Expert Opinion Elicitation for Assisting Deep Learning based Lyme Disease Classifier with Patient Data

Sk Imran Hossain, Jocelyn de Goër de Herve, David Abrial, Richard Emillion, Isabelle Lebert, Yann Frendo, Delphine Martineau, Olivier Lesens, Engelbert Mephu Nguifo

International Journal of Medical Informatics 193 (2025) 105682

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Expert opinion elicitation for assisting deep learning based Lyme disease classifier with patient data

Sk Imran Hossain [a], Jocelyn de Goër de Herve [b,c], David Abrial [b,c], Richard Emilion [d], Isabelle Lebert [b,c], Yann Frendo [a,b], Delphine Martineau [e], Olivier Lesens [f,g], Engelbert Mephu Nguifo [a,*]

[a] Université Clermont Auvergne, Clermont Auvergne INP, CNRS, ENSMSE, LIMOS, France
[b] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, France
[c] Université de Lyon, INRAE, VetAgro Sup, UMR EPIA, France
[d] Université d'Orléans, Institut Denis Poisson, France
[e] Infectious and Tropical Diseases Department, CHU Clermont-Ferrand, France
[f] Infectious and Tropical Diseases Department, CRIOA, CHU Clermont-Ferrand, France
[g] UMR CNRS 6023, Laboratoire Microorganismes: Génome Environnement (LMGE), UCA, France

ARTICLE INFO

Keywords:
Expert opinion elicitation
Erythema migrans
Lyme disease
Formal concept analysis

ABSTRACT

Background: Diagnosing erythema migrans (EM) skin lesion, the most common early symptom of Lyme disease, using deep learning techniques can be effective to prevent long-term complications. Existing works on deep learning based EM recognition only utilizes lesion image due to the lack of a dataset of Lyme disease related images with associated patient data. Doctors rely on patient information about the background of the skin lesion to confirm their diagnosis. To assist deep learning model with a probability score calculated from patient data, this study elicited opinions from fifteen expert doctors. To the best of our knowledge, this is the first expert elicitation work to calculate Lyme disease probability from patient data.

Methods: For the elicitation process, a questionnaire with questions and possible answers related to EM was prepared. Doctors provided relative weights to different answers to the questions. We converted doctors' evaluations to probability scores using Gaussian mixture based density estimation. We exploited formal concept analysis and decision tree for elicited model validation and explanation. We also proposed an algorithm for combining independent probability estimates from multiple modalities, such as merging the EM probability score from a deep learning image classifier with the elicited score from patient data.

Results: We successfully elicited opinions from fifteen expert doctors to create a model for obtaining EM probability scores from patient data.

Conclusions: The elicited probability score and the proposed algorithm can be utilized to make image based deep learning Lyme disease pre-scanners robust. The proposed elicitation and validation process is easy for doctors to follow and can help address related medical diagnosis problems where it is challenging to collect patient data.

1. Introduction

Lyme disease, a common tick-borne illness in Europe and North America, is caused by *Borrelia burgdorferi* sensu lato [1]. It typically presents with erythema migrans (EM) skin lesions in its early stage, which can resolve naturally, but the infection may spread to affect the nervous system, joints, heart, eyes, or skin [2–4]. Early-stage Lyme disease is treatable with antibiotics. Diagnosis in Europe and North America often involves a two-tier serology test to detect antibodies against *Borrelia burgdorferi* sensu lato [5,6]. However, serology is advised only in the absence of EM, as early tests may yield false negatives with low sensitivity (40-60%) [5]. Direct detection methods like culture (the gold standard), microscopy, and polymerase chain reaction have limitations: specialized requirements, infeasibility, and variable sensitivity, respectively [5,6]. Given these constraints, early EM detection is crucial to prevent long-term complications.

* Corresponding author.
E-mail address: engelbert.mephu_nguifo@uca.fr (E. Mephu Nguifo).

https://doi.org/10.1016/j.ijmedinf.2024.105682
Received 27 July 2024; Received in revised form 4 October 2024; Accepted 28 October 2024
Available online 31 October 2024

Recent studies show that incorporating multiple data modalities significantly improves artificial intelligence (AI) based model performance in medical diagnosis compared to single-modality analysis [7–10]. Current AI-based early Lyme disease prediction relies solely on EM lesion images, while doctors suggest including patient data for better accuracy [11,12]. Training a multimodal deep learning model with both images and patient data requires datasets linking lesion images to patient data. Although EM image datasets exist, creating multimodal datasets utilizing Electronic Health Records (EHR) is challenging due to:

- Privacy regulations making it difficult to obtain large-scale EHR data linked to images.
- EHRs not consistently collecting all relevant information for early Lyme disease diagnosis.
- Clinicians often relying on subtle cues and context not captured in EHRs.

Expert opinion elicitation is valuable when high-quality data is scarce [13]. It can include point estimates, uncertainty intervals, or probability distributions [14]. Expert opinion elicitation and aggregation processes can be classified into two categories: behavioral and mathematical approaches [15,14]. The behavioral approach tries to produce group consensus among experts whereas, the mathematical approach combines subjective probabilities from experts using mathematical methods (some form of averaging) [14].

Expert elicitation proved effective for medical diagnosis and decision making. For example, Van Der Gaag et al. [16] created a probabilistic network to describe the oesophageal cancer presentation characteristics and the pathophysiological mechanisms of invasion and metastasis by eliciting opinions from two experts. Saegerman et al. [17] engaged eleven European experts to rank drivers of bovine besnoitiosis emergence. Suleiman et al. [18] incorporated expert guesses to improve diagnosis-related group misclassification detection. Wilson et al. [13] elicited opinions from sixteen experts on untreated melanoma progression probabilities. Cadham et al. [14] provide a comprehensive review of expert elicitation in health research computational modeling.

In this study, we elicited opinions from fifteen expert hospital practitioners to create a model for calculating EM probability from patient data. A questionnaire was prepared with the help of experts, based on the questions doctors ask during EM diagnosis. Traditional expert elicitation process of collecting probability estimates for cases based on the questionnaire is time consuming and it is difficult for doctors to provide probability estimates for cases or distribution parameters. Thus, we used a more flexible approach of assigning relative weights to answers and converted these evaluations into EM probabilities using Gaussian mixture-based density estimation (described in Section 2.1). To validate the elicited probability model and explain its behavior to the experts we utilized formal concept analysis (described in Section 2.3) and decision tree (described in Section 2.4). We also developed an algorithm to combine the EM probability score from a deep learning image classifier with the elicited probability from patient data. The main contributions of this work are:

- First expert elicitation study on EM probability estimation from patient data for early Lyme disease diagnosis.
- Novel elicitation and validation process combining relative weighting, mixture model, and concept lattice.
- Algorithm for combining the probability score from a deep learning image classifier with elicited probability score from patient data.

The paper is organized as follows: Section 2 covers the theoretical background; Section 3 describes the expert elicitation process and results; Section 4 presents the model use case and probability combination strategy; finally, Section 5 provides concluding remarks.

2. Theoretical background

The required theoretical concepts to understand the rest of the paper are briefly described in the following subsections.

