RESEARCH

Facial biotype classification for orthodontic treatment planning using an alternative learning algorithm for tree augmented Naive Bayes

Gonzalo A. Ruz[1,2,3*] , Pamela Araya-Díaz[4] and Pablo A. Henríquez[5]

Abstract

Background: When designing a treatment in orthodontics, especially for children and teenagers, it is crucial to be aware of the changes that occur throughout facial growth because the rate and direction of growth can greatly affect the necessity of using different treatment mechanics. This paper presents a Bayesian network approach for facial biotype classification to classify patients' biotypes into Dolichofacial (long and narrow face), Brachyfacial (short and wide face), and an intermediate kind called Mesofacial, we develop a novel learning technique for tree augmented Naive Bayes (TAN) for this purpose.

Results: The proposed method, on average, outperformed all the other models based on accuracy, precision, recall, F_1-score, and kappa, for the particular dataset analyzed. Moreover, the proposed method presented the lowest dispersion, making this model more stable and robust against different runs.

Conclusions: The proposed method obtained high accuracy values compared to other competitive classifiers. When analyzing a resulting Bayesian network, many of the interactions shown in the network had an orthodontic interpretation. For orthodontists, the Bayesian network classifier can be a helpful decision-making tool.

Keywords: Bayesian networks, Tree augmented Naive Bayes, Evolution strategy, Facial biotypes, Orthodontic treatment planning

Background

In recent years, there has been a rise in the use of machine learning-based tools in medical treatments to aid in decision-making for treatment planning. In particular, the output of these models can be used as a support tool for health personnel who ultimately make decisions. Given the implications for patients on these decisions, the machine learning technique used should be interpretable. An interesting machine learning technique

for this purpose is Bayesian networks (BN) [1], which combines graph theory with probability theory.

In the field of dentistry BN have been applied in diverse areas. For example, in [2] prior to and during the application of a certain orthodontic procedure, BN were employed to describe certain tooth color parameters. To better understand the underlying data structure of the patterns of dental caries in the population, the prevalence of dental caries in the primary dentition of 352 Myanmar schoolchildren was examined at the tooth level using BN in [3]. The effectiveness of BN in the assessment of dental age-related evidence obtained using a geometrical approximation approach of the pulp chamber volume

*Correspondence: gonzalo.ruz@uai.cl

[1] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile
Full list of author information is available at the end of the article

Ruz *et al. BMC Medical Informatics and Decision Making* (2022) 22:316

Page 2 of 10

was examined in [4]. BN are used in [5] for age estimation and classification based on dental evidence, in particular, to the development of third molars. A BN clinical decision support system was designed in [6] to assist general practitioners in determining whether patients with permanent dentition need orthodontic treatment. A Dental Caries Clinical Decision Support System is evaluated in [7] which uses a BN to provide suggestions and represent clinical patterns. The outcomes demonstrated the Bayesian network's accuracy in various cases. In [8], a minimally invasive method for elevating the lateral maxillary sinus was described, and BN was used to determine the link between the parameters involved. The use of BN to MR images to identify temporomandibular disorders was looked at in [9]. The goal was to ascertain how temporomandibular disorders were diagnosed, concentrating on how each discovery affected the other. The findings demonstrated that the BN path condition method was more than 99% accurate when employing resubstitution validation and 10-fold cross-validation. The key benefit of utilizing BN, however, is its ability to express the causal links between various data and assign conditional probabilities, which might subsequently be utilized to interpret the course of temporomandibular disorders. In [10], BN are used to identify and depict the relationships between several Class III malocclusion maxillofacial features during growth and treatment. The authors demonstrate that as compared to individuals undergoing orthodontic treatment with rapid maxillary expansion and facemask therapy, untreated participants exhibit different Class III craniofacial growth patterns. Also, it is important to point out that BN have been used for meta-analysis in several dental research topics [11–16].

BN are probabilistic graphical models representing discrete random variables and conditional dependencies via a directed acyclic graph (DAG). In classification (supervised learning) problems, when using a probabilistic approach, the difficulty is to compute effectively the posterior probability of the class variable Y_k (with $k = 1, \ldots, K$) given an n-dimensional input data point $\mathbf{x} = (x_1, \ldots, x_n)$. This can be carried out using the Bayes rule:

