IP-4 | 1243

MILITARY TECHNICAL COLLEGE

CAIRO - EGYPT

AN APPLICATION OF PATTERN RECOGNITION TO AUTOMATIC DETERMINATION OF LITHOLOGY FROM WELL LOGS

M.M.SYIAM*, A.SHAHIN**, F.SALAEH***

ABSTRACT

Determination of lithology from well logs is the subject of interest of many geologists. The method introduced in this paper gives an automatic determination of lithology from well logs using a 1-nearest neighbor classifier. The 1-nearest neighbor rule (1-NNR) with editing and condensing techniques is used for the design of 1-NN classifier.For this study a training data set has been obtained from a key well in (ABU GARADIG) field. This well has a suitable suite of logs and a continuous core as a geological reference. First, the training data set is edited to obtain a homogeneous clusters of data that improves the performance of the 1-NNR. Basic condensing technique and the ordered condensing technique are applied on the edited data set to obtain a reference patterns for the 1-NN classifier. HOLD-OUT method and ROTATION method are used to estimate the performance of the 1-NN classifier. Both methods partition the edited data set into design data set and test data set. The 1-NN classifier is trained on design data set and tested on the test data set. The estimated recognition rate of the proposed classifier using all patterns in the design data set as a reference patterns is compared with ones using condensed data subset and ordered condensed data subset.The comparison have demonstrated that the 1-NN classifier using ordered condensed data subset as a reference patterns is the optimum one in this application. The designed classifier is tested on other several wells in the same field.The analysis of the experimental results demonstrates that all the output lithologies are in accordance with geological references.

---

* Graduated student, ** Associate professor, Det. of A/C Elec. Equip.& ARM., M.T.C, Cairo, Egypt, *** Professor, Dpt. of Electronics and Communications Engineering, Cairo University, Guiza, Egypt.

I. INTRODUCTION

Classification of objects is an important area of research and application in variety fields. In the presence of full knowledge of the underlaying probabilities, BAYES decision rule gives optimal error rate [1]. However in many pattern recognition problems the available information is just a set of samples representing different classes.The 1-NNR has often been used in these cases [2].

The 1-NNR assigns unclassified sample to the same class as the nearest of N reference, correctly classified samples, each classified by some external source. The most interesting theoritcal property of 1-NNR is that, with an unlimited number of reference samples, the risk in making a decision is less twice the BAYES risk [3]. But as all reference samples must be searched to classify a test sample, the 1-NNR impose large storage and computional requirements. To solve these problems, the 1-NNR has been studied and modified upon by numerous researchers [4-6]. The main modification of basic procedure is, namely, the edited 1-NNR which consists in editing and condensing the training data prior to applying the basic 1-NNR.

The determination of lithology is an important step in the analysis of a reservoir in petroleum prospection.The lithology is normally determined from cores analysis. A problem arises , though where it is not possible to obtain a continuous core. The lithology determination based on core data can not cover all depth levels of the well.

Well logs can contribute a great deal to the determination of the lithology since they respond to a wide range of geological parameters and run in all well levels . With the increase of phisical parameters recorded by modern logging tool , it becomes more obvious that their combination can give a good idea on the lithology [7]. Geologists devote much knowledge and experince to determine lithology from well logs. Yet it is time consuming and can become tedious.This is particularly true analysis of high dimensional well logging data, where instead of data in one dimension, we seek to determine lithology from combined d-dimensional well logging data.

The basic idea in the pattern recognition approach is to represent a set of d log readings at a given depth level as a point in d-dimensional space. A lithology is then represented as a cluster or class in this space, that is an area of relatively high concentration of points, the points are close because the log responses are similar. Our task is to attach to each depth level of the well the lithology to which it belongs using its d log readings.

