Design and Evaluation of Logistic Regression Model for Pattern Recognition Systems

Pranav Rao[†] and Manikandan J[‡]

[†]Department of Telecommunication Engineering, [‡]Department of ECE and Crucible of Research and Innovation (CORI),
PES University, 100-Feet Ring Road, BSK Stage III, Bangalore 560085, Karnataka, INDIA.
{email : pranavrao1234@gmail.com, manikandanj@pes.edu}

***Abstract*— In this paper, an attempt is made to design pattern recognition systems using logistic regression model and few mapping functions are proposed for the same. The performance of proposed logistic regression model and mapping functions are assessed by evaluating the model using design of digital circuits, standard datasets from UMASS database and datasets pertaining to wireless sensor network applications. It is observed that for a majority of cases, the recognition accuracy is enhanced on using proposed mapping functions for both binary and multi class pattern classification problems.**

Keywords- Machine Learning; Logistic Regression; Mapping Functions; Pattern Recognition.

I. INTRODUCTION

Machine learning is employed for various applications ranging from healthcare, biomedical, image and signal processing to data mining based pattern recognition systems. Logistic regression is one among the several pattern recognition algorithms reported in literature and the same has been employed for various applications such as predicting fault-prone code across software projects in [1], predicting fault proneness in software models using three different logistic regression models in [2], 3D scan point classification in [3], hyperspectral image classification in [4], Actigraphy-based scratch detection in [5], speech recognition in [6], cerebral infraction disease detection in [7], identification of human disease genes in [8], EEG classification in [9], diagnostic estimation of obstructive sleep apnea in [10], determining children's working memory ability in [11], evaluation of credit risk for data mining applications in [12] and many more.

In this paper, an attempt is made to propose and evaluate various mapping functions for pattern recognition systems using logistic regression model. The performance of logistic regression model and proposed mapping functions is assessed using digital circuit design and various datasets from University of Massachusetts (UMASS), Labelled Wireless Sensor Network Data Repository (WSNDR) and University of California, Irvine (UCI) database.

The organization of the paper is as follows: Mathematical overview of logistic regression and the proposed mapping functions for input features are given in Section II. Design of digital circuits using logistic regression is discussed in Section III. Application of proposed logistic regression model to pattern recognition systems and their performance evaluation is reported in Section IV. Performance evaluation of proposed model for wireless sensor network applications is reported in Section V, followed by conclusion and references.

II. LOGISTIC REGRESSION

Logistic regression is a probabilistic linear classifier, parameterized by a weight matrix **w** and a bias *b*. It enables the system to estimate categorical results with the help of a group of independent variables. The classifier equation for logistic regression model is given as

$$y = \mathrm{sgn}(\mathbf{w}^T\mathbf{x} + b) \quad (1)$$

where y ϵ {-1,1} denoting the output class recognized for the input **x** fed to the system. The classifier equation in (1) is rewritten using augmented weight matrix $\boldsymbol{\theta}$ as

$$y = \mathrm{sgn}(\boldsymbol{\theta}^T\mathbf{x}) \quad (2)$$

The augmented weight matrix $\boldsymbol{\theta}$ is obtained during the training phase of logistic regression model as explained below. Let $\{x_j, y_j\}$ for j = 1,2,…*m* denote the training data set, where y_j is the target output for training data x_j. The weight matrix is first initialized to 1, i.e., $\boldsymbol{\theta} = \mathbf{1}$. The equation for weight updates is given as

$$\boldsymbol{\theta}_j(\mathbf{n}) = \boldsymbol{\theta}_j(\mathbf{n-1}) + \alpha \cdot \upsilon_j \quad (3)$$

where α is the learning rate of the model and υ_j is given as

$$\upsilon_j = \sum_{i=1}^{m} \left(y^{(i)} - h_\theta\left(x^{(i)}\right)\right) x_j^{(i)}) \quad (4)$$

where *m* denotes the number of samples available in training dataset, j ϵ {1,2,…,m} and $h_\theta(x)$ is the logistic function given as

$$h_\theta(x) = \frac{1}{(1+e^{-\theta^T x})} \quad (5)$$

The average cost function J(θ) for logistic regression model is given as

$$J(\theta) = (1/m) \sum_{i=1}^{m} \left(\mathrm{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) \quad (6)$$

where $\mathrm{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right)$ is given as

$$Cost(h_\theta(x), y) = \begin{cases} -\,log\left(h_\theta(x)\right) & if\ y = 1 \\ -log\ (1 - h_\theta(x)) & if\ y = 0 \end{cases} \quad (7)$$

978-1-5090-3646-2/16/$31.00 ©2016 IEEE

In order to obtain minimum average cost for the logistic regression model designed, the Gradient Descent method is employed to obtain the iterative expression for $\mathrm{J}(\theta)$ as

$$\mathrm{J}(\theta)=\mathrm{J}(\theta)+\left(-\frac{1}{\mathrm{m}}\right)\left(\sum_{i=1}^{m} y^{(i)} \log\left(h_{\theta}\left(x^{(i)}\right)\right)\right. \\ \left.+\left(1-y^{(i)}\right)\log\left(1-h_{\theta}\left(x^{(i)}\right)\right)\right) \quad (8)$$

Let ε denote the acceptable threshold for cost function such that the iterative equation in (8) is terminated when $|\mathrm{J}(\theta)| \le \varepsilon$. In other words, the number of epochs to obtain the optimum weights, $\boldsymbol{\theta}$ is decided by the condition $|\mathrm{J}(\theta)| \le \varepsilon$. Once the optimum weights are obtained, the logistic regression model is ready for testing using (1). The pseudo code for logistic regression model is given in Table I for the ease of implementation. Various mapping functions proposed and considered for the logistic regression model designed are given in Table II for an input with feature size 6.

