



Automated Assessment of Lymphocytes Using Machine Learning Techniques

Afaq Ahmad, Azmat Ullah, Kaleem Nawaz Khan and
Muhammad Salman Khan

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 18, 2021

Automated Assessment of Lymphocytes Using Machine Learning Techniques

Afaq Ahmad

Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
afaqkhan7222@gmail.com

Azmat Ullah

Artificial Intelligence in Healthcare,
IIPL, NCAI
UET Peshawar, Pakistan
azmatullah@uetpeshawar.edu.pk

Kaleem Nawaz Khan

Artificial Intelligence in Healthcare,
IIPL, NCAI
UET Peshawar, Pakistan
kaleemnawaz@uetmardan.edu.pk

Muhammad Salman Khan

Department of Electrical Engineering
Artificial Intelligence in Healthcare,
IIPL, National Center of Artificial Intelligence
University of Engineering and Technology
Jalozai campus, Peshawar, Pakistan
salmankhan@uetpeshawar.edu.pk

Abstract—This paper investigates the performance of machine learning techniques for automated assessment of lymphocytes. A total of four algorithms, i.e., support vector machine (SVM), deep learning for java (DL4J), multi-layer perceptron (MLP) and K* are applied to the lymphocytes dataset. To ensure the robustness, all the four algorithms are evaluated with both pre-processed data and data without prior pre-processing. With the pre-processed data, MLP outperforms all the other techniques and achieve an accuracy of 98.64%. SVM, DL4J and K* achieve accuracies of 97.97%, 96.62% and 97.29%, respectively.

Index Terms—lymphocytes, machine learning, SVM, MLP, DL4J, K*

I. INTRODUCTION

During the last couple of decades a lot of medical diagnosis tools have been evolved and many more to come [1]. With the growth of cancer immunotherapy, these characterizations can provide clinically significant information to further our understanding of the immune response in cancer patients across a wide spectrum of cancer types [2]. Now, computer-aided diagnosis becomes important research topic in medical diagnostics [3]. Despite of other diagnostic techniques, lymphocytes still has a foundation part in diagnosis and care of lymphatic circulatory disorders [4]. Lymphocytes plays a vital role in diagnosis of fatal diseases such as cancer, which makes it an important tool for medical practitioners for different diagnosis purposes.

Tissues, vessels and organs of a body make a network to form lymphatic system. These tissues, vessels and organs operate simultaneously to transfer fluid called lymph. Lymphatic system has a lot of responsibilities, such as safeguarding our body from germs, maintaining body fluid levels, absorbs fats from the digestive tract and eliminate cellular waste [5]. Pivotal duty of lymphatic system is to remove harmful materials from the body [6]. The lymphatic system has comprised of lymph, lymph nodes, lymphatic vessels and collecting

ducts. Lymphatic organs are present in different parts of body such as spleen, thymus, tonsils, appendix and bone marrow. Abnormalities of lymphatic system can cause cancer. Cancer cause by enlargement of lymphocytes is called lymphoma. In lymphoma cancer, lymphocytes grow from its original size.

Medical practitioners and researchers' focus is to propose and investigate automated solutions and improve accuracy of lymphography on different available datasets. In this research, four different machine learning classifiers, i.e., SVM, DL4J, MLP and K* are implemented with a dataset obtained from UCI machine learning database [7]. It consists of 148 instances and 18 attributes. On the same dataset, MLP achieves 98.64% accuracy, 98% recall and 98% precision which are the highest among all the four classifiers.

The paper is organized in six different sections. After the introduction to the problem in section I, section II presents the detailed literature review and existing methods. Proposed methods are discussed in section III. Moreover, section IV presents the experiments. The results comparison and analysis are discussed in section V. Finally the last section conclude the paper and provides future directions.

