hand will not only reduce the complexity of the problem and improve the efficiency of the processing but also simplify significantly the design of the classifier. The FS is one of the essential and frequently used techniques in machine learning. A FS method generates different candidates from the feature space and assesses them based on an evaluation criterion to find the best feature subset [5]. On the basis of the evaluation criterion, FS can be divided into filter methods and wrapper methods. Filters assess the relevance of features by looking only at the intrinsic properties of the data, such as distance, consistency, and correlation [5–7]. These criteria are independent of any inductive learning algorithm. In contrast, the wrapper approach requires one predetermined mining algorithm and uses its performance to evaluate and determine which features are selected [8]. Wrappers often select features that have a higher accuracy; however, they are criticized for their high computational cost and low generality. To take advantage of the above two approaches, a hybrid model was proposed to handle large data sets [9]. Moreover, some methods, known as embedded, use internal information of the classification model to perform FS [10,11].

Based on the generation procedure, FS can be divided into individual feature ranking (FR) and feature subset selection (FSS) [10,12]. FR measures the relevance of each feature to the class and then ranks features by their scores and selects the top-ranked features. These methods are widely used because of their simplicity, scalability, and good empirical success [10,13]. However, FR is criticized because it can capture only the relevance of the features to the target concept, whereas the redundancy and basic interactions between features are not discovered. Additionally, the number of features retained is difficult to determine; as a result, a threshold is required. In contrast, FSS attempts to find a set of features that have good performance. This method integrates the metric for measuring the feature–class relevance and the feature–feature interactions. In [14] Liu and Yu, a large number of selection methods are categorized, in which different algorithms address these issues distinctively. We found different search strategies, namely exhaustive, heuristic and random searches, and combined them with several types of measures to form different algorithms. The time complexity is exponential in terms of the data dimensionality for an exhaustive search, and it is quadratic for a heuristic search. The complexity can be linear with the number of iterations in a random search, but experiments show that, to find the best feature subset, the number of iterations required is usually at least quadratic to the number of features [15]. In this categorization, to handle large data sets, a hybrid model was also proposed to combine the advantages of the FR and FSS techniques. These methods decouple relevance analysis and redundancy analysis, and they have been proven to be more effective than ranking methods and more efficient than subset evaluation methods on many traditional high-dimensional data sets. In this framework, [16] proposed a hybrid search algorithm. Yu and Liu [17] proposed a fast correlation-based filter algorithm (FCBF) that used a correlation measure to obtain relevant features and to remove redundancy. Ding and Peng [18] used mutual information for gene selection, finding maximum relevance with minimal redundancy by solving a simple two-objective optimization.

2.2. Classification with evolutionary product unit neural networks based on a two-stage algorithm

There are several kinds of neural networks, being the single-hidden-layer feed-forward network architecture the most popular one. Multiplicative neural networks contain nodes that multiply their inputs instead of adding them. This class of neural networks comprises such types as sigma-pi networks and product

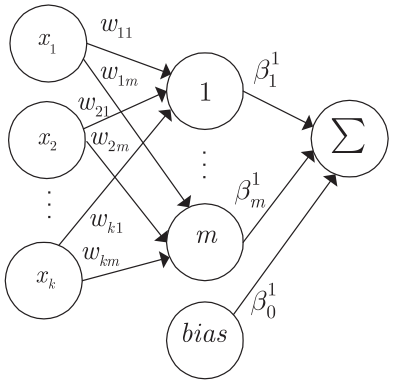

Fig. 1. Structure of a product unit neural network model for a bi-classification problem.

unit networks. The latter type was introduced by R. Durbin and D. Rumelhart [19]. The methodology employed here consists of the use of an EA as a tool for learning the architecture and weights of a PUNN model [20]. More details about PUNNs, such as some of the advantages, the universal approximation theorem, problems and learning methods, can be found in [4,21].

Fig. 1 shows the structure of a PUNN model with a *k:m:*1 architecture for a bi-classification problem; this is a three-layer architecture, that is, k nodes in the input layer, m ones (product units) and a bias one in the hidden layer and one node in the output layer.