$$p(Y_k|\mathbf{x}) = \frac{p(Y_k)p(\mathbf{x}|Y_k)}{p(\mathbf{x})}. \tag{1}$$

The numerator, which comprises the a priori probability of the class variable and the *likelihood* (the joint probability of the input features conditioned to the class variable), is what is important in this case. The calculation of the class variable's a priori probability is simple. It can be determined from the training set's class variable values' relative frequency. However, there are numerous methods for calculating *likelihood*. The usage of Bayesian
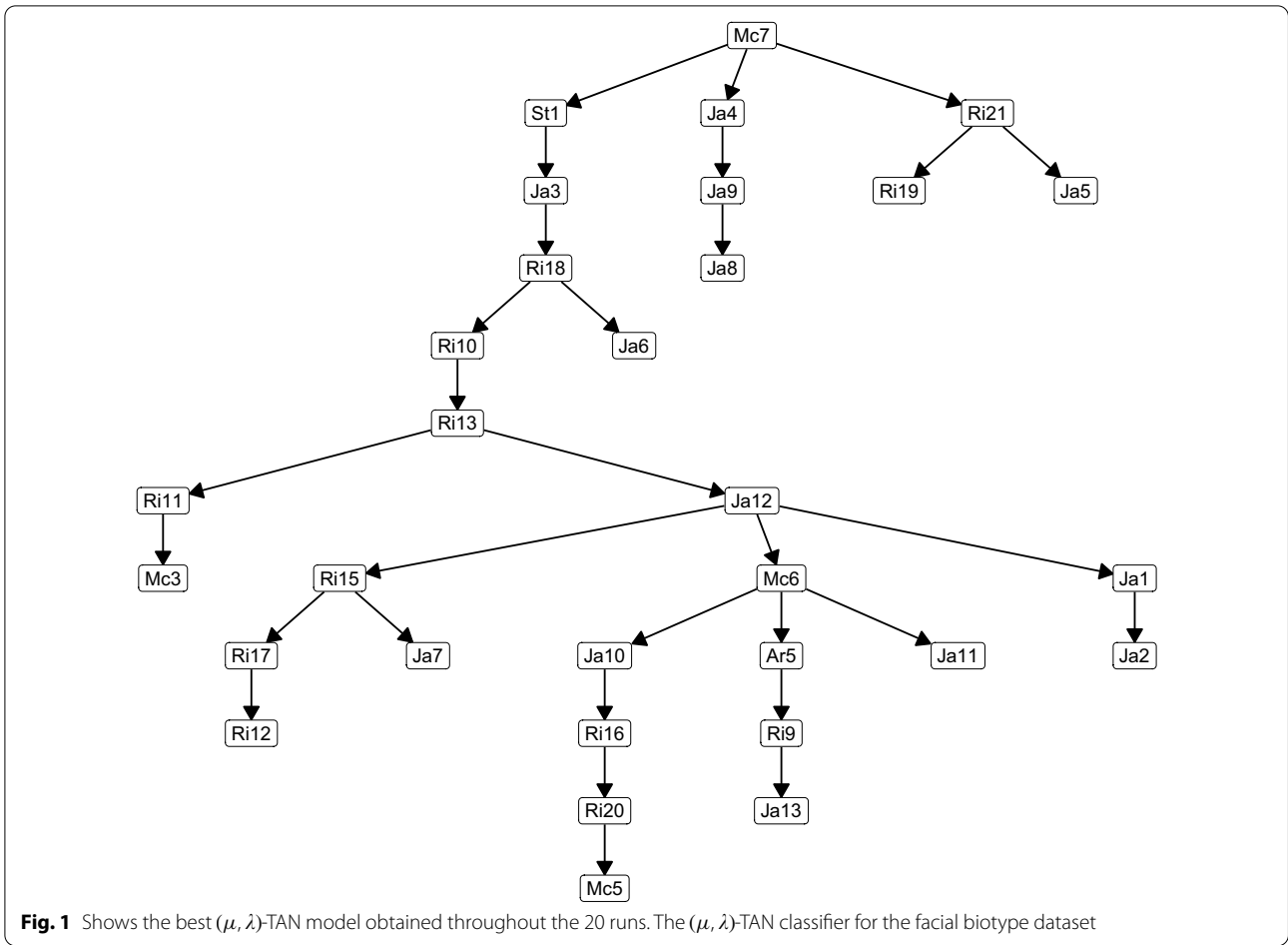
networks, thus, Bayesian network classifiers [17], is one of them.

There are various Bayesian network classifiers [18–23]. However, the two most often used are the tree augmented Naive Bayes (TAN) classifier [17] and the Naive Bayesian network classifier, also known as the Naive Bayes [24]. The Naive Bayes approach computes the likelihood in (1) by assuming conditional independence among the attributes given the class variable. There are no edges between the attributes as a result. As opposed to TAN, which begins by taking into account a fully connected network with weighted edges, it uses the *conditional mutual information* between pairs of attributes to generate these weights. Then, the application of Kruskal's algorithm (the maximum weighted spanning tree (MWST)) to produce a tree structure is carried out, leaving just $n - 1$ edges. Each attribute in this version of the Bayesian network classifier will have an incoming edge from another attribute, with the exception of the selected root attribute node.

The TAN model corrects the naive version's strong assumption of conditional independence. Theoretically, it ought to deliver better outcomes (accuracy) than the Naive Bayes. However, TAN has significant drawbacks, one of which is its difficulty to estimate the conditional mutual information accurately. Two direct difficulties when working with conditional mutual information are: (1) the computational complexity for n nodes and N training samples is $\mathcal{O}(n^2N)$ [25], therefore, for datasets with many attributes the computation becomes very slow, needing more computational power, (2) the conditional mutual information estimate produced when there are not enough training instances in each class to accurately estimate the joint probability distribution and the conditional distributions. This is significant because conditional mutual information is used as weights in the fully connected graph throughout TAN's tree construction technique. The obvious question is: Can the network weights be learned from the data to achieve satisfactory classification results without estimating conditional mutual information?

When preparing a treatment in orthodontics, especially for children and teenagers, it is crucial to be aware of the changes that take place throughout facial growth because the rate and direction of growth can greatly affect the necessity of using different treatment mechanics. The Ricketts' VERT index is one of the most widely used methods for identifying facial biotypes [26]. The biotypes can be divided into Dolichofacial (long and narrow face), Brachyfacial (short and wide face), and an intermediate form known as Mesofacial based on the VERT index.

In this paper, we propose a different approach for learning TAN classifiers without estimating conditional

Fig. 1 Shows the best (μ, λ)-TAN model obtained throughout the 20 runs. The (μ, λ)-TAN classifier for the facial biotype dataset