II. LITERATURE REVIEW

Different solutions have been proposed for automated assessment of lymphocytes using different methodologies. Al-mayyan et al. in [3] used random forest (RF) and particle swarm optimization (PSO) techniques to predict lymph diseases. Simple random sampling technique was used for features selection. Along with RF other classifiers, i.e., k-nearest neighbor (KNN), MLP and C4.5 were also investigated. With feature selection and random over sampling technique highest precision of 0.95 with RF, 0.94 with KNN, 0.93 with MLP and 0.88 with C4.5. In [8], Azar et al. assessed lymph diseases using lymphography database. In this work RF and genetic algorithm (GA) were used on lymphography database. Using

TABLE I: Lymphography Dataset

S.No	Attribute selection	Possible values of attributes	Assigned values
1	Lymphatic	Normal, arched, deformed, displaced	1-4
2	Block of afferent	No, Yes	1-2
3	Block of lymph c (superior and inferior flaps)	No, Yes	1-2
4	Block of lymph s (lazy incision)	No, Yes	1-2
5	By pass	No, Yes	1-2
6	Extravasates (force out of lymph)	No, Yes	1-2
7	Regeneration	No, Yes	1-2
8	Early uptake	No, Yes	1-2
9	Lymph nodes diminish	0-3	0-3
10	Lymph nodes enlarge	1-4	1-4
11	Changes in lymph	Bean, oval, round	1-3
12	Defect in node	No, lacunar, lacunar marginal, lacunar central	1-4
13	Changes in node	No, lacunar, lacunar marginal, lacunar central	1-4
14	Changes in structure	No, grainy, droplike, coarse, diluted, reticular, stripped, faint	1-8
15	Special forms	No, Chalices, vesicles	1-3
16	Dislocation	No, Yes	1-2
17	Exclusion of node	No, Yes	1-2
18	Number of nodes	0-80	1-8
19	Target Class	Normal, metastases, malign lymph, fibrosis	1-4

GA features were decreased from 18 to 6 and achieved accuracy of 92.2%. Madden et al. in [9] compared the performance of three classifiers Naive Bayes (NB), Tree augmented Bayes (TAN) and General Bayesian network (GBN) classifier. They used K2 search and hill-climbing search using lymphography database. Accuracies of 82.16% with NB, 81.07% with TAN, 77.46% were achieved.

Karabulut et al. in [10], evaluated fifteen datasets including lymphography database by applying three classifiers, NB, MLP and J48. Feature selection techniques used in this study were information gain (IF), gain ratio (GR), symmetrical uncertainty (SU), Relief-F and One-R. The best result achieved with MLP using chi-squared feature selection technique and gave accuracy of 84.46%. Abellan and Masegosa in [11], evaluated several datasets including lymphography dataset and proposed credal decision tree (CBT) using imprecise probabilities and uncertainty determine with pruning and without pruning. The suggested decision tree paradigm recorded an accuracy of 79.69% and 77.51% without pruning and with pruning respectively, on lymphography database.

Polat and Gunes [12], proposed a method based on adaptive neuro-fuzzy inference system (ANFIS) classifier. They used principal component analysis (PCA) technique on lymphography dataset and reduced features from 18 to 4. After feature selection, accuracy obtained on ANFIS classifier was 88.83%. Kuse et al. in [13], classified lymphocytes and non-lymphocytes using SVM by taking the digital colour images stained by hematoxyline and eosin (HE) as an identification input lymphocytes. The steps comprise segmentation, classifying, extraction and overlap resolution of the extracellular matrix cells. They obtained 58% sensitivity which means that 58% of the positives are correctly classified.

De Falco et al. in [14], used differential evolution (DE) method on lymphography dataset along with other seven medical databases. In their work, a new approach based on DE for the automated classification of items in medical databases was

TABLE II: Extracted Features

Feature selection technique	Extracted features	Selected features label
Chi-squared attribute	12	1,2,7,8,9,10,11,12,13,14,15,18

proposed which achieved 85.14% accuracy on lymphography database. Arora and Suman in [15], assessed lymphography database on MLP and J48. They obtained F1 score of 0.83 and 0.77 on MLP and j48, respectively. Gutierrez et al. in [16] proposed a two stage evolutionary algorithm and tested with 17 datasets including lymphography dataset and obtained 85.05% accuracy on lymphography database. McSherry et al. in [17] presented a medical and diagnostic approach to conversational case-based reasoning (CCBR). They introduced a CCB algorithm known as iNN (k), which aimed to confirm a target class by selecting features and informs them on a features' discrimination in favor of classification model. The method was tested on two databases, lymphography and SPECT heart obtained from UCI machine learning repository. Accuracies achieved were 86.5% and 84.3% on lymphography and SPECT heart dataset, respectively.