The transfer function of each node in the hidden and output layers is the identity function. Thus, the functional model obtained by each of the nodes in the output layer with J classes is given by:

$$f(x_1,x_2,\ldots,x_k)=\beta_0^l+\sum_{j=1}^{m}\beta_j^l\left(\prod_{i=1}^{k}x_i^{w_{ij}}\right)\ l=1,2,\ldots,J;\ w_{ij}\in\Re \tag{1}$$

Next, we are going to describe briefly the TSEA applied. A full explanation of it and the details about common parameters can be read in Section 3 of [4]. TSEA is used to design the structure and learn the weights of PUNNs in two sequential phases. The population is subjected to the operations of replication and mutation; two types of mutations have been applied: parametric and structural ones. The TSEA pseudo-code for a classification problem appears in Fig. 2. In the first stage, TSEA evolves two populations for a small number of generations. The best half individuals of each one are merged in a new population that follows the full evolutionary cycle. The main parameters of the TSEA are the maximum number of generations (gen) and the maximum number of nodes in the hidden layer (neu). The minimum number of nodes is an unit lower than neu. The remaining parameters will be described further on. At the end of the TSEA, it returns the best PUNN model with a number of nodes between *neu* and *neu+1* in the hidden layer.

We have considered a standard soft-max activation function, associated with the g network model, given by:

$$g_j(\mathbf{x})=\frac{\exp f_j(\mathbf{x})}{\sum_{j=1}^{J}\exp f_j(\mathbf{x})}\qquad j=1,\ldots,J \tag{2}$$

where J is the number of classes in the problem, $f_j(\mathbf{x})$ is the output of node j for pattern $\mathbf{x}$ and $g_j(\mathbf{x})$ is the probability that this pattern belongs to class j.

Given a training set $D=(\boldsymbol{x}_i,\boldsymbol{y}_i)i=1,\ldots,N$, a function of cross-entropy error is used to evaluate a network g with the instances of a problem

Table 5
Results obtained in 18 data sets applying TSEA and TSEAFS.