2.1. Gaussian mixture model

A Gaussian mixture model (GMM) is a probability density function expressed as a weighted sum of Gaussian components [19]. The mixture represents a normally distributed population, while the components represent subpopulations. For one-dimensional data, a GMM with M components is defined as:

$$\hat{f}_{GMM}(x) = \sum_{m=1}^{M} \varnothing_m \mathcal{N}(x|\,\mu_m, \sigma_m) \tag{1}$$

where, $\varnothing_m \geq 0$ is the mixture weight i.e. the probability of m-th component κ_m satisfying $\sum_{m=1}^{M} \varnothing_m = 1$ so that the total probability distribution normalizes to 1, and $\mathcal{N}(x|\,\mu_m, \sigma_m)$ is the distribution of a Gaussian component with mean μ_m and standard deviation σ_m defined as:

$$\mathcal{N}(x|\,\mu_m, \sigma_m) = \frac{1}{\sigma_m\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_m}{\sigma_m}\right)^2} \tag{2}$$

Expectation-Maximization, an iterative unsupervised learning technique can be used to determine the parameters of GMM [20]. Steps involved in Expectation-Maximization for n data points $X = \{x_t | t = 1, \ldots, n\}$ are:

- Guess initial values for GMM parameters denoted by $\hat{\mu}_m, \hat{\sigma}_m$, and $\hat{\varnothing}_m$ respectively.
- Expectation step: calculate $\hat{\gamma}_{t,m}$, the probability of a point x_t being generated by κ_m

$$\hat{\gamma}_{t,m} = \frac{\hat{\varnothing}_m \mathcal{N}\left(x_t \middle|\, \hat{\mu}_m, \hat{\sigma}_m\right)}{\sum_{r=1}^{M} \hat{\varnothing}_r \mathcal{N}\left(x_t \middle|\, \hat{\mu}_r, \hat{\sigma}_r\right)} \tag{3}$$

- Maximization step: Update GMM parameters using the following equations:

$$\hat{\mu}_m = \frac{\sum_{t=1}^{n} \hat{\gamma}_{t,m} x_t}{\sum_{t=1}^{n} \hat{\gamma}_{t,m}} \tag{4}$$

$$\hat{\sigma}_m = \sqrt{\frac{\sum_{t=1}^{n} \hat{\gamma}_{t,m}(x_t - \hat{\mu}_m)^2}{\sum_{t=1}^{n} \hat{\gamma}_{t,m}}} \tag{5}$$

$$\hat{\varnothing}_m = \sum_{t=1}^{n} \frac{\hat{\gamma}_{t,m}}{n} \tag{6}$$

- Repeat Expectation and Maximization steps until the total likelihood L converges, where

$$L = \prod_{t=1}^{n} \hat{f}_{GMM}(x_t) \tag{7}$$

Information criterion tests like Akaike Information Criteria (AIC) [21] and Bayesian Information Criteria (BIC) [22] help select the optimal GMM by penalizing free parameters to avoid overfitting. AIC and BIC are defined as:

$$AIC = 2p + 2\ln L \tag{8}$$

$$BIC = p\ln L + 2\ln L \tag{9}$$

where p epresents the number of free parameters, and ln denotes the natural logarithm. The GMM with the lowest AIC and BIC values is preferred.

2.2. Kernel density estimation

Kernel density estimation (KDE) is a non-parametric method for estimating the probability density function of an independent and identically distributed random variable [23,24]. For n data points $X = \{x_t | t = 1, \ldots, n\}$, KDE is:

$$\hat{f}_{KDE}(x) = \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{x - x_t}{h}\right) \tag{10}$$

where h is bandwidth, and K is kernel function. For a Gaussian kernel, the bandwidth can be chosen via Silverman's rule [25]:

$$h = 0.9min\left(\hat{\sigma}, \frac{IQR}{1.34}\right) n^{\frac{-1}{5}} \tag{11}$$

where IQR is interquartile range, and $\hat{\sigma}$ is the sample standard deviation.

2.3. Formal concept analysis and concept lattice

Formal concept analysis (FCA) generates a formal concept hierarchy from a set of objects and their attributes [26]. FCA is widely applied in machine learning and bioinformatics [27–29]. Each concept in FCA represents objects sharing a specific attribute set. FCA computes concept lattice — a directed acyclic graph ordering all formal concepts derived from tabular data.

The central notion in FCA is the formal context, a triple $\langle O, Y, I \rangle$ where O is the set of objects, Y is the set of attributes, and incidence $I \subseteq O \times Y$ is a binary relation. A pair $\langle A, B \rangle$ is a formal concept of $\langle O, Y, I \rangle$ if $A \subseteq O$, $B \subseteq Y$, $A^{\uparrow} = B$, and $B^{\downarrow} = A$ where

$$A^{\uparrow} = \{y \in Y | for\ each\ o \in A : \langle o, y \rangle \in I\}\ and$$

$$B^{\downarrow} = \{o \in O | for\ each\ y \in B : \langle o, y \rangle \in I\}$$

A is the extent and B is the intent of concept $\langle A, B \rangle$. Formal concepts are ordered by the subconcept-superconcept relation:

$$\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle \iff A_1 \subseteq A_2 (\iff B_2 \subseteq B_1) \tag{12}$$

For a formal context $\langle O, Y, I \rangle$ the set $\mathfrak{B}(O, Y, I)$ of all formal concepts with the ordering in Equation (12) is the concept lattice. The last section of Supplementary Research Data file in Appendix A explains a sample formal context and corresponding concept lattice.

2.4. Decision tree

A decision tree is a supervised learning algorithm used for both regression and classification [30,31]. This study focuses on its use for classification. A decision tree classifies instances by recursively partitioning the instance space based on splitting rules, which are easy to visualize and interpret [30,31]. It is a directed tree with the root having no incoming edges, and each node has one incoming edge. Leaf (or terminal) nodes have no outgoing edges, while internal (or test) nodes divide the instance space into sub-spaces based on input attributes. Each decision node is assigned a class corresponding to the best target value, and instances are classified by navigating from the root to a leaf.

3. Elicitation method

The following subsections detail our expert elicitation process, including expert recruitment, questionnaire design, opinion collection, and elicitation methods.

3.1. Expert selection

The recruited experts are hospital practitioners specializing in infectious diseases or dermatology at France's reference centers for tick-borne diseases, Centres de Référence des Maladies Vectorielles liées aux Tiques (CRMVT) [32]. At a CRMVT steering committee meeting in June 2021, Professor Olivier Lesens (Infectious and Tropical Diseases Department, CRIOA, CHU Clermont-Ferrand, France) emphasized the importance of expert elicitation for calculating EM probability based on patient data. He invited interested experts to participate, and fifteen agreed. No monetary compensation was provided. The Elicitation Survey Data section of the Supplementary Research Data file in Appendix A contains a list of the reference centers and the number of participating experts.

3.2. Questionnaire and experts' evaluation

For EM probability elicitation, a questionnaire was developed based on questions typically asked by physicians about the onset and progression of skin lesions when diagnosing EM. This was inspired by a previous study on EM data collection in rural France [33]. The questionnaire was finalized during several meetings in April 2020, involving CRMVT doctors in Clermont-Ferrand and tick ecology experts from the French national research institute for agriculture, food and the environment (INRAE) [34]. At the June 2021 meeting, experts acknowledged the numerous possible cases from combining questions and answers, making it difficult and time-consuming to estimate probabilities for all cases. Instead, they agreed to assign relative weights independently to each answer, ranging from −1 to +3 (a higher value represents higher contribution of the answer towards the possibility of EM). Experts were emailed detailed instructions to provide weight attributions independently. The Elicitation Survey Data section of the Supplementary Research Data file in Appendix A contains the questions, answers, and expert weight attributions.