Table 1 Performance measures for each model

Algorithm	Accuracy Avg. $\pm$ SD.	Precision Avg. $\pm$ SD.	Recall Avg. $\pm$ SD.	F_1-score Avg. $\pm$ SD.	Kappa Avg. $\pm$ SD.
NB	70.27 $\pm$ 5.21	70.30 $\pm$ 5.92	74.09 $\pm$ 6.66	70.72 $\pm$ 5.55	0.54 $\pm$ 0.11
TAN	71.01 $\pm$ 4.19	70.29 $\pm$ 5.84	74.21 $\pm$ 4.52	70.81 $\pm$ 5.39	0.56 $\pm$ 0.11
SVM	70.63 $\pm$ 4.43	70.31 $\pm$ 5.22	73.68 $\pm$ 5.26	70.55 $\pm$ 5.31	0.55 $\pm$ 0.08
DT	69.27 $\pm$ 7.19	69.93 $\pm$ 5.02	73.16 $\pm$ 4.14	70.72 $\pm$ 4.82	0.52 $\pm$ 0.10
RF	69.07 $\pm$ 4.93	67.70 $\pm$ 4.78	71.08 $\pm$ 7.41	67.30 $\pm$ 5.78	0.51 $\pm$ 0.07
RVFL	70.11 $\pm$ 5.34	70.44 $\pm$ 4.31	74.16 $\pm$ 4.32	71.19 $\pm$ 4.35	0.54 $\pm$ 0.11
ATAN	71.10 $\pm$ 5.77	70.22 $\pm$ 3.56	73.89 $\pm$ 4.29	71.36 $\pm$ 4.36	0.56 $\pm$ 0.09
HC-TAN	70.41 $\pm$ 7.44	69.67 $\pm$ 6.39	73.95 $\pm$ 6.04	70.22 $\pm$ 6.26	0.58 $\pm$ 0.11
HC-SP-TAN	70.81 $\pm$ 6.48	71.63 $\pm$ 4.81	74.98 $\pm$ 5.98	71.98 $\pm$ 5.36	0.56 $\pm$ 0.11
BSEJ	71.09 $\pm$ 4.24	72.09 $\pm$ 5.95	74.28 $\pm$ 5.22	71.09 $\pm$ 5.11	0.55 $\pm$ 0.12
FSSJ	71.69 $\pm$ 3.92	72.03 $\pm$ 3.34	73.88 $\pm$ 5.02	72.27 $\pm$ 4.56	0.58 $\pm$ 0.09
(μ, λ)-TAN	74.09 $\pm$ 3.62	73.89 $\pm$ 2.54	76.88 $\pm$ 2.34	75.14 $\pm$ 3.24	0.59 $\pm$ 0.08

Table 2 Statistical significance test for different simulations in terms of Accuracy

Algorithm	NB	TAN	SVM	DT	RF	RVFL	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ
(μ, λ)-TAN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The ✓ symbol denotes that these two methods are statistically significantly different with $p < 0.05$
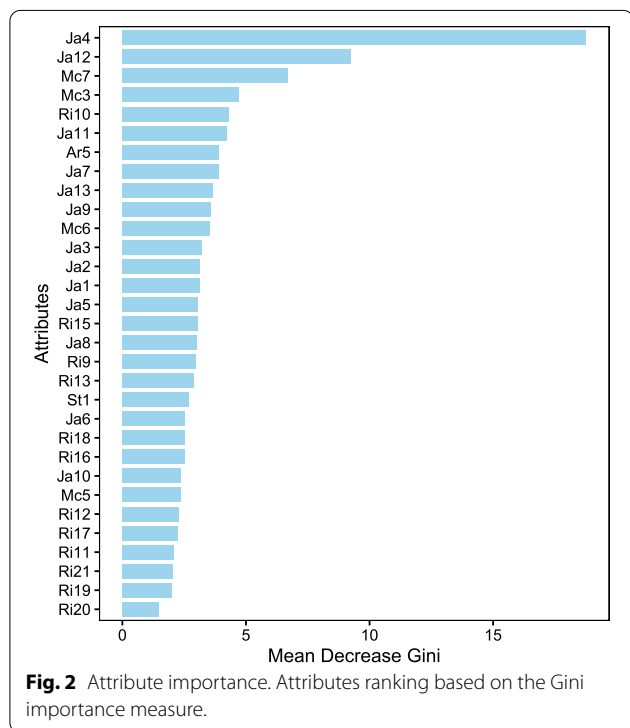
Fig. 2 Attribute importance. Attributes ranking based on the Gini importance measure.

mutual information. Instead, we use an evolution strategy to learn the weights of the networks from the data. Using attributes that are unaffected by the sagittal position of the jaws, we apply the proposed method to automatically classify a patient's biotype, eradicating the inaccuracies shown with the VERT index. In particular, one of the measurements used to calculate the VERT index is the facial depth, which indicates the sagittal relationship between the jaws. When this sagittal relationship is altered, the VERT is also altered. Therefore, a higher VERT is obtained in individuals with a prominent jaw, diagnosing the patient as more Brachyfacial than it is. Conversely, a patient with a mandible positioned further back will appear more Dolichofacial than it is.

Results

The results are shown in Table 1. Overall we notice that (μ, λ)-TAN, on average, outperforms all the other models for the particular dataset analyzed. Moreover, (μ, λ)-TAN

presents the lowest dispersion, making this model more stable and robust against different runs.

Table 2 shows that the results in terms of Accuracy of (μ, λ)-TAN are statistically significantly different to the results obtained by the other methods. Also, it is important to highlight that the results obtained are better than previously published results for the same dataset [27].

The best resulting network using (μ, λ)-TAN is shown in Fig. 1. For better visualization, we have omitted in this figure, the node with the class variable and the edges from this node to all the other nodes.

We used the *importance* function from the randomForest package in R [28] to create a smaller network. Based on the Gini importance, a metric used to assess the node impurity during the tree inference process, this function calculates the importance of each attribute (in decision trees or random forests). The outcome is displayed in Fig. 2.
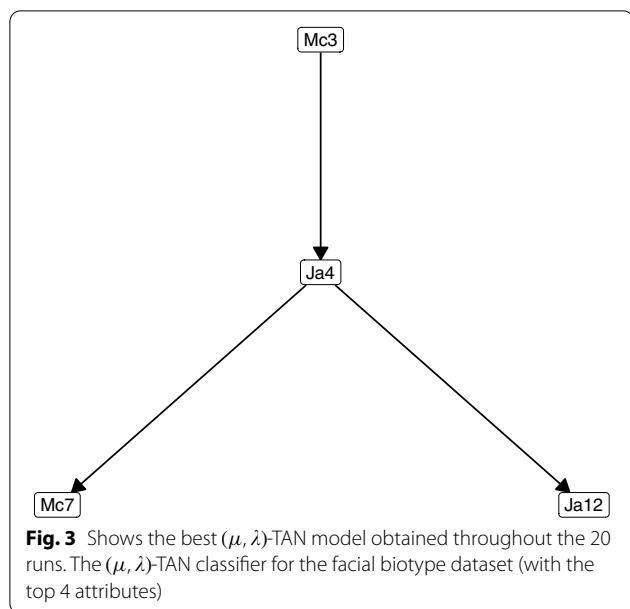
Using the top four attributes from Fig. 2, the outcomes of our repeated experiments are displayed in Tables 3 and 4. We notice similar results as before, with slight improvements in the evaluation measures.

