Original Article

Artificial Intelligence Tools for the Diagnosis of Eosinophilic Esophagitis in Adults Reporting Dysphagia: Development, External Validation, and Software Creation for Point-of-Care Use

Pierfrancesco Visaggi, MD[a,b,*], Giulio Del Corso, PhD[c,*], Federica Baiano Svizzero, MD[a], Matteo Ghisa, MD[d], Serena Bardelli, PhD[e], Arianna Venturini, MD[a], Delio Stefani Donati, MD[a], Brigida Barberio, MD, PhD[d], Emanuele Marciano, MD[f], Massimo Bellini, MD[a], Jason Dunn, MD[b], Terry Wong, MD[b], Nicola de Bortoli, MD[a], Edoardo V. Savarino, MD, PhD[d,‡], and Sebastian Zeki, MD, PhD[b,‡] *Pisa and Padua, Italy; and London, United Kingdom*

What is already known about this topic? Despite increased awareness of eosinophilic esophagitis (EoE), the diagnostic delay has remained stable over the past 3 decades. Current guidelines recommend multiple esophageal biopsies in all patients with dysphagia regardless of the risk of EoE. In addition, there is no consensus on how to assess the individual risk of EoE in patients reporting dysphagia. Accordingly, there is a need to improve the diagnostic performance and optimize resources allocation in the setting of EoE.

What does this article add to our knowledge? Our software-integrated models are highly accurate for the diagnosis of EoE before biopsy collection and can be used for the assessment of the individual risk of EoE in patients reporting dysphagia.

How does this study impact current management guidelines? Our software is available at https://webapplicationing.shinyapps.io/PointOfCare-EoE/ and can be used at point-of-care to improve the diagnostic workup of EoE and optimize resources allocation.

BACKGROUND: Despite increased awareness of eosinophilic esophagitis (EoE), the diagnostic delay has remained stable over the past 3 decades. There is a need to improve the diagnostic performance and optimize resources allocation in the setting of EoE.

OBJECTIVE: We developed and validated 2 point-of-care machine learning (ML) tools to predict a diagnosis of EoE before histology results during office visits.

METHODS: We conducted a multicenter study in 3 European tertiary referral centers for EoE. We built predictive ML models using retrospectively extracted clinical and esophagogastroduodenoscopy (EGDS) data collected from 273 EoE and 55 non-EoE dysphagia patients. We validated the models on an independent cohort of 93 consecutive patients with dysphagia undergoing EGDS with biopsies at 2 different centers. Models' performance was assessed by area under the curve (AUC), sensitivity, specificity, and positive and negative predictive values (PPV and NPV). The models were integrated into a point-of-care software package.

RESULTS: The model trained on clinical data alone showed an AUC of 0.90 and a sensitivity, specificity, PPV, and NPV of 0.90,

[a]Gastroenterology Unit, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
[b]Centre for Oesophageal Diseases, Guy's and St. Thomas Hospital, London, United Kingdom
[c]Institute of Information Science and Technologies "A. Faedo", National Research Council of Italy (CNR), Pisa, Italy
[d]Division of Gastroenterology, Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy
[e]Neonatal Learning and Simulation Centre "NINA", Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy
[f]Endoscopy Unit, Azienda Ospedaliero Universitaria Pisana, Pisa, Italy
*These authors share first authorship.
‡These authors share last authorship.
No funding was received for this work.
Conflicts of interest: J. Dunn received lecture fees from Dr Falk. T. Wong received grant support from Reckitt-Benckiser and lecture fees from AstraZeneca and Dr Falk. N. de Bortoli received lectures fees from Malesci, Reckitt Benckiser, and Dr Falk. E. V. Savarino has served as a speaker for Abbvie, AGPharma, Alfasigma, Dr Falk, EG Stada Group, Fresenius Kabi, Grifols, Janssen, Innovamedica, Malesci, Pfizer, Reckitt Benckiser, Sandoz, SILA, Sofar, Takeda, and Unifarco; has served as a consultant for Alfasigma, Amgen, Biogen, Bristol-Myers Squibb, Celltrion, Diadema Farmaceutici, Dr. Falk, Fresenius Kabi, Janssen, Merck & Co, Reckitt Benckiser, Regeneron, Sanofi, Shire, SILA, Sofar, Synformulas GmbH, Takeda, and Unifarco; and has received research support from Pfizer, Reckitt Benckiser, SILA, Sofar, and Unifarco. S. Zeki has served as a speaker for Dr Falk. The rest of the authors declare that they have no relevant conflicts of interest.
Received for publication August 29, 2023; revised November 23, 2023; accepted for publication December 20, 2023.
Available online ■■
Corresponding author: Nicola de Bortoli, MD, Gastroenterology Unit, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Via Roma 67, 56124 Pisa, Italy. E-mail: nicola.debortoli@unipi.it.
2213-2198

https://doi.org/10.1016/j.jaip.2023.12.031

Abbreviations used
AUC- Area under the curve
CI- Confidence interval
EGDS- Esophagogastroduodenoscopy
EoE- Eosinophilic esophagitis
EREFS- Endoscopic reference score
FAMD- Factor analysis of mixed data
GSTT- Guy's and St Thomas Hospital
IQR- Interquartile range
ML- Machine learning
NPV- Negative predictive value
PPI- Proton pump inhibitor
PPV- Positive predictive value
ROS- Random oversampling

0.75, 0.80, and 0.87, respectively, for the diagnosis of EoE in the external validation cohort. The model trained on a combination of clinical and endoscopic data showed an AUC of 0.94, and a sensitivity, specificity, PPV, and NPV of 0.94, 0.68, 0.77, and 0.91, respectively, in the external validation cohort.

CONCLUSION: Our software-integrated models (https://webapplicationing.shinyapps.io/PointOfCare-EoE/) can be used at point-of-care to improve the diagnostic workup of EoE and optimize resources allocation.

Key words: *Artificial intelligence; Eosinophilic esophagitis; Diagnosis*

Eosinophilic esophagitis (EoE) is a chronic disease of the esophagus triggered by food and inhaled antigens.[1] The disease is characterized by symptoms of esophageal dysfunction and an eosinophil-predominant esophageal infiltrate.[2] Because dysphagia—the hallmark symptom of EoE—can be present in several esophageal diseases,[3-5] a conclusive diagnosis of EoE requires esophageal biopsies showing at least 15 eosinophils in at least 1 high-power field.[6,7] In addition, although patients with EoE may have typical endoscopic findings,[8] the esophagogastroduodenoscopy (EGDS) may be normal in up to 32% of patients, particularly in nontertiary centers.[9,10] Accordingly, although noninvasive biomarkers are being investigated,[11] the distinction between EoE and other causes of dysphagia is currently difficult without histology results. However, esophageal biopsies are frequently omitted by physicians during index endoscopy, leading to a reported diagnostic delay of up to 10 years.[12-14] Moreover, despite increased awareness of the disease and retrospective data showing improvement in the diagnostic workup of EoE in Europe over time,[15] recent data have shown that the diagnostic delay of the disease has remained stable over the past 3 decades.[16] Because diagnostic delay is associated with an increased risk of EoE-related complications, such as esophageal strictures, hospitalization, and episodes of food impaction,[16,17] there is a need to improve the diagnostic performance in the setting of dysphagia.