TABLE I. PSEUDOCODE FOR LOGISTIC REGRESSION MODEL

Training : Step 1 : Set the following : Acceptable threshold for Cost Function, ε. Maximum Number of Epochs, N_{max}. Number of Epochs, N. Initialize Augmented Weight Matrix, $\boldsymbol{\theta} = 1$. Initialize the Cost Function, $\mathrm{J}(\theta) = 0$. Step 2 : Select a Mapping function for input features. Step 3 : Update Augmented Weight Matrix $\boldsymbol{\theta}$ using $\boldsymbol{\theta_j(n) = \theta_j(n-1)} + \alpha\cdot\upsilon_{\mathbf{j}}$ Step 4 : Find the Cost Function or Average Cost $\mathrm{J}(\theta)$ using $\mathrm{J}(\theta)=\mathrm{J}(\theta)+\left(-\frac{1}{\mathrm{m}}\right)\left(\sum_{i=1}^{m} y^{(i)}\log\left(h_{\theta}\left(x^{(i)}\right)\right)+\left(1-y^{(i)}\right)\log\left(1-h_{\theta}\left(x^{(i)}\right)\right)\right)$ Step 5 : If $|\mathrm{J}(\theta)| \le \varepsilon$ (or) $N = N_{max}$ Goto Step 6 Else Goto Step 3 to update the Augmented Weight Matrix $\boldsymbol{\theta}$ Step 6 : Optimum weights are obtained for $\boldsymbol{\theta}$.
Testing : Find the output using optimum weights and test input $y = \mathrm{sgn}(\boldsymbol{\theta}^{T}\mathbf{x})$

TABLE II. MAPPING FUNCTIONS PROPOSED FOR EVALUATION

Function1 (MF1)	[1 X0 X1 X2 … X5]
Function2 (MF2)	[1 $X0^2$ $X1^2$ $X2^2$ … $X5^2$]
Function3 (MF3)	[1 $X0^3$ $X1^3$ $X2^3$ …… $X5^3$]
Function4 (MF4)	[1 X0 X1 X2 … X5 X0·X1 X2·X3 X4·X5]
Function5 (MF5)	[1 $X0^2$ $X1^2$ $X2^2$ … $X5^2$ X0·X1 X2·X3 X4·X5]

III. DESIGN OF DIGITAL CIRCUITS USING PROPOSED LOGISTIC REGRESSION MODEL

Digital circuits are good examples for pattern classification and design of digital circuits using Support Vector Machine (SVM) are reported in [13]. In order to assess the performance of proposed non-linear mapping functions for logistic regression model, various digital circuits are implemented in this section using proposed mapping function and logistic regression model. The digital circuits considered for evaluation are 2-input AND gate, 2-input OR gate, 2-input XOR gate, 8:3 Encoder and 3:8 Decoder which includes both linear and non-linear classification problems. 100% recognition accuracy is a must for designing logistic regression model based digital circuits. Details about the mapping function used to design digital circuits are reported in Table III. The classifier equation obtained for various digital circuits considered in this paper are reported in Table IV along with the minimum number of epochs required to get the cost function $\mathrm{J}(\theta)$ with acceptable threshold set as $\varepsilon = 0.01$ for AND gate, OR gate, XOR gate, encoder, and $\varepsilon = 0.001$ for decoder. The decision boundaries obtained from the logistic regression model for AND gate, OR gate and XOR gate using the classifier equation given in Table IV are shown in Fig. 1(a), 1(b) and 1(c) respectively. The number of epochs vs. cost function $\mathrm{J}(\theta)$ characteristic for AND gate, OR gate and XOR gate are shown in Fig. 2.

The block diagram of proposed logistic regression model is shown in Fig. 3, where X0 to X*n* denote the *n* input features, T0 to T*m* represent the *m* terms obtained at the output of mapping function and C0 to C*m* denotes the coefficients for terms T0 to T*m*. It is observed during the study and analysis that any digital circuit can be designed using Fig. 3 by employing the mapping function that gives 100% recognition accuracy. The same circuit can perform different functionalities by just changing the coefficients and mapping function (if needed).