The best resulting network in this case using (μ, λ)-TAN is shown in Fig. 3.

To evaluate the robustness of the proposed method, we tested (μ, λ)-TAN on high-dimensional datasets chosen from the UCI database [29]. For this, three datasets were considered, as described in Table 5. Table 6 shows the performance of (μ, λ)-TAN and RF. It can be noticed that, in the case of the three datasets, our method achieves better performances on average.

Discussion

From Fig. 1 we notice that Mc7 (Lower anterior facial height) is the parent node of 3 variables, St1 (SNA angle), Ja4 (Lower Gonial angle), and Ri21 (Symphysis length). Mc7 is measured from a point close (anterior nasal spine) to one of the points that constitute the SNA angle (point A) and both points are part of the same structure (maxillary), so the modification of the first one could be accompanied of a modification of St1 as well. On the other hand, the relationship between Mc7, Ja4, and Ri21 is explained given that the three correspond to vertical measurements and the modification of one should be accompanied by the modification of the other two variables.

Fig. 3 Shows the best (μ, λ)-TAN model obtained throughout the 20 runs. The (μ, λ)-TAN classifier for the facial biotype dataset (with the top 4 attributes)

Ja4 is the parent node of Ja9 (Cranial base and Mandibular length ratio), a relationship for which we do not have a satisfactory biological explanation since Ja4 is a vertical measurement and Ja9 is a horizontal one. In turn, Ja9 is the parent node of Ja8 (Mandibular corpus length), which is explained given that Ja8 is one of the measurements that make up Ja9.

Ri21 is the parent node of Ri19 (Condylar height) and Ja5 (Anterior cranial base length), a relationship that does not have an acceptable biological explanation, except that, as they correspond to linear measurements, they are influenced by the volumetric proportionality that exists between the structures given the greater or lesser general size of the skull.

St1 is the parent node of Ja3 (Upper Gonial angle), a relationship that could be explained since both represent sagittal growth, St1 indicates sagittal position of the maxilla with respect to the skull and Ja3 horizontal projection of the mandible; Normally, both structures tend to grow proportionally in the sagittal direction. Ja3 is the parent node of Ri18 (Posterior height), which is explained by the fact that both measurements share a reference point (gonion). Ri18 is the parent node of Ri10 (Maxillary depth angle) and Ja6 (Posterior cranial base length), there is no biological explanation for the relationship between Ri18 and Ri10 since one corresponds to a sagittal measurement and the other is vertical and they are measured in different areas of the face. In the case of Ri18 and Ja6 they use different landmarks but both measure posterior height of the face, so a relationship between both variables is clearly explained.

Ri10 is the parent node of Ri13 (Anterior Cranial length), a relationship that can be explained since both measurements share a reference point (Nasion).

Table 3 Performance measures for each model (with four attributes)

Algorithm	Accuracy Avg. $\pm$ SD.	Precision Avg. $\pm$ SD.	Recall Avg. $\pm$ SD.	F_1-score Avg. $\pm$ SD.	Kappa Avg. $\pm$ SD.
NB	72.27 $\pm$ 5.03	72.36 $\pm$ 5.63	76.23 $\pm$ 5.61	72.84 $\pm$ 5.81	0.57 $\pm$ 0.09
TAN	72.18 $\pm$ 3.53	72.41 $\pm$ 6.66	76.43 $\pm$ 6.16	73.33 $\pm$ 6.95	0.56 $\pm$ 0.05
SVM	66.90 $\pm$ 4.59	67.86 $\pm$ 4.91	71.89 $\pm$ 6.22	69.23 $\pm$ 6.73	0.49 $\pm$ 0.07
DT	69.63 $\pm$ 5.21	69.63 $\pm$ 3.27	72.79 $\pm$ 2.74	70.04 $\pm$ 3.18	0.52 $\pm$ 0.08
RF	70.27 $\pm$ 4.06	71.28 $\pm$ 4.45	73.82 $\pm$ 4.24	71.11 $\pm$ 4.46	0.54 $\pm$ 0.06
RVFL	71.53 $\pm$ 5.67	72.51 $\pm$ 4.56	76.56 $\pm$ 6.21	73.11 $\pm$ 5.55	0.56 $\pm$ 0.09
ATAN	72.21 $\pm$ 4.78	72.25 $\pm$ 4.96	76.43 $\pm$ 6.78	73.01 $\pm$ 5.99	0.58 $\pm$ 0.09
HC-TAN	72.12 $\pm$ 5.01	72.07 $\pm$ 5.73	75.33 $\pm$ 3.71	72.73 $\pm$ 5.11	0.58 $\pm$ 0.10
HC-SP-TAN	71.09 $\pm$ 7.03	71.82 $\pm$ 5.34	74.27 $\pm$ 5.01	72.11 $\pm$ 5.28	0.55 $\pm$ 0.09
BSEJ	71.63 $\pm$ 6.03	71.76 $\pm$ 4.65	74.80 $\pm$ 5.12	72.26 $\pm$ 4.51	0.56 $\pm$ 0.09
FSSJ	72.27 $\pm$ 4.43	72.35 $\pm$ 4.75	76.20 $\pm$ 6.86	72.81 $\pm$ 6.26	0.58 $\pm$ 0.09
(μ, λ)-TAN	75.05 $\pm$ 3.86	74.85 $\pm$ 4.58	77.85 $\pm$ 4.46	75.51 $\pm$ 6.36	0.60 $\pm$ 0.07

Table 4 Statistical significance test for different simulations in terms of Accuracy