III. METHODOLOGY

In this section, the dataset used, pre-processing techniques and methods for the proposed approach are discussed.

A. Database

The lymphography database was acquired from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia [7]. There are 148 instances altogether and there are no missing attributes. There are 18 attributes and four classes, normal, metastases, malign lymph and fibrosis. Description of lymphography database is given in Table I.

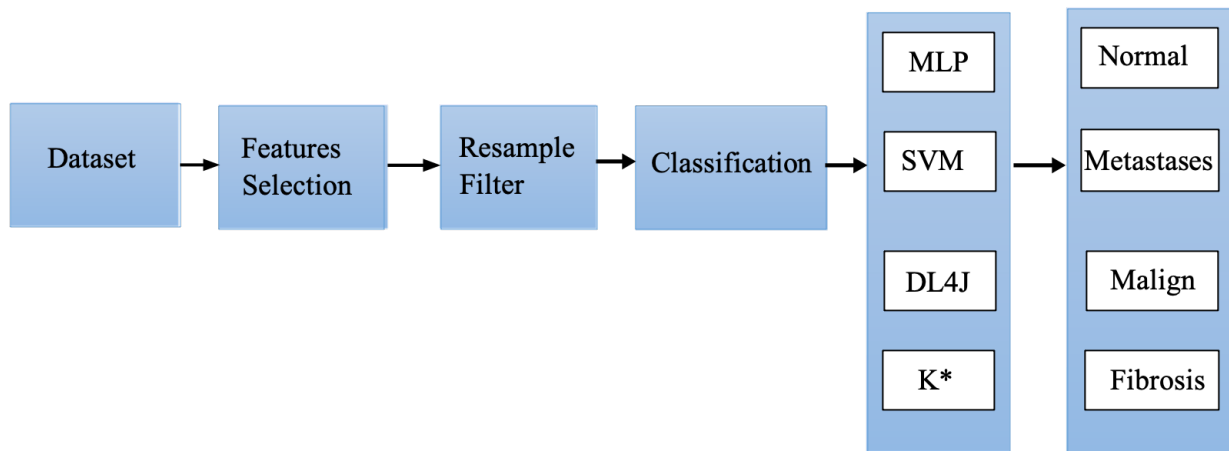


Fig. 1: Proposed Methodology

B. Proposed Approach

The proposed methodology for the assessment of lymphocytes in this research is given in Fig. 1. We carried out this research in two main phases. In first phase we applied SVM, DL4J, MLP and K* algorithms to lymphography database using Waikato environment for knowledge analysis (WEKA) tool. Moreover, for classification, we applied all these algorithms without pre-processing techniques. Second phase itself comprised of two steps. First we use chi-squared technique to extract important features given in Table II and then apply three times resample filter on lymphography dataset. In supervised instance selection filter (resample) the bias to uniform class (B) value is set at 1 while sample size percent value is set at 100. This step is applied before all the classifiers. In the second step we classify the dataset using SVM, MLP, DL4J and K* algorithms.

C. Pre-Processing

Pre-processing is an important step in classification problems. Without pre-processing step it is difficult to achieve an accurate model. For our problem we have used resample filter to balance the dataset. For resample technique we use resample filter provided by WEKA which produces a random subsample of a dataset using either sampling with replacement or without replacement and chi-squared feature selection technique for extracting key attributes.

1) *Resample Filter*: It is a supervised instance based filter. It generates a random sample of dataset using either sampling with replacement or without replacement. We set the bias to uniform class (B) value is set at “1” while sample size percent value is set at 100, which increase number of instances of minority class. We apply resample filter three times on dataset before classification.