To minimize the number of patients with unrecognized EoE, clinical guidelines propose that patients undergoing EGDS for dysphagia should undergo esophageal biopsy sampling to rule out EoE.[6,18] However, the incidence and prevalence of the disease are increasing,[19] and the economic burden of EoE already exceeds that of celiac and inflammatory bowel disease.[20,21] Accordingly, whether performing esophageal biopsies in all patients with dysphagia will be cost-effective in the long term remains uncertain. Similarly, there is currently no consensus on which patients should be considered at risk of EoE and undergo repeat EGDS to rule out EoE after previous failure to collect biopsies or previous nondiagnostic biopsies collected while on proton pump inhibitor (PPI) therapy.[6]

To improve the diagnostic workup and optimize resources allocation in the setting of EoE, it will be useful to have clinically applicable tools to support decision-making in the diagnostic workup of suspected EoE according to patients' individual risk. In this regard, although artificial intelligence is being increasingly used in gastroenterology to support precision medicine,[22-24] only 1 study has developed a clinical machine learning (ML) model for the diagnosis of EoE based on current diagnostic criteria.[25] However, the ML model was not validated on an external and independent cohort of patients, hampering the generalizability of the models to other settings.

In this multicenter and international study, we developed and validated externally in a prospective fashion 2 different ML tools that could be used to predict a diagnosis of EoE before histology results in patients reporting dysphagia. Furthermore, to facilitate implementation in clinical practice, we created user-friendly, freely available software for point-of-care use.

MATERIALS AND METHODS

Study design and patients

This was a 2-step multicenter international study conducted in 3 European tertiary referral centers for EoE: Guy's and St. Thomas' Hospital NHS Foundation Trust (GSTT, London, United Kingdom), Pisa University Hospital (Pisa, Italy), and Padua University Hospital (Padua, Italy) (Figure 1).

In the first phase of the study (predictive model building phase), data from all patients undergoing EGDS with at least 6 esophageal biopsies for dysphagia with or without other upper gastrointestinal symptoms referred at GSTT between January 2012 and December 2020 were extracted from the electronic patient record. The data were used to build a predictive model for the diagnosis of EoE according to current diagnostic criteria.[26-28] Patients with EoE were eligible for inclusion if they were adults (>18 years of age), if dysphagia was one of the reported symptoms, and if the histological diagnosis of EoE was confirmed based on the most recent consensus criteria, that is, the presence of at least 15 eosinophils per high-power field in at least 1 high-power field in at least 1 esophageal biopsy.[26-28] All patients were off-medications potentially interfering with esophageal eosinophil counts for at least 12 weeks (ie, PPIs, steroids, immunosuppressants, and biologic drugs). Non-EoE controls were consecutive adult patients older than 18 years undergoing EGDS for dysphagia with or without other upper gastrointestinal symptoms, in which EoE had been ruled out based on at least 6 esophageal biopsies collected while off-medications potentially interfering with esophageal eosinophil counts for at least 12 weeks. Accordingly, non-EoE controls were required to have less than 15 eosinophils per high-power field in all 6 esophageal biopsies, which is the optimal number of biopsies to diagnose or rule out EoE.[29] Patients were eligible for inclusion if all the following data were available for extraction: age at diagnosis, sex, presence or absence of heartburn, chest pain, dyspepsia, history of food impaction requiring endoscopic removal, allergic

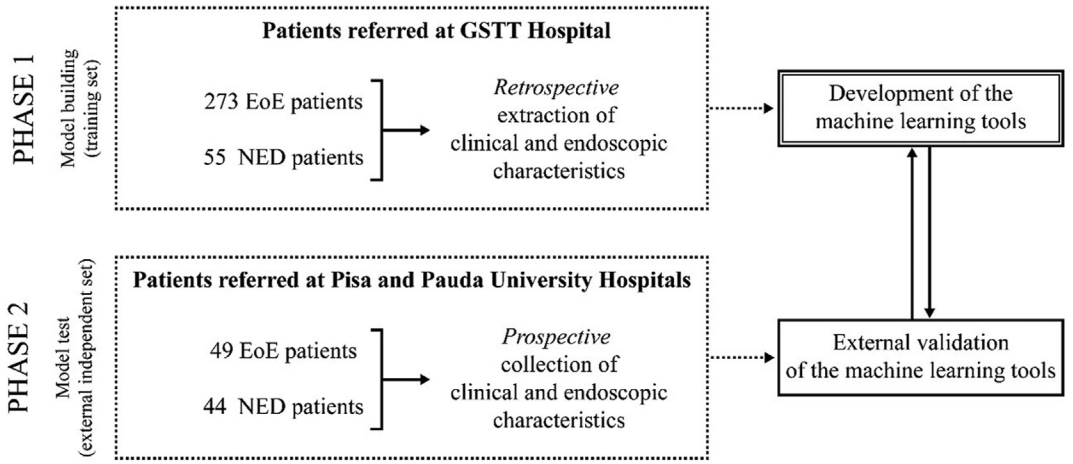

FIGURE 1. The 2-phase process used for the development and the independent external validation of the artificial intelligence models. *AI*, Artificial intelligence; *EoE*, eosinophilic esophagitis; *GSTT*, Guy's and St Thomas' Hospital; *NED*, non-EoE dysphagia.

rhinitis, asthma, atopic dermatitis, nasal polyposis, endoscopic features of EoE (exudates, rings, edema, furrows, and stricture), and presence or absence of esophageal Schatzki ring at endoscopy. We excluded patients with known concomitant non-EoE eosinophilic gastrointestinal disease, esophageal cancer, history of upper gastrointestinal surgery, and in case they had been taking medications potentially interfering with esophageal eosinophil counts during the previous 12 weeks according to their medical records. The retrospective data extraction was performed by 2 investigators at GSTT Hospital (PV and SZ).