After receiving the experts' weight attributions, a meeting was held in November 2021, where the experts agreed that fever, fatigue, faintness, and headache should contribute equally if any of these symptoms were present, with their contribution being the average of the four. As a result, these four answers were consolidated into one, reducing the possible cases from 12,288 to 1,536. This modification is detailed in Table 1.

3.3. Opinion elicitation

Following are some notations used in the rest of the manuscript:

- Set of doctors, $D = \{d_e | e = 1, \ldots, 15\}$
- Set of questions, $Q = \{q_i | i = 1, \ldots, 6\}$
- Set of possible cases, $C = \{c_l | l = 1, \ldots, 1536\}$
- Total number of answers corresponding to q_i question $= n_{q_i}$
- j^{th} answer corresponding to q_i question,

$$a_{j,q_i} = \begin{cases} 1, & \text{if the answer is true} \\ 0, & \text{otherwise} \end{cases}, \text{where } j = 1, \ldots, n_{q_i}$$

- Weight assigned by doctor d_e to a_{j,q_i} answer $= w_{d_e, a_{j,q_i}}$

First, we summarized each of the 1,536 possible cases as a weight sum s_{c_l} as shown in Equation (13).

$$s_{c_l} = \sum_{i=1}^{|Q|} \sum_{j=1}^{n_{q_i}} a_{j,q_i} \times \left(\frac{1}{|D|} \sum_{d=1}^{|D|} w_{d_e, a_{j,q_i}}\right) \tag{13}$$

The set of case weight sum is defined as $S = \{s_{c_l} | l = 1, \ldots, 1536\}$. Then, we normalized each case weight sum with min-max normalization as shown in Equation (14).

$$\tilde{s}_{c_l} = \frac{s_{c_l} - \min(S)}{\max(S) - \min(S)} \tag{14}$$

The set of min-max normalized case weight sum is defined as $\tilde{S} = \{\tilde{s}_{c_l} | l = 1, \ldots, 1536\}$. We proposed three approaches to the experts to

Table 1
Weight modified questionnaire and doctors' weight attribution for erythema migrans. The assigned weight values are in the range −1 to +3 (a higher value represents a higher contribution of the answer towards the possibility of the erythema migrans). d_1 to d_{15} represents the doctors.

Question	Answer	Weight Assigned by Doctors (Doctor's Evaluation)															
		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	Average
Other symptoms observed alongside the skin lesion (q_1)	No (a_{1,q_1})	0	0	3	0	0	1	2	1	2	1	2	1	1	2	3	1.27
	Fever/ Fatigue/ Faintness/ Headache (a_{2,q_1})	-0.25	0.25	0	0.75	0.75	0.25	0.75	0.75	0.75	1	1	0.25	1	0.25	0	0.5
	Joint pain (a_{3,q_1})	1	1	-1	2	2	1	1	1	1	1	1	1	1	0	0	0.87
	Itching (a_{4,q_1})	-1	-1	-1	-1	1	-1	0	0	0	1	-0.5	-1	-1	1	0	-0.3
What was the maximum size of the red rash (q_2)	< 1 cm (a_{1,q_2})	-1	-1	0	-1	-1	-1	-1	0	0	-1	-1	-1	-1	1	-1	-0.67
	1 to 5 cm (a_{2,q_2})	1	1	1	0	1	0	1	1	2	1	1	1	1	2	1	1
	> 5 cm (a_{3,q_2})	3	2	2	2	3	2	3	2	1	3	2	2	3	3	3	2.4
	I do not know (a_{4,q_2})	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0
Is the size of the red rash currently increasing, or has it already increased gradually over time and now stabilized (q_3)	Yes (a_{1,q_3})	3	1	3	3	3	3	3	3	2	3	3	3	3	3	3	2.8
	No (a_{2,q_3})	0	-1	-1	-1	-1	-1	-1	0	-1	0	-1	-1	-1	1	-1	-0.67
	I do not know (a_{3,q_3})	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0.07
Have you seen a tick bite on this red rash in the past 30 days (q_4)	Yes (a_{1,q_4})	3	2	3	2	3	1	3	3	2	3	2	3	1	3	3	2.47
	No (a_{2,q_4})	0	0	0	0	0	1	0	1	0	0	-0.5	-1	0	1	0	0.1
Frequency of tick bites in the last 30 days before the appearance of the red rash (q_5)	Never (a_{1,q_5})	-1	-1	0	0	-1	0	0	0	0	0	-1	-1	0	-1	0	-0.4
	1 time (a_{2,q_5})	0	0	2	1	1	1	1	1	2	1	1	1	1	2	1	1.07
	2 to 5 times (a_{3,q_5})	1	1	3	1	1	1	1	2	2	1	2	1	1	3	1	1.47
	> 5 times (a_{4,q_5})	2	2	1	2	2	1	1	2	2	2	3	1	1	3	2	1.8
Outdoor activities in the last 30 days before the onset of the red rash (q_6)	Yes (a_{1,q_6})	1	1	2	2	1	1	2	2	2	2	2	2	1	3	2	1.73
	No (a_{2,q_6})	-1	-1	-1	-1	-1	0	-1	1	-1	-1	-1	-1	0	-1	0	-0.67

Table 2
Parameters of Gaussian Mixture Model used to model the density of min-max normalized weight sum of erythema migrans cases. $\varnothing$, μ, and σ represent mixture weight, mean and standard deviation respectively.

Parameter name	Value
Components	2
$\varnothing_1$	0.364801
$\varnothing_2$	0.635199
μ_1	0.359548
μ_2	0.572878
σ_1	0.128782
σ_2	0.156241

convert the normalized case weight sum to a probability score for EM. The following subsections explain the three approaches.

3.3.1. Cumulative probability from density estimate based on GMM

We modeled the density of our normalized weight sum data using a two-component GMM. This choice was based on the intuition that the data contains two subpopulations: one ill and one not ill. The selection was further supported by AIC and BIC values. Table 2 lists the selected GMM parameters. The blue curve in Fig. 1 shows the estimated density function using GMM. We defined the cumulative probability [35] of a normalized case weight sum from the GMM density estimate as the probability of EM as shown in Equation (15).

$$\hat{F}_{GMM}(x) = \int_{-\infty}^{x} \left(\sum_{m=1}^{2} \varnothing_m \mathcal{N}(x|\,\mu_m, \sigma_m) \right) dx \tag{15}$$

3.3.2. Posterior probability of a case belonging to the ill subpopulation of GMM

The first and second components of our GMM are shown in Fig. 1 with green and orange dotted lines, respectively. Assuming the second component represents the ill subpopulation, the posterior probability of a normalized case weight sum belonging to this component [19] can be defined as the EM probability, as shown in Equation (16).

$$p(\kappa_2|x) = \frac{\varnothing_2 \mathcal{N}(x|\,\mu_2, \sigma_2)}{\sum_{m=1}^{2} \varnothing_m \mathcal{N}(x|\,\mu_m, \sigma_m)} \tag{16}$$

3.3.3. Cumulative probability from density estimate based on KDE

We used a Gaussian kernel with bandwidth, $h = 0.03676$ on our 1,536 data points for the probability density estimation of the normalized weight sum variable as shown in Equation (17).