The application of pattern recognition to this problem has been studied in [8]. BAYES decision rule is used simply as a statistical procedure for classification. It was assumed that the log data for each lithology has d-variate gaussian

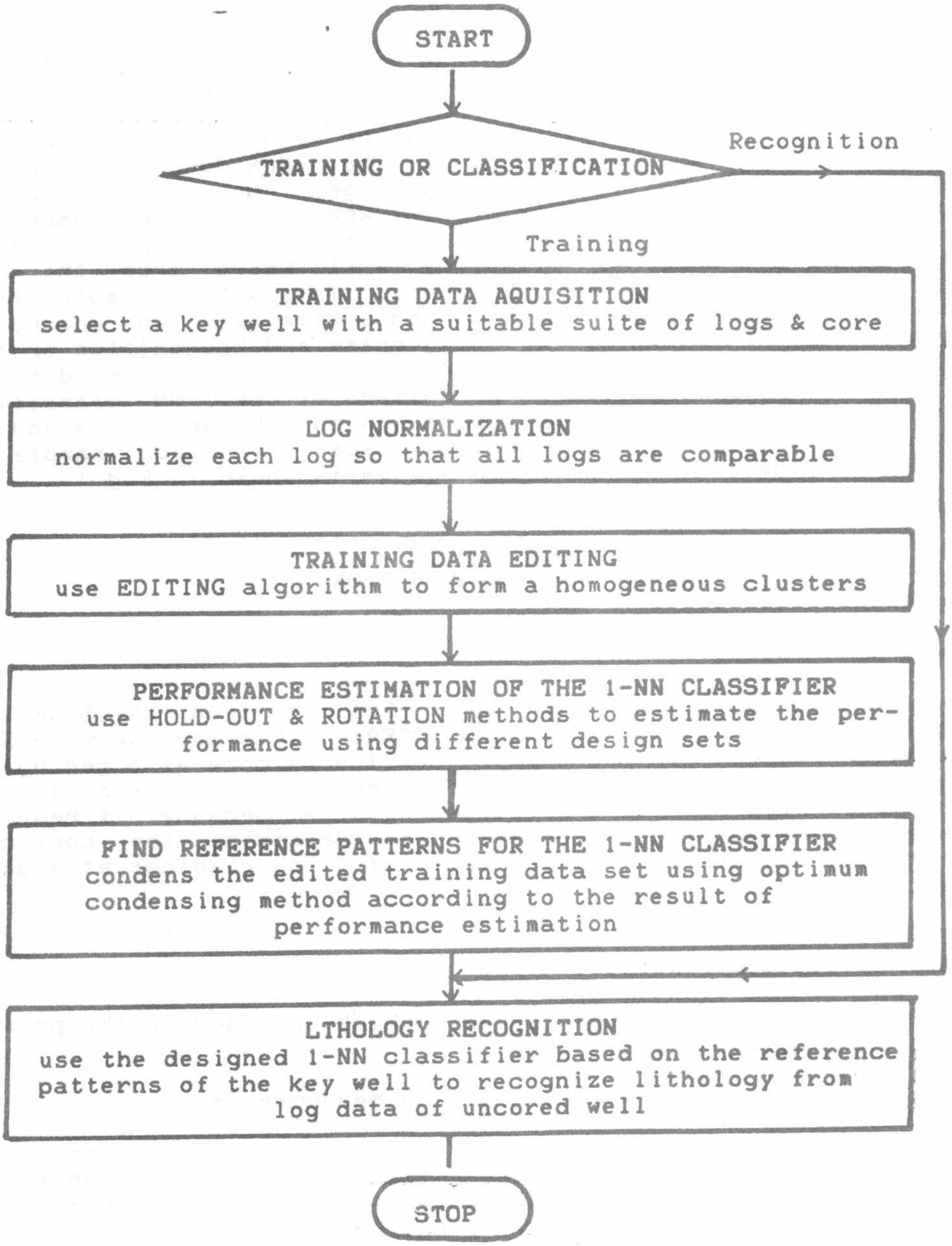

Fig.1. Flow chart of the proposed lithology recognition system

distrbution, of course if the log data are not gaussian distrbuted these statistics can give a very missleading description of data and may lead to poor or meaningless results.

By contrast, this paper presents the method of the 1-NNR with editing and condensing techniques. The 1-NNR does not use the statistical distrbutions and exchanges the need to know the underlaying distrbutions for that of knowing a large number of samples representing different classes. Each sample has d log readings and a correctly lithology label known from the core analysis and geologists experience. This set of samples forms the training data set. The proposed system has two phases of operation, training or learning phase and recognition or test phase. The input to the training phase is the training data set and the output is the design parameters of the 1-NN classifier. The input to recognition phase is an unknown sample and the output is classification lithology label for such sample.The main flow chart of the proposed system is shown in Fig.1.

