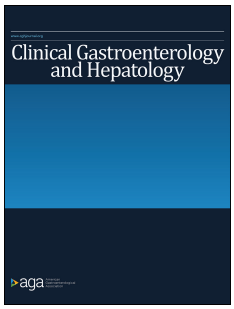
Journal Pre-proof



A MACHINE-BASED LEARNING MODEL FOR RECURRENCE PREDICTION AND TIMING AFTER ENDOSCOPIC ERADICATION THERAPY FOR BARRETT'S ESOPHAGUS

Venkata Akshintala, Samuel Han, Yukun Yan, Raf Bisschops, Case Brennan, Yanna Cai, D. Chamil Codipilly, Cary Cotton, Dayna Early, Steven A. Edmundowicz, Swathi Eluri, Jazmyne Gallegos, Rohit Goyal, Hazem T. Hammad, Rehan Haidry, Thomas Hollander, Khalid Husain, Prasad G. Iyer, Justeena Jojo, Vladimir Kushnir, Srinadh Komanduri, Cadman Leggett, Laurence Lovat, V. Raman Muthusamy, Amit Rastogi, Stefan Seewald, Nicholas J. Shaheen, Adarsh Thaker, Kornpong Vantanasiri, Craig Jones, Sachin Wani

PII: S1542-3565(26)00236-3
DOI: https://doi.org/10.1016/j.cgh.2026.03.026
Reference: YJCGH 60355

To appear in: *Clinical Gastroenterology and Hepatology*
Accepted Date: 20 March 2026

Please cite this article as: Akshintala V, Han S, Yan Y, Bisschops R, Brennan C, Cai Y, Codipilly DC, Cotton C, Early D, Edmundowicz SA, Eluri S, Gallegos J, Goyal R, Hammad HT, Haidry R, Hollander T, Husain K, Iyer PG, Jojo J, Kushnir V, Komanduri S, Leggett C, Lovat L, Muthusamy VR, Rastogi A, Seewald S, Shaheen NJ, Thaker A, Vantanasiri K, Jones C, Wani S, A MACHINE-BASED LEARNING MODEL FOR RECURRENCE PREDICTION AND TIMING AFTER ENDOSCOPIC ERADICATION THERAPY FOR BARRETT'S ESOPHAGUS, *Clinical Gastroenterology and Hepatology* (2026), doi: https://doi.org/10.1016/j.cgh.2026.03.026.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 American Gastroenterological Association Institute. Published by Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

A MACHINE-BASED LEARNING MODEL FOR RECURRENCE PREDICTION AND TIMING AFTER ENDOSCOPIC ERADICATION THERAPY FOR BARRETT'S ESOPHAGUS

Journal Pre-proof

Study Population

2, 511

Patients with BE-related neoplasia achieving CE-IM post-EET



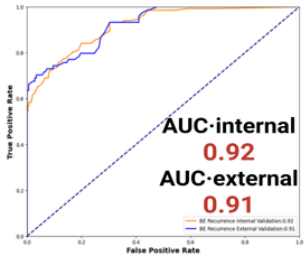
3 Prospective US Databases
Development cohort

US RFA Registry
148 institutions
External validation

Model Development & Validation

Random Forest + Cox Hazards Model
Predicts recurrence risk & timing at 1-, 3-, 5-year intervals
Calculator Development

BE Recurrence



Recurrence Timing

C-index at 1 year
0.70
C-index at 3 year
0.68
C-index at 5 year
0.66

Model Explanation



Clinical Gastroenterology and Hepatology

Journal Pre-proof

A MACHINE-BASED LEARNING MODEL FOR RECURRENCE PREDICTION AND TIMING AFTER ENDOSCOPIC ERADICATION THERAPY FOR BARRETT'S ESOPHAGUS

Venkata Akshintala[1]*, Samuel Han[2]*, Yukun Yan[3], Raf Bisschops[4], Case Brennan[5], Yanna Cai[1], D. Chamil Codipilly[2], Cary Cotton[6], Dayna Early[7], Steven A. Edmundowicz[5], Swathi Eluri[8], Jazmyne Gallegos[5], Rohit Goyal[2], Hazem T. Hammad[5], Rehan Haidry[9], Thomas Hollander[7], Khalid Husain[1], Prasad G. Iyer[10], Justeena Jojo[11], Vladimir Kushnir[7], Srinadh Komanduri[11], Cadman Leggett[2], Laurence Lovat[12], V. Raman Muthusamy[13], Amit Rastogi[14], Stefan Seewald[15], Nicholas J. Shaheen[6], Adarsh Thaker[13], Kornpong Vantanasiri[2], Craig Jones[3], Sachin Wani[6]

1. Division of Gastroenterology, Johns Hopkins Medical Institutions, Baltimore, MD, USA

2. Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

3. Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

4. Department of Gastroenterology and Hepatology, UZ Leuven, Leuven, Belgium

5. Division of Gastroenterology and Hepatology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

6. Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

7. Division of Gastroenterology, Washington University School of Medicine, St. Louis, MO, USA

8. Division of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, FL, USA

9. Department of Gastroenterology, Cleveland Clinic London, London, UK

10. Division of Gastroenterology and Hepatology, Mayo Clinic, Scottsdale, AZ, USA

11. Division of Gastroenterology and Hepatology, Northwestern Feinberg School of Medicine, Chicago, IL, USA

12. Division of Surgery and Interventional Sciences, University College London, London, UK

13. Vatche and Tamar Manoukian Division of Digestive Diseases, University of California, Los Angles, Los Angeles, CA, USA

14. Division of Gastroenterology and Hepatology, University of Kansas, Kansas City, Kansas, USA

Journal Pre-proof

Journal Pre-proof

14. Center for Gastroenterology, Hirlanden Clinic Zurich, Zurich, Switzerland

***co-first author**

Corresponding Author:

Sachin Wani, MD
Division of Gastroenterology and Hepatology
University of Colorado Anschutz Medical Campus
Mail Stop F735
1635 Aurora Ct. Rm. 2.03
Aurora, CO 80045
Email: Sachin.Wani@cuanschutz.edu

Results from this manuscript were presented as an oral presentation at Digestive Disease Week 2024 in Washington, DC

Funding: Supported by the Research Investment in the Scientific Enterprise (RISE) Award, Department of Medicine, University of Colorado Anschutz Medical Campus (SW)

Conflicts of Interest:

