Proceedings of Machine Learning Research 343:629–645, 2026 Probabilistic Graphical Models (PGM)

Gradient-Based Fine-Tuning of Expert Bayesian Network Classifiers: Application in Forensic Porcine Wound Age Assessment

Joseph Mietkiewicz JOTOMI@CS.AAU.DK
Department of Computer Science, Aalborg University, Denmark. School of Food Science & Environmental Health, TU Dublin, Dublin, Ireland

Anders L Madsen ALM@CS.AAU.DK
Department of Computer Science, Aalborg University, and HUGIN EXPERT A/S, Denmark

Thomas Dyhre Nielsen TDN@CS.AAU.DK
Department of Computer Science, Aalborg University, Denmark

Cecilie Bækgård CECB@SUND.KU.DK
Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

Henrik Elvang Jensen ELVANG@SUND.KU.DK
Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

Kristiane Barington KRISB@SUND.KU.DK
Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

Editors: Gustau Camps-Valls, Manuele Leonelli and Gherardo Varando

Abstract

Bayesian network classifiers built from expert knowledge encode valuable domain understanding but may be improved when labelled data become available. We propose EB-GFTBN, a Gradient-based Fine-Tuning algorithm for Bayesian Networks with Expert-Bounded parameter updates. Building on the fine-tuning framework of El Hindi (2014) and the Brier-score variant of Mietkiewicz and Madsen (2022), we derive natural-gradient updates from the derivative of the Brier score and extend the algorithm to mixed discrete–continuous Bayesian networks. To remain faithful to the expert specification, each parameter is constrained to a relative tolerance around its expert elicited value. The methodology is evaluated in a case study regarding forensic age assessment of porcine wounds, where the parameters of a Bayesian network classifier were elicited from veterinary experts. Across 100 cross-validation folds, EB-GFTBN improves both accuracy and Brier score over the expert baseline and the penalised EM alternative, while keeping every parameter inside the expert-allowed window.

Keywords: Bayesian network classifiers, fine-tuning, Brier score, natural gradient, expert elicitation, forensic wound age assessment.

© 2026 J. Mietkiewicz, A.L. Madsen, T.D. Nielsen, C. Bækgård, H.E. Jensen & K. Barington.

1. Introduction

Bayesian network classifiers are attractive in domains where the structure can be elicited from subject-matter experts and the parameters carry an interpretable probabilistic meaning. When data subsequently become available, combining the expert assessment with the data is traditionally done generatively, by weighting the elicited parameters against a maximum-likelihood or Bayesian estimate. We propose instead to blend expert knowledge and data discriminatively. Discriminative parameter learning can yield better classification performance than generative estimation when the data are limited or the structure does not reflect the true distribution (Ng and Jordan, 2002; Su et al., 2008; Pernkopf et al., 2012). In particular, the ELR algorithm of Greiner et al. (2005) performs gradient ascent on the conditional log-likelihood to optimise the parameters of a Bayesian network with a fixed structure. Our work follows the same idea of improving elicited parameters by gradient descent on a classification-oriented score, with the emphasis placed on combining expert knowledge with data, on covering conditional linear Gaussian nodes alongside discrete conditional probability tables, and on the fine-tuning framework of El Hindi (2014) as the algorithmic template. The original fine-tuning algorithm for naive Bayes (FTNB) proposed a heuristic correction that increases the probability terms of the true class and decreases those of the predicted class when an instance is misclassified, iterating over training instances and renormalising each table row after every update; it has since been extended in several directions (Alhussan and El Hindi, 2015, 2016; El Hindi et al., 2018, 2020; Zhang and Jiang, 2022; Diab and El Hindi, 2017; Mietkiewicz and Madsen, 2022).

In this paper we revisit fine-tuning from the perspective of probabilistic gradient descent. Starting from the exact derivative of the Brier score with respect to each Bayesian network parameter, we derive a closed-form natural-gradient update that applies uniformly to discrete CPT entries, Gaussian means, and Gaussian variances. To prevent the data-driven updates from drifting too far from the expert prior, we introduce a per-CPT-entry (and per Gaussian parameter) relative-budget constraint that bounds each parameter to a domain-specified tolerance around its expert elicited value. The resulting algorithm, EB-GFTBN (Expert-Bounded-Gradient-based Fine-Tuning for Bayesian Networks), is evaluated on a porcine wound age classification problem where the parameters of a naive Bayes model were elicited from veterinary experts.

The motivating application is wound age estimation, an important but challenging task within veterinary forensic pathology (Ressel et al., 2016; Barington et al., 2018). Forensic wound age assessment may, in a judicial setting, be used to assess the degree of animal neglect, that is, for how long the animal suffered from the wound (Pankoke et al., 2023; Barington et al., 2016). In animals, age estimations of wounds are carried out by veterinary pathologists and are based on macroscopical and microscopical assessment of variables such as the presence and thickness of granulation tissue and the degree of infiltration by different inflammatory cells (Barington et al., 2018, 2016). However, such expert assessments are influenced by the experience of the pathologist and are not entirely objective (Bækgård et al., 2026). Therefore, more objective methods are warranted, and machine learning models could most likely aid in improving the precision of wound age assessments. A Bayesian probability model based on experimental ovine bruises was able to age bruises as either 1–20 h or 24–72 h (Thornton and Jolly, 1986), and the model was superior to

subjective interpretation. However, changes in bruises and wounds are not comparable, and wounds in veterinary forensic cases are more often several days to weeks old (Barington et al., 2016, 2023).

The contributions of this paper are: (i) a unified natural-gradient fine-tuning rule for Bayesian network classifiers, applicable to discrete CPT entries and to the parameters of conditional linear Gaussian densities; (ii) an expert-bounded projection scheme that enforces relative tolerances on parameter deviation; (iii) the algorithm EB-GFTBN and an empirical study of its behaviour on a forensic wound age assessment case study.

The remainder of the paper is organised as follows. Section 2 introduces Bayesian network notation, the Brier score, and the original fine-tuning framework. Section 3 presents the gradient-based fine-tuning methodology with the per-family update rules and the expert-bounded constraints. Section 4 presents the wound age case study, the data, and the expert-elicited model. Section 5 reports on the experimental analysis. Section 6 discusses the results and limitations, and Section 7 concludes the paper and outlines directions for future work.