II. TRAINING DATA ACQUISITION

Area Under Study

A training data set has been obtained from AG-25 well in ABU-GARADIG field in the WESTERN DESERT. This key well has a suitable suite of logs and a continuous core as a geological reference. The construction of the training data is the critical step in the developed system because it requires careful core to log correlation to determine the correctly lithology label for each sample of d log readings at a given depth.

Core To Log Correlation

The available logs in the key well which are used in the present study are shown in Table 1.

Table 1 List Of Log Measurements

Log No	Symbole	definition	units
1	RHOB	formation bulk density	g/cc
2	NPHI	neutron log	%
3	DT	interval transit time (SONIC)	usec/ft
4	RD	deep resistivity	ohm-m
5	RM	medium resistivity	ohm-m
6	GR	gamma ray measure	API
7	CL	calipre measure	inch.
8	SP	spontaneous potenial	milivolt

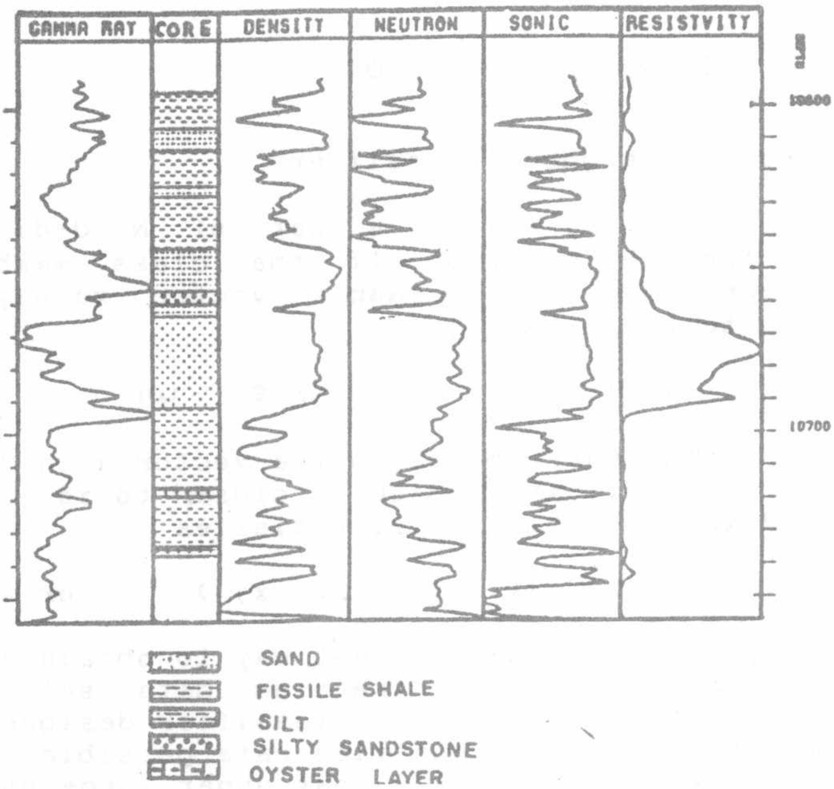

Fig.2. Manual core to log correlation

Fig. 2 shows the manual core to log correlation to determine the lithology label for each log sample of d log reading at a given depth in the cored interval of the key well [9].

The training data log samples are subdivided into five classes (lithologies) on the basis of core to log correlation. The number of samples for each lithology is shown in Table 2.

Table 2 List of lithologies in the key well

Lithology No	Lithology Name	Number of samples
1	Sandstone	28
2	Fissile Shale	53
3	Silt	17
4	Siltysandstone	13
5	Oyster-layers	8
Total		191

Data Normalization

The comparison of different logs in units is arrived by normalizing each log to mean zero and standard deviation one.