TABLE III. PERFORMANCE EVALUATION FOR DIGITAL CIRCUIT DESIGN USING LOGISTIC REGRESSION MODEL

Digital Circuit	Input Features	Mapping Function	Type of Mapping Function	Recognition Accuracy
AND gate	X(1:0)	[1 X0 X1]	Linear	100%
OR gate	X(1:0)	[1 X0 X1]	Linear	100%
XOR gate	X(1:0)	[1 X0 X1 X0·X1]	Non-Linear	100%
8:3 Encoder	X(7:0)	[1 X0…… X7]	Linear	100%
3:8 Decoder	X(2:0)	[1 X0 X1 X2]	Linear	100%

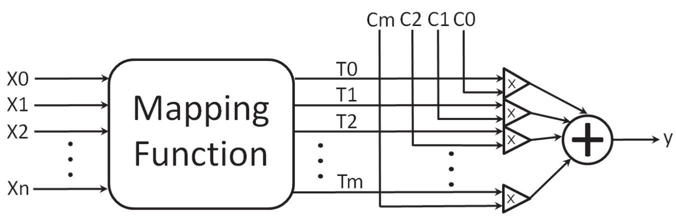


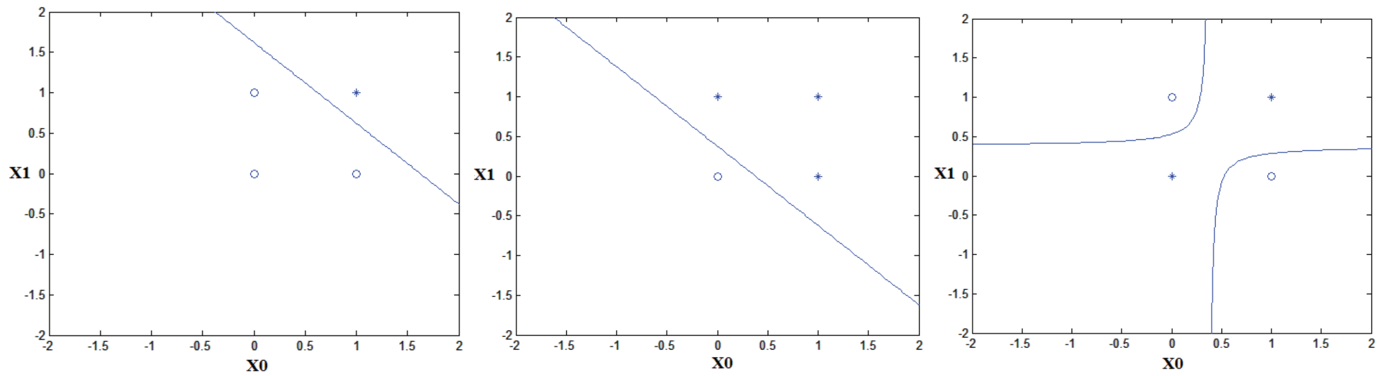
Fig. 3. Block diagram of proposed Logistic Regression Model

IV. PERFORMANCE EVALUATION OF PROPOSED MODEL FOR UMASS DATASETS

In order to further assess the performance of the logistic regression model using proposed mapping functions, standard datasets from University of Massachusetts, Amherst (UMASS) database [14] are considered. Details about the datasets considered for performance evaluation are given in Table V and the results of their performance evaluation are reported in Table VI. The block diagram shown in Fig. 3 is used to build logistic regression model for pattern recognition

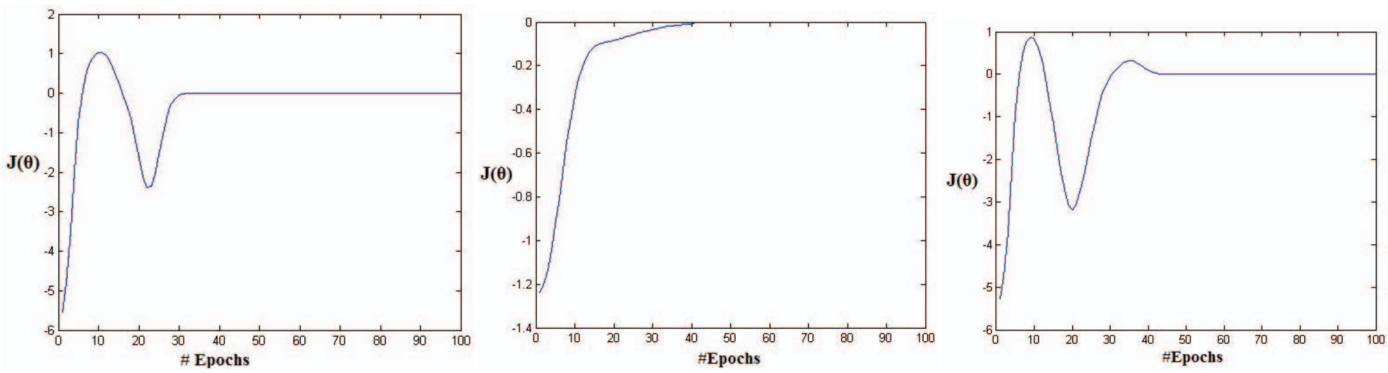
TABLE IV. CLASSIFIER EQUATION OBTAINED FOR DIGITAL CIRCUITS DESIGNED USING LOGISTIC REGRESSION MODEL