These authors disclose the following: Venkata Akshintala is Co-founder of Origin Endoscopc Inc and Sol Endotherapy Inc. Raf Bisschops is a consultant for Medtronic, Pentax Medical, ERBE, Olympus, and Fujifilm. Steven Edmundowicz is a consultant for Olympus, Medpace, Fractyl, Allurion, and Bariendo. Swathi Eluri has received research funding from Steris. Rehan Haidry has received research funding from Cook Medical, Odin Vision, Boston Scientific, Fractyl, Pentax Medical, Merit Medical and Aqua Medical. Samuel Han is a consultant for Boston Scientific. Prasad Iyer is a consultant for Exact Sciences, Castle Biosciences, CDx Medical and Pentax Medical and has received research funding from Exact Sciences, Castle Biosciences, CDx Medical and Pentax Medical. Srinadh Komanduri is a consultant for Boston Scientific, Medtronic, Virgo Video, and Lucid Diagnostics. V. Raman Muthusamy is a Consultant for Medtronic, Boston Scientific, Castle Biosciences, and Pentax Medical, an advisory board member for Merit Medical and has stock options in CapsoVision. Nicholas Shaheen is a consultant for Exact Sciences, Phathom Pharmaceuticals, and Previse and has received research funding from CDx Medical, Steris, Pentax, GIE Medical, Lucid Diagnostics and Aqua. Adarsh Thaker is a consultant for Boston Scientific, Fractyl, Endogenex, Steris, and Neptune Medical. Sachin Wani is a consultant for Exact Sciences, Castle Biosciences, Enterotracker, Cyted, and Lucid and has received research funding from Lucid, Exact Sciences, and CDx Diagnostics. All other authors report no competing interests.

Author Contributions:

Conceptualization: Authors VA, SH, and SW contributed to the conceptualization.
Methodology: Authors VA, SH, YY, YC, CJ, and SW contributed to the methodology.
Software: Authors VA, YY, YC, and CJ contributed to the software
Validation: Authors VA, SH, YY, YC, RB, RH, LL, SS, CJ, and SW contributed to the validation
Formal analysis: Authors VA, YY, YC, and CH contributed to the formal analysis
Investigation: Authors VA, SH, YY, YC, KH, CC, CB, DE, SAE, SE, JG, RG, HTH, VK, VRM, TH, JJ, SK, PGI, CL, DCC, KV, RB, AT, RH, LL, SS, NJS CJ, and SW contributed to the investigation.
Resources: Authors VA, SH, YY, YC, CC, CB, DE, SAE, SE, JG, RG, HTH, VK, VRM, TH, JJ, SK, PGI, CL, DCC, KV, RB, AT, RH, LL, SS, NJS CJ, and SW contributed to the resources.
Data Curation: Authors VA, YY, YC, and CJ contributed to the data curation
Writing – original draft: Authors VA, SH, YY and SW contributed writing the original draft
Writing – review and editing: Authors VA, SH, YY, YC, KH, CC, CB, DE, SAE, SE, JG, RG, HTH, VK, VRM, TH, JJ, SK, PGI, CL, DCC, KV, RB, AT, RH, LL, SS, NJS CJ, and SW contributed to reviewing and editing the final draft
Visualization: Authors VA, YY, YC, and CJ contributed to data visualization.
Supervision: Author SW led the supervision of this study.
Project administration: Author SW led the project administration
Funding acquisition: Author SW led funding acquisition for this project.

Journal Pre-proof

Abbreviations:

EET: endoscopic eradication therapy

BE: Barrett's esophagus

EAC: esophageal adenocarcinoma

IMC: intramucosal carcinoma

HGD: high-grade dysplasia

LGD: low-grade dysplasia

NDBE: non-dysplastic Barrett's esophagus

EMR: endoscopic mucosal resection

ESD: endoscopic submucosal dissection

RFA: radiofrequency ablation

CE-IM: complete eradication of intestinal metaplasia

ML: machine learning

TREAT-BE: Treatment with Resection and Endoscopic Ablation Techniques for Barrett's Esophagus

US RFA: United States Radiofrequency Ablation

BMI: body mass index

PPI: proton pump inhibitors

H2RA: histamine-2 receptor antagonist

IM: intestinal metaplasia

MICE: Multivariate Imputation by Chained Equations

SVM: Support Vector Machines

AUC: area under the receiver operating characteristic curve

C-index: concordance index

SHAP: Shapley Additive exPlanations

SD: standard deviation

Journal Pre-proof

Journal Pre-proof

Abstract

Background and Aims: Tools that can predict recurrence in patients with Barrett's esophagus (BE)-related neoplasia treated with endoscopic eradication therapies (EET) to guide surveillance decisions are needed. We aimed to develop and validate a machine learning (ML)-based prediction tool to predict the risk and timing of recurrence status post EET.

Methods: Three prospective US databases of patients who underwent EET for BE-related neoplasia and achieved complete eradication of intestinal metaplasia (CE-IM, n=1114) were utilized to develop and internally validate a ML-based prediction tool using the Random Forest model and imputation techniques. Predictors incorporated in this model included demographics, endoscopy and pathology results and EET details. A Cox proportional hazards model was utilized to predict the time to recurrence. External validation was performed using the US Radiofrequency Ablation database (n=1397).

Results: BE recurrence occurred in 29.2% (n=734) of patients and BE-related neoplasia recurrence in 10.6% (n=265) with a mean time to recurrence of 21.3 months (mean follow-up 37.7 months). The top predictors for recurrence included BE length, body mass index, age, sessions needed to achieve CE-IM, and baseline histology. The model was well-calibrated and area under the receiver operating characteristic curve (AUC) was 0.92 (95% CI 0.85-0.95) on internal validation and 0.91 (0.87-0.95) on external validation for BE recurrence. For BE-related neoplasia recurrence, the AUC was 0.90 (0.88-0.93). The model had moderate discriminative performance to predict timing of recurrence with a C-index of 0.701 at 1 year (AUC 0.71), 0.68 (AUC 0.69) at 3 years and 0.66 (AUC 0.69) at 5 years.

Conclusions: This US-based externally validated tool accurately predicts BE and BE-related neoplasia recurrence and timing post EET. This practical tool may help provide a personalized approach to surveillance strategies.

Keywords: Barrett's esophagus; endoscopic eradication therapy; recurrence; surveillance

Journal Pre-proof

INTRODUCTION

Endoscopic eradication therapy (EET) has fundamentally altered the management of patients with Barrett's esophagus (BE); a practice endorsed by all gastroenterology society guidelines.[1,2] EET is recommended in patients with BE at the highest risk of progression to invasive esophageal adenocarcinoma (EAC), including patients with intramucosal EAC (IMC), high-grade dysplasia (HGD), and select cases with low-grade dysplasia (LGD). The effectiveness and safety of these interventions is well-established and offers a minimally-invasive treatment approach that avoids the morbidity and mortality associated with esophagectomy. [3]

A contemporary approach to EET includes endoscopic mucosal resection (EMR) or endoscopic submucosal dissection (ESD)] of any visible lesion followed by ablation of the remaining flat BE using radiofrequency ablation (RFA) or cryotherapy to achieve complete eradication of intestinal metaplasia (CE-IM); an *immediate effectiveness goal* of EET.[1] The ability to achieve this endpoint has been demonstrated in randomized controlled trials and several large observational studies.[4] The *intermediate goal* is the detection and management of BE and BE-related neoplasia recurrence with the *long-term goal* of reduction in mortality rates related to EAC.[1] There is a growing body of evidence describing the durability of EET and risk of BE and BE-related neoplasia after achieving CE-IM.[2] Variable rates of recurrence have been reported ranging from 4.9-10.8 per 100-person-years for recurrent IM, 1.0-2.8 per 100-person years for any neoplastic recurrence and 0.7-1.6 per 100-person-years for recurrent HGD/EAC.[5] The variability in reported rates of recurrence may be attributed to study heterogeneity including differences in definitions of CE-IM and recurrence, small sample size, limited follow-up, and varying expertise. Based on these observations, endoscopic surveillance is recommended after patients

achieve CE-IM to identify and treat recurrences at early stages and prevent progression to advanced EAC and related mortality.