$$\hat{f}_{KDE}(x) = \frac{1}{1536 \times 0.03676} \sum_{l=1}^{1536} \frac{1}{2\pi} e^{-0.5\left(\frac{x-\tilde{x}_l}{0.03676}\right)^2} \tag{17}$$

The red curve in Fig. 1 shows the estimated density function. We defined the cumulative probability of a normalized case weight sum as the probability of having EM as shown in Equation (18).

$$\hat{F}_{KDE}(x) = \int_{-\infty}^{x} \hat{f}_{KDE}(x)\, dx \tag{18}$$

3.4. Elicitation result and analysis

We calculated the EM probability score for all cases using the three approaches described in Sections 3.3.1, 3.3.2, and 3.3.3, and presented the results to experts in a May 2022 meeting. Fig. 2 shows the EM probability plot for all cases using these approaches. The blue and red lines represent probability scores from the Gaussian mixture model (approach 1) and kernel density estimate (approach 2), respectively. The orange line shows the posterior probability of a case belonging to the ill subpopulation (approach 3). Since both approaches 1 and 2 rely on density estimates, their results are similar, while scores from approach 3 are consistently higher. The experts reached a consensus on using approach 1 (Section 3.3.1) due to its smoother density estimate compared to approach 2 (Section 3.3.3).

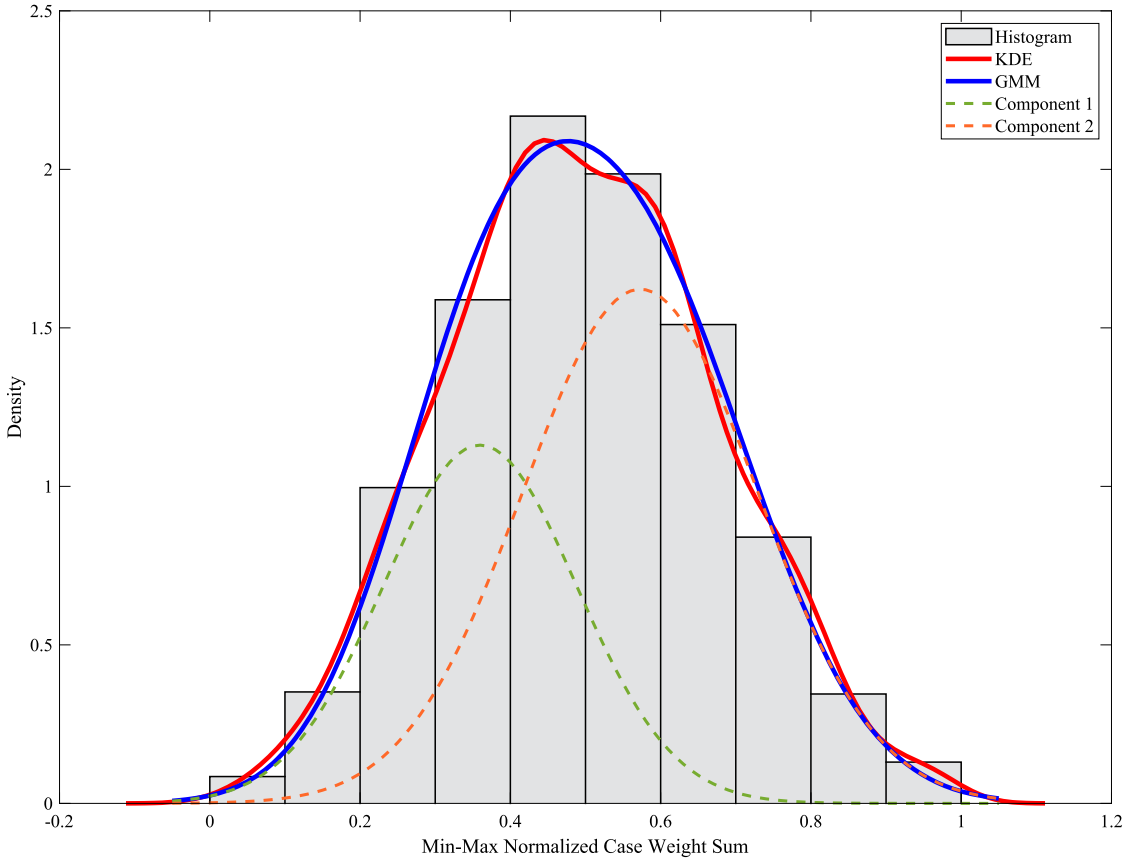


Fig. 1. Proposed approaches for expert opinion elicitation. GMM and KDE stand for Gaussian mixture model and kernel density estimation respectively. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

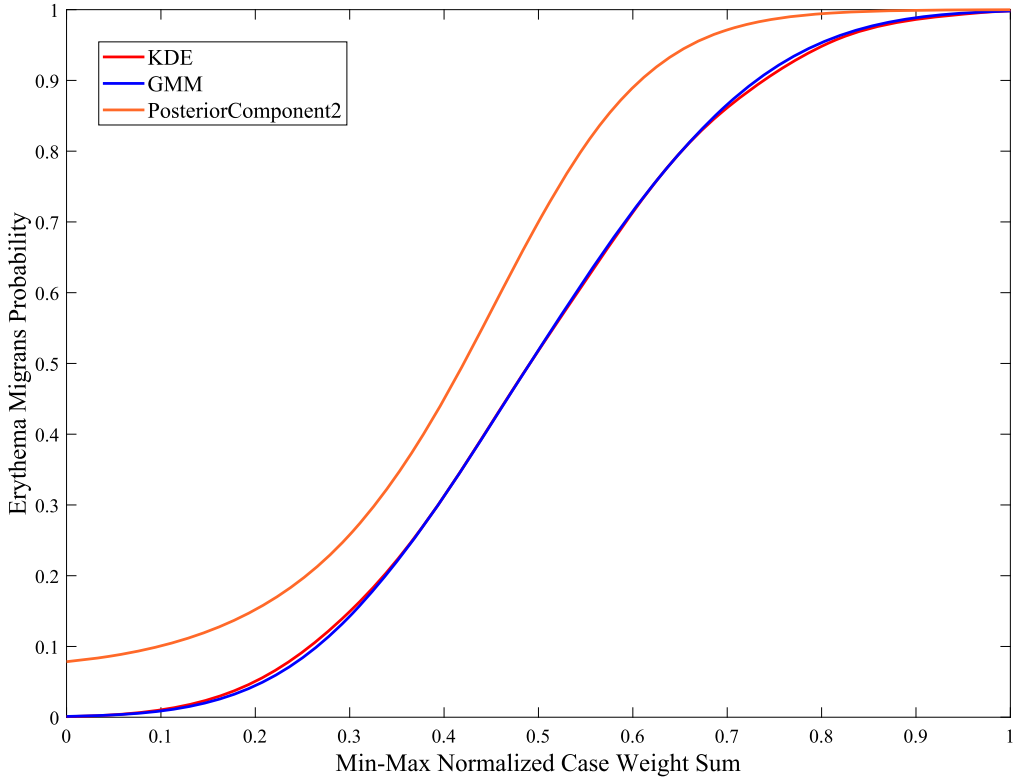


Fig. 2. Elicited erythema migrans probability plot. Blue and red lines represent the probability scores based on density estimates from Gaussian mixture model and kernel density estimate respectively. Orange line represents probability scores based on the posterior probability of a case belonging to the second component i.e. the ill subpopulation of the Gaussian mixture model. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