Digital Circuit	Decision Boundary/ Classifier Equation	Mapping Fn Used	#Epochs	J(θ)
AND gate	y = –15.7732 + 9.7250·X0 + 9.7250·X1	MF1	32	–0.0077
OR gate	y = –4.8557 +12.8460·X0 +12.8460·X1	MF1	40	–0.0088
XOR gate	y = –5.0135 + 9.4844·X0 + 9.4844·X1 – 25.1558· (X1·X2)	MF4	44	0.0043
8:3 Encoder	y2 = –1.1657 + 3.5375·X0 + 3.5373·X1 + 3.5373·X2 + 3.5373·X3 – 2.0787·X4 – 2.0787·X5 – 2.0787·X6 – 2.0787·X7	MF1	11	–0.0055
	y1 = –1.1657 + 3.5375·X0 + 3.5373·X1 – 2.0787·X2 – 2.0787·X3 + 3.5373·X4 +3.5373·X5 – 2.0787·X6 – 2.0787·X7	MF1	11	–0.0055
	y0 = –1.1657 + 3.5375·X0 – 2.0787·X1 + 3.5373·X2 – 2.0787·X3 + 3.5373·X4 – 2.0787·X5 + 3.5373·X6 – 2.0787·X7	MF1	11	–0.0055
3:8 Decoder	y7= –38.3804 + 15.2677·X0 +15.2677·X1 +15.2677·X2	MF1	47	0.0008
	y6= –27.6505 + 17.1574·X0 +17.1574·X1 –33.7761·X2	MF1	41	–0.0002
	y5= –27.6505 + 17.1574·X0 – 33.7761·X1 +17.1574·X2	MF1	41	–0.0002
	y4= –6.2350 + 13.4903·X0 – 20.4070·X1 –20.4070·X2	MF1	26	–0.0003
	y3= –27.6505 –33.7761·X0 + 17.1574·X1+ 17.1574·X2	MF1	41	–0.0002
	y2= –6.2350 –20.4070·X0 + 13.4903·X1– 20.4070·X2	MF1	26	–0.0003
	y1= –6.2350 –20.4070·X0 – 20.4070·X1+ 13.4903·X2	MF1	26	–0.0003
	y0= 7.2221 –27.5200·X0 – 27.5200·X1– 27.5200·X2	MF1	35	0.0003



(a) AND gate (b) OR gate (c) XOR gate

Fig. 1. Decision boundaries obtained for digital gates designed using proposed Logistic Regression model



(a) AND gate (b) OR gate (c) XOR gate

Fig. 2. Number of Epochs vs. Cost Function J(**θ**) for digital gates designed using proposed Logistic Regression model

TABLE V. UMASS DATASETS CONSIDERED FOR EVALUATION OF PROPOSED MODEL

Datasets	N_A	N_{TR}	N_T	Output (Yes/No)
Myopia Study	15	309	309	Myopia within the first five years of follow-up
Prostate Cancer	7	98	97	Tumour penetration of Prostatic Capsule
Low Birthweight	9	96	93	Giving birth to a low birth weight baby
UMASS AIDS Research Unit	7	300	275	Subject remained drug-free for 12 months
Poly- Pharmacy	11	1750	1750	Taking drugs from more than three different classes

N_A : Number of Attributes/ Input Features ; N_{TR} : Number of Training Samples ; N_T : Number of Testing Samples systems designed to evaluate the datasets. The bold face values represent the best accuracy obtained for a particular mapping function and dataset employed by the logistic regression model. The classifier equation obtained for various datasets considered in this paper for experimentation and analysis are reported in Table VII for the mapping function with best recognition accuracy.