2) *Chi-squared Feature Selection Technique*: Chi squared feature selection technique is a widely utilized technique used [18]. Chi-squared attribute evaluation, evaluate the quality of a feature by computing the value of the chi-squared statistic with respect to the class using the formula given

in Equation 1. Where, observed frequency is the number of observations of a class and expected frequency is the number of expected observations of the class if there was no relationship between a feature and target.

$$\chi^2 = \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \quad (1)$$

IV. EXPERIMENTAL SETUP

After the pre-processing step and features extraction phase, the extracted features are first fed to the SVM, MLP, DL4J and then K*. Different parameters for all the four classifiers are given in Table III.

A. SVM

It is a supervised machine learning classification algorithm used to identify classifies data into distinct marked classes. The aim of the SVM calculation is to discover a hyper plane in an N-dimensional space (N-number of features). Hyper planes are decision boundaries that help classify the data points. It uses different parameters for classification. For our experimentation we have used 10 folds with batch size = 100, poly kernel with random seed = 1.

B. MLP

MLP is a type of feed forward artificial neural network (ANN). It comprises of three layers of nodes: input layer, hidden layer output layer. MLP generate an efficient ANN training in combination along with optimization strategies such as gradient descent. For experimentation we have used 10 folds, batch size = 100, learning rate 0.3 with gradient decent.

C. DL4J

DL4J is the main DL programmable library written in Java. It is a model with wide help for profound learning algorithms. The DL4J classifier corroborates convolutional networks, fully connected feed forward networks and recurrent networks. It has the series of different neural networks layers to design sophisticated models such as: convolution layer, dense layer, sub sampling layer, batch normalization, output layer, and

TABLE III: Configuration of Algorithms

S.No	Classifier	Configuration Setting
1	SVM	10 folds batch size = 100, kernel: poly kernel, random seed = 1
2	MLP	10 folds, batch size = 100, learning rate 0.3 random seed = 1, training time=500
3	DL4J	10 folds, batch size = 100, epochs = 20 Random seed=1, optimization algorithm= stochastic gradient descent
4	K*	10 folds, batch size = 100, global blend=20

global pooling layer. For this study DL4J is used with 10 folds, batch size = 100, epochs = 20 random seed=1 and stochastic gradient descent as an optimization algorithm.

D. K*

K* is an instance based learner classifier. Instance based learner are those classifiers which try not to make model but make use of training samples for prediction. By doing comparison, unseen instance with examples of training set instance based learners classify an instance. K* utilizes an entropy-based distance function, founded on probability of conversion an instance into the other by arbitrarily picking between all possible conversions. We have used K* with 10 folds, batch size = 100 and a global blend=20.

V. RESULTS AND COMPARISONS

For performance evaluation and comparison five different evaluation metrics are calculated for all the four classifiers. Accuracy, precision, recall, F-measure and Matthews's correlation coefficient (MCC) are calculated using the Equations (2) to (6).

$$\text{Sensitivity/Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity/Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{F-measure} = 2 \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

$$\text{MCC} = \frac{TP \times TN - FN \times FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

The performance of SVM, MLP, DL4J and K* without pre-processing techniques can be seen in Table IV. The results are also depicted in Fig. 2. After balancing the number of instances by applying resample filter and reduced the features from 18 to 12 by chi-squared technique, a significant improvement can be seen, as shown in Table V and depicted in Fig. 3. We can see from the findings of Table V where MLP achieves 98.64%

TABLE IV: Result before pre-processing

Algorithm	Accuracy	Precision	Recall	F-measure	MCC
SVM	86.48%	0.86	0.86	0.86	0.74
MLP	84.45%	0.84	0.84	0.83	0.70
DL4J	83.10%	0.84	0.83	0.83	0.69
K*	85.13%	0.85	0.85	0.84	0.71

TABLE V: Result after pre-processing

Algorithm	Accuracy	Precision	Recall	F-measure	MCC
SVM	97.97%	0.98	0.98	0.98	0.96
MLP	98.64%	0.98	0.98	0.98	0.98
DL4J	96.62%	0.96	0.96	0.96	0.95
K*	97.29%	0.97	0.97	0.97	0.96