2. Preliminaries and Notation

2.1. Bayesian Network Classifiers

A Bayesian network $\mathcal{B} = (\mathcal{G}, \boldsymbol{\Theta})$ consists of a directed acyclic graph $\mathcal{G} = (V, E)$, whose vertices V index the random variables $\{X_v\}_{v \in V}$, and parameters $\boldsymbol{\Theta}$. It encodes the joint distribution

$$P(\mathcal{X}) \;=\; \prod_{v \in V} P\big(X_v \mid X_{\mathrm{pa}(v)}\big), \tag{1}$$

where $\mathrm{pa}(v)$ are the parents of v in $\mathcal{G}$. We consider classifiers that contain a distinguished class variable C taking values $\{c_1, \ldots, c_L\}$ and feature variables $X_1, \ldots, X_m$ that may be either discrete or continuous. The predicted class for instance $\mathbf{x} = (x_1, \ldots, x_m)$ is obtained by maximum a posteriori inference,

$$c_{\text{predicted}} \;=\; \arg\max_{c \in C} \; P(c \mid x_1, \ldots, x_m). \tag{2}$$

The parameters $\boldsymbol{\Theta}$ represent every tunable quantity: conditional probability table (CPT) entries $\theta_{i,a,j} = P(X_i = a \mid \mathrm{pa}(X_i) = j)$ for discrete nodes, and means μ_{ij} and variances σ^2_{ij} for nodes with conditional linear Gaussian densities (Lauritzen and Jensen, 2001).

Given an instance $\mathbf{x}_k$ of a data set with n instances, with true class c_{actual}, propagation in $\mathcal{B}$ yields the posterior $p_{kl}(\boldsymbol{\Theta}) = P_{\mathcal{B}}(C = c_l \mid \mathbf{x}_k; \boldsymbol{\Theta})$, $l = 1, \ldots, L$. We write $E_{kl} = \mathbf{1}[c_l = c_{\text{actual}}]$ for the one-hot truth vector and define the residual $r_{kl} = p_{kl} - E_{kl}$.

2.2. The Brier Score

The Brier score (Brier, 1950) is a strictly proper, differentiable scoring rule for probabilistic forecasts over n instances:

$$B(\boldsymbol{\Theta}) \;=\; \frac{1}{n} \sum_{k=1}^{n} B_k, \qquad B_k \;=\; \sum_{l=1}^{L} r_{kl}^2. \tag{3}$$

We have $B \in [0, 2]$, lower is better, and B is differentiable in $\boldsymbol{\Theta}$, which makes it a natural objective for gradient-based fine-tuning. A useful auxiliary quantity is the weighted mean residual

$$w_k \;=\; \sum_{l=1}^{L} r_{kl}\, p_{kl}, \tag{4}$$

which arises repeatedly in the gradient.

2.3. Fine-Tuning Naive Bayes (FTNB) and the Brier-Score Update

El Hindi (2014) introduced FTNB, which iterates over misclassified training instances and applies a heuristic update with learning rate η that increases the CPT entries of the true class and decreases those of the predicted class. The size of each update is shaped by the gap between the entry and the dominant (or smallest) entry of the same conditional, so that small true-class probabilities receive larger upward pushes and large predicted-class probabilities larger downward pushes. Mietkiewicz and Madsen (2022) kept this framework but replaced the heuristic attribute-dependent factors by a Brier-score-driven update with only the learning rate retained, showing improvement over the initial approach. The next section extends this idea by deriving the exact derivative of the Brier score with respect to a generic Bayesian network parameter, and uses natural-gradient preconditioning to obtain closed-form updates for discrete CPT entries, Gaussian means, and Gaussian variances.

3. Gradient-Based Fine-Tuning

This section develops the methodology of the paper. We first present the gradient of the Brier score at a high level, then specialise it to the parameter families that arise in practice, introduce the natural-gradient preconditioning that removes the parameter-dependent denominators, and finally describe the expert-bounded constraint that keeps the fine-tuned model close to the expert prior. For simplicity we assume complete data and continuous nodes with only discrete parents, so that each parameter enters a single conditional factor for one class. Both assumptions can be relaxed without major modification.

3.1. High-Level Update Rule

For each parameter $\phi \in \boldsymbol{\Theta}$ we replace the heuristic update of FTNB by the exact gradient descent step on the per-instance Brier score:

$$\phi \;\leftarrow\; \phi \;-\; \eta\, \frac{\partial B_k}{\partial \phi}. \tag{5}$$

The whole methodology reduces to computing $\partial B_k / \partial \phi$ for each kind of parameter that the network contains. Throughout, $\eta > 0$ is a learning rate; in our experiments we use $\eta = 10^{-2}$. The derivative factorises compactly into two pieces:

$$\frac{\partial B_k}{\partial \phi} \;=\; \underbrace{2\, p_{kj}\, (r_{kj} - w_k)}_{\text{Brier signal } \beta_k(j)} \cdot \underbrace{s_\phi}_{\text{score function}}, \tag{6}$$

Table 1: Score function and natural gradient per parameter family. Every entry shares the Brier signal $\beta_k = 2\,p_{kj}(r_{kj} - w_k)$. The natural gradient $\tilde{g}_\phi$ is multiplied by the learning rate η to obtain the parameter update.

Parameter ϕ	**Score** s_ϕ	**Natural gradient** $\tilde{g}_\phi$
Discrete $\theta_{i,a,j}$	$1/\theta_{i,a,j}$	β_k
Gaussian mean μ_{ij}	$(x_{ki} - \mu_{ij})/\sigma^2_{ij}$	$\beta_k\,(x_{ki} - \mu_{ij})$
Gaussian variance σ^2_{ij}	$\big[(x_{ki} - \mu_{ij})^2 - \sigma^2_{ij}\big]/(2\sigma^4_{ij})$	$\beta_k\big[(x_{ki} - \mu_{ij})^2 - \sigma^2_{ij}\big]$

where $j = j(\phi)$ is the class index of the conditional that contains ϕ, the Brier signal $\beta_k(j)$ depends only on posteriors and residuals, and the score function $s_\phi = \partial \log p(x_{ki} \mid \mathrm{pa}; \mathbf{\Theta})/\partial\phi$ depends only on the family (discrete, Gaussian) of the density that contains ϕ. The detailed calculation can be seen in section A.

