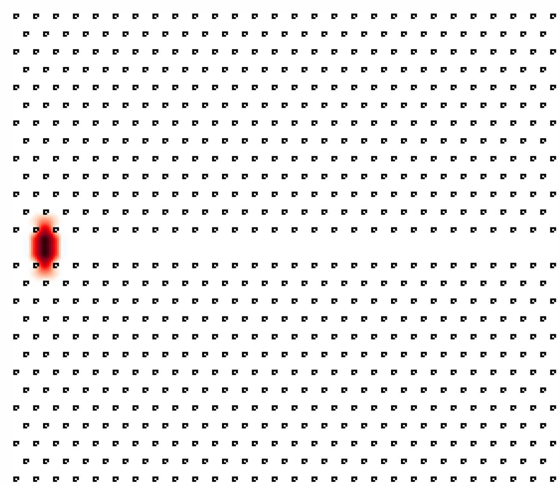


Fig. 1. Line defect for rods in air in Hexagonal Structure

III. SENSOR OUTPUT

The transmission spectrum for the normal and diabetic patient is obtained. The Normal patient spectrum shows that the peak amplitude of 17Hz will be at 0.295 center frequency. The Diabetic patient spectrum shows that the peak amplitude of 16Hz will be at 0.31 center frequency and here the wavelength shift will be around 0.015. This shows that the change in minute level can also be detected.

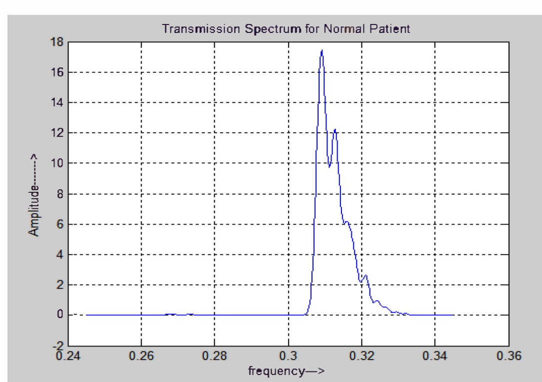


Fig. 2. Transmission Spectrum for Normal Patient.

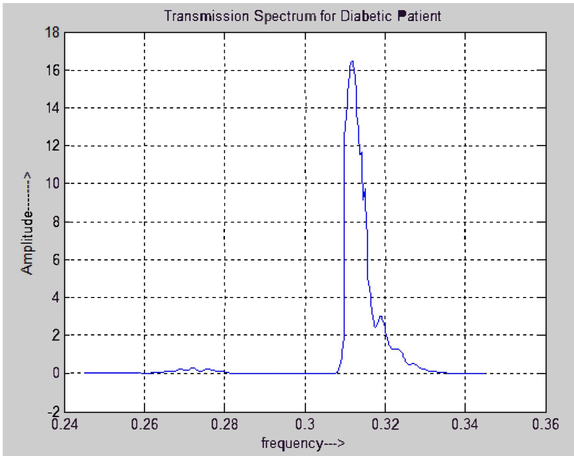


Fig. 3. Transmission Spectrum for Diabetic Patient.

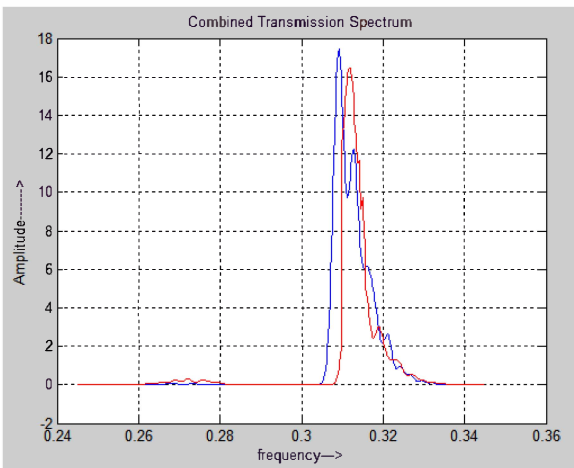


Fig. 4. Combined Transmission Spectrum for Normal and Diabetic Patient

IV. ADAPTIVE BOOSTING (ADABOOST)

The adaptive boosting can be shorted as AdaBoost, where boosting means combination. This implies that the AdaBoost contains many classifiers in it. The weak learner used here is the decision tree which plots the output of sensors. The classifier is done by using the MATLAB AdaBoost toolbox.

The input data is plotted based on the two dimensional das {-1, +1} here the algorithm will classify this two labels.