Current recommendations for timing of surveillance endoscopies after achieving CE-IM were based on a threshold approach of determining surveillance intervals that yielded a 0.1% risk of recurrence with invasive adenocarcinoma.[6] However, these recommendations do not account for variables other than the highest baseline grade of neoplasia prior to EET. US-based guidelines suggest an endoscopic surveillance strategy at 3, 6, 12 months, and annually thereafter for those with baseline HGD/IMC and at 1, 3 and every 2 years thereafter for those with baseline LGD.[2, 7] European guidelines recommend endoscopic surveillance at 1, 2, 3, 4, 5, 7, and 10 years in patients with baseline HGD/IMC and at 1, 3, and 5 years in patients with LGD.[8] The timing of recurrence, however, varies across studies with recent studies suggesting that recurrence peaks between 1-2 years after achieving CE-IM and remains constant thereafter. [9]

There is a critical need to develop and validate a prediction tool to identify patients at the highest risk for recurrent BE and BE-related neoplasia after achieving CE-IM. The aims of this study were (i) to develop and validate a ML-based decision-making tool that incorporates patient-level data to predict the risk of recurrence and (ii) the timing of recurrence among BE patients status post EET and have achieved CE-IM.

Journal Pre-proof

METHODS

The ML model was built using several US-based databases and registries that included prospective data collection for patients with BE undergoing EET (**Supplement**). External validation was performed using the United States Radiofrequency Ablation (US RFA) Patient Registry; a multicenter registry consisting of patients with BE treated with RFA at 148 institutions (113 community-based and 35 academic-affiliated).[10] Institutional Review Board approval and data user agreements were obtained for each site.

Study Population and Data Collection

Using standardized histological grading, patients with Barrett's neoplasia (indefinite for dysplasia, LGD, HGD, or EAC) who had endoscopic and histologic confirmation of disease, completed EET, and achieved CE-IM were included, while those with non-dysplastic BE were excluded from both internal and external validation cohorts. Specific variables collected within these databases are described in the **Supplement**. Data integrity, relevance, and imputation strategies for missing data are described in the **Supplementary Material**. Based on these results, the combination of mean imputation and the Random Forest classifier was identified as the optimal approach for downstream analyses.

Endoscopic Eradication Therapy and Surveillance Protocol

The EET protocol was not standardized across all individual registries. However, the studies shared several procedural characteristics. These included performing EET under high-dose PPI

Journal Pre-proof

therapy and EMR of visible lesions, followed by ablation of remaining flat BE using RFA every 2-3 months until CE-IM was achieved. Patients were then enrolled in endoscopic surveillance programs after achieving CE-IM. Sampling during endoscopic surveillance post-CE-IM is expected to have evolved based on published guidelines, with patients included earlier in these cohorts undergoing sampling in a 4-quadrant fashion every 1-2 cm through the entire pretreatment length of BE, gastric cardia, and target biopsies from visible lesions. Patients enrolled later in these cohorts underwent sampling using a similar strategy, with biopsies obtained in the distal 2-3 cm of the neosquamous epithelium compared to the entire pretreatment length of BE. Similarly, surveillance endoscopies were conducted more frequently among patients included earlier in these databases and with a baseline diagnosis of HGD/EAC.

Study Definitions and Potential Predictors of Recurrence

CE-IM was defined as no visible BE and no intestinal metaplasia on biopsies at the first post-EET endoscopy, while recurrence was defined as any subsequent detection of intestinal metaplasia, dysplasia, or EAC on surveillance biopsies or EMR; location determined by endoscopist.

We included 24 patient and treatment characteristics that may be associated with BE recurrence. Key demographic variables included age at index EET, race, sex, and body mass index (BMI). Clinical predictors included anti-reflux therapy and family history of BE and EAC. Medication data included use of aspirin, proton pump inhibitors (PPIs), and histamine-2 receptor antagonists (H2Ras). Key clinical factors associated with disease progression included BE duration, hiatal hernia presence and size, and dysplasia grade. Other important predictors such as length of BE,

presence of a visible lesion, erosive esophagitis, and reflux symptoms were also included in the model. Procedural variables included EET modality and the number of sessions required to achieve CE-IM.

Model Development and Statistical Analysis

The experimental setup for predicting the risk of recurrence was designed to address two primary endpoints related to BE recurrence. We first aimed to predict the general recurrence risk of BE and then focus on the recurrence of BE-related neoplasia specifically (see **Supplementary Material**). To analyze the timing of recurrence after CE-IM, the time to BE recurrence was classified into intervals of 6 months, extending up to 5 years post-EET, with particular focus on the 1-, 3-, and 5-year time points. This classification approach ensures a comprehensive assessment of dysplasia progression, recurrence patterns, and timing, while simplifying the predictive modeling process.

Journal Pre-proof

Separate Random Forest models were utilized to predict recurrence risk of BE and BE-related neoplasia. The decision threshold was tuned to maximize AUC and balance sensitivity and specificity for predicting recurrence and its timing. Decision curve analysis was used to assess the clinical utility of predictive models by evaluating net benefits across various threshold probabilities (**Supplementary Material)**. 80% of the dataset was allocated for training and 20% for internal validation. For each predictor, the mean was calculated using the combined training and internal validation datasets, and imputed values were applied across the training, internal validation, and external validation subsets. This preprocessing strategy ensured consistency

across datasets and preserved the integrity of the predictive modeling process. To assess the robustness of model performance to missing data and potential imputation-related optimism, we conducted a prespecified sensitivity analysis excluding candidate predictors with greater than 40% missingness. Model derivation, internal validation, and external validation were repeated using this reduced predictor set, and discrimination and calibration metrics were compared with the primary model.