III. CLASSIFICATION

The 1-Nearest Neighbor Classifier

Let $(x1, \ldots, xN)$ be a set of N d-dimensional random vectors from M classes. If the class memberships of these samples are known, a new sample vector xo may be classified according to the 1-NNR [1],

$$xo \in \omega i \quad \text{if} \quad xNN \in \omega i \qquad (1)$$

The vector xNN corresponds to the vector x in the training data set $(x1, \ldots, xN)$ which is closest to xo for the distance measure d(xo,xj), that is, xl satisfies

$$d(xo, xl) < d(xo, xj) \qquad \text{for all } j \neq l \qquad (2)$$

Obviously, this is the simplest way to obtain a 1-NN classifier where all samples in the training data set are used as a reference patterns. A 1-NN classifier designed in this fash achieves the highest recognition rate possible for the training data set.It was proved in[1] that under large number of training samples assum ion, the 1-NNR error rate E1 is bounded from above by twice the optimum BAYES error rate E*

$$E1 \leqslant 2E* \qquad (3)$$

However this 1-NN classifier has one major drawback, that is to cla sify an unknown sample xo, it requires the computation of distances between xo , and all samples in the training data set. Therefore while maintaining the highest possible recognition rate we would like to use a small number of reference patterns. In this paper we shall use a method of editing and condensing to solve this problem.

Training Data Editing

Given training data set $(x1, \ldots, xN)$, with known classification, it will be some times the case that some of samples from one class will lie in a region where most of observations are from another class. It such a case, it may be possible to improve the 1-NNR performance by removing from the training data these samples which are surrounded by samples from a different class. This problem was discussed in [4] and it was proved that after editing of the training data set, the 1-NNR performance is improved as follows

$$E* \leqslant E1 \leqslant 1.62\, E* \qquad (4)$$

So the MULTI EDIT algorithm in [5] is applied on the present training data set.

MULTI EDIT Algorithm

step 1, Diffusion : make a random pool of the available training data set

step 2, Classification : classify each sample in the training data set using Leave One Out method, where each sample is classified using all other samples as a referece patterns

step 3, Editing : discard all samples that were missclassified in step 2

step 4, Confusion : poll all the remaining data samples to constitute a new set

step 5, Termination : if the last I iteration produced no editing, exit with final solution, else go to step 1

Subsequently classification is performed using 1-NNR with the remaining data as a design set.

Table 3 The results of MULTI EDIT algorithm

class code	No of samples before editing	No of sapmles discarded phase 1	phase 2	phase 3	samples retained
1	28	1	1	0	26
2	53	1	0	0	52
3	17	1	0	0	16
4	13	0	0	0	13
5	8	1	0	0	7
Total	119	4	1	0	114

Table 3 shows the results of applying MULTI EDIT algorithm on the training data set. Figures 3 and 4 illustrate the performance of the algorithm for the training data set in a 2-dimensional normalized space (Resistivity & Sonic).

Training Data Condensing

The multi edit algorithm has created homogeneous clusters of samples. The 1-NNR using the edited data implements a piecewise linear decision boundary which is the sample-based approximation of the BAYES optimal decision boundary. The sample based decision boundary is defined by a small number of samples belonging to outer envelopes of the clusters. The samples which do not contribute to define boundary, for example these samples which deeply imbeded within the clusters, may be discarded with no effect on the subsequent performance of the 1-NNR. This is the basic idea in the condensing technique. Two condensing techniques are considered in the study, the basic condensing algorithm and modified condensing algorithm[6].