TABLE VI. PERFORMANCE EVALUATION FOR UMASS DATASETS

Datasets	#Epochs	Recognition Accuracy (%)				
		MF1	MF2	MF3	MF4	MF5
Myopia Study	100	32.69	**86.73**	74.43	84.47	78.96
	400	88.03	85.76	78.96	**89.32**	85.44
	500	86.73	86.73	60.19	86.41	83.50
	1000	88.35	86.73	81.88	87.06	87.38
	1300	**89.97**	85.76	80.91	87.06	86.73
	1500	85.44	80.91	86.41	73.46	86.08
	1600	89.00	79.94	79.94	83.50	**88.03**
	2000	87.70	82.85	44.98	83.82	85.44
	2300	86.41	80.91	**87.06**	87.70	84.14
	2500	87.70	83.82	85.76	79.29	84.47
	3000	84.14	86.73	70.87	86.73	81.23
Prostate Cancer	100	65.26	44.21	38.42	65.26	39.47
	500	72.63	38.95	56.84	61.58	65.26
	1000	**75.79**	38.95	66.84	67.89	73.68
	1500	71.05	41.05	37.89	63.68	73.68
	1700	68.42	68.42	38.42	**71.58**	69.47
	2000	43.68	49.47	65.79	67.37	71.05
	2500	64.21	68.42	38.42	69.47	73.68
	2600	68.95	60.53	**67.89**	68.42	63.68
	2800	73.68	69.47	38.95	68.95	**74.21**
	2900	71.58	**71.05**	37.89	64.21	56.84
	3000	67.89	69.47	38.42	71.58	70.00
Low Birthweight	60	**90.32**	82.80	72.04	83.87	78.49
	80	89.25	**86.02**	73.12	**95.70**	72.04
	90	66.67	74.19	72.04	74.19	**87.10**
	100	73.12	86.02	**83.87**	84.95	53.76
	500	73.12	80.65	83.87	61.29	70.97
	1000	62.37	81.72	75.27	72.04	76.34
	1500	65.59	77.42	65.59	72.04	62.37
	2000	64.52	86.02	74.19	72.04	70.97
	2500	65.59	81.72	68.82	72.04	75.27
	3000	68.82	77.42	70.97	70.97	76.34
UMASS Aids Research Unit	70	72.36	**72.73**	**72.73**	33.45	**72.73**
	90	70.91	72.73	67.64	72.73	72.36
	100	69.45	32.73	66.55	69.09	36.36
	500	69.45	53.09	68.73	68.73	63.64
	800	72.00	72.73	68.00	36.73	71.64
	1000	61.45	70.91	66.55	72.73	65.09
	1500	72.73	69.09	70.18	72.73	67.27
	1700	51.27	72.73	30.18	**73.45**	66.18
	2000	56.73	72.36	70.18	70.18	61.45
	2500	72.73	70.55	65.82	69.45	**72.73**
	2800	**73.82**	64.36	69.45	72.73	69.45
	3000	73.09	62.55	61.82	55.27	64.00
Poly Pharmacy	80	**76.57**	49.77	73.03	75.49	72.40
	100	46.63	65.09	66.74	76.46	74.57
	500	59.66	71.77	65.66	50.63	69.14
	600	75.54	74.69	**75.37**	66.00	70.63
	700	44.29	**75.66**	70.97	56.40	72.11
	1000	52.63	62.51	47.94	56.74	59.31
	1300	76.34	58.29	74.97	**76.69**	69.26
	1500	75.31	66.00	70.06	75.54	72.80
	2000	75.94	66.57	39.83	74.00	58.74
	2300	76.06	75.37	58.40	74.91	**76.40**
	2500	57.94	51.83	73.71	66.91	61.49
	3000	74.91	75.03	75.20	61.43	65.20

TABLE VII. CLASSIFIER EQUATION OBTAINED FOR UMASS DATASETS

Datasets	Mapping Function	Decision Boundary / Classifier Equation
Myopia Study	MF1	y = –0.505–0.2195•X0+0.0209•X1–3.4608•X2+ 0.1747•X3+ 1.0000•X4 –0.5456•X5– 0.2783•X6–0.0147•X7+0.0736•X8+0.0243•X9 – 0.0854•X10 – 0.0246•X11 – 0.0117•X12 +0.7251•X13+ 0.2182•X14
Prostate Cancer	MF1	y = –0.9494–0.0799•X0–2.6928•X1+ 1.0000•X2+ 0.7492•X3+ 0.0691•X4 – 0.0148•X5+ 0.7189•X6
Low Birthweight	MF4	y = 1.0000+ 0.2551•X0+ 0.5082•X1+ 0.0353•X2+ 0.3256•X3+ 0.0004•X4+0.4221•X5 + 0.3530•X6 –0.0322•X7 –2.6957•X8+ 0.0571•X0•X1 –0.1098•X2•X3+ 0.1526•X4•X5 –0.2951•X6•X7
UMASS Aids Research Unit	MF1	y = –0.6275+ 1.0000•X0+ 0.1203•X1 – 0.1484•X2 –0.6747•X3 –0.1811•X4+ 0.0706•X5+ 0.0009•X6
Poly-Pharmacy	MF4	y = –0.9334–0.0362•X0–0.3223•X1– 0.0994•X2+0.1205•X3–0.5737•X4+0.8377•X5– 1.0353•X6+0.0406•X7–0.0804•X8+0.0002•X9+ 1.0000X10 +0.5880•X0•X1–0.1231•X2•X3+ 0.3104•X4•X5 +0.1754•X6•X7+0.0002•X8•X9

The number of epochs vs. cost function J(θ) characteristic for Myopia datasets of the UMASS database is shown in Fig. 4 and similar characteristics are obtained for other datasets of UMASS database too.