Data set	Method	Topology	Mean ± SD	
			Config 1*/1*#	Config 2*/2*#
Appendicitis	FS0	7:[4,5]:1	**81.66 ± 4.24**	80.51 ± 2.84
	FS1	4:[4,5]:1	**82.82 ± 2.19**	81.02 ± 2.65
	FS2	2:[4,5]:1	80.89 ± 0.70	80.76 ± 0.00
	FS3	2:[4,5]:1	80.38 ± 2.73	79.61 ± 3.21
	FS4	5:[4,5]:1	81.79 ± 2.00	81.66 ± 2.97
Breast	FS0	15:[9,10]:1	**65.96 ± 2.89**	62.76 ± 3.08
	FS1	4:[9,10]:1	**69.85 ± 1.50**	68.21 ± 1.08
	FS2	2:[9,10]:1	69.01 ± 0.00	69.01 ± 0.00
	FS3	3:[9,10]:1	68.92 ± 0.73	69.10 ± 0.36
	FS4	4:[9,10]:1	69.01 ± 0.00	69.01 ± 0.00
Breast-t	FS0	9:[5,6]:5	54.53 ± 7.89	**55.33 ± 9.16**
	FS1	6:[5,6]:5	54.40 ± 6.77	48.93 ± 8.83
	FS2	6:[5,6]:5	55.73 ± 8.72	57.87 ± 6.87
	FS3	4:[5,6]:5	**60.93 ± 4.77**	56.53 ± 7.03
	FS4	6:[5,6]:5	59.47 ± 8.90	56.00 ± 6.96
Cardiotocography	FS0	31:[6,7]:2	**81.69 ± 3.56**	81.55 ± 2.90
	FS1	9:[5,6]:2	**85.26 ± 2.27**	84.88 ± 2.11
	FS2	21:[5,6]:2	71.20 ± 2.55	76.71 ± 1.04
	FS3	8:[5,6]:2	81.55 ± 1.69	81.12 ± 1.80
	FS4	7:[5,6]:2	81.58 ± 2.48	82.38 ± 2.42
Heart	FS0	13:[6,7]:1	76.62 ± 2.33	**77.45 ± 3.09**
	FS1	7:[4,5]:1	77.45 ± 2.16	77.69 ± 2.28
	FS2	9:[4,5]:1	**78.57 ± 1.99**	77.79 ± 1.60
	FS3	7:[4,5]:1	75.24 ± 2.70	75.34 ± 2.80
	FS4	7:[4,5]:1	77.45 ± 2.16	77.69 ± 2.28
Hepatitis	FS0	19:[3,4]:1	82.10 ± 4.44	**87.01 ± 3.78**
	FS1	10: [3,4]:1	90.78 ± 1.79	89.29 ± 1.53
	FS2	5: [3,4]:1	86.14 ± 1.81	87.45 ± 1.49
	FS3	6: [3,4]:1	85.00 ± 1.56	**91.05 ± 2.55**
	FS4	10: [3,4]:1	90.78 ± 1.79	89.29 ± 1.53
Labor	FS0	29:[6,7]:1	85.24 ± 8.78	**86.90 ± 5.96**
	FS1	7:[5,6]:1	93.09 ± 4.39	**96.19 ± 4.08**
	FS2	5:[5,6]:1	87.62 ± 4.16	88.33 ± 4.39
	FS3	8:[5,6]:1	90.95 ± 5.28	90.48 ± 5.73
	FS4	8:[5,6]:1	89.76 ± 6.41	89.76 ± 5.20
Led24	FS0	24:[8,9]:9	50.29 ± 6.59	**51.03 ± 5.58**