Algorithm	NB	TAN	SVM	DT	RF	RVFL	ATAN	HC-TAN	HC-SP-TAN	BSEJ	FSSJ
(μ, λ)-TAN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The ✓ symbol denotes that these two methods are statistically significantly different with $p < 0.05$ (with four attributes)
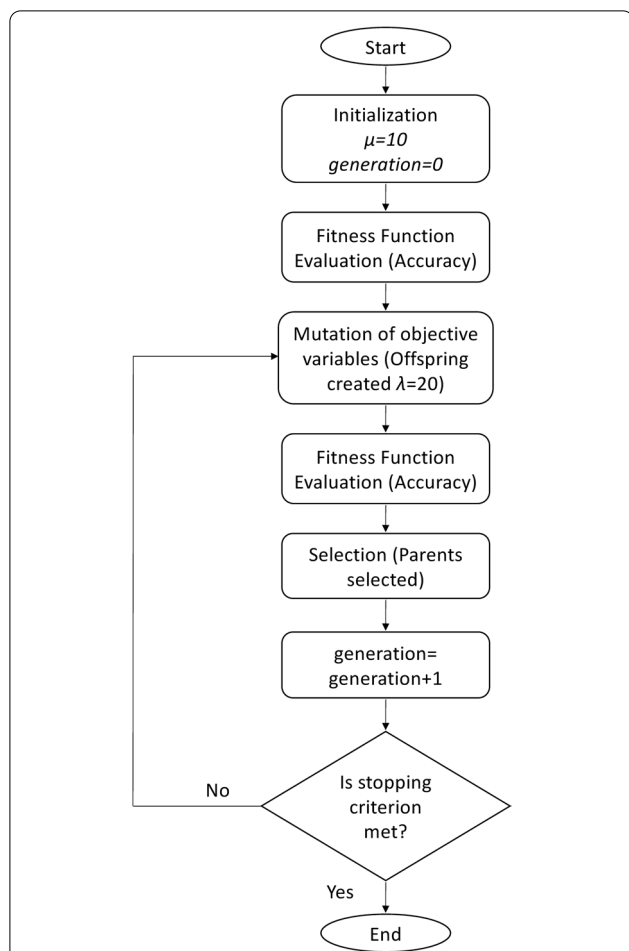
Fig. 4 Alternative learning algorithm for tree augmented Naive Bayes. Flowchart of the evolution strategy

Ri13 is the parent node of Ri11 (Palatal plane angle) and Ja12 (Jarabak's ratio), Ri13 and Ja12 share a reference point (Nasion) and all three correspond to vertical measurements, so the relationship between them is justifiable.

Ri11 is the parent node of Mc3 (Linear distance from point A to nasion perpendicular), which is explained because they share a reference point in the maxilla (Nasion) and the modification of this point would produce a change in both variables.

Ja12 (Jarabak's ratio) is the parent node of 3 variables Ri15 (Mandibular corpus axis), Mc6 (Maxillary length), and Ja1 (Saddle angle). Regarding this relationship, Ja12 and Ri15 correspond to measures indicative of the magnitude of vertical growth, Ja1 is part of Ja12, and with Mc6 instead, it cannot be explained biologically in a satisfactory way.

Ja1 (Saddle angle) is the parent node of Ja2 (Articular angle), which is explained given that both are contiguous angles that tend to compensate each other, that is, the tendency is that if one angle increases, the other tends to decrease in post of maintaining the proportionality of the face.

Ri15 (Mandibular corpus axis) is the parent node of Ja7 (Ramus height) and Ri17 (Mandibular ramus position), a relationship that is explained by the fact that Ri15 and Ri17 share a reference point (Xi), and that the three measurements correspond to vertical variables.

Ri17 is the parent node of Ri12 (Cranial deflection), a relationship that is explained by the fact that both measurements contain the Porion-Orbitale line.

Mc6 (Maxillary length) is the parent node of Ja10 (Posterior facial height), Ar5 (Nasolabial angle), and Ja11 (Anterior facial height), a relationship that does not have an acceptable biological explanation, except for the volumetric proportionality that exists between the structures that contain the landmarks corresponding to MC6, Ja10, and Ja11, given the greater or lesser general size of the skull and that Ar5 can be influenced by Mc6 since the upper lip rests on the maxilla, although this relationship is not direct since it depends mainly on the

Table 5 Information of the high-dimensional datasets

Datasets	Training	Testing	Attributes	Classes
Parkinson's disease	529	227	754	2
Diabetes 130-US hospitals	71236	30530	50	2
Digital colposcopies	67	30	69	2

Table 6 Performance measures for proposed method using high-dimensional data

Algorithm	Datasets	Accuracy Avg. ± SD.	Precision Avg. ± SD.	Recall Avg. ± SD.	F_T-score Avg. ± SD.	Kappa Avg. ± SD.
RF	Parkinson's disease	83.87 ± 1.39	74.58 ± 3.05	90.49 ± 3.97	81.72 ± 2.66	0.67 ± 0.05
(μ, λ)-TAN	Parkinson's disease	85.02 ± 2.81	77.56 ± 4.12	90.26 ± 2.63	83.37 ± 2.79	0.69 ± 0.05
RF	Diabetes	87.03 ± 1.08	79.03 ± 4.11	91.90 ± 2.90	85.34 ± 3.09	0.73 ± 0.05
(μ, λ)-TAN	Diabetes	88.28 ± 3.42	81.36 ± 6.34	92.92 ± 5.24	86.54 ± 3.81	0.76 ± 0.06
RF	Digital colposcopies	84.33 ± 5.88	75.77 ± 2.56	90.23 ± 3.12	82.56 ± 2.99	0.68 ± 0.05
(μ, λ)-TAN	Digital colposcopies	85.91 ± 2.99	78.12 ± 3.45	90.88 ± 2.55	83.96 ± 2.81	0.69 ± 0.06

Ruz *et al. BMC Medical Informatics and Decision Making* (2022) 22:316

Page 7 of 10

sagittal position of the maxilla. Ar5 is the parent node of Ri9 (Maxillary height angle), which could be explained by the fact that, as in the previous case, the position of the upper lip can be modified given the position of the maxilla, although this relationship is not direct since Ri9 corresponds to a indicative measure of the vertical and not sagittal position of the maxilla. Ri9 is in turn the parent node of Ja13 (posterior cranial base ratio to ramus height), a relationship for which we do not have a satisfactory biological explanation since, although both are vertical measurements, they correspond to different areas of the face.