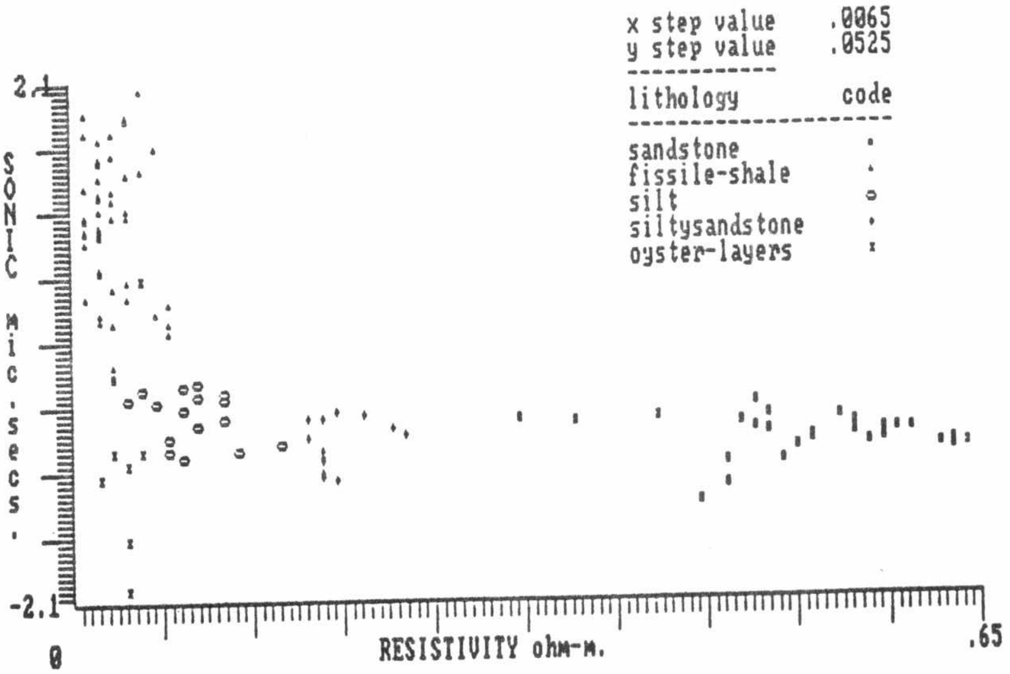

Fig.3. Training data before editing

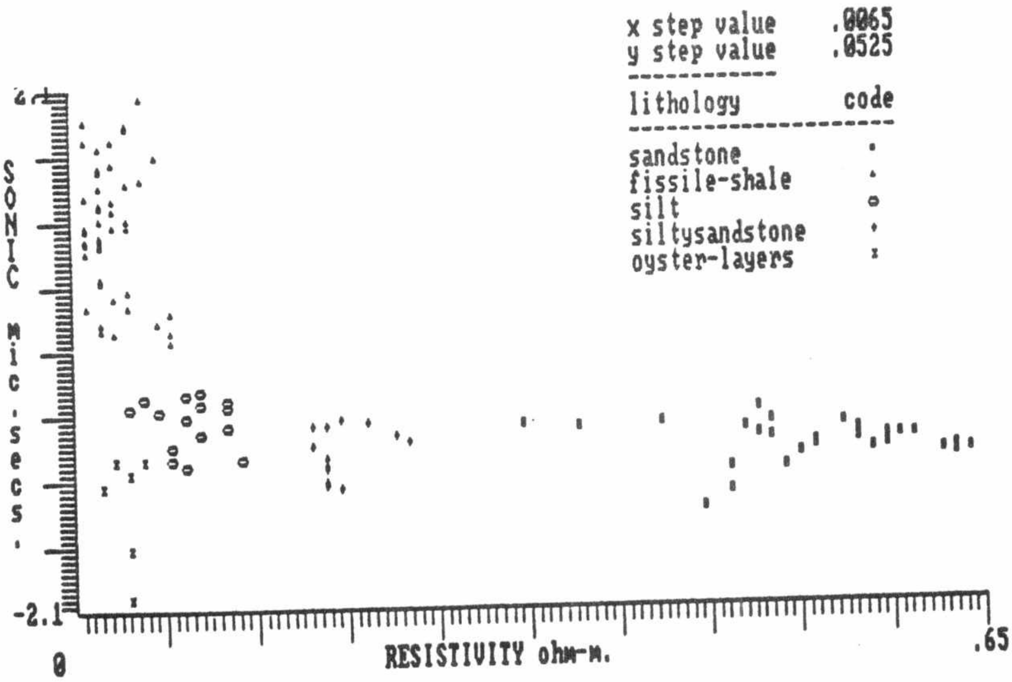

Fig.4. Training data after editing

Basic condensing algorithm

step 1 : set up bins called STORE and GARBBAGE, the first sample in the edited training data set is placed in STORE, all other samples in GARBBAGE.

step 2 : use 1-NNR with current content of STORE to classify the ith sample from GARRABAGE, IF classified correctly the sample is returned to GARBBAGE, otherwise it is placed in STORE. Repeat this procedure for all samples in GARBBAGE.

step 3 : if one complete pass is made through step 2 with no transfer from GARBBAGE to STORE or the GARBBAGE is exahausted, then terminate else go to step 2.