Equation (6) factorises the gradient of the Brier score into an instance-dependent part that is shared across all parameters and a family-dependent part that is closed-form. The update for every parameter then has the same shape; only s_ϕ changes. After a step, the row is renormalised by proportional covariation of the siblings, preserving the simplex constraint $\sum_a \theta_{i,a,j} = 1$ and $\theta_{i,a,j} \geq 0$

3.2. Specialisation by Parameter Family

The score function s_ϕ is known in closed form for every standard conditional density. Substituting it into (6) gives the per-parameter Brier gradient. To avoid the small denominators that arise from s_ϕ (rare CPT entries, small variances), we apply natural-gradient preconditioning (Amari, 1998): each gradient is rescaled by the inverse Fisher information $I(\phi)^{-1}$. The resulting natural gradient is

$$\tilde{g}_\phi \;=\; I(\phi)^{-1}\,\frac{\partial B_k}{\partial \phi} \;=\; I(\phi)^{-1}\,\beta_k(j)\,s_\phi. \tag{7}$$

For each of the three parameter families considered, the product $I(\phi)^{-1}s_\phi$ cancels the parameter-dependent denominator of s_ϕ. The result is summarised in Table 1. The detailed calculation can be seen in section A.

3.3. Expert-Bounded Constraints

The unconstrained gradient step of (5)–(7) can move parameters arbitrarily far from their initial expert value if the data are noisy or scarce. In a situation where parameters were elicited from subject-matter experts, this is undesirable for two reasons. First, the expert specification carries information beyond what the small training set can convey. Second, the experts can themselves expressed an explicit desire that the fine-tuned values remain inside an expert-defined range. We enforce this by augmenting the gradient step with a projection onto an expert-bounded interval at the end of every epoch.

The feasibility set is defined per parameter as a relative box around the expert value. Let $\theta^0_{i,a,j}$, μ^0_{ij}, $\sigma^{2,0}_{ij}$ denote the expert parameters. With user-chosen budgets $\delta_\theta, \delta_\mu, \delta_{\sigma^2} \geq 0$, and

let ε denote a small positive constant used to prevent parameters from reaching boundary values of 0 or 1:

$$\begin{aligned}\theta_{i,a,j} &\in \left[\theta^0_{i,a,j}(1-\delta_\theta),\ \theta^0_{i,a,j}(1+\delta_\theta)\right] \cap [\varepsilon, 1-\varepsilon],\\ \mu_{ij} &\in \left[\mu^0_{ij} - \delta_\mu\,\sigma^0_{ij},\ \mu^0_{ij} + \delta_\mu\,\sigma^0_{ij}\right],\\ \sigma^2_{ij} &\in \left[\sigma^{2,0}_{ij}(1-\delta_{\sigma^2}),\ \sigma^{2,0}_{ij}(1+\delta_{\sigma^2})\right] \cap [\varepsilon, \infty).\end{aligned}$$

The Gaussian mean is bounded in units of the expert standard deviation σ^0_{ij} so that δ_μ has a consistent interpretation across features that live on very different scales.

After every epoch each CPT row is projected onto the box-and-simplex feasibility set using an exact Euclidean projection, and each Gaussian parameter is clamped to its interval. The projection has no interaction with the unconstrained update until the bound is hit.

3.4. Stopping Criterion and Algorithm

EB-GFTBN proceeds iteratively over all instances in the training data, restarting at each epoch until either a maximum epoch budget T is exhausted or the training-set Brier score has not improved by more than a relative threshold δ_B over P consecutive epochs. The best-Brier model is stored at each improving epoch and returned at the end. The complete procedure is given in Algorithm 1.

Computational cost. Because the factorisation (6) shares the Brier signal $\beta_k(j)$ across all parameters of class j, a single propagation per instance suffices for the whole gradient, and each natural gradient of Table 1 is then evaluated in constant time; the preconditioning itself is free at run time, since the Fisher information is scalar within each family and cancels the score analytically. The projection onto the expert-bounded set is performed once per epoch and is negligible in comparison. The total cost of EB-GFTBN is therefore $O(T\,n\,|\Theta|)$, linear in the number of epochs, instances and parameters.

4. Case Study: Porcine Wound Age Assessment

The methodology was motivated and validated by a case study in veterinary forensic pathology. The case study concerns the assessment of the age of porcine wounds: a dataset with a limited number of cases was available, while subject-matter experts have years of experience performing this task. The aim of the case study is not to provide exact estimates on the wound age, but to serve as a vehicle for the fine-tuning algorithm and as a baseline for future model developments.

4.1. Experimental Setup and Data

To study wound healing over time, an experimental porcine wound model was developed (Bækgård et al., 2026). In summary, four experimental wounds were incised on the back of 47 fully anaesthetised pigs and left to heal without treatment. The pigs were euthanised after 5, 10, 15, 20, 25, 30, and 35 days. During the study period, a macroscopic assessment of the wounds was carried out every fifth day. After euthanasia, the wounds were

Algorithm 1: EB-GFTBN: gradient-based fine-tuning with expert-bounded constraints.
Input: Network $\mathcal{B}$ with expert parameters $\boldsymbol{\Theta}^0$; training data $D = \{(\mathbf{x}_k, y_k)\}_{k=1}^n$; learning rate $\eta > 0$; relative budgets $(\delta_\theta, \delta_\mu, \delta_{\sigma^2})$; maximum epochs T, patience P; minimum relative Brier improvement δ_B.
Output: Fine-tuned parameters $\boldsymbol{\Theta}^*$.

`// Initialise from the expert prior`
$\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta}^0$ $\boldsymbol{\Theta}^* \leftarrow \boldsymbol{\Theta}^0$ $B^* \leftarrow B(\boldsymbol{\Theta}, D)$ patience $\leftarrow 0$

for epoch $= 1, \ldots, T$ **do**
  `// Gradient pass over all training instances`
  foreach *instance* $\mathbf{x}_k \in D$ **do**
    Compute posteriors p_{kl}, residuals r_{kl}, and w_k
    foreach *parameter* ϕ *active for* $\mathbf{x}_k$ *with class state* j **do**
      $\beta_k(j) \leftarrow 2\, p_{kj} \left(r_{kj} - w_k\right)$
      $\phi \leftarrow \phi - \eta\, \tilde{g}_\phi$ (with $\tilde{g}_\phi$ from Table 1)
    end
  end
  `// Projection onto the expert-bounded feasibility set`
  Project each parameter using the per-family budgets $(\delta_\theta, \delta_\mu, \delta_{\sigma^2})$ $B \leftarrow B(\boldsymbol{\Theta}, D)$
  `// Early stopping on training Brier`
  if $(B^* - B)/B^* \geq \delta_B$ **then**
    $B^* \leftarrow B$ $\boldsymbol{\Theta}^* \leftarrow \boldsymbol{\Theta}$ P $\leftarrow 0$
  else
    P $\leftarrow$ P $+ 1$
    if P $\geq P_{\max}$ **then return** $\boldsymbol{\Theta}^*$
  end
end
return $\boldsymbol{\Theta}^*$

cross-sectioned, and the thickness of the granulation tissue (the reparative tissue formed in the wound bed) was measured. Moreover, wound tissue was sampled and assessed microscopically (Bækgård et al., 2026). The experimental procedure was approved by the Danish Animal Experiments Inspectorate (2023-15-0201-01363) and several time-dependent variables were presented recently by Bækgård et al. (2026).