Ja10 (Posterior facial height) is the parent node of Ri16 (Articular cavity position: Porion to Ptv), however, there is no direct biological explanation for this relationship; Ri16 is in turn the parent node of Ri20 (Condylar neck length), although the condyle is in relation to the joint cavity, we did not find an explanation for the relationship between the sagittal position of the joint cavity (Ri16) and the length of the neck of the condyle; Ri20 is the parent node of Mc5 (mandibular length), a relationship that could be explained by the fact that Mc5 has a reference point in the condyle and this is related to the joint cavity and both are sagittal measurements.

When analyzing the importance of each variable shown in Fig. 2, we notice that the four variables that turned out to have the greatest discriminatory power are: Ja4 (Lower Gonial angle), Ja12 (Jarabak's ratio), Mc7 (Lower anterior facial height), Mc3 (Linear distance from point A to nasion perpendicular). In particular, the first 3 variables are the measurements that account for the direction of vertical growth of the mandible, which is the main determinant in the pattern of facial growth and it is therefore logical that they appear as the most important. On the other hand, the Mc3 variable is indicative of the sagittal position of the maxilla with respect to the skull, which is not considered a determinant of the pattern of facial growth, however it could be related, since the rotation of the mandible generally in normally, it is accompanied by a rotation of the maxilla in the same direction and magnitude.

In the case of Fig. 3, it is observed that Mc3 (Linear distance from point A to nasion perpendicular) is the parent node of Ja4 (Lower Gonial angle), however, there is no direct biological explanation to explain this relationship since Mc3 is a sagittal measurement of the maxilla and Ja4 a vertical measurement of the mandible. In turn, Ja4 is the parent node of the variables Ja12 (Jarabak's ratio) and Mc7 (Lower anterior facial height), which can be explained because the three variables correspond to measures indicative of vertical growth, so that when increasing or decreasing a of them, the others also increase or decrease respectively proportionally.

Conclusion

In this paper, we have presented an alternative learning method based on an evolution strategy to learn the weights for constructing the TAN classifier. We applied this method to the facial biotype classification problem, obtaining high accuracy values compared to other competitive classifiers. When analyzing a resulting BN from (μ, λ)-TAN, many of the interactions shown in the network had an orthodontic interpretation, nevertheless, there were a few which did not have a satisfactory biological explanation. Future research will consider more benchmark datasets as well as other medical applications.

Methods
Dataset description

We use the [27] dataset, which comprises 182 lateral teleradiographies taken from patients in Chile. 31 continuous attributes that describe the craniofacial morphology were computed for each one using cephalometric analysis (see Table 7). Orthodontists have personally classified and validated each lateral teleradiograph into one of the three categories (Brachyfacial, Dolichofacial, and Mesofacial).

Alternative learning algorithm for tree augmented Naive Bayes

We propose an evolution strategy (ES) for learning TAN classifiers. The standard versions of the ES are denoted by [30]

$$(\mu, \lambda) - ES \quad \text{and} \quad (\mu + \lambda) - ES, \tag{2}$$

where λ represents the number of offspring and μ the number of parents. From the multi-set of either the offspring, known as comma-selection ($\mu < \lambda$ must hold), or both the parents and offspring, known as plus-selection, the parents are deterministically chosen (i.e., deterministic survivor selection). Selection is based on the ranking of the individuals' fitness, taking the μ best individuals (also referred to as truncation selection).

In this study, we generate weights for the TAN model that produce good facial biotype classification results without estimating the conditional mutual information by using the deterministic survivor selection (μ, λ) technique.

In order to do this, a candidate solution (an individual) is encoded as an m-dimensional vector that holds the m weight values of a network. We must define $m = n(n-1)/2$ weights for a network of n nodes. Consequently, we must locate 465 weights for $n = 31$. (parameters). We proceed as follows in order to determine the right values for these weights.

Table 7 A description of the attributes [27]

Attribute	Description
Mc3	Linear distance from point A to nasion perpendicular
Mc5	Mandibular length (Condylion to Gnathion)
Mc6	Maxillary length (Condylion to Point A)
Mc7	Lower anterior facial height (Anterior nasal spine to menton)
St1	SNA angle (Sella-Nasion-A)
Ja1	Saddle angle (Nasion-Sella-Articulare)
Ja2	Articular angle (Sella-Articulare-Gonion)
Ja3	Upper Gonial angle (Articulare-Gonion-Nasion)
Ja4	Lower Gonial angle (Nasion-Gonion-Menton)
Ja5	Anterior cranial base length (Sella to Nasion)
Ja6	Posterior cranial base length (Sella to Articulare)
Ja7	Ramus height (Articulate to Gonion)
Ja8	Mandibular corpus length (Gonion to Gnathion)
Ja9	Cranial base and Mandibular length ratio (Sella-Nasion/Gonion-Gnathion)
Ja10	Posterior facial height (Sella to Gonion)
Ja11	Anterior facial height (Nasion to Menton)
Ja12	Jarabak's ratio (Posterior facial height/Anterior facial height)
Ja13	Posterior cranial base and ramus height ratio (Sella-Articulare/Articulare-Gonion)
Ri9	Maxillary height angle (Nasion-Center of Face-A)
Ri10	Maxillary depth angle (Porion-Orbitale and Nasion-A)
Ri11	Palatal plane angle (Porion-Orbitale/anterior nasal spine-posterior nasal spine)
Ri12	Cranial deflection (Basion-Nasion/Porion-Orbitale)
Ri13	Anterior Cranial length (Center of Cranium to Nasion)
Ri15	Mandibular corpus axis (point Xi to point protuberance menti or Pm)
Ri16	Articular cavity position: Porion to Ptv (intersection of the distal outline of pterigomaxillary fissure perpendicular to the porion-orbitale plane)
Ri17	Mandibular ramus position (Porion-Orbitale/Center of Face-point Xi)
Ri18	Posterior height (Gonion to Center of Face)
Ri19	Condylar height
Ri20	Condylar neck length
Ri21	Symphysis length
Ar5	Nasolabial angle (Columella-Subnasale-upper lip)

By evenly distributing all the weight values in each individual's unit hypercube at random, we create an initial population of μ individuals. The procedure is then repeated a specified number of times. Each iteration starts with a population-wide evaluation of each proposed solution. A flowchart that briefly explains how the ES algorithm functions is shown in Fig. 4. The reader is directed to [31] for further information on how ES functions.