TABLE VI: Comparison with other related work

Citation	Classification technique	Pre-processing techniques	Accuracy	Precision
Almayyan [3]	RF	IGR for features selection		0.95
Karabulut et al. [10]	MLP	chi-squared	84.46%	
Proposed	MLP, SVM, DL4J	resampling+ chi-squared	97.97%, 98.64%, 96.62%	0.98, 0.98, 0.96,
	K*		97.29%	0.97

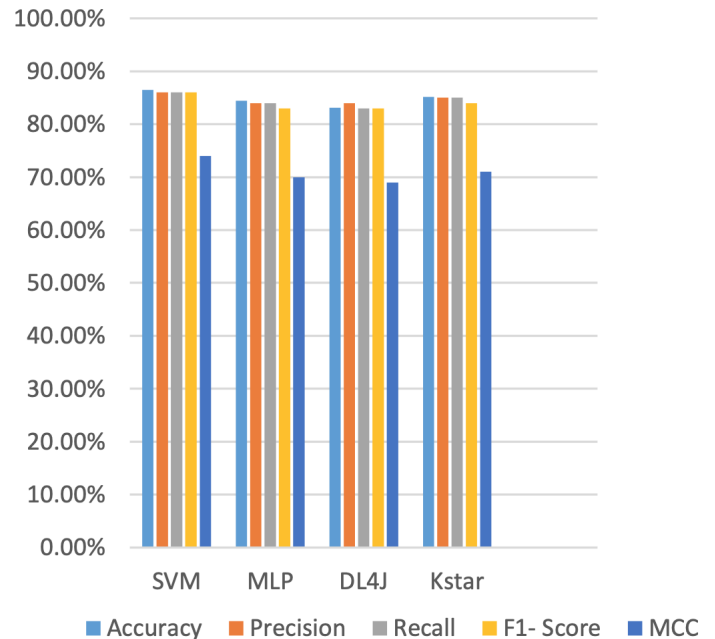


Fig. 2: Classifiers response without data pre-processing

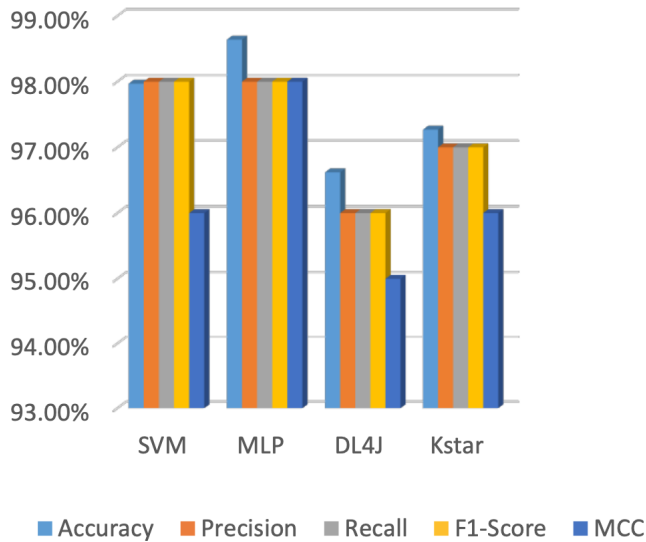


Fig. 3: Classifiers response with data pre-processing

accuracy. The second highest accuracy is obtained by SVM which is 97.97%. After this, K* and DL4J likewise accomplished accuracy of 97.29% and 96.62% respectively. After all, the performance of the proposed method is contrasted with other relevant works on the same dataset as shown in Table VI. We trained the classifiers with the lymphography dataset using chi-squared attribute feature selection technique and balanced the dataset using oversampling technique. The results are compared with Almayyan [3] and Karabulut et al. [10]. Our results are substantially improved in terms of accuracy and precision.