The data originating from the experimental porcine model (Bækgård et al., 2026) consisted of three groups of variables: (1) macroscopic ("gross examination") assessments carried out by naked eye while the pigs were still alive — level of wound tissue relative to surrounding skin (Ordinal), hyperaemia (Binary), oedema (Binary), exudate (Binary), demarcation (Binary) and the presence of granulation tissue (Binary), presence of epithelialisation (Binary); (2) the thickness of the granulation tissue (Continuous), measured using a calliper after euthanasia; (3) microscopic ("histological examination") assessments obtained after euthanasia — presence of neutrophils (Binary), eosinophils (Binary), giant cells (Binary), and haemosiderophages (Binary), per-margin epithelialisation length (Continuous) and uncovered length (Continuous), and microscopic granulation-tissue thickness (Continuous). The wound age, the target variable, was originally observed at seven time points: 5,

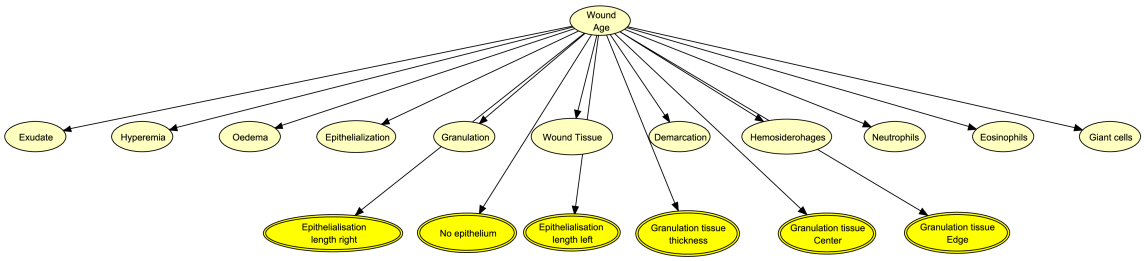


Figure 1: Naive Bayes structure of the expert-elicited Bayesian network classifier. The class variable is wound age; each feature variable, gross-examination or histological, is conditionally independent of the others given the class.

10, 15, 20, 25, 30, and 35 days. To balance the training set, the longest two time points (30 and 35 days) were merged with the 25-day class. Wounds from the same pig were considered to share information that could leak between training and test sets if the splits were drawn at random across wounds. The cross-validation scheme of Section 5 is therefore stratified at the pig level: every fold places all four wounds of each pig either in the training set or in the test set, never in both.

4.2. Expert-Elicited Bayesian Network Model

A naive Bayes structure was used for the Bayesian network classifier (Figure 1). The naive Bayes structure has the advantage that only parameters must be assessed by subject-matter experts: each feature variable is conditionally independent of the others given the class, so the elicitation reduces to one conditional density per feature. The parameters of the network, the conditional probability tables for discrete variables and the means and variances for the continuous nodes, were defined through expert elicitation. Three domain experts with direct experience in veterinary pathology and wound healing participated in the process. As the experts had been involved in the data collection, their judgements were informed by first-hand observation of wound progression across the time points rather than being blind to the experimental outcomes.

The elicitation was carried out over multiple workshops, each considering a different set of features, following the anchoring approach of Leva and Hansen (2017), also used by Briwa et al. (2026). Parameter values were established through group discussion: the three experts deliberated collectively on each variable, expressing their beliefs as odds ratios subsequently converted to conditional probabilities, until a consensus value was reached. No formal aggregation procedure was applied; the final parameters reflect the shared judgement of the group. In addition to providing initial parameter assessments, the experts defined an acceptable level of change per parameter — the fine-tuning should not move a parameter outside the corresponding expert-defined window. The relative budgets $\delta_\theta, \delta_\mu, \delta_{\sigma^2}$ of Section 3.3 capture this expert intent.

5. Experiments

This section evaluates EB-GFTBN on the wound age case study. We describe the cross-validation protocol, the hyperparameter grid, the metrics, and the baselines, and we report performance and parameter-deviation statistics.

5.1. Protocol

Splits. The dataset contains four wounds per pig and five wound-age classes (Section 4.1). To avoid information leakage at the pig level, the cross-validation protocol is pig-disjoint and class-stratified: every fold draws two pigs per class uniformly at random and places all four wounds of each selected pig into the test set. This produces a test set of two pigs per class — eight wounds per class, forty wounds in total — and a training set containing the remaining pigs. The protocol is repeated for $K = 100$ folds; reported metrics are the mean and standard deviation across folds.

Hyperparameters. The learning rate is $\eta = 10^{-2}$ throughout. We evaluate EB-GFTBN at five expert-fidelity budgets parameterised by the triple $(\delta_\theta, \delta_\mu, \delta_{\sigma^2})$: $(0.05, 0.5, 0.05)$, $(0.10, 1.0, 0.10)$, $(0.20, 1.5, 0.20)$, $(0.30, 2.0, 0.30)$, and unconstrained. The first triple is the tightest constraint (parameters may not deviate more than $\pm 5\%$, or ± 0.5 expert standard deviations); the last is the loosest. Early-stopping uses $\delta_B = 10^{-4}$ with patience $P = 50$ and a hard cap of $T = 1000$ epochs. The variance floor is $\varepsilon = 10^{-6}$.

Metrics. We report accuracy, Brier score, and the one-vs-rest AUC with macro-averaged. We also report the maximum observed per-parameter deviations $|\Delta\theta|$, $|\Delta\mu|/\sigma^0$, and $|\Delta\sigma^2|/\sigma^{2,0}$ from the expert values at convergence, which empirically verifies that the relative budgets are respected. Paired t-tests between fine-tuned and expert baseline results are reported with the corresponding p-values where relevant.

Baselines. We compare EB-GFTBN to two baselines: (i) the expert-elicited Bayesian network model (Section 4.2) used as is, and (ii) an expert-data combination (Penalised EM) in which the expert and the data parameter estimates receive equal weight.