To predict the time to recurrence, we employed a Cox proportional hazards model to predict the timing of BE recurrence. The Cox model is a multivariable survival analysis approach that evaluates the relationship between predictors and the hazard rate, enabling the estimation of time-to-event outcomes. In this Aim, patients who experienced recurrence during follow-up were treated as events, while patients without recurrence were censored at last follow-up (events occurred at specific follow-up times), while those who remained event-free until the last follow-up were treated as censored. The observable survival time for each patient was determined based on either the event time if an event occurred or the censoring time if no event was observed. Specifically, the Cox model integrated recurrence and timing information into a combined variable represented as (T/F, Time), where "True" indicated recurrence with the associated time to recurrence (in months), and "False" denoted no recurrence, with the total follow-up time recorded. Patients with zero follow-up time were excluded. The training process used all 24 predictors with mean imputation to handle missing data. The proportional hazards assumption was formally evaluated using Schoenfeld residuals, and no significant violations were detected (global test $p = 0.32$). Model performance was evaluated using Harrell's concordance index (C-index) and AUC, with values ranging between 0 and 1. These metrics provided robust

assessments of the model's predictive accuracy and its ability to distinguish between recurrence and non-recurrence over time. This modeling framework allowed the integration of censored data and provided robust estimates of the timing of recurrence, contributing to a comprehensive understanding of the temporal patterns of BE recurrence (See **Supplementary Material** for further details). To account for the possibility that early detection of recurrence may represent undetected disease rather than true biological recurrence, a sensitivity analysis excluding recurrence detected within 3 months of CE-IM confirmation was performed.

External Validation of Model

To validate the ML model's generalizability, we performed external validation using data from the US RFA Registry, which included 1,397 patients. The model's predictions were applied to an independent cohort from the registry, assessing its performance in identifying recurrence risk.

Methodological details regarding model assessment, discriminative ability (Gain Plot Analysis) and model interpretability using SHAP (Shapley Additive exPlanations) analysis are provided in the **Supplementary Material**.

RESULTS

Study Cohort

A total of 2511 patients (1114 in the development and 1397 in the external validation) who underwent EET and achieved CE-IM were included in our model. The Sankey diagrams (**Figure**

Journal Pre-proof

1) mapped the temporal trajectories of patient inclusion and BE recurrence patterns over a quinquennial observational span. In terms of baseline characteristics (**Table 1**), the mean age was 65 years [standard deviation (SD) 10], and the majority were White (92.5%) and males (80.5%). In all, 1068 (42.5%) patients had LGD, 1152 (45.9%) had HGD and 291 (11.6%) had IMC with a mean BE length of 5.1 (SD 3.6) cm. Nearly all patients received RFA (99.6%), and EMR was performed in 30.1% of patients. Achievement of CE-IM required a median of 3.3 treatment sessions (IQR 2-4). The mean follow-up period was 37.7 months (SD 70; range 1–150 months), during which a mean of 3.9 surveillance endoscopies were performed per patient. No patients were lost to follow-up within the first year following CE-IM confirmation; 18% were lost by year 3. **Supplementary Tables 1-2** display baseline characteristics of the development cohort and the external validation cohort.

Journal Pre-proof

Recurrence of intestinal metaplasia and neoplasia

Recurrence of BE occurred in 734 (29.2%) of patients, with BE-neoplasia recurrence detected in 265 (10.6%) patients (**Table 1**). The worst histologic grade of dysplastic recurrences were LGD/indefinite for dysplasia (n=152), HGD (n=69) or EAC (n=44). The mean time to recurrence was 21.3 months, with the distribution of time to first detected recurrence following CE-IM confirmation displayed in **Figure 2**. **Figure 2 also** provides a non-parametric estimation of recurrence over time. In **Figure 2a**, the result illustrates a steady increase in the probability of BE recurrence throughout the follow-up period, indicating a continuous accumulation of recurrence risk. Initially, recurrence is infrequent, as shown by the low probability in the early months. However, the risk gradually rises over time, exceeding 40% at approximately 100 months. Additionally, the CI expands over time, indicating greater uncertainty in later follow-up

periods due to a decreasing number of subjects at risk. **Figure 2b** highlights the probability for BE-neoplasia recurrence. The probability of BE-related neoplasia recurrence is significantly lower and also increases with time. Even at the longest follow-up durations, the probability remains below 20%, suggesting a comparatively reduced risk of BE-related neoplasia recurrence. Compared to BE recurrence, the CI remains narrower, indicating slightly more precision in the estimation. Exclusion of recurrence detected within 3 months of CE-IM confirmation did not alter recurrence patterns or model performance, suggesting that early events did not disproportionately influence predictor associations.

Development and Validation of Machine Learning Model

A total of 24 potential predictors (demographic variables, disease characteristics, family history, presence of hiatal hernia, and EET details) were incorporated into the ML model. The ML model for predicting BE recurrence demonstrated strong discriminatory power. On the internal validation set, the model maintained robust performance, achieving an accuracy of 0.87 (95% CI: 0.82–0.90), a sensitivity of 0.73 (95% CI: 0.68–0.77), a specificity of 0.93 (95% CI: 0.89–0.95), and an AUC of 0.92 (95% CI: 0.85–0.95). Further testing on the external validation set yielded an accuracy of 0.81 (95% CI: 0.77–0.85), with a sensitivity of 0.78 (95% CI: 0.73–0.82), a specificity of 0.83 (95% CI: 0.80–0.85), and an AUC of 0.91 (95% CI: 0.87–0.95), indicating consistent generalizability of the model. The corresponding AUC curves for internal and external validation are presented in **Figure 3**. In sensitivity analyses excluding predictors with greater than 40% missingness, model discrimination was modestly attenuated but remained clinically meaningful. In the external validation cohort, the reduced model demonstrated an AUC of 0.87 for BE recurrence and 0.84 for BE-related neoplasia recurrence, with preserved calibration

across risk strata. These findings indicate that the predictive signal of the model is not driven solely by highly imputed variables.

Explanatory Analysis

The gain plot (**Supplementary Figure 1**) presented the importance scores of the predictors in the BE recurrence predictive model, ranking them based on their contribution to model performance. The top seven predictors, which account for 69.7% of the model's contribution, are as follows: BE length, BMI, age, sessions needed to achieve CE-IM, baseline histology, duration of BE and treatment with RFA.

To complement the gain-based importance, we employed SHAP analysis. The bar plot of mean absolute SHAP values (**Supplementary Figure 2**) ranked predictors by their average contribution to model output, identifying BE length, number of sessions needed to achieve CE-IM, and treatment with RFA as the most impactful predictors. These predictors exerted the largest influence on the predicted recurrence risk, aligning closely with the top-ranked predictors in the gain plot. The beeswarm plot (**Figure 4**) further illustrated the distribution and directionality of SHAP values for each feature across the population. For example, longer BE length and a higher number of CE-IM sessions was consistently associated with higher predicted risk. Given that both the gain plot and SHAP analysis identified BE length as the most influential predictor in the ML model for BE recurrence, we stratified the development cohort into two subgroups based on BE segment length: <5 cm and ≥5 cm. The BE recurrence predictive model was retrained accordingly, and in the external validation cohort, it achieved an accuracy of 0.76

Journal Pre-proof

(95% CI: 0.71–0.82), a sensitivity of 0.73 (95% CI: 0.70–0.78), a specificity of 0.80 (95% CI: 0.76–0.84), and an AUC of 0.86 (95% CI: 0.83–0.91).