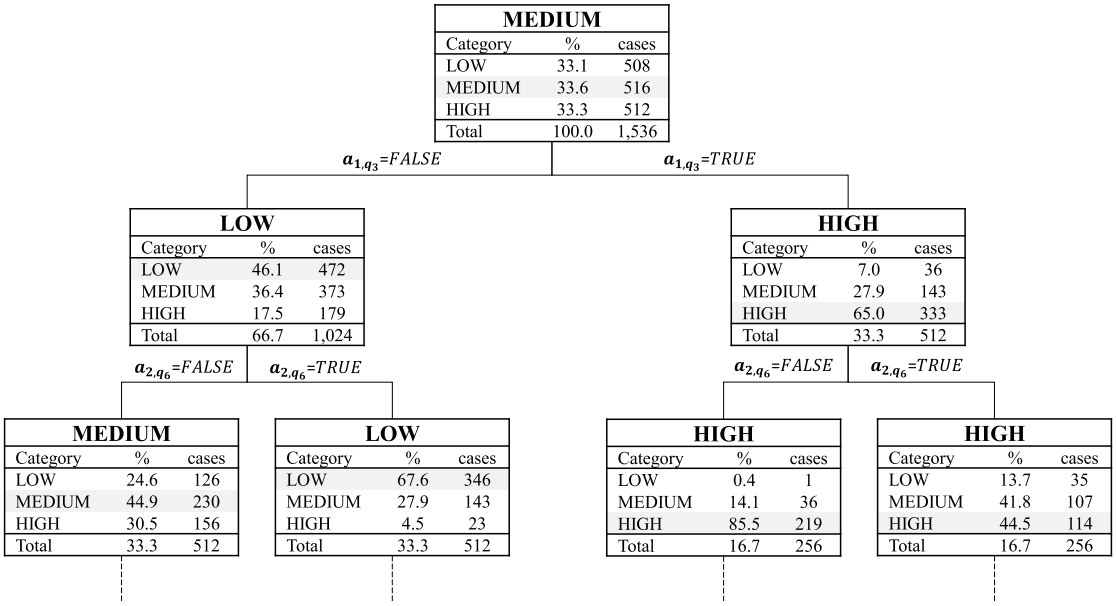


Fig. 3. Pruned decision tree explaining elicited erythema migrans probability model behavior. Each node shows the majority category along with percentage and number of cases belonging to each category. Refer to Table 1 for details about the questions and answers.

To validate the elicited model and explain its behavior to the experts, we first used decision trees. We categorized the calculated EM probability scores into three groups: LOW ($[0, 0.33)$), MEDIUM ($[0.33, 0.68)$), and HIGH ($[0.68, 1]$). Fig. 3 shows a pruned decision tree for approach 1, where each node displays the majority category, along with the percentage and number of cases in each category. The tree shows that the model assigns a HIGH EM probability whenever the first answer, "yes" to the third question, "*Is the size of the red rash currently increasing, or has it already increased gradually over time and now stabilized*", a_{1,q_3} is true. This supports the doctors' opinion, as most gave the highest weight to this answer.

To further explain the behavior of the model we utilized formal concept analysis (FCA) to find out questions and answers important for different probability groups. Fig. 4 shows a simplified FCA lattice view for the 162 cases belonging to the lowest probability score group in the range $[0, 0.1)$ obtained from approach 1. In the figure, the top box of a node represents an attribute (answer) or a number of attributes, which are connected by lines, and the bottom box represents how many objects (cases) contain the corresponding attribute shown in the top box. In Fig. 4, we start with 162 cases in the root node. At the first level, the number inside the bottom box of a node represents how many cases out of 162 cases contain the corresponding answer shown in the top box. For example, the "*no*" answer to the question "*Outdoor activities in the last 30 days before the onset of the red rash*", a_{2,q_6} is present in 145 cases. At the second level, each node represents how many cases contain two answers connected by a line. For example, a_{2,q_4} and a_{2,q_6} are jointly true in 128 cases. The rest of the FCA lattice is organized similarly. We can see from the figure that the answers common to most of these cases are the ones having lowest assigned weights or the opposites of the answers having highest assigned weights by most of the doctors.

The elicited EM probability scores for all possible cases, detailed decision tree, and FCA context files for different probability score groups are available at the link stated in the first section of the Supplementary Research Data file in Appendix A.

4. Model use case: combining probabilities from image and patient data

Our EM probability model, derived from patient data, is designed to complement deep learning-based EM image classifiers. This model will assist in developing effective tools for the early diagnosis of Lyme disease when patients present with suspicious skin lesions. It can be particularly useful in primary care or telemedicine settings, guiding clinicians in determining whether further testing or referral is necessary. Hossain et al. [12] trained 23 deep convolutional neural networks (CNNs) on an EM image dataset annotated by multiple expert dermatologists. Upon reviewing images misclassified by most CNNs, dermatologists identified errors in some initial annotations. This indicates that certain images can be challenging to classify, even for experts, without additional context. Additionally, uncalibrated CNNs may produce high confidence scores even for incorrect classifications. Experts recommend that the EM probability from CNNs should not be prioritized over patient data and probability from patient data should have veto power over image data. Let p_{image} and p_{data} represent the probabilities from image and patient data, respectively. The combined probability, $p_{combined}$, calculated using the geometric mean, $\sqrt{p_{image} \times p_{data}}$, ensures veto power for both. However, based on expert guidance, we aim to retain veto power only for patient data. To achieve this, we adjusted p_{image} in the lower probability range using the transformation in Equation (19). This transformation is popular in the literature of forecast probability aggregation for making the forecasts less or more extreme [36–38].

$$\tilde{p}_{image} = \frac{{p_{image}}^{\vartheta_{image}}}{{p_{image}}^{\vartheta_{image}} + \left(1 - p_{image}\right)^{\vartheta_{image}}} \tag{19}$$

The adjustment factor ϑ_{image} was set to 0.2 so that a very low value of p_{image} does not pull down $p_{combined}$ too much. This value was selected based on expert's suggestion to ensure that $p_{combined}$ will be at least 50% if $p_{data} \geq 90\%$. The adjustment of p_{image} is shown in Fig. 5c. The plot of $p_{combined}$ after the adjustment of p_{image} is shown in Fig. 5b. From the figure, we can see that the veto power was retained for p_{data} while effectively revoking it from p_{image}. As geometric mean uses multiplication we replaced a zero value of p_{image} or p_{data} with a small value of 0.1 to avoid a zero value of $p_{combined}$.

The generalized steps involved in our strategy for combining probabilities from image and patient data are shown in Algorithm 1. The notations, inputs, and outputs are listed at the beginning of the algorithm. First, a zero value of probability from image p_{image} or patient data p_{data} is replaced by a small value ϵ to make sure the combined probability $p_{combined}$ does not become zero because of the geometric mean. Then, p_{image} and p_{data} are transformed using the transform function.