The second phase of the study consisted in the prospective external validation of the predictive model on consecutive adult patients (>18 years of age) undergoing EGDS for dysphagia with or without other upper gastrointestinal symptoms at Pisa and Padua University Hospitals between January 2021 and April 2023. Standardized report forms and questionnaires were used for each patient to collect data regarding age at diagnosis, sex, upper gastrointestinal symptoms, history of food impaction requiring endoscopic removal, allergic rhinitis, asthma, atopic dermatitis, nasal polyposis, presence of endoscopic features of EoE according to the EoE endoscopic reference score (EREFS),[8] and esophageal Schatzki ring at endoscopy. The prospective data collection was performed by 4 investigators at Pisa University Hospital (FBS, AV, DSD, and NdB) and 2 investigators at Padua University Hospital (MG and EVS). All patients underwent EGDS with at least 6 esophageal biopsies while off-medications potentially interfering with esophageal eosinophil counts for at least 12 weeks. EoE was diagnosed based on the most recent consensus guidelines.[26-28] Non-EoE controls were consecutive patients reporting dysphagia in whom upper gastrointestinal cancer and EoE had been ruled out based on at least 6 esophageal biopsies. At each institution, investigators were blind to the results of the data collection performed at the other institution. The first phase of the study was considered a review of clinical practice, and ethical approval was not required.[30] In the second phase of the study, all patients were part of each institution's institutional review board–approved data collection, only deidentified data were shared across the participating institutions with no links to the original patients, and repeat institutional review board approval was not deemed necessary.

Predictive model development and validation

The analysis performed involved the definition and training of 2 distinct predictive models to classify cases of EoE. Variables under examination were either continuous or categorical and were classified into clinical (ie, sex, age, food impaction, heartburn, chest pain, and dyspepsia/postprandial nausea) or endoscopic (ie, exudates, rings, edema, furrows, stricture, and Schatzki ring). The candidate predictors (independent variables) of the 2 models were selected from the clinical variables for the training of the clinical history model and from both clinical and endoscopic variables for the training of the clinical-endoscopic model, as reported in Table I.

Collinear variables were excluded from the model after correlation analysis. In addition, a factor analysis of mixed data (FAMD) was performed as an extension to classical principal component analysis to take into account both numerical and categorical variables, and the amount of variation explained by each component was compared to detect higher order interactions.[31]

Given the independent variables and the corresponding n principal components obtained by FAMD transformation, if the percentage of the explained variance of at least one of the principal components was less than a threshold value (set at $n/2$), the model was further reduced. A random forest model was used as a classifier for both the clinical history and clinical-endoscopic cases between 50 and 200 trees.[32] This method was chosen to allow stability at a large number of categorical variables, the possibility of optimizing a hyperparameter (number of trees), and the ability to determine the most influential independent variables.[33] Accordingly, to evaluate the importance of the parameters, we estimated the mean decrease accuracy of each independent variable. The mean decrease accuracy is calculated on out-of-bag data as the averaged (over all trees) difference between the error rate on the out-of-bag portion and the error rate obtained by permutating each predictor variable, normalized by the standard deviation of the differences. This score expresses how much of the accuracy of the model is lost when each variable is excluded.

The dataset was divided into 2 subsets: GSTT dataset (training and k-fold validation) and Pisa-Padova dataset (independent test). The GSTT dataset was used to train the model and choose the optimal hyperparameters (number of trees) by means of a k-fold

TABLE I. Characteristics of EoE and non-EoE patients in the training set (Guys' and St Thomas' Hospital)

Characteristic	London EoE (n = 273)	London non-EoE dysphagia (n = 55)	Statistical comparison (*P* value)
Clinical data, n (%)			
Sex (F/M)	78-195 (28.6-71.4)	36-19 (65.5-34.5)	<.001
Age at diagnosis (range) (y)	34 (28-47)	52 (41-61)	<.001
Food impaction	106 (38.8)	6 (10.9)	<.001
Heartburn/regurgitation	124 (45.4)	25 (45.5)	1
Chest pain	47 (17.2)	9 (16.4)	.87
Dyspepsia	29 (10.6)	12 (21.8)	.02
Rhinitis	183 (69.8)	3 (5.5)	<.001
Asthma	112 (42.7)	17 (30.9)	.1
Atopic dermatitis	80 (30.5)	4 (7.2)	<.001
Endoscopic data, n (%)			
Exudates	54 (80.2)	1 (1.8)	.001
Rings	126 (46.1)	0 (0)	<.001
Edema	32 (11.7)	3 (5.4)	.17
Furrows	102 (37.4)	0 (0)	<.001
Stricture	46 (16.8)	0 (0)	.001
Schatzki	37 (13.5)	2 (3.6)	.04

EoE, Eosinophilic esophagitis.

cross-validation, whereas the Pisa-Padova dataset was used as an independent test set to verify the generalization capabilities of the model. The thresholds used to estimate the model on the test set are calibrated to the training data set itself to avoid inflation of results and produce unbiased estimates.

Scores are reported for the potentially positively biased training data set and for *k*-fold cross-validation ($k = 10$). The *k*-fold scores are a low-bias estimate of the generalization capability on similar datasets, while the external test set can provide information on how the model performs on data collected by different investigators.

To minimize possible issues due to unbalanced classes (ie, EoE/controls), the training dataset was randomly resampled (random oversampling [ROS]) to balance the least represented class (ie, controls). This approach is equivalent to the application of a class-balancing loss to prevent model hyperspecialization on the EoE class.[34] However, because several accuracy metrics depend on prevalence, the reported training *k*-fold cross-validation and test set scores are calculated on the original dataset without ROS.[35]

Predictive model evaluation

The performance of the models was described by the area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for both the training set and the independent test set. Values for *k*-fold cross-validation were reported as the median value for the *k* assessments and the corresponding interquartile range (IQR). The threshold maximizing the sum of sensitivity and specificity was chosen as optimal. Random forest models were also used to determine the subset of variables with the greatest predictive power by comparing their mean decrease accuracy values.

RESULTS

Baseline characteristics of included patients and model building

The first phase of the study (training of the model) included 328 patients. Of these, 273 had a diagnosis of EoE (83.2%) and 55 (16.8%) had non-EoE dysphagia and were used as controls. EoE and control patients recruited in the predictive model building phase of the study were used as a training set and as *k*-fold cross-validation ($k = 10$) for the optimization of the number of trees (ie, hyperparameter). Following the χ^2 test and FAMD results in an explained variance ranging from 18.3% to 34.8% for the clinical variables and from 5.5% to 23.2%, collinear variables were identified and excluded from the model, while the following clinical variables were found to be noncollinear and thus relevant for the predictive model: age at diagnosis, sex, history of food impaction, allergic rhinitis, atopic dermatitis, and dyspepsia, whereas edema, rings, exudates, strictures, and esophageal Schatzki rings were found to be noncollinear among endoscopic findings and were included in the predictive model.