In addition, we also performed gain plot and SHAP analysis on the ML model for BE-related neoplasia recurrence. Regardless of ranking by gain or average absolute SHAP values, seven features were among the top 13 predictors across all models and included age at first EET, BE length, sessions needed to achieve CE-IM, RFA, baseline histology, visible lesion, and duration of BE. The accompanying **Supplementary Figure 3** provides a comprehensive overview of the average absolute SHAP values and gain values of these seven predictors across different models. Details regarding model calibration assessment are included in the **Supplementary Figure 4 and Material**.

Predicting Recurrence of Barrett's related neoplasia

The BE-neoplasia recurrence model demonstrated a prediction accuracy of 0.85 (95% CI: 0.81–0.88) on the internal validation set, with a sensitivity of 0.40 (95% CI: 0.36–0.43) and specificity of 0.90 (95% CI: 0.88–0.93). On the external validation set, the accuracy was 0.83, with a sensitivity of 0.37 (95% CI: 0.34–0.41) and specificity of 0.87 (95% CI: 0.83–0.90) (**Supplementary Figure 5**). The corresponding AUC curves for the validation and testing set are displayed in **Supplementary Figure 6.**

Journal Pre-proof

Predicting Timing of Recurrent Barrett's Esophagus

In terms of predicting the timing of recurrent BE, the corresponding AUC curves are displayed in **Figure 5**. The AUC was 0.71 (95% CI: 0.67–0.73) for recurrence at 1 year, 0.69 (95% CI: 0.65–0.72) for recurrence at 3 years, and 0.69 (95% CI: 0.65–0.73) for recurrence at 5 years. At 1 year, the sensitivity was 0.38 (95% CI: 0.32–0.45) and the specificity was 0.84 (95% CI: 0.80–0.87). At 2 years, the sensitivity was 0.36 (95% CI: 0.32–0.44) and the specificity was 0.82 (95% CI: 0.78–0.84), while at 3 years the sensitivity was 0.35 (95% CI: 0.30–0.43) and the specificity was 0.82 (95% CI: 0.77–0.84). The C-index for was 0.701 at 1 year, 0.676 at 3 years, and 0.664 at 5 years, indicating moderate discriminative performance across all time points.

Calculator Development

Utilizing the Random Forest model with data from the first 2 stages, an online risk prediction tool was made that incorporates 20 predictors that provide the overall risk of recurrence while showing the probability of recurrence by time interval after CE-IM (**Supplementary Figure 7**). This is available on https://be-calculator.onrender.com.

DISCUSSION

This study developed and externally validated a ML-based model for predicting both recurrence (BE and BE-related neoplasia) and the timing of recurrence in patients with BE who achieved CE-IM after undergoing EET. This prediction tool serves an important function in providing

Journal Pre-proof

personalized care on when to perform surveillance after successful EET incorporating patient-specific data.

A recent study by van Munster et al., using the Dutch Barrett Expert Center registry and validated in Swiss and Belgian cohorts, achieved a C-index of 0.92 (95% CI, 0.86–0.94), demonstrating excellent model discrimination for recurrence of visible dysplastic BE. Our study differs in several key points that reflect variations in US and Dutch practice patterns: our study required histologic CE-IM confirmation (vs. visible CE-BE), incorporated both random and targeted biopsies (vs. visible recurrences only), included data from expert and community centers (vs. expert centers alone), and examined recurrence of both BE and BE-related neoplasia (vs. neoplasia alone). Therefore, our study is aligned with current US-based guidelines and quality indicators which recommend the target goal of CE-IM and surveillance endoscopies thereafter with random biopsies of the distal 2-3 cm of neosquamous epithelium and target sampling of any visible recurrence.[2, 4, 11-13] Given that fear of recurrence represents a serious concern for patients undergoing EET, identifying recurrence, regardless of the level of dysplasia, is imperative.

Identifying risk factors for recurrence remains critical in developing prediction models for BE recurrence. Multivariate modeling from the Dutch prediction model had found younger age, male sex, HGD/EAC at baseline, length of BE, number of EMR treatments, and visible lesions to be significantly associated with recurrence.[14] The US-based TREAT-BE database had also previously found baseline HGD/EAC to be associated with recurrence with additional risk factors including presence of reflux symptoms and number of treatment sessions required to

achieve CE-IM.[9] In the current prediction model, similar risk factors for BE and BE-related neoplasia recurrence were identified. Obesity is a well-known risk factor for BE and greater disease duration has been associated with the development of dysplastic BE, which may translate to a greater rate of recurrence.[7, 15]

Within this US cohort, recurrence occurred at mean interval of 21.3 months after CE-IM, which is line with prior prospective cohorts from the US and the United Kingdom where recurrence appears to occur most often within the two years.[16-18] Nevertheless, we did find a gradual overall increase in recurrence risk with an approximate risk of 40% at 100 months post-EET (**Figure 1**), which supports the recommendations of current guidelines of continued surveillance after EET.[3, 7] Our prediction model demonstrated moderate discriminative performance at 1, 3, and 5 years post-EET with the C-index decreasing from 0.701 at 1 year to 0.664 at 5 years, which likely reflects the decreasing data at time intervals further out from the time of CE-IM. Furthermore, our model's performance on external validation demonstrates that it is fairly accurate in predicting recurrence.

Several limitations of this study warrant further discussion, the foremost being missing data. Although mean imputation combined with a Random Forest approach was selected to mitigate overfitting, we acknowledge that models excluding predictors with high levels of missingness demonstrated modestly attenuated but consistent performance, suggesting that imputation improved discrimination without solely driving predictive accuracy. Although the model demonstrated strong discrimination for both BE and BE-related neoplasia recurrence,

sensitivities at conventional probability thresholds were modest. This reflects the distinction between discrimination and classification performance. The area under the curve quantifies the model's ability to rank patients by relative recurrence risk, whereas sensitivity is threshold-dependent and influenced by the low incidence of recurrence following EET. Given the heterogeneous mechanisms of recurrence including true biological recurrence, missed disease, and sampling variability, even well-performing models are unlikely to achieve high sensitivity without substantial loss of specificity. Importantly, the clinical intent of this model is not to exclude recurrence, but to stratify patients by relative risk in order to individualize surveillance intensity. Another limitation is the limited and variable follow-up duration for the included patients and the exclusion of patients who received EET and had no follow-up. The CI for the probability of recurrence was noted to expand over time, indicating greater uncertainty in the later follow-up periods due to a decreasing number of patients with longer follow-up. This likely also accounts for the decreasing C-index of our model in predicting the timing of recurrence at increasing time intervals after CE-IM. The authors also acknowledge that surveillance protocols changed over time and were not standardized for this study, which may have resulted in reduced number of surveillance endoscopies with sampling impacting the number and timing of recurrences. The impact of surveillance biopsy protocol, resection strategy (EMR vs. ESD), and anti-reflux surgery on detection of recurrence could not be assessed in this study. Lastly, limitations in the datasets precluded inclusion of potential drivers of carcinogenesis, such as biomarkers or markers of central obesity, into the model.