VI. CONCLUSIONS

The advancement of automated solutions and emergence of machine learning (ML) related tools, effect almost all the domain of research and development, medical field is one of the targeted domain. Decision-making requires a lot of expertise where automated solutions may help in the decision making process. Medical data can be assessed in shorter time and in more detail by classification systems used in diagnostics. In the research study a novel medical decision making system is introduced by using four ML algorithms and two pre-processing techniques, resample filter and chi-squared features technique for assessment of lymphocytes. Evaluation were done on the lymphography dataset consist of four classes to be classified. Our proposed model gives propitious results and comparatively better classification accuracy, i.e., 98.64%. The advantage of this technique is to support the medical practitioners to make the final determination with less possible doubt.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of Artificial Intelligence in Healthcare, Intelligent Information Processing Lab,

National Center of Artificial Intelligence, UET Peshawar for the necessary research infrastructure and support we are also grateful of NVIDIA Corporation for supporting this research by providing Titan X Pascal high computing facility

REFERENCES

- [1] Benjamin H Kann, Reid Thompson, Charles R Thomas Jr, Adam Dicker, and Sanjay Aneja. Artificial intelligence in oncology: current applications and future directions. *Oncology*, 33(2), 2019.
- [2] Shahira Abousamra, Le Hou, Rajarsi Gupta, Chao Chen, Dimitris Samaras, Tahsin Kurc, Rebecca Batiste, Tianhao Zhao, Shroyer Kenneth, and Joel Saltz. Learning from thresholds: fully automated classification of tumor infiltrating lymphocytes for multiple cancer types. *arXiv preprint arXiv:1907.03960*, 2019.
- [3] Waheeda Almayyan et al. Lymph diseases prediction using random forest and particle swarm optimization. *Journal of Intelligent Learning Systems and Applications*, 8(03):51, 2016.
- [4] Ali Guermazi, Pauline Brice, Christophe Hennequin, and Emile Sarfati. Lymphography: an old technique retains its usefulness. *Radiographics*, 23(6):1541–1558, 2003.
- [5] Melody A Swartz. The physiology of the lymphatic system. *Advanced drug delivery reviews*, 50(1-2):3–20, 2001.
- [6] Gwendalyn J Randolph, Stoyan Ivanov, Bernd H Zinselmeyer, and Joshua P Scallan. The lymphatic system: integral roles in immunity. *Annual review of immunology*, 35:31–52, 2017.
- [7] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [8] Ahmad Taher Azar, Hanaa Ismail Elshazly, Aboul Ella Hassanien, and Abeer Mohamed Elkorany. A random forest classifier for lymph diseases. *Computer methods and programs in biomedicine*, 113(2):465–473, 2014.
- [9] Michael G Madden. On the classification performance of tan and general bayesian networks. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 3–16. Springer, 2008.
- [10] Esra Mahsereci Karabulut, Selma Ayse Ozel, and Turgay Ibricki. A comparative study on the effect of feature selection on classification accuracy. *Procedia Technology*, 1:323–327, 2012.
- [11] Joaquin Abellan and Andres R Masegosa. Bagging schemes on the presence of class noise in classification. *Expert Systems with Applications*, 39(8):6827–6837, 2012.
- [12] Kemal Polat and Salih Güneş. A novel hybrid intelligent method based on c4. 5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, 36(2):1587–1592, 2009.
- [13] Manohar Kuse, Tanuj Sharma, and Sudhir Gupta. A classification scheme for lymphocyte segmentation in h&e stained histology images. In *International Conference on Pattern Recognition*, pages 235–243. Springer, 2010.
- [14] Ivanoe De Falco. Differential evolution for automatic rule extraction from medical databases. *Applied soft computing*, 13(2):1265–1283, 2013.
- [15] Rohit Arora. Comparative analysis of classification algorithms on different datasets using weka. *International Journal of Computer Applications*, 54(13), 2012.
- [16] Juan Carlos Fernandez Caballero, Francisco Jose Martinez, Cesar Hervas, and Pedro Antonio Gutierrez. Sensitivity versus accuracy in multiclass problems using memetic pareto evolutionary neural networks. *IEEE Transactions on Neural Networks*, 21(5):750–770, 2010.
- [17] David McSherry. Conversational case-based reasoning in medical decision making. *Artificial Intelligence in Medicine*, 52(2):59–66, 2011.
- [18] Huan Liu and Rudy Setiono. Chi2: Feature selection and discretization of numeric attributes. In *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, pages 388–391. IEEE, 1995.