The second phase of the study was the prospective external validation of the model and included 93 patients, of whom 49 (52.7%) had EoE, whereas 44 (47.3%) had non-EoE dysphagia and were considered controls. EoE and control patients recruited in the external validation phase provided the independent test set of patients to assess the generalization ability of the model, which represents the ability of a model to be successfully applied on new previously unseen data. The characteristics of the patients included in the training and test phase are reported in Tables I and II.

ML model based on the clinical history

The ML model trained on clinical data alone (random decision tree, number of trees = 110) showed an AUC of 0.97 (95% DeLong confidence interval [CI]: 0.96-0.99) on the training set and a median AUC of 0.95 (IQR: 0.88-0.97) on the *k*-fold cross-validation set for the diagnosis of EoE. On the training set, for the diagnosis of EoE, the model showed a sensitivity, specificity, PPV, and NPV of 0.89, 0.98, 0.99, and 0.66, respectively (Figure 2 and Table III).

When tested externally on the independent test set of patients, the predictive model showed an AUC of 0.90 (95% DeLong CI: 0.84-0.96), with a sensitivity, specificity, PPV, and NPV of 0.89, 0.75, 0.80, 0.87, respectively, for the diagnosis of EoE (Figure 2). Among the variables that were chosen by the model

TABLE II. Characteristics of EoE and Non-EoE patients in the external validation set (Pisa and Padua University Hospitals)

Characteristic	EoE (n = 49)	Non-EoE dysphagia (n = 44)	Statistical comparison (*P* value)
Clinical data, n (%)			
Sex (F/M)	8-41 (16.3-83.7)	23-21 (52.3-47.7)	<.001
Age at diagnosis (range) (y)	36 (25-45)	51 (41-65)	<.001
Food impaction	36 (73.5)	14 (31.8)	<.001
Heartburn/regurgitation	14 (28.6)	21 (47.7)	.06
Chest pain	13 (26.5)	14 (31.8)	.57
Dyspepsia	6 (12.2)	10 (22.7)	.18
Rhinitis	34 (69.4)	4 (9.1)	<.001
Asthma	19 (38.8)	3 (6.8)	<.001
Atopic dermatitis	7 (14.3)	4 (9.1)	.43
Endoscopic data, n (%)			
Exudates	17 (34.7)	2 (4.5)	<.001
Rings	33 (67.3)	2 (4.5)	<.001
Edema	18 (36.7)	0 (0)	<.001
Furrows	17 (34.7)	1 (2.3)	<.001
Stricture	7 (14.3)	1 (2.3)	.04
Schatzki	5 (10.2)	0 (0)	.03

EoE, Eosinophilic esophagitis.

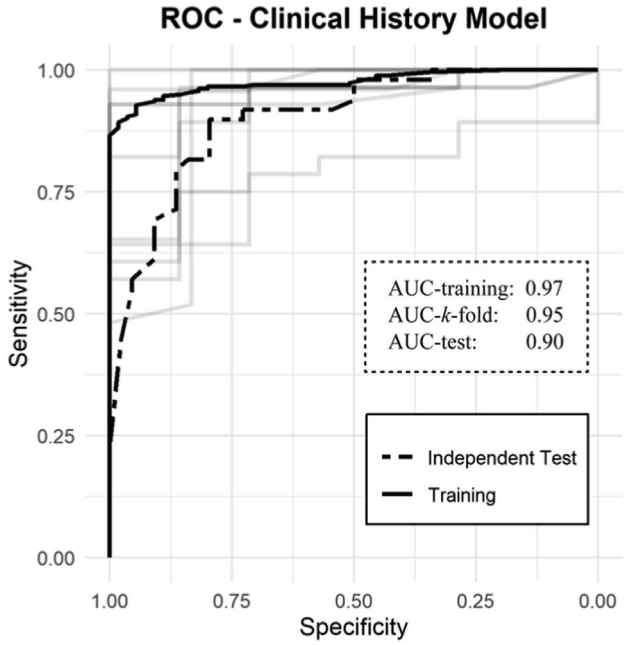

FIGURE 2. Area under the curve (AUC) of the 3 ROC analyses performed for the clinical history model evaluated on the training, *k*-fold cross-validation, and test sets. ROC curves for *k*-fold validation are shown in gray. *ROC*, Receiver operating characteristic.

to predict a diagnosis of EoE based on clinical data alone, the presence of rhinitis, history of food impaction, the patients' age at diagnosis, and sex had the highest predictive ability (Figure E1, available in this article's Online Repository at www.jaci-inpractice.org).

On the basis of these findings, we developed a submodel evaluating only the most relevant variables identified by the model (ie, rhinitis, history of food impaction, age at diagnosis, sex, and atopic dermatitis), which showed an AUC of 0.97 (95% DeLong CI: 0.96-0.99) on the training set and an AUC of 0.91 (95% DeLong CI: 0.84-0.97) on the independent test set for the diagnosis of EoE.

ML model based on clinical and endoscopic data

The ML model trained on a combination of clinical and endoscopic data (random decision tree, number of trees = 60) showed an AUC of 0.99 (95% DeLong CI: 0.99-1) on the training set and a median AUC of 0.98 (IQR: 0.95-0.99) on the *k*-fold cross-validation set. On the training set, the model showed a sensitivity, specificity, PPV, and NPV of 0.97, 1.0, 1.0, and 0.89, respectively (Figure 3 and Table III).

When prospectively tested on the independent test set of patients, the predictive model showed an AUC of 0.94 (95% DeLong CI: 0.89-0.99), with a sensitivity, specificity, PPV, and NPV of 0.94, 0.68, 0.77, and 0.91, respectively (Figure 3 and Table III). Among the variables that were used by the model to predict a diagnosis of EoE based on the combination of clinical and endoscopic data, the presence of rhinitis, the patients' age at diagnosis, esophageal rings at endoscopy, sex, and atopic dermatitis had the highest predictive ability (Figure E2, available in this article's Online Repository at www.jaci-inpractice.org).

On the basis of these findings, we developed a submodel evaluating only the most relevant variables identified by the model (ie, rhinitis, the patients' age at diagnosis, esophageal rings at endoscopy, sex, and atopic dermatitis), which showed an AUC of 0.99 (95% DeLong CI: 0.98-0.99) on the training set and an AUC of 0.92 (95% DeLong CI: 0.86-0.99) on the independent test set for the diagnosis of EoE with a sensitivity, specificity, PPV, and NPV of 0.94, 0.73, 0.79, and 0.91, respectively.