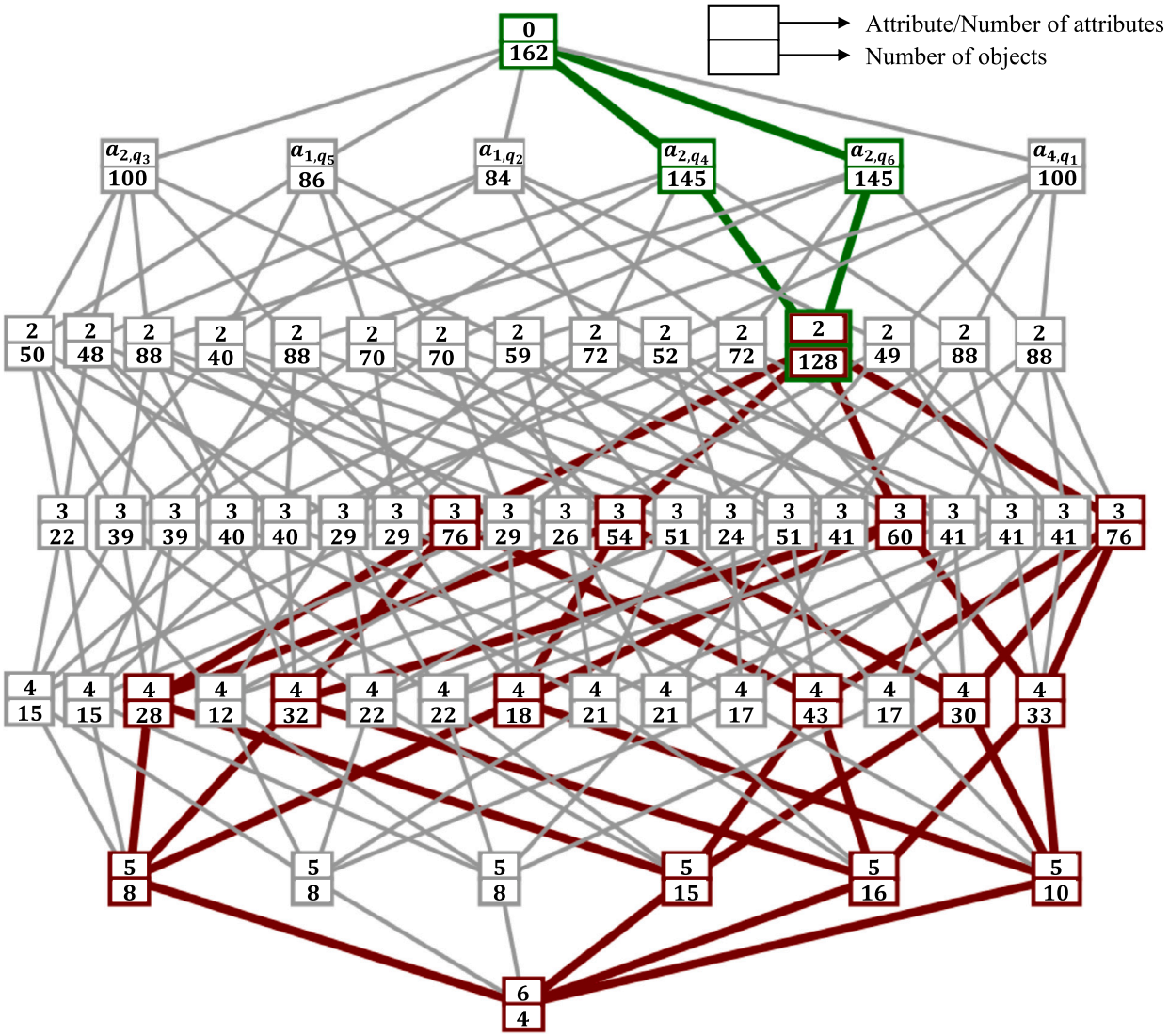


Fig. 4. Formal concept lattice view for 162 very low probability score cases in the range $[0, 0.1)$. The top box of a node represents an attribute (answer) or a number of attributes, which are connected by lines, and the bottom box represents how many objects (cases) contain the corresponding attribute shown in the top box. Refer to Table 1 for details about the questions and answers.

The transform function uses Equation (19) to make the input probability less or more extreme based on the transforming factor if the input probability falls within the user-defined range. Finally, the combined probability $p_{combined}$ is calculated using the geometric mean of transformed probabilities from image $\tilde{p}_{image}$ and patient data $\tilde{p}_{data}$. Geometric mean ensures veto power for the modalities which can be adjusted using the transformation with suggestions from domain experts.

5. Conclusion

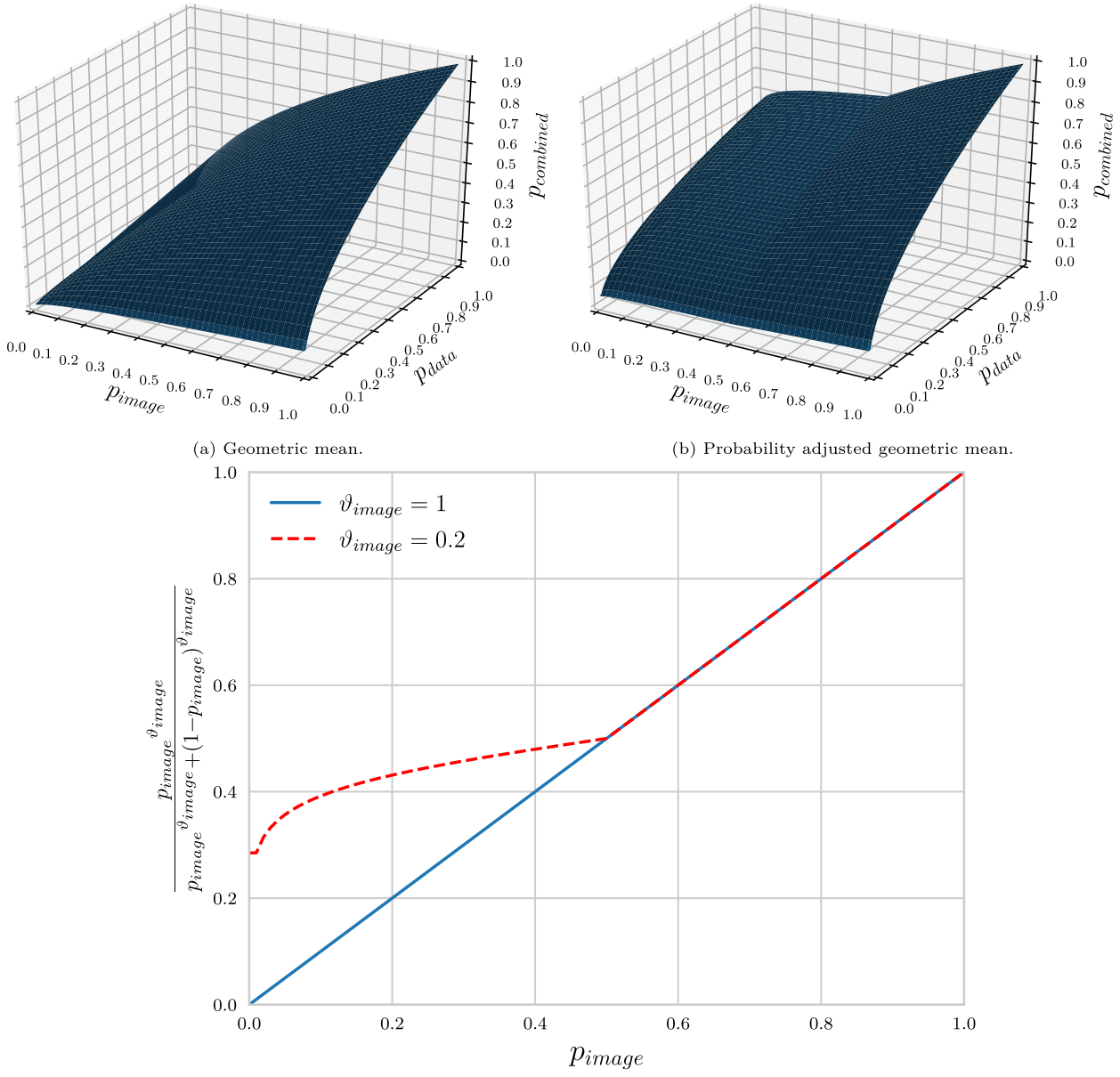
In this study, we successfully elicited opinions from fifteen expert doctors to create a model for obtaining EM probability scores from patient data. By incorporating expert opinions through a carefully developed questionnaire and adapting the survey based on expert feedback, we integrated qualitative insights into our model development. This combination of qualitative and quantitative approaches strengthens the methodology and makes it particularly effective for clinical care applications. Additionally, we proposed a strategy of combining EM probabilities from both image and patient data to address data scarcity, aiding in the creation of an effective Lyme disease pre-scanner system. The proposed techniques of questionnaire based opinion elicitation and combining probabilities from image and patient data will be useful for other diseases with similar requirements.