5.2. Results

Table 2 reports the per-fold paired improvement $\Delta = \text{method} - \text{expert}$, averaged across the $K = 100$ pig-disjoint folds. EB-GFTBN improves accuracy and Brier score over the expert baseline at every tested budget, with both improvements significant at $p < 10^{-3}$ in every configuration (paired t-test). The accuracy gain ranges from $+9.0$ to $+12.3$ percentage points and the Brier gain from 0.15 to 0.22. At the elicited budget $(0.10, 1.0, 0.10)$ the gains are $+11.7$ percentage points and -0.221; the largest accuracy gain, $+12.3$, is obtained one level looser at $(0.20, 1.5, 0.20)$, and the largest AUC gain at the elicited budget itself. Removing the constraint entirely gives the weakest bounded-to-unbounded comparison: $+10.0$ percentage points, -0.150 Brier, and a negative average ΔAUC.

The three rightmost columns report the win rate, the percentage of folds in which the fine-tuned model strictly beats the expert on the corresponding metric. At the bounded budgets the algorithm improves accuracy and Brier score in 96–99% of folds, so that even in folds where absolute accuracy is low the fine-tuned model almost always outperforms the

Table 2: Per-fold paired improvement of Penalized EM (Pen-EM) and EB-GFTBN (Ours) over the expert baseline (Expert) at five relative-budget levels, across $K = 100$ disjoint folds. Paired differences (Δ = method − expert per fold) reported. Best ΔAcc, ΔBrier and ΔAUC in bold.

	Fidelity budget			**Paired improvement over expert**			**Win rate**		
Method	δ_θ	δ_μ	δ_{σ^2}	ΔAcc	ΔAUC	ΔBrier	Acc	AUC	Brier
Expert	–	–	–	53.06%	0.854	0.850	–	–	–
Pen-EM	–	–	–	+5.54±5.76	−0.0001±0.0182	−0.127±0.095	78%	44%	95%
EB-GFTBN	Unconstrained			+10.00±6.31	−0.0049±0.0248	−0.150±0.133	93%	46%	86%
	0.30	2.0	0.30	+11.74±6.49	+0.0057±0.0239	−0.206±0.110	96%	56%	97%
	0.20	1.5	0.20	**+12.33**±5.86	+0.0089±0.0222	−0.220±0.104	**99**%	70%	**99**%
	0.10	1.0	0.10	+11.67±6.05	**+0.0128**±0.0207	**−0.221**±0.103	97%	**72**%	**99**%
	0.05	0.5	0.05	+9.05±5.18	+0.0077±0.0156	−0.171±0.093	96%	70%	98%

All accuracy and Brier improvements are significant at $p < 10^{-3}$ (paired t-test, $K = 100$).

expert model evaluated on the same data. The AUC win rate is more modest (46–72%), reflecting small and noisy absolute differences, but the average paired ΔAUC is positive at every budget except the unconstrained one. The Penalised EM baseline also improves on the expert (ΔAcc = +5.54, ΔBrier = −0.127, with 78% and 95% win rates), but delivers roughly half the accuracy gain and also the lower Brier improvement compared to the EB-GFTBN and does not improve AUC on average ($-10^{-4} \pm 0.018$).

Table 3 shows the mean deviation of the maximum deviation in each fold, absolute for the discrete variables and relative for the continuous ones. At every moderate EB-GFTBN budget, the per-fold maximum CPT and variance deviations equal the budget exactly in all 100 folds; the near-zero standard deviation simply reflects that the budget is active on every run rather than fluctuating around it. The Gaussian-mean shift saturates the budget at the tighter δ_μ levels and is close to saturation at $\delta_\mu = 2.0$, indicating that the means would benefit from a larger window. The unconstrained EB-GFTBN row shows the contrast: average CPT deviation reaches 0.34, individual cells move by up to 0.50, and the relative variance deviation exceeds 4.6. The Penalised EM produces shifts of a similar magnitude (CPT ~ 0.32, relative variance ~ 3.5, worst-case 4.59), moving the parameters substantially away from the expert specification on every fold without any mechanism to control the drift.

Figure 2 shows the confusion matrix before and after fine-tuning at the expert-defined budget $(0.10, 1.0, 0.10)$.

6. Discussion

At the budget specified by the experts, the algorithm improves accuracy by more than 10 percentage points and reduces the Brier score by more than 20% relative to the expert baseline, while every parameter remains inside the elicited window. The additional budget of Section 5.2 serve to characterise how the algorithm behaves as the constraint is tightened or relaxed around it. The pattern is consistent. Tightening the window further leaves part of the data-driven signal unused. Loosening it degrades the results in both directions

Table 3: Expert-fidelity and computational cost of EB-GFTBN across $K = 100$ folds. For each parameter family we report the maximum per-fold maximum deviation from the expert specification, averaged across folds (mean ± std), together with the worst-case fold maximum.

	$\lvert\Delta\theta\rvert$		$\lvert\Delta\mu\rvert/\sigma^0$		$\lvert\Delta\sigma^2\rvert/\sigma^{2,0}$		**Epochs**
Budget $(\delta_\theta, \delta_\mu, \delta_{\sigma^2})$	mean	max	mean	max	mean	max	per fold
Penalized EM	0.321±0.002	0.325	1.09±0.06	1.24	3.471±0.722	4.589	–
Unconstrained	0.337±0.073	0.499	1.20±0.40	1.94	2.030±0.777	4.662	164±72
$(0.30, 2.0, 0.30)$	0.149±0.003	0.150	1.62±0.39	2.00	0.300±0.000	0.300	197±87
$(0.20, 1.5, 0.20)$	0.100±0.001	0.100	1.43±0.15	1.50	0.200±0.000	0.200	177±77
$(0.10, 1.0, 0.10)$	0.050±0.000	0.050	1.00±0.02	1.00	0.100±0.000	0.100	144±60
$(0.05, 0.5, 0.05)$	0.025±0.000	0.025	0.50±0.00	0.50	0.050±0.000	0.050	96±40

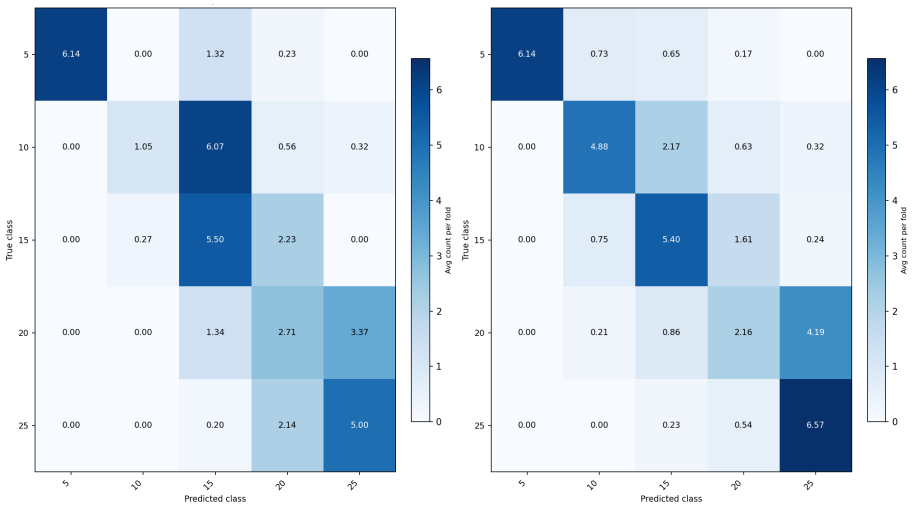


Figure 2: Confusion matrices before (left, expert model) and after (right, EB-GFTBN) fine-tuning at the expert define accuracy budget $(\delta_\theta, \delta_\mu, \delta_{\sigma^2}) = (0.10, 1.0, 0.10)$.