The strengths of this study include its application of a ML model, which provided a novel approach to incorporate extensive patient-level data to accurately provide risk stratification for

recurrence. Patient-level data from prospective EET databases within the US provided a large sample size with a relatively heterogeneous population. The inclusion of academic and community-based practices across the US also adds to the generalizability of these results. This model assessed a wide range of clinically relevant predictors and reported on the endpoint of BE and BE-related neoplasia recurrence. In analyzing the time to recurrence, we were also able to incorporate the risk of recurrence at each point of follow-up, which adds to the practical value of our model. Lastly, from a methodological standpoint, utilization of a Random Forest model avoided overfitting to help create a model that would work accurately in clinical practice.

Future studies will need to externally validate and refine this model that includes clinical variables and biomarkers using large prospective international cohorts of patients with BE undergoing EET in a standardized manner. Future modeling studies will need to define acceptable thresholds for prediction models that can change clinical practice. These future endeavors need to account for diagnostic characteristics (sensitivity, specificity, AUC), impact on cancer deaths averted, reduction in unnecessary upper endoscopies, feasibility regarding implementation and incorporation into electronic medical records. Recent US-based guidelines recommend that practices and endoscopists performing EET maintain and monitor data regarding key quality metrics such as CE-IM and recurrence rates; a recommendation that should aide in implementing this prediction tool in clinical practice.[2, 7] Future modeling studies and surveillance guidelines for post-EET will need to account for mortality risks unrelated to EAC. A recent Dutch study showed that the risk of mortality from unrelated (non-EAC) causes was greater than 40 times higher than the risk of EAC-related mortality.[19] Future prediction models should also incorporate tools such as Charlson Comorbidity Index to help determine patients

most likely to benefit from surveillance after EET. Finally, this study raises another important question regarding the management of recurrent NDBE that should be addressed in future modeling studies – should recurrent NDBE or only recurrent neoplasia be the trigger for repeat EET?

In conclusion, this study developed and externally validated a ML-based prediction tool for BE and BE-related neoplasia using prospective US-based registries that included a large dataset of patients with BE-related neoplasia who successfully underwent EET and achieved CE-IM. This tool also provides insight into the timing of recurrence also using clinical variables associated with recurrence. This tool may be implemented easily, guide decision making and has the potential to personalize surveillance practices in this patient population.

Journal Pre-proof

References

1. Wani S. Personalized Approach to the Post-Endoscopic Eradication Therapy Barrett's Esophagus Patient. Gastroenterology 2022;163:39-42.
2. Rubenstein JH, Sawas T, Wani S, et al. AGA Clinical Practice Guideline on Endoscopic Eradication Therapy of Barrett's Esophagus and Related Neoplasia. Gastroenterology 2024;166:1020-1055.
3. Standards of Practice C, Wani S, Qumseya B, et al. Endoscopic eradication therapy for patients with Barrett's esophagus-associated dysplasia and intramucosal cancer. Gastrointest Endosc 2018;87:907-931 e9.
4. Wani S, Muthusamy VR, Shaheen NJ, et al. Development of Quality Indicators for Endoscopic Eradication Therapies in Barrett's Esophagus: The TREAT-BE (Treatment With Resection and Endoscopic Ablation Techniques for Barrett's Esophagus) Consortium. Am J Gastroenterol 2017;112:1032-1048.
5. Kolb JM, Davis C, Wani S. Durability of Endoscopic Eradication Therapy for Barrett's Esophagus-Related Neoplasia: A Call for Centralized Care. Gastroenterology 2022;162:343-345.
6. Cotton CC, Haidry R, Thrift AP, et al. Development of Evidence-Based Surveillance Intervals After Radiofrequency Ablation of Barrett's Esophagus. Gastroenterology 2018;155:316-326.e6.
7. Shaheen NJ, Falk GW, Iyer PG, et al. Diagnosis and Management of Barrett's Esophagus: An Updated ACG Guideline. Am J Gastroenterol 2022;117:559-587.
8. Weusten B, Bisschops R, Dinis-Ribeiro M, et al. Diagnosis and management of Barrett esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 2023;55:1124-1146.
9. Wani S, Han S, Kushnir V, et al. Recurrence Is Rare Following Complete Eradication of Intestinal Metaplasia in Patients With Barrett's Esophagus and Peaks at 18 Months. Clin Gastroenterol Hepatol 2020;18:2609-2617 e2.
10. Wolf WA, Pasricha S, Cotton C, et al. Incidence of Esophageal Adenocarcinoma and Causes of Mortality After Radiofrequency Ablation of Barrett's Esophagus. Gastroenterology 2015;149:1752-1761 e1.
11. Sharma P, Katzka DA, Gupta N, et al. Quality indicators for the management of Barrett's esophagus, dysplasia, and esophageal adenocarcinoma: international consensus recommendations from the American Gastroenterological Association Symposium. Gastroenterology 2015;149:1599-606.
12. Wolf WA, Pasricha S, Cotton C, et al. Incidence of Esophageal Adenocarcinoma and Causes of Mortality After Radiofrequency Ablation of Barrett's Esophagus. Gastroenterology 2015;149:1752-1761.e1.
13. Sawas T, Alsawas M, Bazerbachi F, et al. Persistent intestinal metaplasia after endoscopic eradication therapy of neoplastic Barrett's esophagus increases the risk of dysplasia recurrence: meta-analysis. Gastrointest Endosc 2019;89:913-925.e6.

14. van Munster SN, Nieuwenhuis E, Bisschops R, et al. Dysplastic Recurrence After Successful Treatment for Early Barrett's Neoplasia: Development and Validation of a Prediction Model. Gastroenterology 2022;163:285-294.
15. Gatenby P, Bhattacharjee S, Wall C, et al. Risk stratification for malignant progression in Barrett's esophagus: Gender, age, duration and year of surveillance. World J Gastroenterol 2016;22:10592-10600.
16. Kahn A, Crook J, Heckman MG, et al. Optimized Surveillance Intervals Following Endoscopic Eradication of Dysplastic Barrett's Esophagus: An International Cohort Study. Clin Gastroenterol Hepatol 2022;20:2763-2771.e3.
17. Wolfson P, Ho KMA, Wilson A, et al. Endoscopic eradication therapy for Barrett's Esophagus related neoplasia. A final 10 year report from the United Kingdom National Halo Radiofrequency Ablation Registry. Gastrointest Endosc 2022.
18. Cotton CC, Wolf WA, Overholt BF, et al. Late Recurrence of Barrett's Esophagus After Complete Eradication of Intestinal Metaplasia is Rare: Final Report From Ablation in Intestinal Metaplasia Containing Dysplasia Trial. Gastroenterology 2017;153:681-688.e2.
19. van Munster SN, Verheij EPD, Ozdemir O, et al. Incidence and Prediction of Unrelated Mortality After Successful Endoscopic Eradication Therapy for Barrett's Neoplasia. Gastroenterology 2024;166:1058-1068.