DISCUSSION

In this multicenter international study, we developed and validated prospectively 2 ML tools to predict a diagnosis of EoE

TABLE III. Performance of the 2 machine learning models in the internal development set, on the *k*-fold cross-validation sets, and in the independent external validation test set

Model	Setting	AUC (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Clinical	Internal development	97.5	89.3	98.2	99.6	65.8
	k-fold cross-validation	95.2 (88-97)	85.7 (74-96)	92.8 (89-96)	75 (64-87)	93.5 (81-96)
	Independent external validation	90.0	89.8	75.0	80.0	86.8
Clinical (reduced variables)	Internal development	97.2	89.7	96.3	99.2	66.2
	Independent external validation	90.5	93.9	65.9	75.4	90.6
Clinical—endoscopic	Internal development	99.7	97.3	100	100	88.7
	k-fold cross-validation	97.8 (95-99)	85.7 (74-96)	92.2 (88-96)	76.4 (65-85)	93.6 (91-96)
	Independent external validation	94.2	93.9	68.2	76.7	90.9
Clinical—endoscopic (reduced variables)	Internal development	98.5	93.1	98.2	99.6	75.0
	Independent external validation	92.4	93.9	72.7	79.3	91.4

AUC, Area under the curve; *NPV*, negative predictive value; *PPV*, positive predictive value.

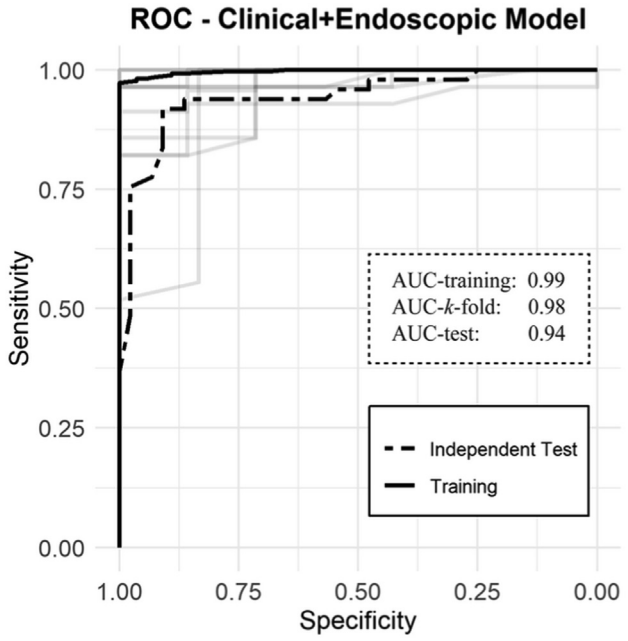

FIGURE 3. Area under the curve (AUC) of the 3 ROC analyses performed for the clinical and endoscopic model evaluated on the training, *k*-fold cross-validation, and test sets. ROC curves for *k*-fold validation are shown in gray. *ROC*, Receiver operating characteristic.

in adults reporting dysphagia. The model using information available before endoscopy achieved an AUC of 0.90, whereas the model using information available at the time of the diagnostic endoscopy before collecting biopsies achieved an AUC of 0.94. Of note, the external validation performed in this study demonstrates the generalizability of our findings and makes it possible to apply our models to other settings. Therefore, our ML tools can be useful in clinical practice to predict the risk of EoE before the results of esophageal biopsies. For this purpose, we made the ML tools freely available online at https://webapplicationing.shinyapps.io/PointOfCare-EoE/. Of note, the online ML tool was retrained on the entire dataset to mitigate the effect of data collection variability among centers.

The strengths of this study include the 2-step process of internal development and external prospective validation of the ML tools, as well as the multicenter international setting (Figure 1). Both models were built based on data from patients enrolled at a tertiary referral center in the United Kingdom and were tested externally on an independent cohort of patients enrolled at 2 different tertiary referral centers in Italy. In this regard, ML models are highly susceptible to selection bias because their performance depends entirely on the original development dataset.[36]

However, in this study, we performed an external validation on an independent international cohort of patients and demonstrated that both models could effectively segregate EoE from other causes of dysphagia outside of the development setting (Figures 2 and 3). It must be acknowledged, however, that there is the possibility that our ML tool may overcall EoE in some instances. In this regard, the presence of a difference between the *k*-fold cross-validation scores and the external test results suggests that additional fine-tuning of our ML models on a broader set of data could have further mitigated the effect of epistemic variability on data collection.

This study has limitations that should be mentioned: for the development and validation of the models, we included only adult patients reporting dysphagia with or without other upper gastrointestinal complaints. Accordingly, our findings cannot be generalized to children, adolescents, or to patients reporting upper gastrointestinal symptoms without dysphagia. However, dysphagia represents the most common symptom in adults with EoE.[2,19] In such clinical scenario, multiple esophageal biopsies should be taken according to all the international guidelines in the field, even when the esophagus shows no endoscopic abnormalities.[6,7,26,27] Therefore, we developed ML models to assess the individual risk of EoE and predict the utility of taking biopsies even in the absence of endoscopic findings (Table IV). It

TABLE IV. Clinical scenarios and outcomes of point-of-care support from the machine learning tools developed and validated in this study

Patient scenario	Example	Which point-of-care tool could be useful	Software-generated probability of EoE (%)	Outcome plan
Biopsies omitted at previous EGDS	73-year-old woman with dysphagia and heartburn. Has asthma. Previous EGDS showed edema and Schatzki ring. Biopsies were omitted.	Clinical-endoscopic tool	20	Avoid repeat EGDS and proceed to other investigations
	35-year-old man with dysphagia. No history of food impaction. Has allergic rhinitis. Does not have esophageal cancer, but the report of his previous EGDS is unavailable. Biopsies were omitted.	Clinical tool	97	Repeat EGDS with multiple esophageal biopsies while off PPIs
Nondiagnostic biopsies for EoE at previous EGDS performed while on PPI	25-year-old woman with dysphagia, heartburn, and dyspepsia. History of bolus impaction and asthma. Previous EGDS performed on PPI showed rings and erosive esophagitis. Biopsies showed 5 eos/hpf.	Clinical-endoscopic tool	78	Repeat EGDS with multiple esophageal biopsies while off PPIs
	74-year-old man with dysphagia, heartburn, and regurgitation. No allergic conditions. Previous EGDS performed on PPI showed edema. Biopsies showed 6 eos/hpf.		10	Avoid repeat EGDS and proceed to other investigations
First EGDS for dysphagia shows no endoscopic abnormalities or is unavailable	23-year-old woman with dysphagia and heartburn. Never had bolus impaction. Has allergic rhinitis and asthma. Does not have esophageal cancer, but the report of his previous EGDS is unavailable.	Clinical tool	97	Perform multiple esophageal biopsies to rule out EoE
	45-year-old man with dysphagia, history of previous bolus impaction, and allergic rhinitis. EGDS shows no mucosal abnormalities.	Clinical-endoscopic tool	98	Perform multiple esophageal biopsies to rule out EoE
	80-year-old woman with dysphagia. EGDS shows no mucosal abnormalities.	Clinical-endoscopic tool	0	Avoid multiple esophageal biopsies as EoE is unlikely, proceed to other investigations
	70-year-old man with dysphagia. Never had bolus impaction. No allergic comorbidities. Does not have esophageal cancer, but the report of his previous EGDS is unavailable.	Clinical tool	14	Avoid multiple esophageal biopsies as EoE is unlikely, proceed to other investigations