at once: performance falls back towards the unconstrained case while the parameters drift substantially away from the expert-elicited values, so the additional freedom mostly enables overfitting rather than better generalisation. The projection is therefore not a cost paid for interpretability: over this range it improves generalisation and preserves the elicited specification at the same time. The comparison with the Penalised EM baseline is of the same kind. A fixed combination of the expert and maximum-likelihood estimates uses the data less efficiently than following the Brier-score gradient, and offers no mechanism to control how far the result drifts from the expert model.

In veterinary forensic investigations, wound age is currently assessed manually by pathologists from macroscopical and microscopical changes (Ressel et al., 2016; Barington et al.,

2018; Pankoke et al., 2023; Barington et al., 2016; Bækgård et al., 2026), with only granulation tissue thickness measured directly using a ruler or caliper; the remaining variables are interpreted visually and therefore depend on the experience of the individual pathologist. In a legal context this subjectivity is problematic: wound age estimates are used to characterise the duration and severity of animal neglect, and may have serious consequences for individuals accused of violating the animal protection act. Multi-variable machine learning models such as the Bayesian network fine-tuned here provide a more objective and reproducible basis for these estimates, and the expert-bounded fine-tuning ensures that the model remains anchored in the consensus of domain experts. The result is a trustworthy and performant decision-support tool for wound age assessment, grounded in both expert knowledge and data.

The early-stopping criterion uses the training-set Brier score, so some overfitting is unavoidable. A held-out validation set would mitigate this, but the dataset is too small to afford one in addition to the test set. The saturation of the Gaussian-mean budget at every level suggests that the means would benefit from a wider window than the one elicited. EB-GFTBN has also been developed for a single structure (naive Bayes); further validation on additional datasets and on TAN and general DAGs will be needed to confirm the methodology. The two-factor form of (6) assumes only that each parameter enters a single conditional factor of one class, which continues to hold for TAN, so Table 1 applies unchanged. While the Brier score is a strictly proper scoring rule and avoids the plateaus of the original FTNB update, other objectives need testing, in particular the conditional log-likelihood used by ELR (Greiner et al., 2005).

7. Conclusion and Future Work

We have presented EB-GFTBN, a gradient-based fine-tuning algorithm for Bayesian network classifiers built from expert knowledge. The algorithm replaces the heuristic update of the original FTNB family by the exact derivative of the Brier score with natural-gradient preconditioning, and adds an expert-bounded projection that constrains each parameter to a relative window around its expert value. The methodology applies uniformly to discrete CPT entries and to the means and variances of conditional linear Gaussian densities. On a porcine wound age case study with 100-fold pig-disjoint cross-validation, the algorithm improves accuracy and Brier score over the expert baseline and the penalised EM baseline, while keeping every parameter inside its expert-defined window. The algorithm provides a trustworthy and performant decision-support model for wound age assessment, anchored in both expert knowledge and data.

A natural extension replaces the hard projection of Section 3.3 with a soft KL-divergence regulariser between the current and expert conditional densities, turning fine-tuning into a single unconstrained optimisation in which the strength of the regulariser plays the role of the budget. A second extension would make the budget parameter-specific rather than global: parameters supported by few observations, or elicited with high expert confidence, would be allowed to move less than those with abundant data or uncertain expert values.

Acknowledgments

The study was funded by The Independent Research Fund Denmark, grant ID 10.46540/2067-00003B.

References

Amel Alhussan and Khalil El Hindi. Fine tuning the tree augmented naïve Bayes (FTTAN) learning algorithm. In *SAI Intelligent Systems Conference*, pages 171–177, 2015.

Amel Alhussan and Khalil El Hindi. Selectively fine-tuning Bayesian network learning algorithm. *International Journal of Pattern Recognition and Artificial Intelligence*, 30 (8):1651005, 2016.

Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10 (2):251–276, 1998.

C Bækgård, L Andresen, MH Hansen, HE Jensen, K Freude, and K Barington. A multidisciplinary approach for the study of age-dependent expression patterns in porcine experimental wounds. *Forensic Sci Med Pathol*, 2026.

K Barington, K Dich-Jorgensen, and HE Jensen. A retrospective study of forensic cases of skin ulcerations in Danish pigs from 2000 to 2014. *Acta Vet Scand*, 58(1):48, 2016.

K Barington, K Dich-Jorgensen, and HE Jensen. A porcine model for pathomorphological age assessment of surgically excised skin wounds. *Acta Vet Scand*, 60(1):33, 2018.

K Barington, TB Nielsen, RS Andersen, and HE Jensen. Forensic necropsies of cattle: a study of Danish cases from 2010 to 2021 and a guideline for forensic examination of cattle. *J Comp Pathol*, 201:57–62, 2023.

Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

Houda Briwa, Anders L. Madsen, and Maria Chiara Leva. Predicting operators reliability for control room alarm management using knowledge-based Bayesian networks. *Reliability Engineering & System Safety*, 273:112261, 2026. ISSN 0951-8320. doi: 10.1016/j.ress.2026.112261.

Dima M. Diab and Khalil M. El Hindi. Using differential evolution for fine tuning naïve Bayesian classifiers and its application for text classification. *Applied Soft Computing*, 54:183–199, 2017.

Khalil El Hindi. Fine tuning the naïve Bayesian learning algorithm. *AI Communications*, 27(2):133–141, 2014.

Khalil El Hindi, Hussien AlSalman, Safwan Qasem, and Saad AlAhmadi. Building an ensemble of fine-tuned naïve Bayesian classifiers for text classification. *Entropy*, 20(11): 857, 2018.

Khalil El Hindi, Reem R. Aljulaidan, and Hussien AlSalman. Lazy fine-tuning algorithms for naïve Bayesian text classification. *Applied Soft Computing*, 96:106652, 2020.