The final contents of STORE constitute the condensed subset to be used with 1-NNR as a reference patterns and the contents of the GARBBAGE are discarded. The basic algorithm processes the samples randomly, so internal rather than boundary samples are occasionally retained. This means that the condensed subset is larger than necessary and contains interior samples which could be eleminated completely without changing the 1-NNR performance.

The modification algorithm would work essentially as basic algorithm but would only arrange the original samples in some order such that samples close to decision boundary would be only used.

Modified condensing algorithm

step 1 : for each sample x of edited training data set , find the nearest neighbor xNN from opposite class. Also record the distance between x and xNN and associate it with xNN .

step 2 : using the results from step 1, order the samples according to associated distances in an ascending order.

step 3 : continue with the basic algorithm.

The final contents of STORE in this case constitute the ordered condensed subset.

Performance Estimation

The edited training data set is partitioned into two sets, design set and test set, in different ways using HOLD-OUT method and ROTATION method [10]. The 1-NN classifier is trained on design set and tested on the test set. The estimated recognition rates of the 1-NN classifier using all samples in the design set as a reference patterns are compared with ones using condensed subset and ordered condensed subset of the design set as a reference patterns.The comparison results are shown in Table 4. and Table 5.

Table 4 The estimated recognition rates using HOLD-OUT method

Classifierfs	Reference patterns	Design set (57 samples) recogn. rate %	Test set (57 samples) recogn. rate %
1-NN classifier using initial design set	57	100	98.70
1-NN classifier using condensed subset	10	100	95.68
1-NN classifier using ordered condensed subset	9	100	96.57

Table 5 The estimated recognition rate using ROTATION method

Classifiers	Reference patterns	Design set (78 samples) recogn. rate %	Test set (38 samples) recogn. rate %
1-NN classifier using initial design set	78	100	98.25
1-NN classifier using condensed subset	11	100	97.25
1-NN classifier using ordered subset	9	100	97.53

From these tables, the 1-NN classifier using ordered condensed subset as a reference patterns is the optimum one because it still gives high recognition rates for both the design set (100 %) and the test set (97.53 %) even though the number of reference patterns is only (9) patterns.

So modified condensing algorithm is applied on all samples of the edited training data set to obtain the reference patterns of proposed 1-NN lithology classifier. Fig.5 shows the reference patterns for the edited data in Fig.4 after applying modified condensing algorithm.