The calculation of error is given below by the following equation which gives the probability of output that are not mismatched with the weak learners

$$err_t = \sum_{t=1}^{m} D_t(i)[\, y_i \neq h_t(i)] \ldots\ldots.. (2)$$

Where,

err_t ……… Error function
D_t (i) …….. Initial probability
y_i ………… output

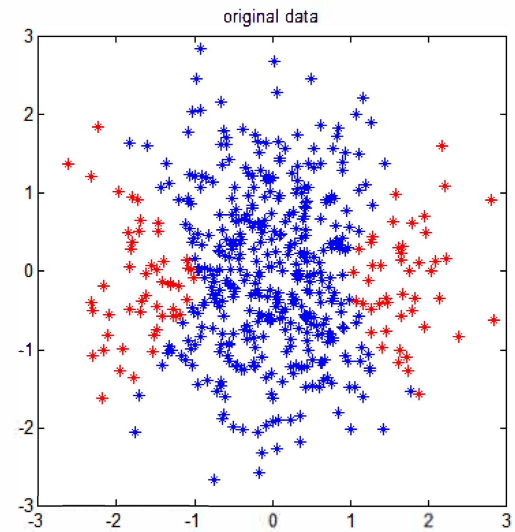


Fig. 5. The plotted sensor output for the input of algorithm.

The weight of each classifier is calculated by using the error function as given above and the logarithmic of the above equation will give weight of the each weak classifiers given below

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-err_t}{err_t}\right) \ldots\ldots.. (3)$$

Where,

α_t … … … Weight of weakclassifier

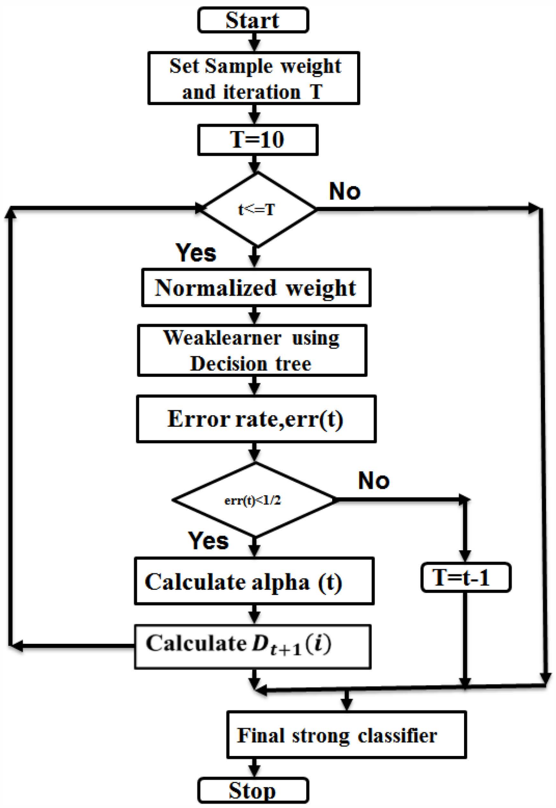
The probability distribution is updated by the following equation which gives the wrong and correctly matched probability. By using this equation the probability is updated after each weak learner.

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t} \ldots\ldots.. (4)$$

The final strong classifier H(x) is obtained by combining all weak classifier and their weights as given below

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t\, h_t(x)) \ldots\ldots.. (5)$$

The flow chart of the AdaBoost algorithm is given below by



V. CONFUSION MATRIX

The result is expressed in terms of two-dimensional confusion matrix. This matrix is also called as error matrix which tells the matched and mismatched label percentage. The confusion matrix is given below as

- The True Positive (TP) is the number of matched prediction in the class of normal patient.
- The True Negative (TN) is the number of matched prediction in the class of diabetic patient.

		PREDICTED CLASS	
		Normal	Diabetic
OUTCOME	Normal	TP	FN
	Diabetic	FP	TN

Table 1: The Confusion Matrix

- The False Positive (FP) is the number of mismatched prediction in the class of Normal patient.
- The False Negative (FN) is the number of mismatched prediction in the class of Diabetic patient.

The result of confusion matrix can be verified in many ways by calculating

- Accuracy
- Sensitivity
- Specificity
- Precision
- Recall
- Fscore

The Accuracy is calculated by the following formula which gives the percentage of the normal and diabetic patient among the given number of patient.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \text{........ (6)}$$

The Sensitivity is the probability of matched set among the normal patient and it is given below as

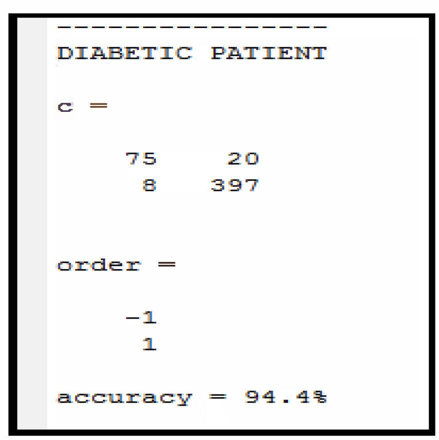
$$Sensitivity = \frac{TP}{TP+FP} \text{........ (7)}$$