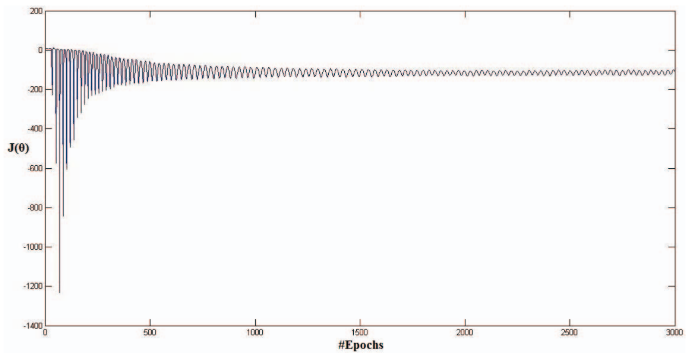


Fig. 4. Number of Epochs vs. Cost Function J(**θ**) for Myopia datasets

V. PERFORMANCE EVALUATION OF PROPOSED MODEL FOR WIRELESS SENSOR NETWORK DATASETS

The proposed logistic regression model and mapping functions are evaluated for applications pertaining to wireless sensor networks too using the standard datasets from Labelled Wireless Sensor Network Data Repository (LWSNDR)[15], University of Carolina and UCI Machine Learning Repository, University of California, Irvine. Details about the datasets are given in Table VIII and the performance evaluation of logistic regression model with various mapping functions are reported in Table IX and Table X for binary-class and multi-class problems respectively. It is observed from Table IX and X that the mapping function MF5 gave best recognition accuracy for multi-class problem and for some cases in binary-class problem. The comparison between results obtained on using proposed model with best mapping function and results reported in literature are given in Table XI and Table XII.

TABLE VIII. DATASETS PERTAINING TO WIRELESS SENSOR NETWORKS CONSIDERED FOR EVALUATION OF PROPOSED MODEL

Datasets	N_A	N_{TR}	N_T	Classification Requirements and Other Details	Ref.
Single hop (Indoor)	2	2208	2209	Anomaly detection of single-hop networks with sensors located indoor Data collection using TelosB motes with humidity and temperature sensor.	[15]
Single hop (Outdoor)	2	2520	2519	Anomaly detection of single-hop networks with sensors located outdoor Data collection using TelosB motes with humidity and temperature sensor.	[15]
Multi hop (Indoor)	2	2345	2345	Anomaly detection of multi-hop networks with sensors located indoor Data collection using TelosB motes with humidity and temperature sensor.	[15]
Multi hop (Outdoor)	2	2345	2345	Anomaly detection of multi-hop networks with sensors located outdoor Data collection using TelosB motes with humidity and temperature sensor.	[15]
Occupancy-1	5	8143	2665	Occupancy detection of an office room with room doors opened Sensors used: Temperature, RH, Light, CO_2 and Humidity Ratio.	[16]
Occupancy-2	5	8143	9752	Occupancy detection of an office room with room doors closed Sensors used: Temperature, RH, Light, CO_2 and Humidity Ratio.	[16]
EEG eye state	14	7490	7490	Prediction of the state of eye as Open or Closed Data collected using Emotiv EEG Neuroheadset.	[17]
Wilt	5	4339	500	Detection of diseased Pine and Oak trees from segmented images Data collected using Quickbird Multi spectral Imagery System.	[18]
Human Activity	12	5000	5000	Human activity recognition from Accelerometer Readings. Activities : Sitting down, standing up, standing, walking and sitting.	[19]

N_A : Number of Attributes/ Input Features ; N_{TR} : Number of Training Samples ; N_T : Number of Testing Samples

TABLE IX. BINARY CLASS : DATASETS PERTAINING TO WIRELESS SENSOR NETWORKS CONSIDERED FOR EVALUATION OF PROPOSED MODEL

Datasets	#Epochs	Recognition Accuracy (%)				
		MF1	MF2	MF3	MF4	MF5
Single hop Indoor	500	99.46	99.86	99.73	96.83	95.97
	1100	99.37	99.86	99.73	98.96	99.73
	1300	99.50	99.86	99.68	96.83	**99.91**
	1900	99.77	99.86	99.37	96.83	99.68
	2400	99.77	98.82	99.50	98.55	99.55
	2600	99.77	99.68	99.55	96.83	74.78
Single hop Outdoor	200	99.48	99.48	**99.84**	0.52	93.33
	900	**99.84**	99.64	99.64	99.48	99.76
	1500	96.71	98.49	99.76	99.48	99.84
	2500	99.72	**99.84**	99.33	99.48	**99.84**
	2600	99.84	99.76	99.76	99.80	98.69
Multi hop Indoor	500	99.15	**99.83**	99.62	2.35	99.66
	1100	99.45	99.62	99.66	2.35	99.83
	1300	98.72	98.85	99.79	98.64	99.45
	1900	99.53	99.83	**99.83**	2.35	99.74
	2000	99.62	99.28	99.83	2.35	**99.83**
	2200	**99.83**	99.70	99.10	2.35	99.79
Multi hop Outdoor	600	87.04	98.85	**99.32**	1.15	99.06
	1100	86.78	99.28	98.85	1.15	**99.32**
	1500	97.53	**99.32**	98.85	1.15	99.32
	2000	99.02	98.85	99.32	1.15	99.15
	2800	**99.32**	99.23	99.32	1.15	98.85
Occupancy-1	70	97.90	97.07	36.47	97.07	97.52
	1100	93.32	36.47	91.14	97.07	97.90
	1200	97.82	**97.94**	97.90	97.07	96.89
	3000	96.29	89.49	36.47	97.22	95.95
Occupancy-2	700	**99.38**	94.78	21.01	83.23	89.21
	1100	87.18	40.54	94.44	85.26	94.91
	1800	94.62	95.01	95.05	89.73	90.55
	2900	96.91	95.44	83.06	90.34	88.65
	3000	90.58	94.77	21.01	96.20	88.21
EEG eye state	30	63.78	63.63	63.63	36.36	36.36
	700	63.66	37.54	36.36	64.26	36.37
	1000	63.66	51.92	63.75	**64.26**	36.37
	2100	63.78	37.22	37.62	63.15	36.85
	2300	36.65	38.61	62.27	37.21	63.72
Wilt	10	62.60	62.60	62.60	62.60	62.60
	1300	75.40	62.20	62.20	57.40	66.60
	2300	62.60	62.00	62.60	62.60	**76.40**