Journal Pre-proof

FIGURE LEGENDS:

Journal Pre-proof

Figure 1: Sankey diagram of patient inclusion and 5-year Barrett's esophagus recurrence patterns for the entire cohort (Figure 1a), training and internal validation set (Figure 1b) and external validation cohort (Figure 1c).

Figure 2: Kaplan-Meier and cumulative incidence curves of Barrett's esophagus (BE) recurrence (a) and dysplastic BE recurrence (b) and distribution of time (each bar representing a 4-month time interval) to recurrence of BE (c) and dysplastic BE recurrence (d) following complete eradication of intestinal metaplasia (CE-IM).

Figure 3: Predictive performance of the model in predicting the recurrence of Barrett's esophagus

Figure 4: Beeswarm plot of SHAP values for each variable.

Figure 5: Performance of the model for predicting the timing of recurrence of Barrett's esophagus represented by the receiver operating characteristic (ROC) curve.

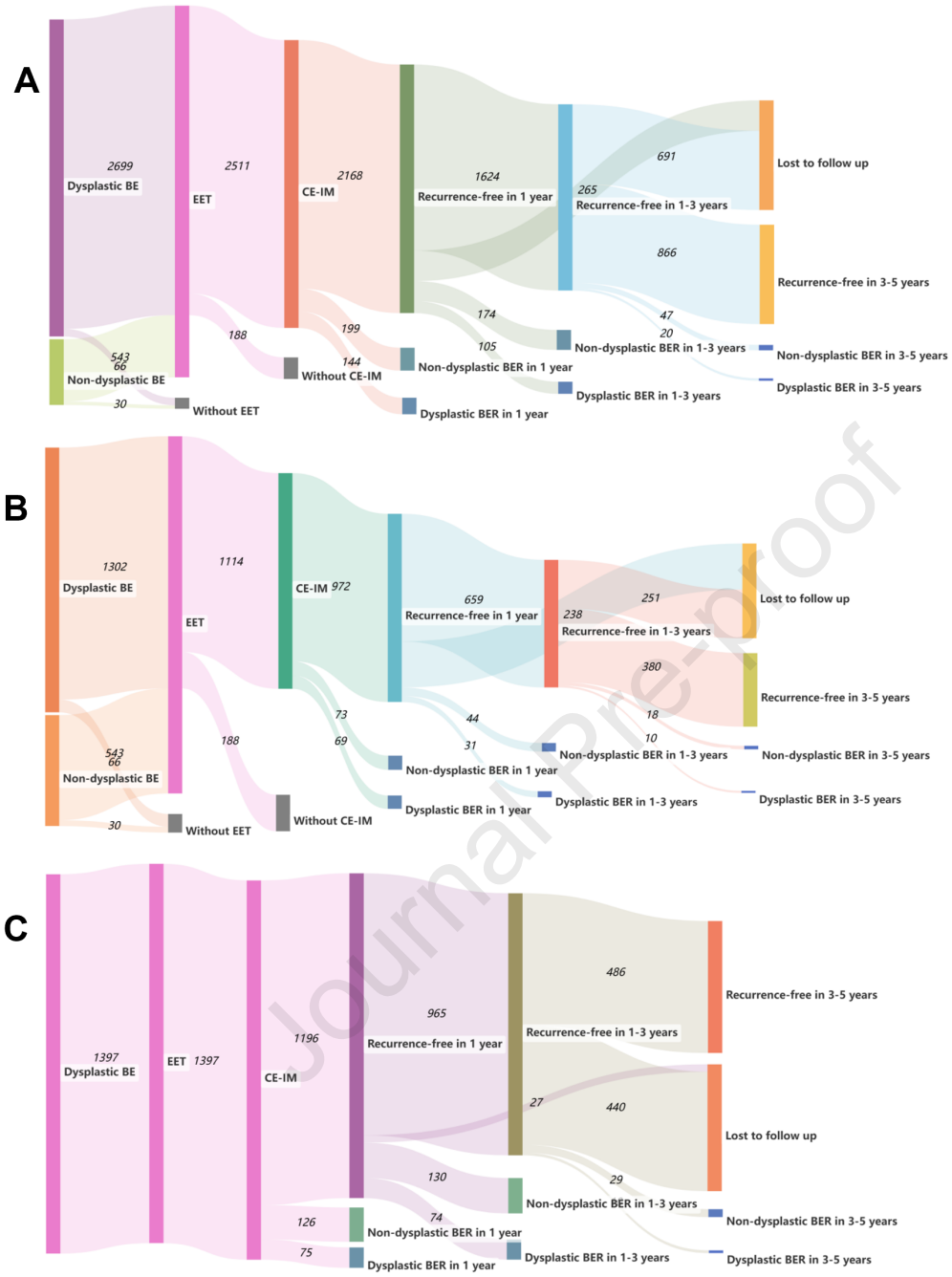
Journal Pre-proof

Table 1: **Baseline patient and disease characteristics and endoscopic eradication therapy details of overall cohort (n=2511)**