The accuracy, or the percentage of correctly classified instances in the training set, is then used to calculate the fitness function.

The highest scoring μ parents are then chosen. We perform the subsequent actions:

- The parent population consists of $\mu = 10$ individuals.
- The number of offspring produced per each iteration is indicated by $\lambda = 20$.
- Individuals die out after one iteration step (we use 1000 iterations) and only the offspring (the youngest individuals) survive to the following generation. Then, μ parents are chosen from λ offspring via environmental selection.

These hyperparameters were selected through simulations of trial and error.

Model performance assessment

We evaluated four metrics: precision (*Prc*), recall (*Rec*), accuracy (*Acc*), and F_1-score. This is how these measurements are calculated:

$$Prc = \frac{TP}{TP + FP} \times 100 \tag{3}$$

$$Rec = \frac{TP}{TP + FN} \times 100 \tag{4}$$

$$Acc = \frac{\text{number of correctly classified instances}}{\text{total number of instances}} \times 100 \tag{5}$$

$$F_1\text{-score} = 2 \times \frac{Prc \times Rec}{Prc + Rec} \tag{6}$$

where *TP*, *TN*, *FP*, and *FN* stand for true positive, true negative, false positive, and false negative, respectively. Since we are dealing with a multiclass problem, we compute *Prc*, *Rec*, and F_1-score for each individual class, and then report the average.

Additionally, we calculate the Kappa statistic, which contrasts the trained model's Acc (in the test set) with a random model's accuracy. We utilize the classification suggested in [32] to interpret the Kappa value: values ≤ 0 indicate poor agreement, $0 - 0.2$ indicate slight, $0.21 - 0.4$ indicate fair, $0.41 - 0.6$ indicate moderate, $0.61 - 0.8$ indicate substantial, and $0.81 - 1$ indicate practically perfect agreement.

Ruz *et al. BMC Medical Informatics and Decision Making* (2022) 22:316

Page 9 of 10

Experimental setup

The continuous features were discretized using Fayyad and Irani's Minimum Description Length method [33], which has been shown to have a positive effect on the classifiers' performance [34]. We compared the performance of the Naive Bayesian network classifier (NB), TAN, support vector machine (SVM) [35], decision tree (DT) [36], Random Forest (RF) [37], random vector functional link neural network [38] (RVFL), Averaged TAN (ATAN) [39] and the proposed method (μ, λ)-TAN. Four greedy hill-climbing algorithms were also used as a basis for learning Bayesian network classifiers:

- Hill-climbing tree augmented Naive Bayes (HC-TAN) [40].
- Hill-climbing super-parent tree augmented Naive Bayes (HC-SP-TAN) [40].
- Backward sequential elimination and joining (BSEJ) [18].
- Forward sequential selection and joining (FSSJ) [18].

The HC-TAN and HC-SP-TAN algorithms begin with a Naive Bayes structure and continue to add edges until the network score does not increase. Beginning with a Naive Bayes structure, BSEJ adds augmenting edges before removing features from the model until there is no longer any increase in the network score. On the other hand, FSSJ begins with a structure that only has the class node and builds upon it by adding features and enhancing edges.

Averages and standard deviations were recorded after 20 times of doing each experiment run. We divided the dataset into 70% for training and 30% for testing for each run, with the division being done at random.

We partitioned the original training set into 70% for training and the remaining 30% to evaluate different hyperparameter configurations through grid search for all the algorithms that needed hyperparameter tuning.

The open-source R software environment for statistical computation was used for all of the simulations. We used the test set's kappa statistic and accuracy metric to assess classification performance. Additionally a statistical significance test, the paired sample t-test, for different simulations in terms of accuracy was conducted.

Abbreviations

BN: Bayesian networks; ES: Evolution strategy; NB: Naive Bayesian network classifier; TAN: Tree augmented Naive Bayes; SVM: Support vector machine; DT: Decision tree; RVFL: Random vector functional link neural network; ATAN: Averaged tree augmented Naive Bayes; (μ, λ)-TAN: Evolution strategy deterministic survivor selection tree augmented Naive Bayes; HC-TAN: Hill-climbing tree augmented Naive Bayes; HC-SP-TAN: Hill-climbing super-parent tree augmented Naive Bayes; BSEJ: Backward sequential elimination and joining; FSSJ: Forward sequential selection and joining.

Declarations

Ethics approval and consent to participate

The teleradiographies used in this work were obtained from a file of an orthodontic clinic in Chile, which was taken previously as part of the routine protocol at the beginning of orthodontic treatment. Informed consent was obtained for all subjects.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile. [2]Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile. [3]Data Observatory Foundation, Santiago, Chile. [4]Departamento del Niño y Adolescente, Área de Ortodoncia, Facultad de Odontología, Universidad Andres Bello, Santiago, Chile. [5]Facultad de Administración y Economía, Universidad Diego Portales, Santiago, Chile.

References

1. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco: Morgan Kaufmann; 1988.
2. Mesaros A-S, Sava S, Mitrea D, Gasparik C, Alb C, Mesaros M, Badea M, Dudea D. In vitro assessment of tooth color changes due to orthodontic treatment using knowledge discovery methods. J Adhes Sci Technol. 2015;29(20):2256–79.
3. Nomura Y, Otsuka R, Wint WY, Okada A, Hasegawa R, Hanada N. Tooth-level analysis of dental caries in primary dentition in Myanmar children. Int J Environ Res Public Health. 2020;17(20):7613.
4. Sironi E, Taroni F, Baldinotti C, Nardi C, Norelli G-A, Gallidabino M, Pinchi V. Age estimation by assessment of pulp chamber volume: a Bayesian network for the evaluation of dental evidence. Int J Legal Med. 2018;132(4):1125–38.
5. Sironi E, Pinchi V, Pradella F, Focardi M, Bozza S, Taroni F. Bayesian networks of age estimation and classification based on dental evidence: a study on the third molar mineralization. J Forensic Legal Med. 2018;55:23–32.
6. Bhornsawan T. Bayesian-based decision support system for assessing the needs for orthodontic treatment. Healthc Inform Res. 2018;24(1):22–8.
7. Bessani M, de Lima DR, Cleiton Cabral Correia Lins E, Maciel CD. Evaluation of a dental caries clinical decision support system. In: Proceedings of the 10th international joint conference on biomedical engineering systems and technologies—BIOSIGNALS, (BIOSTEC 2017), 2017. pp. 198–204.