EGDS, Esophagogastroduodenoscopy; *EoE*, eosinophilic esophagitis; *PPI*, proton pump inhibitor.

must be acknowledged, however, that adult patients with EoE may infrequently present with symptoms that fall within the gastroesophageal reflux disease spectrum in the absence of dysphagia. Our models are not applicable in such instances. However, this scenario has already been investigated by Cotton et al,[25] whose model achieved good performance in the diagnosis of EoE without dysphagia. Another criticism that could be leveled at this study is that, although we extracted patients' data on the presence or absence of all the endoscopic features included in the EoE EREFS,[8] we did not include the grade of each endoscopic finding according to the EREFS score. However, assessing the presence or absence of each finding may be easier than grading each finding and calculate the full EREFS score for both experts and trainees.[37] In addition, this allowed us to identify the presence of esophageal rings as the most important endoscopic finding predictive of EoE (Figure E2, available in this article's Online Repository at www.jaci-inpractice.org). On a final note, in this study, we did not perform a cost saving analysis. Therefore, we are unable to estimate the economic impact of the use of our ML tools at present, and this should be further investigated in future studies. Moreover, future studies should aim at providing comparisons among different model alternatives including Gaussian Process Regression, Support Vector Machine, and K-nearest neighbors to explore their relative performance.

There are several clinical scenarios in which the use of our ML tools would be useful to optimize resources allocation in the diagnostic workup of suspected EoE (Table IV). It has been shown that EoE can be misdiagnosed and pass unrecognized even in patients reporting dysphagia because esophageal biopsies are omitted when patients undergo EGDS.[13,38,39] Accordingly, recent clinical guidelines suggest repeat endoscopy with adequate biopsy sampling in patients in whom there is a high level of suspicion for EoE, both when biopsies have been omitted during the previous EGDS and when previous biopsies show a number of eosinophils below the diagnostic threshold for EoE.[6] In this regard, although experienced physicians may have confidence in the identification of high-risk patients for repeat upper endoscopy with biopsies, there is currently no consensus on how to assess the risk of EoE in patients reporting dysphagia. Our ML tools could serve the purpose of assessing the risk of EoE and support decision-making on performing repeat EGDS with biopsies. Another possible scenario is when patients reporting dysphagia have nondiagnostic levels of eosinophils on esophageal biopsies collected while on PPI treatment. Recent clinical guidelines acknowledged that EoE remains a possible diagnosis when esophageal biopsies are taken while on PPIs or within 3 weeks from PPI withdrawal, and recommended repeat endoscopy with biopsies in such instances.[6] However, PPIs are among the most commonly used medications in the world,[40] and it is likely that many patients with upper gastrointestinal symptoms will have EGDS and biopsies while on PPI treatment,[6,41] making the diagnostic workup of EoE more cumbersome. In this setting, both our ML tools could be used during office visits to identify high-risk patients who may deserve repeat EGDS after PPI withdrawal and avoid repeat procedures in very low risk patients, who may proceed to other investigations. Finally, our ML tool could support the choice of avoiding esophageal biopsies when patients undergoing EGDS have a very low risk of having EoE, although we acknowledge that a high degree of confidence will be required in such instances.

Other studies have demonstrated that ML can be used to diagnose EoE based on clinical characteristics.[25,42] However, one of these studies[42] built a predictive model based on outdated criteria for the diagnosis of EoE and is therefore not applicable to current clinical practice. Another study developed 2 models achieving an AUC of 0.84 and 0.92 for the diagnosis of EoE based on clinical data or on clinical-endoscopic data, respectively.[25] However, the study did not perform an external validation of the models' performance, hampering the generalizability of the results to other settings. In this regard, to assess the real-world diagnostic performance of an ML model, it is essential to test the model on an external and independent dataset that derives from a source that is different from the development data and not used in the training of the model.[43] Accordingly, in this study, the models were validated in an international setting on an independent cohort of patients and demonstrated a consistently high performance.

Another important aspect that emerged from this study is that among the variables that were chosen by the model to predict a diagnosis of EoE, clinical data alone may predict EoE diagnosis regardless of the endoscopic appearance with a high degree of accuracy (AUC = 0.90). It is well known that the esophageal mucosa may appear normal in a proportion of patients with EoE ranging from 5% to 32%,[9,44] increasing the risk that an EoE passes unrecognized. In this regard, data from a recent survey highlighted that, in some European Countries, the percentage of gastroenterologists taking biopsies in patients with dysphagia showing no endoscopic abnormalities is lower than 60%.[39] In this setting, our clinical model may help clinicians to identify high-risk patients requiring multiple esophageal biopsies regardless of the endoscopic appearance.

In conclusion, we built 2 ML tools that could be used in different settings to support the diagnostic workup of EoE and start early appropriate treatment without delay.[45] The rigorous methodology used in this study allows generalizability of our findings to other centers. For this purpose, we developed a user-friendly version of the models that can be used in clinical practice at point-of-care.

Acknowledgments

P. Visaggi, G. Del Corso, N. de Bortoli, E. V. Savarino, and S. Zeki conceived and drafted the study. P. Visaggi, N. de Bortoli, F. Baiano Svizzero, M. Ghisa, A. Venturini, D. Stefani Donati, E. V. Savarino, and S. Zeki collected all data. P. Visaggi and G. Del Corso analyzed all data. G. Del Corso and S. Bardelli contributed to software development. P. Visaggi, G. Del Corso, J. Dunn, T. Wong, N. de Bortoli, E. V. Savarino, and S. Zeki were involved in writing, review, and editing. All authors commented on the drafts of the paper.