	FS1	6:[8,9]:9	67.26 ± 1.46	**68.30 ± 0.57**
	FS2	6:[8,9]:9	67.26 ± 1.46	**68.30 ± 0.57**
	FS3	6:[8,9]:9	67.26 ± 1.46	**68.30 ± 0.57**
	FS4	6:[8,9]:9	67.26 ± 1.46	**68.30 ± 0.57**
Lymphography	FS0	38:[6,7]:3	**79.37 ± 4.73**	78.73 ± 4.79
	FS1	11:[6,7]:3	79.09 ± 5.71	78.55 ± 4.42
	FS2	9:[6,7]:3	80.18 ± 3.27	80.36 ± 4.54
	FS3	8:[6,7]:3	79.18 ± 5.17	80.61 ± 3.12
	FS4	12:[6,7]:3	78.19 ± 3.88	**80.90 ± 5.71**
Parkinsons	FS0	22:[6,7]:1	73.94 ± 2.43	**78.09 ± 3.51**
	FS1	5:[6,7]:1	78.36 ± 2.86	78.77 ± 1.66
	FS2	6:[6,7]:1	80.13 ± 2.26	80.06 ± 3.73
	FS3	4:[6,7]:1	82.52 ± 2.92	**82.79 ± 2.50**
	FS4	6:[6,7]:1	79.25 ± 2.15	76.05 ± 3.47
Pima	FS0	8:[4,5]:1	78.38 ± 1.59	**79.21 ± 1.53**
	FS1	3:[4,5]:1	79.35 ± 1.09	**79.72 ± 1.08**
	FS2	5:[4,5]:1	78.52 ± 0.80	78.54 ± 1.37
	FS3	4:[4,5]:1	78.42 ± 1.35	79.53 ± 0.98
	FS4	4:[4,5]:1	78.42 ± 1.35	79.53 ± 0.98
Plates	FS0	27:[6,7]:6	50.74 ± 4.24	**51.46 ± 3.03**
	FS1	16:[6,7]:6	53.81 ± 3.99	53.38 ± 4.17
	FS2	21:[6,7]:6	**56.93 ± 2.43**	54.40 ± 4.95
	FS3	6:[6,7]:6	50.87 ± 4.75	51.87 ± 3.06
	FS4	10:[6,7]:6	48.84 ± 4.60	48.53 ± 2.58
Promoter	FS0	114:[11,12]:1	65.76 ± 8.99	**68.20 ± 9.52**
	FS1	7:[6,7]:1	83.84 ± 3.83	**85.64 ± 4.03**
	FS2	7:[6,7]:1	80.00 ± 2.74	76.30 ± 4.10
	FS3	11:[6,7]:1	73.66 ± 6.77	75.12 ± 4.48
	FS4	10:[6,7]:1	74.74 ± 5.11	73.97 ± 3.73
SPECTF	FS0	44:[6,7]:1	60.17 ± 4.15	**61.56 ± 4.97**
	FS1	12:[6,7]:1	73.20 ± 2.18	73.85 ± 2.71
	FS2	9:[6,7]:1	72.07 ± 1.16	71.64 ± 1.56
	FS3	6:[6,7]:1	**73.99 ± 1.30**	70.60 ± 1.84
	FS4	12:[6,7]:1	72.35 ± 1.69	73.76 ± 1.02
Vowel	FS0	11:[6,7]:10	45.04 ± 2.93	**47.18 ± 4.03**
	FS1	3:[6,7]:10	48.07 ± 3.11	**54.31 ± 2.29**
	FS2	9:[6,7]:10	47.65 ± 5.01	46.80 ± 4.27