Russell Greiner, Xiaoyuan Su, Bin Shen, and Wei Zhou. Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers. *Machine Learning*, 59(3):297–322, 2005.

Steffen L. Lauritzen and Frank Jensen. Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, 11(2):191–203, April 2001. doi: 10.1023/A:1008935617754.

M. C. Leva and P. F. Hansen. Use of Bayesian network for human reliability modelling: Possible benefits and an example of application. *Human Factors and Reliability Engineering for Safety and Security in Critical Infrastructures*, pages 189–223, 2017.

Joseph Mietkiewicz and Anders L. Madsen. Improvement of the fine-tuning algorithm for Bayesian network classifiers. In *Proceedings BNAIC/BeNeLearn, Joint International Scientific Conferences on AI and Machine Learning*, 2022.

Andrew Y. Ng and Michael I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *Advances in Neural Information Processing Systems 2*, 2002.

K Pankoke, SS Nielsen, BM Jørgensen, HE Jensen, and K Barington. Immunohistochemical study of CD31 and α-SMA expression for age estimation of porcine skin wounds. *J Comp Pathol*, 206:22–31, 2023.

Franz Pernkopf, Michael Wohlmayr, and Sebastian Tschiatschek. Maximum margin Bayesian network classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):521–532, 2012.

L Ressel, U Hetzel, and E Ricci. Blunt force trauma in veterinary forensic pathology. *Vet Pathol*, 53:941–961, 2016. doi: 10.1177/0300985816653988.

Jiang Su, Harry Zhang, Charles X. Ling, and Stan Matwin. Discriminative parameter learning for Bayesian networks. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 1016–1023, 2008.

RN Thornton and RD Jolly. The objective interpretation of histopathological data: an application to the ageing of ovine bruises. *Forensic Sci Int*, 31:225–239, 1986.

Huan Zhang and Liangxiao Jiang. Fine tuning attribute weighted naïve Bayes. *Neurocomputing*, 488:402–411, 2022.

Appendix A. Full Derivation of the Natural-Gradient Updates

This appendix derives in full the natural-gradient updates of Table 1 for both parameter families: the discrete CPT entries $\theta_{i,a,j}$ (Section A.2) and the means μ_{ij} and variances σ^2_{ij} of

the conditional linear Gaussian nodes (Section A.3). Both families share the two-factor form of the Brier gradient, which we derive first (Section A.1); each specialisation then reduces to substituting the score function and Fisher information of the relevant conditional factor. Throughout we use the instance-first posterior indexing $p_{kl} = P_{\mathcal{B}}(C = c_l \mid \mathbf{x}_k)$ of Section 2.

A.1. The two-factor Brier gradient

Differentiating the per-instance Brier score $B_k = \sum_{l=1}^{L} r_{kl}^2$ of (3), with residual $r_{kl} = p_{kl} - E_{kl}$, through the posteriors gives

$$\frac{\partial B_k}{\partial \phi} \;=\; 2 \sum_{l=1}^{L} r_{kl}\, \frac{\partial p_{kl}}{\partial \phi}. \tag{8}$$

Let $N_l = P(c_l) \prod_i p(x_{ki} \mid \mathrm{pa}_i; \boldsymbol{\Theta}_l)$ be the unnormalised posterior numerator and $Z = \sum_{l'} N_{l'}$, so $p_{kl} = N_l / Z$. With complete data and continuous nodes that have only discrete parents, every tunable parameter ϕ, a discrete CPT entry or a Gaussian mean or variance, enters exactly one factor of the single numerator N_j, where j indexes the class whose conditional factor contains ϕ. Hence

$$\frac{\partial N_j}{\partial \phi} = N_j\, s_\phi, \qquad \frac{\partial N_l}{\partial \phi} = 0 \;\; (l \neq j), \qquad s_\phi := \frac{\partial}{\partial \phi} \log p(x_{ki} \mid \mathrm{pa}_i; \boldsymbol{\Theta}_j), \tag{9}$$

where s_ϕ is the score function of that conditional factor. As only N_j depends on ϕ, the normaliser inherits $\partial Z / \partial \phi = N_j s_\phi$, and the quotient rule yields

$$\frac{\partial p_{kj}}{\partial \phi} = \frac{(N_j s_\phi) Z - N_j (N_j s_\phi)}{Z^2} = p_{kj}(1 - p_{kj})\, s_\phi, \tag{10}$$

$$\frac{\partial p_{kl}}{\partial \phi} = \frac{0 \cdot Z - N_l (N_j s_\phi)}{Z^2} = -\, p_{kl}\, p_{kj}\, s_\phi \quad (l \neq j). \tag{11}$$

Substituting (10)–(11) into (8) and factoring out p_{kj},

$$\frac{\partial B_k}{\partial \phi} = 2\, s_\phi \Big[r_{kj}\, p_{kj}(1 - p_{kj}) - p_{kj} \sum_{l \neq j} r_{kl}\, p_{kl} \Big] = \underbrace{2\, p_{kj}\, (r_{kj} - w_k)}_{\beta_k(j)}\, s_\phi, \tag{12}$$

where $w_k = \sum_l r_{kl}\, p_{kl}$ is the weighted mean residual of (4) and the last equality absorbs $r_{kj}\, p_{kj}$ into the sum. Equation (12) is the two-factor gradient (6): the instance-dependent Brier signal $\beta_k(j)$ times the family-dependent score s_ϕ.

A.2. Discrete CPT entries

For a discrete child X_i with discrete parents, the tunable parameter is a single CPT entry $\theta_{i,a,j} = P(X_i = a \mid \mathrm{pa}_i, C = c_j)$. When the instance has $x_{ki} = a$, this entry is the class-j conditional factor in N_j, so its log-factor is $\log \theta_{i,a,j}$ and the score is

$$s_\theta \;=\; \frac{\partial}{\partial \theta_{i,a,j}} \log \theta_{i,a,j} \;=\; \frac{1}{\theta_{i,a,j}}. \tag{13}$$

The categorical Fisher information of the active entry, with the expectation taken over $X_i \sim \theta_{i,\cdot,j}$, is

$$I(\theta_{i,a,j}) \;=\; \mathbb{E}\big[s_\theta^2\big] \;=\; \sum_{a'} \theta_{i,a',j} \left(\frac{\mathbf{1}[a'=a]}{\theta_{i,a,j}} \right)^2 \;=\; \frac{\theta_{i,a,j}}{\theta_{i,a,j}^2} \;=\; \frac{1}{\theta_{i,a,j}}. \tag{14}$$