Summary Table

What is already known	What this study adds
• Existing deep learning based Image only analysis without patient data is not sufficient for early Lyme disease diagnosis. Moreover, collecting patient data is challenging.	• The elicited Lyme disease probability model and proposed algorithm for combining probabilities from multiple modalities can address the lack of patient data and make deep learning based diagnosis robust.
• It is difficult and time consuming for doctors to provide probability estimates for all possible cases in questionnaire based opinion elicitation process.	• Our proposed approach of relative weight assignment to the answers to the questions is easier for the experts.
• Intuitive explanation and validation of the questionnaire based elicited model are important.	• Formal concept lattice view can be an effective tool for explaining and validating questionnaire based opinion elicited model.

CRediT authorship contribution statement

Sk Imran Hossain: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Jocelyn de Goër**



(a) Geometric mean.

(b) Probability adjusted geometric mean.

(c) Transformed image probability. ϑ_{image} is the transformation factor.

Fig. 5. Combining erythema migrans probabilities from image and patient data. p_{image} and p_{data} represent probabilities from image and patient data.

de Herve: Writing – review & editing, Resources, Conceptualization. **David Abrial:** Software, Methodology, Formal analysis. **Richard Emilion:** Writing – review & editing, Methodology. **Isabelle Lebert:** Writing – review & editing, Project administration, Data curation, Conceptualization. **Yann Frendo:** Software, Methodology. **Delphine Martineau:** Validation, Project administration, Data curation, Conceptualization. **Olivier Lesens:** Validation, Supervision, Resources, Funding acquisition, Data curation, Conceptualization. **Engelbert Mephu Nguifo:** Writing – review & editing, Methodology, Supervision, Resources, Conceptualization, Funding acquisition.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the experts who participated in the expert opinion elicitation process. This work is supported by the European Regional Development Fund, project DAPPEM-AV0021029. The DAPPEM (Développement d'une APPlication d'identification des Erythèmes Migrants à partir de photographies) project, coordinated by Olivier Lesens, was carried out under the Call for Proposal 'Pack Ambition Research' from the Auvergne-Rhône-Alpes region, France. Mutualité Sociale Agricole (MSA), France also partially funded this research work.

Algorithm 1: Combining probabilities from image and patient data.

Input :
Probability estimate from lesion image: $p_{image} \in [0,1]$
Probability estimate from patient data: $p_{data} \in [0,1]$
Factor for transforming p_{image}: ϑ_{image}
Factor for transforming p_{data}: ϑ_{data}
Value used to avoid zero probability: $\epsilon \in (0,1]$
Range beginning for transforming p_{image}: $b_{image} \in [0,1]$
Range end for transforming p_{image}: $e_{image} \in [0,1]$
Range beginning for transforming p_{data}: $b_{data} \in [0,1]$
Range end for transforming p_{data}: $e_{data} \in [0,1]$
Output:
Combined probability: $p_{combined} \in [0,1]$
begin
  if $p_{image} = 0$ **then**
    $p_{image} \leftarrow \epsilon$ `// avoiding zero probability from image modality`
  if $p_{data} = 0$ **then**
    $p_{data} \leftarrow \epsilon$ `// avoiding zero probability from patient data modality`
  $\tilde{p}_{image} \leftarrow$ `transform` $(p_{image}, \vartheta_{image}, b_{image}, e_{image})$ `// transform` p_{image}
  $\tilde{p}_{data} \leftarrow$ `transform` $(p_{data}, \vartheta_{data}, b_{data}, e_{data})$ `// transform` p_{data}
  $p_{combined} \leftarrow \sqrt{\tilde{p}_{image} \times \tilde{p}_{data}}$ `// geometric mean`
  return $p_{combined}$
Function *transform* (p, ϑ, b, e)
  if $p \geq= b$ *and* $p \leq= e$ **then**
    $p \leftarrow \frac{p^{\vartheta}}{p^{\vartheta}+(1-p)^{\vartheta}}$ `// transformation in specified range`
  return p

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2024.105682.