Table 7
Results obtained in 18 data sets for several classifiers with and without feature selection.

Data set	Method	C4.5	1-NN	SVM	PART	MLP	RBF	TSEAFS
Appendicitis	FS0	73.08	69.23	84.62	73.08	76.92	74.67	81.66
	FS1	80.77	69.23	76.92	80.77	78.85	80.00	82.82
	FS2	76.92	57.69	76.92	76.92	77.95	77.05	80.89
	FS3	80.77	80.77	80.77	80.77	80.77	74.49	80.38
	FS4	80.77	65.38	76.92	80.77	79.23	79.36	81.79
Breast	FS0	70.42	64.79	64.79	69.01	60.80	68.78	65.96
	FS1	69.01	70.42	66.20	71.83	69.01	67.46	69.85
	FS2	69.01	70.42	64.79	69.01	69.01	69.01	69.01
	FS3	69.01	70.42	64.79	69.01	69.53	67.65	69.10
	FS4	69.01	70.42	66.20	71.83	69.01	67.46	69.01
Breast-t	FS0	52.00	60.00	52.00	44.00	63.20	61.20	55.33
	FS1	56.00	52.00	60.00	44.00	65.33	58.67	54.40
	FS2	52.00	52.00	64.00	52.00	67.20	61.20	57.87
	FS3	48.00	48.00	56.00	48.00	65.60	60.40	60.93
	FS4	68.00	56.00	60.00	56.00	65.47	60.67	59.47
Cardiotocography	FS0	82.71	76.32	83.65	82.52	80.75	81.80	81.69
	FS1	77.07	81.77	81.20	82.52	81.94	83.40	85.26
	FS2	75.19	63.91	75.19	75.00	68.29	65.91	76.71
	FS3	77.82	81.20	81.20	77.26	80.13	80.50	81.55
	FS4	78.38	80.45	81.39	81.20	80.86	84.12	82.38
Heart	FS0	70.59	73.53	76.47	73.53	74.85	78.53	77.45
	FS1	73.53	73.53	76.47	77.94	72.50	78.24	77.69
	FS2	72.06	75.00	76.47	75.00	74.85	77.60	78.57
	FS3	73.53	70.59	77.94	75.00	74.90	76.37	75.34
	FS4	73.53	73.53	76.47	77.94	72.50	78.53	77.69
Hepatitis	FS0	84.21	86.84	89.47	81.58	84.73	89.30	87.01
	FS1	84.21	89.47	86.84	84.21	87.28	89.30	90.78
	FS2	89.47	84.21	89.47	84.21	84.21	88.42	87.45
	FS3	89.47	84.21	89.47	86.84	87.72	90.79	91.05
	FS4	84.21	89.47	86.84	84.21	87.28	89.30	90.78
Labor	FS0	85.71	71.43	78.57	85.71	69.52	71.67	86.90
	FS1	85.71	71.43	78.57	85.71	64.29	71.43	96.19
	FS2	85.71	64.28	78.57	78.57	78.57	64.29	88.33
	FS3	85.71	78.57	71.43	78.57	71.43	64.29	90.95
	FS4	85.71	64.29	71.43	85.71	57.62	71.43	89.76
Led24	FS0	65.67	39.43	58.97	55.80	57.48	55.14	51.03
	FS1	68.10	67.90	67.93	68.50	68.44	67.42	68.30
	FS2	68.10	67.90	67.93	68.50	68.44	67.42	68.30
	FS3	68.10	67.90	67.93	68.50	68.44	67.42	68.30
	FS4	68.10	67.90	67.93	68.50	68.44	67.42	68.30
Lymphography	FS0	75.68	83.78	91.89	75.68	86.58	70.99	79.37
	FS1	88.29	78.38	83.78	70.27	73.24	68.92	79.09
	FS2	75.68	70.27	78.38	64.86	71.89	75.77	80.36
	FS3	81.08	75.68	81.08	70.27	74.50	69.64	80.61
	FS4	81.08	81.08	81.08	64.86	80.45	69.16	80.90
Parkinsons	FS0	71.43	77.55	75.51	75.51	77.62	70.27	78.09
	FS1	75.51	79.59	75.51	77.55	81.56	77.75	78.77
	FS2	79.59	79.59	75.51	81.63	75.65	73.47	80.13
	FS3	81.63	73.47	79.59	77.55	84.83	80.27	82.79
	FS4	73.47	81.63	75.51	79.59	83.13	77.55	79.25
Pima	FS0	74.48	73.96	78.13	74.48	75.94	77.34	79.21
	FS1	76.04	74.48	77.60	76.04	78.18	79.17	79.72
	FS2	74.48	67.19	78.65	74.48	76.89	75.64	78.54
	FS3	76.04	67.71	79.17	76.04	79.01	80.28	79.53
	FS4	76.04	67.71	79.17	76.04	78.73	80.28	79.53
Plates	FS0	39.05	49.17	57.02	46.69	53.50	59.94	51.46
	FS1	40.50	51.24	51.03	46.90	56.71	64.08	53.81
	FS2	38.22	50.62	55.17	44.63	55.24	62.17	56.93
	FS3	44.63	43.18	45.04	49.79	52.85	55.88	51.87
	FS4	54.75	47.31	51.65	51.65	57.33	59.88	48.84
Promoter	FS0	69.23	65.38	88.46	53.85	86.03	79.36	68.20
	FS1	73.08	57.69	84.62	80.77	84.49	83.46	85.64
	FS2	80.77	57.69	84.62	76.92	75.64	85.00	80.00
	FS3	73.08	76.92	73.08	80.77	78.21	79.74	75.12
	FS4	73.08	69.23	73.08	80.77	76.28	80.00	74.74
SPECTF	FS0	67.91	61.50	72.19	70.59	71.28	76.19	61.56
	FS1	66.84	59.36	72.19	72.19	73.67	76.24	73.85
	FS2	65.78	60.96	70.05	65.78	70.02	74.60	72.07
	FS3	67.91	59.36	65.24	64.71	69.57	74.58	73.99
	FS4	66.84	57.75	73.26	70.05	72.26	74.63	73.76
Vowel	FS0	39.39	48.48	45.45	38.53	45.87	47.25	47.18
	FS1	45.24	46.54	54.33	44.59	52.79	43.12	54.31
	FS2	38.53	51.52	48.48	40.04	52.05	44.73	47.65
	FS3	41.56	46.97	41.34	36.58	44.97	46.95	49.45
	FS4	45.24	46.54	54.33	44.59	52.79	43.12	54.31