The Specificity is inverse of the above, it gives the probability of matched set among the diabetic patient, and it is given below as

$$Specificity = \frac{TN}{TN+FN} \text{........ (8)}$$

VI. RESULTS

The confusion matrix which we got as the output of the algorithm is given below by the following image which is obtained from the Mat lab output and the order of the confusion matrix is also given below in fig 5,



The accuracy percentage is about to 95%, sensitivity and specificity as 78% and 98%

```
>> X=ans.sensitivity

X =

    0.7895
    0.9802

>> Y=ans.specificity

Y =

    0.9802
    0.7895

>> Z=ans.precision

Z =

    0.9036
    0.9520
```

Fig 5: Output of AdaBoost algorithm.

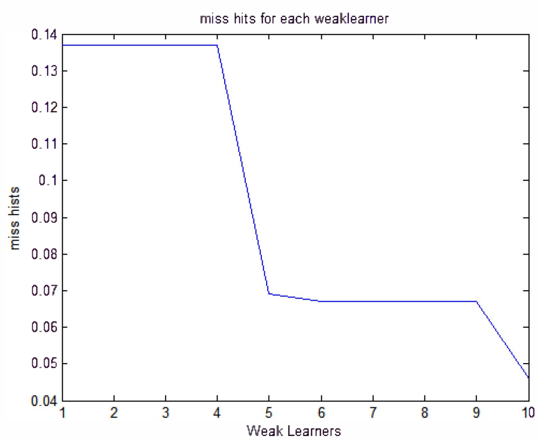


Fig 6: Miss hits Percentage for each weak learners

VII. CONCLUSION AND FUTURE WORK

We have designed a two-dimensional photonic crystal based optic sensor which is used to detect the diabetic patient and the spectrum of the sensor output is given to the AdaBoost algorithm. The data from the sensor is matched with the input data which clearly notify whether a person is a diabetic or not and this process is done automatically without any human assistance. The accuracy, sensitivity and specificity of algorithm is above 90 percentage which helps to improve the quality of data set. In future the algorithm can be modified to improve the data quality and by integrating the algorithm with other weak learners the sensitivity, accuracy and specificity of models can be improved. The future work also includes the implementation of GUI and in turn making the entire work automated.

REFERENCES

[1] E. L. Allwein, R. E. Schapire, and Y. Singer. Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of MachineLearning Research, 1:113–141, 2000.

[2] Englander HR, Jeffay AI, Fuller JB, Chauncey HH (1963) glucose concentrations in saliva plasma and parotid saliva of individuals with and without diabetes mellitus. J Dent Res 42: 1246.

[3] Schapire, Robert; Singer, Yoram (1999). “Improved Boosting Algorithms Using Confidence-rated Predictions”. CiteSeerX: 10.1.1.33.4002.

[4] Yamaguchi M, Mitsumori M, Kano Y (1998). A new approach to noninvasive measurement of saliva glucose using saliva, 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 20: 4.

[5] MATLAB AdaBoost toolbox. Includes Real AdaBoost, Gentle AdaBoost and Modest AdaBoost implementations.

[6] Reim M, Lax F, Lichte H, and Turss R: Steady state levels of glucose in the different layers of the cornea, aqueous humor, saliva and tears in vivo. Ophthalmologica 154:39, 1967.

[7] S.Hinterstoisser, V. Lepetit, S. Ilic, P. Fua and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects," IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[8] Zheng, Z. & Webb, G. I. (1998). Multiple boosting: A combination of boosting and bagging. In Proceedings of the 1998 International Conference on Parallel and Distributed Processing Techniques and Applications (pp. 1133–1140). CSREA Press.

[9] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," IEEE Conference on Computer Vision and Pattern Recognition, pp. 9-12, 2007.

[10] P. Viola, M. J. Jones y D. Snow, «Detecting pedestrians using patterns of motion and appearance,» Proceedings of the 9th International Conference on Computer Vision, vol. 1, pp. 734-741, 2003.

[11] X. Wang, T. Han y S. Yan, «An HOG-LBP human detector with partial occlusion handling,» IEEE 12th International Conference on Computer Vision, pp. 32-39, 2009.

[12] J. Whitehill y C. W. Omlin, «Haar features for FACS AU recognition, » Proc. IEEE International Conference ON Automatic Face and Gesture Recognition (AFGR06), pp. 217-222, 2006.

[13] S. M. Hanif, L. Prevost, and P. A. Negri. A cascade detector For text detection in natural scene images. ICPR, pages 1–4, 2008.

[14] R. Rifkin and A. Klautau (2004) In defense of one-vs-all classification, Journal of Machine Learning Research, 5: 101–141.