Premultiplying by the inverse Fisher information, the parameter-dependent factors cancel exactly,

$$\tilde{g}_\theta \;=\; I(\theta_{i,a,j})^{-1}\, \beta_k(j)\, s_\theta \;=\; \theta_{i,a,j}\, \beta_k(j)\, \frac{1}{\theta_{i,a,j}} \;=\; \beta_k(j), \tag{15}$$

so the natural gradient is the bare Brier signal of Table 1. The gradient-descent update with learning rate $\eta > 0$ is

$$\theta_{i,a,j} \;\leftarrow\; \mathrm{clip}\big(\theta_{i,a,j} - \eta\, \beta_k(j),\; \varepsilon,\; 1-\varepsilon\big), \tag{16}$$

after which the remaining entries of the same conditional distribution are rescaled proportionally,

$$\theta_{i,a',j} \;\leftarrow\; \theta_{i,a',j}\, \frac{1-\theta_{i,a,j}^{\mathrm{new}}}{1-\theta_{i,a,j}^{\mathrm{old}}} \qquad (a' \neq a), \tag{17}$$

to restore $\sum_{a'} \theta_{i,a',j} = 1$ while preserving their relative proportions. The clip bound $\varepsilon > 0$ keeps every entry in $(0,1)$.

A.3. Gaussian parameters

For a conditional linear Gaussian child X_i with discrete parents, the class-j factor is the density

$$p(x_{ki} \mid \mathrm{pa}_i; \boldsymbol{\Theta}_j) \;=\; \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{(x_{ki}-\mu_{ij})^2}{2\sigma_{ij}^2} \right), \tag{18}$$

with log-density, dropping the additive constant,

$$\ell_{ij} \;=\; \log p(x_{ki} \mid \mathrm{pa}_i; \boldsymbol{\Theta}_j) \;=\; -\tfrac{1}{2}\log\sigma_{ij}^2 - \frac{(x_{ki}-\mu_{ij})^2}{2\sigma_{ij}^2} + \text{const}. \tag{19}$$

The two parameters μ_{ij} and σ_{ij}^2 are tuned through their score functions and Fisher information.

Score functions. Differentiating the log-density (19) with respect to the mean,

$$s_\mu \;=\; \frac{\partial \ell_{ij}}{\partial \mu_{ij}} \;=\; \frac{x_{ki}-\mu_{ij}}{\sigma_{ij}^2}. \tag{20}$$

Treating the variance σ_{ij}^2 as the differentiation variable in (19),

$$s_{\sigma^2} \;=\; \frac{\partial \ell_{ij}}{\partial \sigma_{ij}^2} \;=\; -\frac{1}{2\sigma_{ij}^2} + \frac{(x_{ki}-\mu_{ij})^2}{2\sigma_{ij}^4} \;=\; \frac{(x_{ki}-\mu_{ij})^2 - \sigma_{ij}^2}{2\sigma_{ij}^4}. \tag{21}$$

Fisher information. The per-coordinate Fisher information is $I(\phi) = \mathbb{E}\big[s_\phi^2\big]$, with the expectation taken over $X \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$. For the mean, using $\mathbb{E}\big[(X-\mu_{ij})^2\big] = \sigma_{ij}^2$,

$$I(\mu_{ij}) \;=\; \mathbb{E}\left[\frac{(X-\mu_{ij})^2}{\sigma_{ij}^4}\right] \;=\; \frac{\sigma_{ij}^2}{\sigma_{ij}^4} \;=\; \frac{1}{\sigma_{ij}^2}. \tag{22}$$

For the variance, the numerator of (21) has zero mean, and using the Gaussian fourth central moment $\mathbb{E}\big[(X-\mu_{ij})^4\big] = 3\sigma_{ij}^4$,

$$\mathbb{E}\Big[\big((X-\mu_{ij})^2 - \sigma_{ij}^2\big)^2\Big] \;=\; \mathbb{E}\big[(X-\mu_{ij})^4\big] - \sigma_{ij}^4 \;=\; 3\sigma_{ij}^4 - \sigma_{ij}^4 \;=\; 2\sigma_{ij}^4, \tag{23}$$

so that

$$I(\sigma_{ij}^2) \;=\; \frac{\mathbb{E}\Big[\big((X-\mu_{ij})^2 - \sigma_{ij}^2\big)^2\Big]}{4\sigma_{ij}^8} \;=\; \frac{2\sigma_{ij}^4}{4\sigma_{ij}^8} \;=\; \frac{1}{2\sigma_{ij}^4}. \tag{24}$$

Natural-gradient updates. The natural gradient (7) premultiplies the Euclidean gradient by the inverse Fisher information, $\tilde{g}_\phi = I(\phi)^{-1}\,\beta_k(j)\,s_\phi$, and the product $I(\phi)^{-1}s_\phi$ cancels the variance-dependent denominator of the score exactly. For the mean, combining (20) and (22),

$$\tilde{g}_\mu \;=\; \sigma_{ij}^2\;\beta_k(j)\;\frac{x_{ki}-\mu_{ij}}{\sigma_{ij}^2} \;=\; \beta_k(j)\,(x_{ki}-\mu_{ij}), \tag{25}$$

and for the variance, combining (21) and (24),

$$\tilde{g}_{\sigma^2} \;=\; 2\sigma_{ij}^4\;\beta_k(j)\;\frac{(x_{ki}-\mu_{ij})^2 - \sigma_{ij}^2}{2\sigma_{ij}^4} \;=\; \beta_k(j)\big[(x_{ki}-\mu_{ij})^2 - \sigma_{ij}^2\big]. \tag{26}$$

Equations (25)–(26) are the right-hand column of Table 1. The resulting gradient-descent updates with learning rate $\eta > 0$ are

$$\mu_{ij} \;\leftarrow\; \mu_{ij} - \eta\,\beta_k(j)\,(x_{ki}-\mu_{ij}), \tag{27}$$

$$\sigma_{ij}^2 \;\leftarrow\; \max\Big(\sigma_{ij}^2 - \eta\,\beta_k(j)\big[(x_{ki}-\mu_{ij})^2 - \sigma_{ij}^2\big],\;\varepsilon\Big), \tag{28}$$

where the variance floor $\varepsilon > 0$ keeps σ_{ij}^2 strictly positive. The mean is pulled toward the observed value x_{ki} in proportion to the miscalibration $\beta_k(j)$ of class j on the instance; the variance grows when the squared deviation $(x_{ki}-\mu_{ij})^2$ exceeds the current variance and shrinks otherwise. Neither update involves a small-denominator division, which is the purpose of the natural-gradient preconditioning (Amari, 1998).