References

[1] A.R. Marques, F. Strle, G.P. Wormser, Comparison of lyme disease in the United States and Europe, Emerg. Infect. Dis. 27 (8) (2021) 2017–2024, https://doi.org/10.3201/eid2708.204763.
[2] E.D. Shapiro, Clinical practice. Lyme disease, N. Engl. J. Med. 370 (18) (2014) 1724–1731, https://doi.org/10.1056/NEJMcp1314325.
[3] P. Burlina, N. Joshi, E. Ng, S. Billings, A. Rebman, J. Aucott, Skin Image Analysis for Erythema Migrans Detection and Automated Lyme Disease Referral, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, vol. 11041, 2018, pp. 244–251.
[4] F. Strle, G. Stanek, Clinical Manifestations and Diagnosis of Lyme Borreliosis, Current Problems in Dermatology, vol. 37, KARGER, Basel, 2009, pp. 51–110, https://www.karger.com/Article/FullText/213070.
[5] C. Eldin, A. Raffetin, K. Bouiller, Y. Hansmann, F. Roblot, D. Raoult, P. Parola, Review of European and American guidelines for the diagnosis of Lyme borreliosis, Méd. Mal. Infect. 49 (2) (2019) 121–132, https://doi.org/10.1016/j.medmal.2018.11.011.
[6] G. Trevisan, S. Bonin, M. Ruscio, A practical approach to the diagnosis of Lyme Borreliosis: from clinical heterogeneity to laboratory methods, Front. Med. 7 (2020) 265, https://doi.org/10.3389/fmed.2020.00265.
[7] A.G.C. Pacheco, R.A. Krohling, An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification, IEEE J. Biomed. Health Inform. 25 (9) (2021) 3554–3563, https://doi.org/10.1109/JBHI.2021.3062002.
[8] Q. Chen, M. Li, C. Chen, P. Zhou, X. Lv, C. Chen, MDFNet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification, J. Cancer Res. Clin. Oncol. (2022), https://doi.org/10.1007/s00432-022-04180-1.
[9] D. Li, J. Vaidya, M. Wang, B. Bush, C. Lu, M. Kollef, T. Bailey, Feasibility study of monitoring deterioration of outpatients using multimodal data collected by wearables, ACM Trans. Comput. Healthc. 1 (1) (mar 2020), https://doi.org/10.1145/3344256.
[10] H. Senaratne, S. Oviatt, K. Ellis, G. Melvin, A critical review of multimodal-multisensor analytics for anxiety assessment, ACM Trans. Comput. Healthc. 3 (4) (nov 2022), https://doi.org/10.1145/3556980.
[11] P.M. Burlina, N.J. Joshi, P.A. Mathew, W. Paul, A.W. Rebman, J.N. Aucott, AI-based detection of erythema migrans and disambiguation against other skin lesions, Comput. Biol. Med. 125 (2020) 103977, https://doi.org/10.1016/j.compbiomed.2020.103977, https://linkinghub.elsevier.com/retrieve/pii/S0010482520303085.
[12] S.I. Hossain, J. de Goër de Herve, M.S. Hassan, D. Martineau, E. Petrosyan, V. Corbin, J. Beytout, I. Lebert, J. Durand, I. Carravieri, A. Brun-Jacob, P. Frey-Klett, E. Baux, C. Cazorla, C. Eldin, Y. Hansmann, S. Patrat-Delon, T. Prazuck, A. Raffetin, P. Tattevin, G. Vourc'h, O. Lesens, E. Mephu Nguifo, Exploring convolutional neural networks with transfer learning for diagnosing Lyme disease from skin lesion images, Comput. Methods Programs Biomed. 215 (2022) 106624, https://doi.org/10.1016/j.cmpb.2022.106624, https://linkinghub.elsevier.com/retrieve/pii/S0169260722000098.
[13] E.C. Wilson, J.A. Usher-Smith, J. Emery, P.G. Corrie, F.M. Walter, Expert elicitation of multinomial probabilities for decision-analytic modeling: an application to rates of disease progression in undiagnosed and untreated melanoma, Value Health 21 (6) (2018) 669–676, https://doi.org/10.1016/j.jval.2017.10.009.
[14] C.J. Cadham, M. Knoll, L.M. Sánchez-Romero, K.M. Cummings, C.E. Douglas, A. Liber, D. Mendez, R. Meza, R. Mistry, A. Sertkaya, N. Travis, D.T. Levy, The use of expert elicitation among computational modeling studies in health research: a systematic review, Med. Decis. Mak. 42 (5) (2022) 684–703, https://doi.org/10.1177/0272989X211053794.
[15] R.T. Clemen, R.L. Winkler, Combining probability distributions from experts in risk analysis, Risk Anal. 19 (2) (1999) 187–203, https://doi.org/10.1111/j.1539-6924.1999.tb00399.x.
[16] L.C. Van Der Gaag, S. Renooij, C.L. Witteman, B.M. Aleman, B.G. Taal, Probabilities for a probabilistic network: a case study in oesophageal cancer, Artif. Intell. Med. 25 (2) (2002) 123–148, https://doi.org/10.1016/S0933-3657(02)00012-X.
[17] C. Saegerman, J. Evrard, J.-Y. Houtain, J.-P. Alzieu, J. Bianchini, S.E. Mpouam, G. Schares, E. Liénard, P. Jacquiet, L. Villa, G. Álvarez-García, A.L. Gazzonis, A. Gentile, L. Delooz, First expert elicitation of knowledge on drivers of emergence of bovine besnoitiosis in Europe, Pathogens (Basel, Switzerland) 11 (7) (jul 2022), https://doi.org/10.3390/pathogens11070753, http://www.ncbi.nlm.nih.gov/pubmed/35889998, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC9323894.
[18] M. Suleiman, H. Demirhan, L. Boyd, F. Girosi, V. Aksakalli, Incorporation of expert knowledge in the statistical detection of diagnosis related group misclassification, Int. J. Med. Inform. 136 (2020) 104086, https://doi.org/10.1016/j.ijmedinf.2020.104086.
[19] D. Reynolds, Gaussian mixture models, in: S.Z. Li, A. Jain (Eds.), Encyclopedia of Biometrics, Springer US, Boston, MA, 2009, pp. 659–663.
[20] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc., Ser. B, Methodol. 39 (1) (1977) 1–22, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x, https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1977.tb01600.x.
[21] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705, http://ieeexplore.ieee.org/document/1100705/.
[22] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (mar 2007), https://doi.org/10.1214/aos/1176344136, https://projecteuclid.org/journals/annals-of-statistics/volume-6/issue-2/Estimating-the-Dimension-of-a-Model/10.1214/aos/1176344136.full.
[23] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat. 33 (3) (1962) 1065–1076, https://doi.org/10.1214/aoms/1177704472.
[24] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Stat. 27 (3) (1956) 832–837, https://doi.org/10.1214/aoms/1177728190.
[25] B.W. Silverman, Density Estimation: For Statistics and Data Analysis, Chapman & Hall, London, 2018.
[26] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: I. Rival (Ed.), Ordered Sets, Springer, Netherlands, Dordrecht, 1982, pp. 445–470.
[27] S. Motameny, B. Versmold, R. Schmutzler, Formal concept analysis for the identification of combinatorial biomarkers in breast cancer, in: R. Medina, S. Obiedkov (Eds.), Formal Concept Analysis, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 229–240.
[28] E. Mephu Nguifo, P. Njiwoua, Using lattice-based framework as a tool for feature extraction, in: C. Nédellec, C. Rouveirol (Eds.), Machine Learning: ECML-98, Springer, Berlin Heidelberg, Berlin, Heidelberg, 1998, pp. 304–309.
[29] E. Mephu Nguifo, J. Sallantin, Prediction of primate splice junction gene sequences with a cooperative knowledge acquisition system, in: ISMB, 1993, pp. 292–300.
[30] J.R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106, https://doi.org/10.1007/BF00116251.
[31] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth and Brooks, Monterey, CA, 1984.
[32] CRMVT, Centre de référence des maladies vectorielles liées aux tiques - crmvt, https://crmvt.fr/, 2021.
[33] P. Letertre-Gibert, G. Vourc'h, I. Lebert, M. Rene-Martellet, V. Corbin-Valdenaire, D. Portal-Martineau, J. Beytout, O. Lesens, Lyme snap: a feasibility study of on-line declarations of erythema migrans in a rural area of France, Ticks Tick-Borne Dis. 11 (1) (jan 2020), https://doi.org/10.1016/j.ttbdis.2019.101301.
[34] INRAE, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement - inrae, https://www.inrae.fr/en, 2023.
[35] D.B. Carr, Graphics in the physical sciences, Encycl. Phys. Sci. Technol. (2003) 1–14, https://doi.org/10.1016/B0-12-227410-5/00297-0.

[36] J. Baron, B.A. Mellers, P.E. Tetlock, E. Stone, L.H. Ungar, Two reasons to make aggregated probability forecasts more extreme, Decis. Anal. 11 (2) (2014) 133–145, https://doi.org/10.1287/deca.2014.0293.
[37] U.S. Karmarkar, Subjectively weighted utility: a descriptive extension of the expected utility model, Organ. Behav. Hum. Perform. 21 (1) (1978) 61–72, https://doi.org/10.1016/0030-5073(78)90039-9.
[38] Y. Shlomi, T.S. Wallsten, Subjective recalibration of advisors' probability estimates, Psychon. Bull. Rev. 17 (2010) 492–498, https://doi.org/10.3758/PBR.17.4.492.