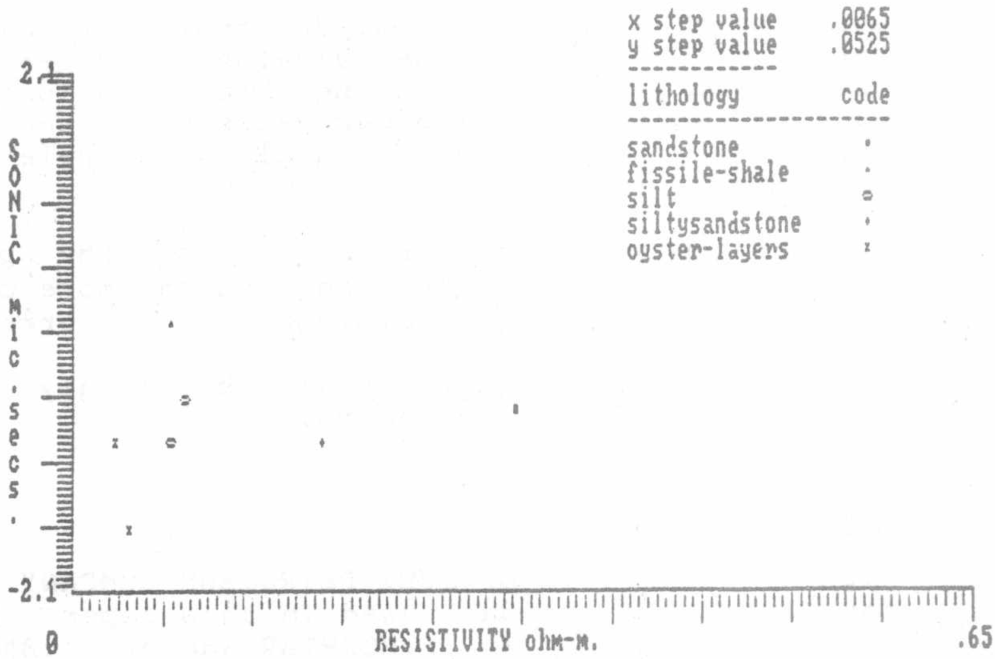

Fig.5. Reference patterns of edited data

V. EXPERIMENTAL RESULTS AND CONCLUSION

For the purpose of testing the developed classifier, it is applied on several test examples. Three wells in the same field (AG-13, AG-22 and AG-28) are selected as test examples, (314) log samples in these wells are classified using the developed classifier. The classification results of these test samples are compared with the core descriptions and geologists interpretation. The comparison results indicates that (276) samples i.e (87.8 %) percent are correctly classified. This shown in Table 6.

Table 6 Classification results of test wells

test well	no of test samples	no of correctly classified samples	percentage % of correctly classified samp.
AG-13	132	120	90.9
AG-22	54	46	85.1
AG-28	128	110	85.9
Total	314	276	87.8

From the obtained results we can conclude that the developed classifier works satisfactorily.The developed system has the advantage of being fully automatic as the interpretation work is done once (training phase) for all other wells. But the quality of results critically depends on the quality and relevance of the training data set.

In order to further improve the performance of the proposed approach, more training samples are required and more carefull are need for depth matching and manual logs to core correlation.

Future developments in the present work will be mainly in the direction of the feature selection analysis.

ACKNOWLEDGMENT

The authors wish to thank GULF OF SUEZ PETROLEUM COMPANY (GUPCO) management for releasing the data used in this paper. Special thanks also to Dr. M.HASSOUBA, Mr. A.MOKHTAR and Mr. A.BADAWI of GUPCO for their valuable and useful comments.

REFERENCES

1. Duda, R.O. and HART, P.E., " Pattern classification and scene analysis", NEW YORK:Willey, (1973).
2. Cover, T.M. and Hart, P.E., "Nearest neighbor pattern classification", IEEE Trans. Inform.Theory, vol.IT.13, pp. 21-27, Jan. (1967).
3. Devijver, P.A. and Kittler, J., " Pattern Recognition: A Statistical Approach ", New Jersy: Prentice/Hall Intrnational, (1981).
4. Wilson, D.L., " Asymptotic properties of nearest neighbor rule using edited data ", IEEE Trans. Syst. Man Cybern., vol. smc-2, pp. 408-420, July (1972).
5. Devijver, P.A. and Kittler, J., " On the edited nearest neighbor rule ", Proc. 6th Int. Conf. Pattern Recognition, pp. 72-80, (1980).
6. Tomek, I., " Two modifications of CNN ", IEEE Trans. Syst. Man.Cyber., vlo. smc-6, pp. 769-772, Nov. (1976).
7. Serra, O. and Abort, H., " Tht contribution of logging data to sedimentology and stratigraphy ", 55th Ann. fall Tech. Conf. SPE OF AIME, Dallas, pp. 90-102, (1982).
8. Serra, O. and Delfiner, P.C., " Automatic determination of lithology from well logs ", SPE 13290, PP. 1-8, (1985).
9. Wasfi, S., Hassouba, M., Mokhatar, A., and Badawi, A. " ABU GARADIG "C" sandstone environmental model for hydrcarbon exploration ", E.G.P.C. eight exploration conf., Nov. (1986).
10. Toussaint, G.T., " Bibigoraphy on estimation of miss classification ", Machine Recognition Of Patterns, IEEE press, pp. 184-191, (1976).