TABLE X. MULTI CLASS : DATASETS PERTAINING TO WIRELESS SENSOR NETWORKS CONSIDERED FOR EVALUATION OF PROPOSED MODEL

Datasets	#Epochs	Recognition Accuracy (%)				
		MF1	MF2	MF3	MF4	MF5
Human Activity (Sitting)	1700	99.92	99.44	97.88	78.66	**100.0**
	1800	99.92	99.76	96.78	90.34	99.96
	2100	99.80	99.84	98.28	81.60	100.0
	2600	99.92	99.82	98.92	83.30	100.0
	2700	99.88	99.80	99.58	88.54	99.98
	2800	99.88	99.84	97.46	96.10	99.96
Human Activity (Sitting down)	300	92.36	92.34	89.66	87.18	92.82
	1300	91.38	93.96	90.90	86.40	96.56
	2200	88.76	93.24	89.72	89.48	97.30
	2600	91.58	94.56	90.22	87.22	**97.74**
	2900	90.40	95.14	90.10	88.72	97.48
Human Activity (Standing)	80	92.86	83.76	77.28	92.18	94.54
	1200	87.78	97.16	78.68	82.92	**99.06**
	2300	90.08	97.60	78.44	74.64	98.76
	2600	90.30	92.62	90.32	77.92	98.96
	2700	86.42	95.56	79.50	92.26	98.94
Human Activity (Standing up)	600	97.56	93.80	90.90	90.60	97.74
	1100	97.74	91.20	89.86	91.36	**99.26**
	1200	97.86	92.36	91.60	88.46	99.02
	2700	97.74	93.16	92.42	57.08	99.02
	3000	97.84	92.60	90.78	93.38	99.00
Human Activity (Walking)	1600	90.64	95.16	92.48	84.08	95.16
	2300	68.38	92.80	92.98	82.32	91.28
	2500	64.94	95.88	88.22	88.20	97.48
	2900	87.56	93.16	87.66	70.48	**97.52**

TABLE XI. COMPARISON OF SINGLE AND MULTI HOP DATASET RESULTS

Method	SH(I)	SH(O)	MH(I)	MH(O)
k-NN [20]	97.41	99.36	97.90	98.77
LOF [20]	97.30	99.36	97.43	98.76
aLOCI [20]	97.41	99.36	97.90	98.77
LOOP [20]	97.28	99.36	97.83	98.75
INFLO [20]	97.29	99.36	97.83	98.76
CBLOF [20]	97.35	99.36	97.84	98.76
LDCOF [20]	97.41	99.36	97.91	98.76
HBOS [20]	97.41	99.36	97.91	98.76
LIBSVM [20]	97.40	99.36	97.90	98.77
Proposed Model	**99.91**	**99.84**	**99.83**	**99.32**

TABLE XII. COMPARISON OF OTHER DATASET RESULTS

Dataset	Literature	Proposed Model
Occupancy	99.00% [16]	99.38%
Human Activity (Sitting)	99.90% [21]	100.0%
Human Activity (Sitting down)	96.90% [21]	97.74%
Human Activity (Standing)	99.80% [21]	99.06%
Human Activity (Standing up)	96.90% [21]	99.26%
Human Activity (Walking)	99.80% [21]	97.52%

It is observed from Table XI that proposed model gave best recognition accuracy for LWSNDR datasets when compared with the results reported in the literature. Also a maximum enhancement in recognition accuracy by 2.63% is observed on using proposed model. Similarly proposed model gave better results for datasets reported in Table XII, except for one case. The results confirm that the performance of proposed model is in par with the pattern classification techniques reported in literature.

VI. CONCLUSION

In this paper, design and evaluation of pattern recognition systems using logistic regression model and various mapping functions is proposed. The performance of proposed model is in par with other techniques reported in literature and it is observed that a single circuit can be used to perform multiple functionalities by reconfiguring the coefficients and mapping function. The proposed model is evaluated for different functionalities and can be easily adapted for various other pattern recognition applications too.

ACKNOWLEDGEMENTS

The evaluation of proposed model for wireless sensor network datasets is carried out as a part of research work, funded by the All India Council for Technical Education (AICTE) under the Research Promotion Scheme (RPS) vide project grant No. 8-138/RIFD/RPS/POLICY-3/2013-14.

REFERENCES

[1] Ana Erika Camargo Cruz, “Towards logistic regression models for predicting fault-prone code across software projects”, 3rd International Symposium on Empirical Software Engineering and Measurement, 2009, pp.460-463, Lake Buena Vista, FL.