Ruz *et al. BMC Medical Informatics and Decision Making*   (2022) 22:316

Page 10 of 10

8. Merli M, Moscatelli M, Mariotti G, Pagliaro U, Bernardelli F, Nieri M. A minimally invasive technique for lateral maxillary sinus floor elevation: a Bayesian network study. Clin Oral Implants Res. 2016;27(3):273–81.

9. Iwasaki H. Bayesian belief network analysis applied to determine the progression of temporomandibular disorders using mri. Dentomaxillofac Radiol. 2015;44(4):20140279.

10. Scutari M, Auconi P, Caldarelli G, Franchi L. Bayesian networks analysis of malocclusion data. Sci Rep. 2017;7(1):15236.

11. Liang M, Lian Q, Kotsakis GA, Michalowicz BS, John MT, Chu H. Bayesian network meta-analysis of multiple outcomes in dental research. J Evid Based Dent Pract. 2020;20(1):101403.

12. Hu S, An K, Peng Y. Comparative efficacy of the bone-anchored maxillary protraction protocols for orthopaedic treatment in skeletal class iii malocclusion: A bayesian network meta-analysis. Orthod Craniofac Res. 2021;25(2):243–50.

13. Aldhohrah T, Mashrah MA, Wang Y. Effect of 2-implant mandibular overdenture with different attachments and loading protocols on peri-implant health and prosthetic complications: a systematic review and network meta-analysis. J Prosthet Dent. 2021;126(6):832–44.

14. Zhao P, Song X, Nie L, Wang Q, Zhang P, Ding Y, Wang Q. Efficacy of adjunctive photodynamic therapy and lasers in the non-surgical periodontal treatment: a Bayesian network meta-analysis. Photodiagn Photodyn Ther. 2020;32:101969.

15. Wu Z, Zhang X, Li Z, Liu Y, Jin H, Chen Q, Guo J. A Bayesian network meta-analysis of orthopaedic treatment in class iii malocclusion: Maxillary protraction with skeletal anchorage or a rapid maxillary expander. Orthod Craniofac Res. 2020;23(1):1–15.

16. Machado V, Botelho J, Mascarenhas P, Mendes JJ, Delgado A. A systematic review and meta-analysis on Bolton's ratios: normal occlusion and malocclusion. J Orthod. 2020;47(1):7–29.

17. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn. 1997;29(2):131–63.

18. Pazzani MJ. Constructive induction of Cartesian product attributes. In: Liu H, Motoda H (editors). Feature extraction, construction and selection. The Springer International Series in Engineering and Computer Science, vol. 453. Boston: Springer; 1998. p. 341–54.

19. Provan GM, Singh M. In: Fisher D, Lenz H-J (editors) Learning Bayesian networks using feature selection. New York: Springer, 1996. p. 291–300.

20. Sahami M. Learning limited dependence Bayesian classifiers. In: Proceedings of the second international conference on knowledge discovery and data mining. KDD'96, 1996. p. 335–8.

21. Margaritis D, Thrun S. Bayesian network induction via local neighborhoods. In: Solla SA, Leen TK, Müller K, editors. Advances in neural information processing systems, vol. 12. Cambridge: MIT Press; 1999. p. 505–11.

22. Ruz GA, Pham DT. Building Bayesian network classifiers through a Bayesian complexity monitoring system. Proc IMechE Part C J Mech Eng Sci. 2009;223:743–55.

23. Bielza C, Larrañaga P. Discrete Bayesian network classifiers: a survey. ACM Comput Surv. 2014;47:5–1543.

24. Duda RO, Hart PE. Pattern Classif Scene Anal. New York: John Wiley & Sons; 1973.

25. Pham DT, Ruz GA. Unsupervised training of Bayesian networks for data clustering. Proc R Soc A Math Phys Eng Sci. 2009;465(2109):2927–48.

26. Ricketts RM, Roth RH, Chaconas SJ, Schulhof RJ, Engel GA. Orthodontic diagnosis and planning: their roles in preventive and rehabilitative dentistry. Pacific Palisades: Rock Mountain Data Systems; 1982.

27. Ruz GA, Araya-Díaz P. Predicting facial biotypes using continuous Bayesian network classifiers. Complexity. 2018;2018:4075656.

28. Liaw A, Wiener M. Classification and regression by Randomforest. R News. 2002;2(3):18–22.

29. Dua D, Graff C. UCI machine learning repository 2017. http://archive.ics.uci.edu/ml.

30. Alrashdi Z, Sayyafzadeh M. $(\mu + \lambda)$ evolution strategy algorithm in well placement, trajectory, control and joint optimisation. J Petrol Sci Eng. 2019;177:1042–58.

31. Back T. Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford: Oxford University Press; 1996.

32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.

33. Fayyad UM, Irani KB. Multi-interval discretization of continuous-valued attributes for classification learning. In: IJCAI, 1993. pp. 1022–1029.

34. Dougherty J, Kohavi R, Sahami M. Supervised and unsupervised discretization of continuous features. In: Prieditis A, Russell S, editors. Machine learning proceedings 1995. San Francisco (CA): Morgan Kaufmann; 1995. p. 194–202.

35. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.

36. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. London: Chapman and Hall/CR; 1984.

37. Breiman L. Random forests. Mach Learn. 2001;45:5–32.

38. Henríquez PA, Ruz GA. A non-iterative method for pruning hidden neurons in neural networks with random weights. Appl Soft Comput. 2018;70:1109–21.

39. Jiang L, Cai Z, Wang D, Zhang H. Improving tree augmented Naive Bayes for class probability estimation. Knowl Based Syst. 2012;26:239–45. https://doi.org/10.1016/j.knosys.2011.08.010.

40. Keogh EJ, Pazzani MJ. Learning the structure of augmented Bayesian classifiers. Int J Artif Intell Tools. 2002;11(04):587–601.

Publisher's Note