REFERENCES

1. Visaggi P, Savarino E, Del Corso G, Hunter H, Svizzero FB, Till SJ, et al. Six-food elimination diet is less effective during pollen season in adults with eosinophilic esophagitis sensitized to pollens. Am J Gastroenterol 2023;118:1957-62.
2. Visaggi P, Savarino E, Sciume G, Chio TD, Bronzini F, Tolone S, et al. Eosinophilic esophagitis: clinical, endoscopic, histologic and therapeutic differences and similarities between children and adults. Therap Adv Gastroenterol 2021;14:1756284820980860.
3. Visaggi P, Ghisa M, Del Corso G, Baiano Svizzero F, Mariani L, Tolone S, et al. Chicago classification v4.0 protocol improves specificity and accuracy of

diagnosis of oesophagogastric junction outflow obstruction. Aliment Pharmacol Ther 2022;56:606-13.
4. Visaggi P, Ghisa M, Marabotto E, Venturini A, Stefani Donati D, Bellini M, et al. Esophageal dysmotility in patients with eosinophilic esophagitis: pathogenesis, assessment tools, manometric characteristics, and clinical implications. Esophagus 2023;20:29-38.
5. Visaggi P, Barberio B, Ghisa M, Ribolsi M, Savarino V, Fassan M, et al. Modern diagnosis of early esophageal cancer: from blood biomarkers to advanced endoscopy and artificial intelligence. Cancers 2021;13:3162.
6. Dhar A, Haboubi HN, Attwood SE, Auth MKH, Dunn JM, Sweis R, et al. British Society of Gastroenterology (BSG) and British Society of Paediatric Gastroenterology, Hepatology and Nutrition (BSPGHAN) joint consensus guidelines on the diagnosis and management of eosinophilic oesophagitis in children and adults. Gut 2022;71:1459-87.
7. Hirano I, Chan ES, Rank MA, Sharaf RN, Stollman NH, Stukus DR, et al. AGA institute and the joint task force on allergy-immunology practice parameters clinical guidelines for the management of eosinophilic esophagitis. Gastroenterology 2020;158:1776-86.
8. Hirano I, Moy N, Heckman MG, Thomas CS, Gonsalves N, Achem SR. Endoscopic assessment of the oesophageal features of eosinophilic oesophagitis: validation of a novel classification and grading system. Gut 2013;62:489-95.
9. Liacouras CA, Spergel JM, Ruchelli E, Verma R, Mascarenhas M, Semeao E, et al. Eosinophilic esophagitis: a 10-year experience in 381 children. Clin Gastroenterol Hepatol 2005;3:1198-206.
10. Mackenzie SH, Go M, Chadwick B, Thomas K, Fang J, Kuwada S, et al. Eosinophilic oesophagitis in patients presenting with dysphagia—a prospective analysis. Aliment pharmacol Ther 2008;28:1140-6.
11. Visaggi P, Solinas I, Baiano Svizzero F, Bottari A, Barberio B, Lorenzon G, et al. Non-invasive and minimally invasive biomarkers for the management of eosinophilic esophagitis beyond peak eosinophil counts: filling the gap in clinical practice. Diagnostics (Basel) 2023;13:2806.
12. Melgaard D, Westmark S, Laurberg PT, Krarup AL. A diagnostic delay of 10 years in the DanEoE cohort calls for focus on education—a population-based cross-sectional study of incidence, diagnostic process and complications of eosinophilic oesophagitis in the North Denmark Region. United European Gastroenterol J 2021;9:688-98.
13. Lenti MV, Savarino E, Mauro A, Penagini R, Racca F, Ghisa M, et al. Diagnostic delay and misdiagnosis in eosinophilic oesophagitis. Dig Liver Dis 2021;53:1632-9.
14. Laserna-Mendieta EJ, Navarro P, Casabona-Francés S, Savarino EV, Pérez-Martínez I, Guagnozzi D, et al. Differences between childhood- and adulthood-onset eosinophilic esophagitis: an analysis from the EoE connect registry. Dig Liver Dis 2023;55:350-9.
15. Navarro P, Laserna-Mendieta EJ, Casabona S, Savarino E, Pérez-Fernández MT, Ghisa M, et al. Accurate and timely diagnosis of eosinophilic esophagitis improves over time in Europe. An analysis of the EoE CONNECT Registry. United European Gastroenterol J 2022;10:507-17.
16. Murray FR, Kreienbuehl AS, Greuter T, Nennstiel S, Safroneeva E, Saner C, et al. Diagnostic delay in patients with eosinophilic esophagitis has not changed since the first description 30 years ago: diagnostic delay in eosinophilic esophagitis. Am J Gastroenterol 2022;117:1772-9.
17. Bon L, Safroneeva E, Bussmann C, Biedermann L, Schreiner P, Vavricka SR, et al. Close follow-up is associated with fewer stricture formation and results in earlier detection of histological relapse in the long-term management of eosinophilic esophagitis. United European Gastroenterol J 2022;10:308-18.
18. Aceves SS, Alexander JA, Baron TH, Bredenoord AJ, Day L, Dellon ES, et al. Endoscopic approach to eosinophilic esophagitis: American Society for Gastrointestinal Endoscopy Consensus Conference. Gastrointest Endosc 2022;96:576-592.e1.
19. Dellon ES, Hirano I. Epidemiology and natural history of eosinophilic esophagitis. Gastroenterology 2018;154:319-332.e3.
20. Mukkada V, Falk GW, Eichinger CS, King D, Todorova L, Shaheen NJ. Health-related quality of life and costs associated with eosinophilic esophagitis: a systematic review. Clin Gastroenterol Hepatol 2018;16:495-503.e8.
21. Anderson J, Moonie S, Hogan MB, Scherr R, Labus B, Word J. Cost of chronic inflammatory disease: the impact of eosinophilic esophagitis in Nevada. J Dig Dis 2020;21:12-9.
22. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci 2021;14:86-93.
23. Visaggi P, Barberio B, Gregori D, Azzolina D, Martinato M, Hassan C, et al. Systematic review with meta-analysis: artificial intelligence in the diagnosis of oesophageal diseases. Aliment Pharmacol Ther 2022;55:528-40.
24. Visaggi P, de Bortoli N, Barberio B, Savarino V, Oleas R, Rosi EM, et al. Artificial intelligence in the diagnosis of upper gastrointestinal diseases. J Clin Gastroenterol 2022;56:23-35.
25. Cotton CC, Betancourt R, Randall C, Perjar I, Bookhout C, Woosley JT, et al. A model using clinical and endoscopic characteristics identifies patients at risk for eosinophilic esophagitis according to updated diagnostic guidelines. Clin Gastroenterol Hepatol 2021;19:1824-1834.e2.
26. de Bortoli N, Penagini R, Savarino E, Marchi S. Eosinophilic esophagitis: update in diagnosis and management. Position paper by the Italian Society of Gastroenterology and Gastrointestinal Endoscopy (SIGE). Dig Liver Dis 2017;49:254-60.
27. Lucendo AJ, Molina-Infante J, Arias A, von Arnim U, Bredenoord AJ, Bussmann C, et al. Guidelines on eosinophilic esophagitis: evidence-based statements and recommendations for diagnosis and management in children and adults. United European Gastroenterol J 2017;5:335-58.
28. Dellon ES, Liacouras CA, Molina-Infante J, Furuta GT, Spergel JM, Zevit N, et al. Updated international consensus diagnostic criteria for eosinophilic esophagitis: proceedings of the AGREE conference. Gastroenterology 2018;155:1022-1033.e10.
29. Nielsen JA, Lager DJ, Lewin M, Rendon G, Roberts CA. The optimal number of biopsy fragments to establish a morphologic diagnosis of eosinophilic esophagitis. Am J Gastroenterol 2014;109:515-20.
30. HRADR Accessed September 3, 2019. http://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2017-1.pdf
31. Pagès J. Multiple factor analysis by example using R. 1st ed. Chapman and Hall/CRC; 2014.
32. Random decision forests. August 14-16. In: Tin Kam H, editor. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995. Accessed September 3, 2019, http://vision.cse.psu.edu/seminars/talks/2009/random_tff/odt.pdf
33. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 2018;19:270.
34. Class-balanced loss based on effective number of samples. June 15-20. In: Cui Y, Jia M, Lin TY, Song Y, Belongie S, editors. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. Accessed September 3, 2019, https://ieeexplore.ieee.org/document/8953804
35. Khushi M, Shaukat K, Alam TM, Hameed IA, Uddin S, Luo S, et al. A comparative performance analysis of data resampling methods on imbalance medical data. IEEE Access 2021;9:109960-75.
36. Yu AC, Eng J. One algorithm may not fit all: how selection bias affects machine learning performance. Radiographics 2020;40:1932-7.
37. van Rhijn BD, Warners MJ, Curvers WL, van Lent AU, Bekkali NL, Takkenberg RB, et al. Evaluating the endoscopic reference score for eosinophilic esophagitis: moderate to substantial intra- and interobserver reliability. Endoscopy 2014;46:1049-55.
38. Ravi K, Talley NJ, Smyrk TC, Katzka DA, Kryzer L, Romero Y, et al. Low grade esophageal eosinophilia in adults: an unrecognized part of the spectrum of eosinophilic esophagitis? Dig Dis Sci 2011;56:1981-6.
39. Tourlamain G, Garcia-Puig R, Gutiérrez-Junquera C, Papadopoulou A, Roma E, Kalach N, et al. Differences in management of eosinophilic esophagitis in Europe: an assessment of current practice. J Pediatr Gastroenterol Nutr 2020;71:83-90.
40. Targownik LE, Fisher DA, Saini SD. AGA clinical practice update on deprescribing of proton pump inhibitors: expert review. Gastroenterology 2022;162:1334-42.
41. Visaggi P, Del Corso G, Gyawali CP, Ghisa M, Svizzero FB, Donati DS, et al. Ambulatory pH-impedance findings confirm that grade B esophagitis provides objective diagnosis of gastroesophageal reflux disease. Am J Gastroenterol 2023;118:794-801.
42. Dellon ES, Rusin S, Gebhart JH, Covey S, Speck O, Woodward K, et al. A clinical prediction tool identifies cases of eosinophilic esophagitis without endoscopic biopsy: a prospective study. Am J Gastroenterol 2015;110:1347-54.
43. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiol Artif Intell 2022;4:e210064.
44. Kim H, Vance R, Shaheen N, Dellon E. The prevalence and diagnostic utility of endoscopic features of eosinophilic esophagitis: a meta-analysis. Clin Gastroenterol Hepatol 2012;10:988-996.e5.
45. Visaggi P, Barberio B, Del Corso G, de Bortoli N, Black CJ, Ford AC, et al. Comparison of drugs for active eosinophilic oesophagitis: systematic review and network meta-analysis. Gut 2023;72:2019-30.