[2] G. Mauša, T. Galinac Grbac and B. D. Bašić, “Multivariate logistic regression prediction of fault-proneness in software modules”, MIPRO, 2012 Proceedings of the 35th International Convention, Opatija, 2012, pp. 698-703.

[3] J. Behley, K. Kersting, D. Schulz, V. Steinhage and A. B. Cremers, “Learning to hash logistic regression for fast 3D scan point classification”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taipei, 2010, pp. 5960-5965.

[4] Z. Wu, Q. Wang, A. Plaza, J. Li, L. Sun and Z. Wei, “Real-Time Implementation of the Sparse Multinomial Logistic Regression for Hyperspectral Image Classification on GPUs”, in IEEE Geoscience and Remote Sensing Letters, Vol. 12, No. 7, July 2015, pp. 1456-1460.

[5] Petersen J, Austin D, Sack R, Hayes TL, “Actigraphy-based scratch detection using logistic regression”, IEEE Journal of Biomedical and Health Informatics, 2013, 17(2), pp.277-283.

[6] O. Birkenes, T. Matsui, K. Tanabe, S. M. Siniscalchi, T. A. Myrvoll and M. H. Johnsen, “Penalized Logistic Regression With HMM Log-Likelihood Regressors for Speech Recognition”, in IEEE Transactions on Audio, Speech, and Language Processing, Vol. 18, No. 6, Aug. 2010, pp. 1440-1454.

[7] Li Yang, “An application of logistic regression in cerebral infraction disease detection based on association rules with pre-rough classifier”, 17th IEEE International Conference on Industrial Engineering and Engineering Management (IE&EM), 2010, pp. 304-306, Xiamen.

[8] Bolin Chen, “A logistic regression based algorithm for identifying human disease genes”, Bioinformatics and Biomedicine(BIBM), 2014, pp. 197-200, Belfast.

[9] P. D. Prasad, H. N. Halahalli, J. P. John and K. K. Majumdar, “Single-Trial EEG Classification Using Logistic Regression Based on Ensemble Synchronization”, in IEEE Journal of Biomedical and Health Informatics, Vol. 18, No. 3, pp. 1074-1080, May 2014.

[10] Yılmaz Kaya, Diagnostic estimation of OSAS using binary mixture logistic regression, 20th Signal Processing and Communications Applications Conference (SIU), 2012, pp. 1-4, Mugla.

[11] Siti Zubaidah Mohd Tumari, “Logistic regression of working memory impairments in children based on single-trial ERP features”, Technology, Informatics, Management, Engineering, and Environment (TIME-E), 2014, pp. 264-269, Bandung.

[12] S. P. Rahayu, “Applying Kernel Logistic Regression in data mining to classify credit risk”, International Symposium of Information Technology, 2008, pp.1-6, Kuala Lumpur, Malaysia.

[13] Manikandan J and Agrawal VK, “Digital Circuit Design Using Support Vector Machines”, Communication Systems and Network Technologies, April 2014, pp. 1114 –1118, Bhopal.

[14] Logistic Regression Datasets, Statistical Software Information, University of Massachusetts, Amherst. http://www.umass.edu/statdata/statdata/stat-logistic.html.

[15] Shan Suthaharan, Mohammed Alzahrani, Sutharshan Rajasegarar, Christopher Leckie and Marimuthu Palaniswami, “Labelled Data Collection for Anomaly Detection in Wireless Sensor Networks”, in Proceedings of the Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2010), Brisbane, Australia, Dec 2010.

[16] Luis M. Candanedo, Veronique Feldheim, “Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models”, Energy and Buildings, Volume 112, January 2016, pp. 28-39.

[17] Oliver Roesler, “EEG Eye State Data Set, UCI Machine Learning Repository”, Baden-Wuerttemberg Cooperative State University, Stuggart, Germany, June 2013.

[18] Johnson, B., Tateishi, R., Hoan, N., 2013. A hybrid pan sharpening approach and multiscale object-based image analysis for mapping diseased pine and oak trees. International Journal of Remote Sensing, Vol. 34, Issue 20, June 2013, pp. 6969-6982.

[19] Velloso E., Bulling A., Gellersen H., Ugulino W., Fuks H., “Qualitative Activity Recognition of Weight Lifting Exercises”, Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13). Stuttgart, Germany: ACM SIGCHI, 2013.

[20] Mohiuddin Ahmed, Adnan Anwar, Abdun Naser Mahmood, Zubair Shah, Michael J. Maher, “An Invesigation of Performance Analysis of Anomaly Detection Techniques for Big Data in SCADA Systems”, EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, Vol. 2, Issue 3, May 2015.

[21] Ugulino W., Cardador D., Vega K., Velloso E., Milidiu R., Fuks H., “Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements”, Proceedings of 21st Brazilian Symposium on Artificial Intelligence, Advances in Artificial Intelligence - SBIA 2012, In: Lecture Notes in Computer Science, Curitiba, PR: Springer Berlin / Heidelberg, 2012, pp. 52-61.