ONLINE REPOSITORY

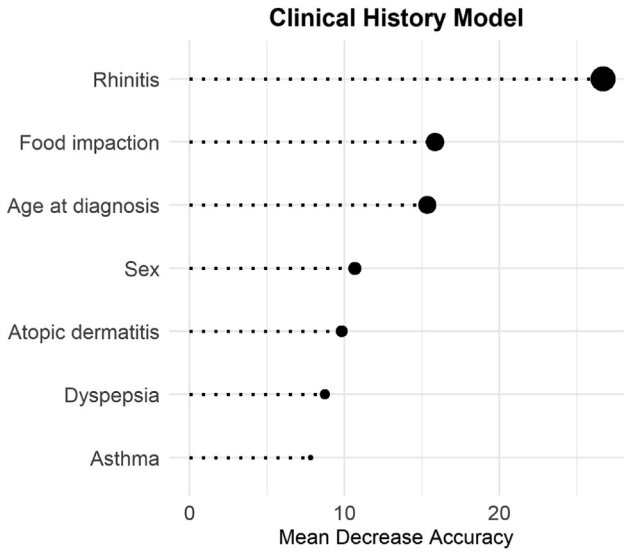

FIGURE E1. Ranking of clinical data for the assessment of the risk of eosinophilic esophagitis.

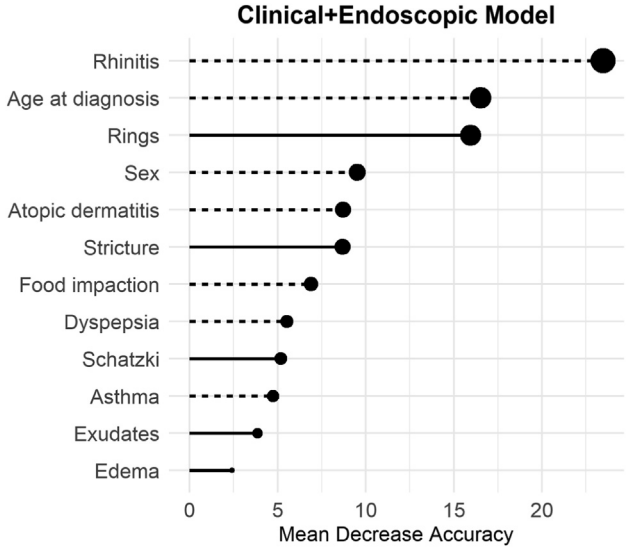

FIGURE E2. Ranking of clinical and endoscopic data for the assessment of the risk of eosinophilic esophagitis.