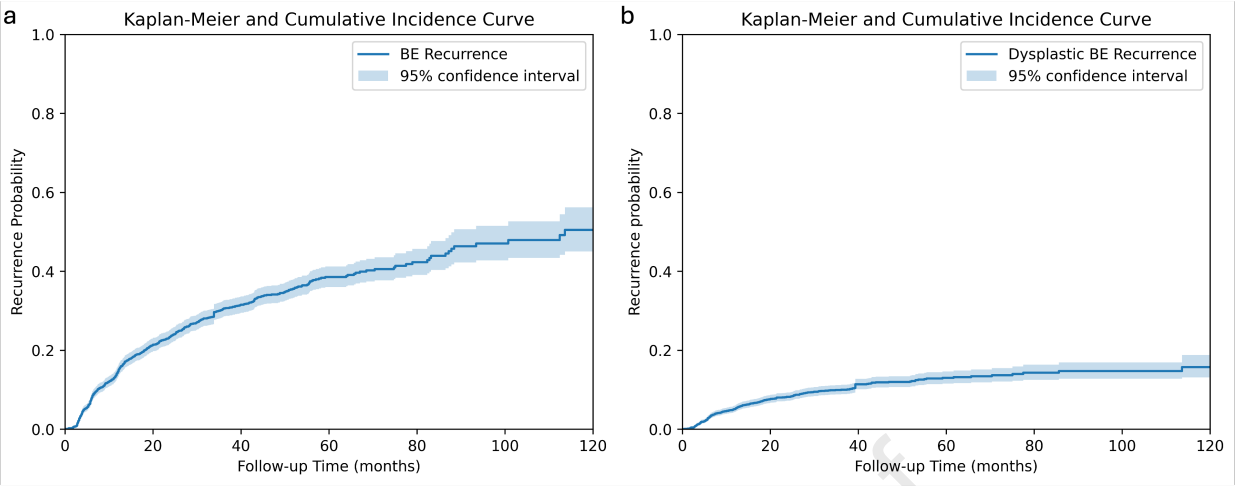
Patient and Disease Characteristics	Mean and Standard deviation or n (%)
Age at index EET	65±10 years
Race	White: 2322 (92.5%), Others: 77 (3.1%)
Sex	Male: 2020 (80.5%), Female: 489 (19.5%)
Body Mass Index	29.9 ±5.9 kg/m^2
Anti-reflux Surgery	124 (5.0%)
Family history of Barrett's esophagus	11 (0.4%)
Family history of esophageal adenocarcinoma	15 (0.6%)
Barrett's esophagus disease duration	18.7±24.4 months
Reflux symptoms	170 (6.8%)
Endoscopy Findings, Treatment and Recurrence Details	
Hiatal Hernia	506 (20.2%)
Hiatal Hernia length	3.5±1.6 cm
Grade of Dysplasia	Low-grade dysplasia: 1068 (42.5%) High-grade dysplasia: 1152 (45.9%) Intramucosal adenocarcinoma: 291 (11.6%)
Length of Barrett's esophagus	5.1±3.6 cm
Visible Lesion	661 (26.3%)
Erosive Esophagitis	258(10.3%)
EET Treatments (may be >1 therapy/patient)	
APC	238 (9.5%)
Cryotherapy	60 (2.4%)
RFA	2503 (99.6%)
EMR	756 (30.1%)
ESD	70 (2.8%)
Treatment sessions to achieve CE-IM	Median: 3.3 [Q1 2.0, Q3 4.0]
Recurrence of BE (with or without neoplasia)	734 (29.2%)
Recurrence of neoplastic BE	265 (10.6%) IND 40, LGD 112, HGD 69, IMC 44

EET: endoscopic eradication therapy; APC: argon plasma coagulation; RFA: radiofrequency ablation; EMR: endoscopic mucosal resection; ESD: endoscopic submucosal dissection; CE-IM: complete eradication of intestinal metaplasia; IND: indefinite for dysplasia; LGD: low-grade dysplasia; HGD: high-grade dysplasia; IMC: intramucosal cancer

Journal Pre-proof

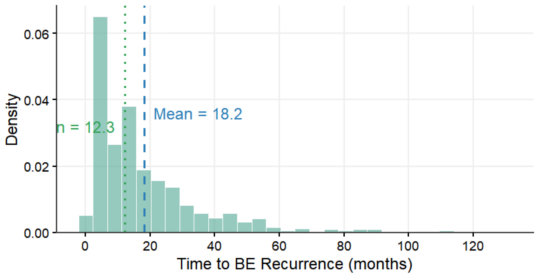


Journal Pre-proof

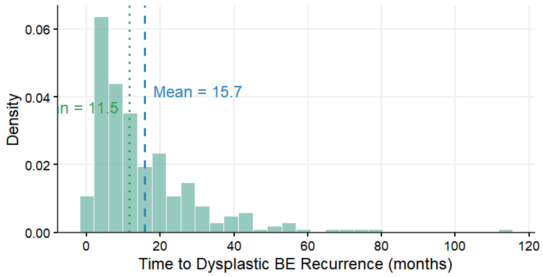


Journal Pre-proof

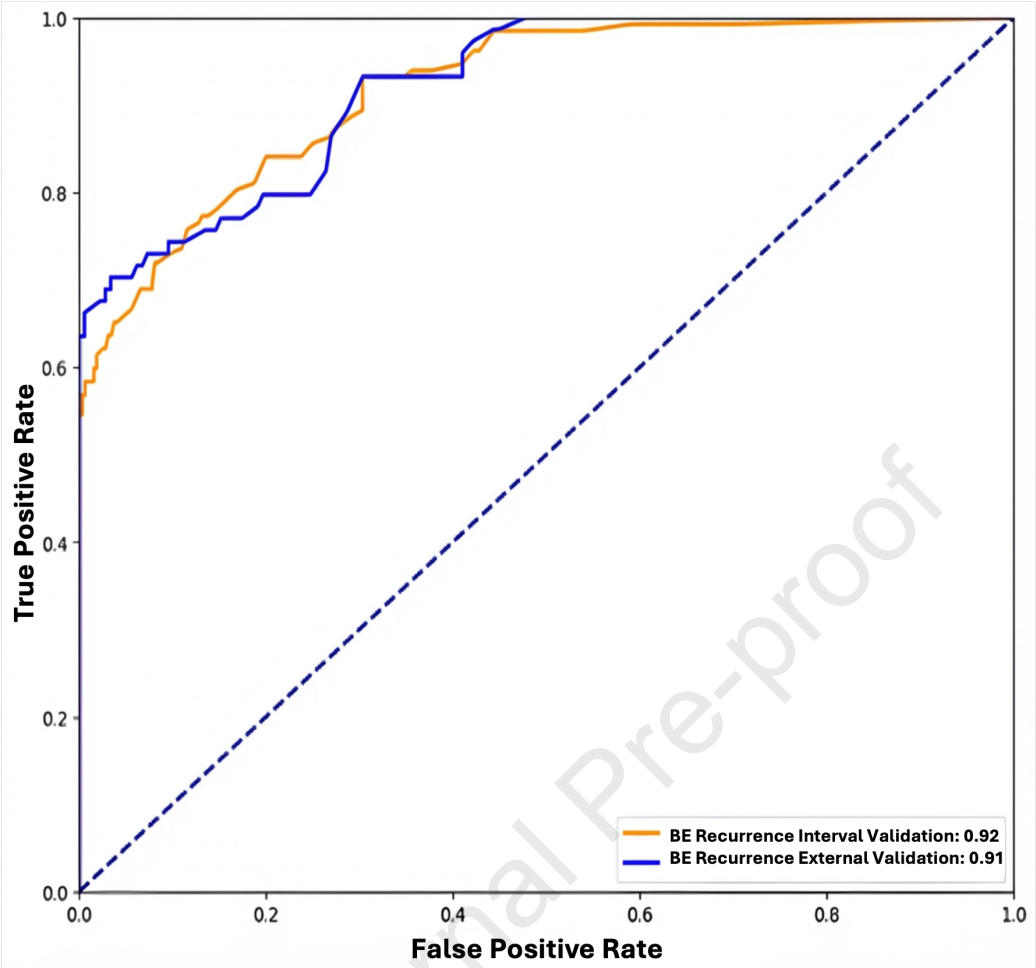
A. Distribution of time to recurrence of all Barrett's esophagus



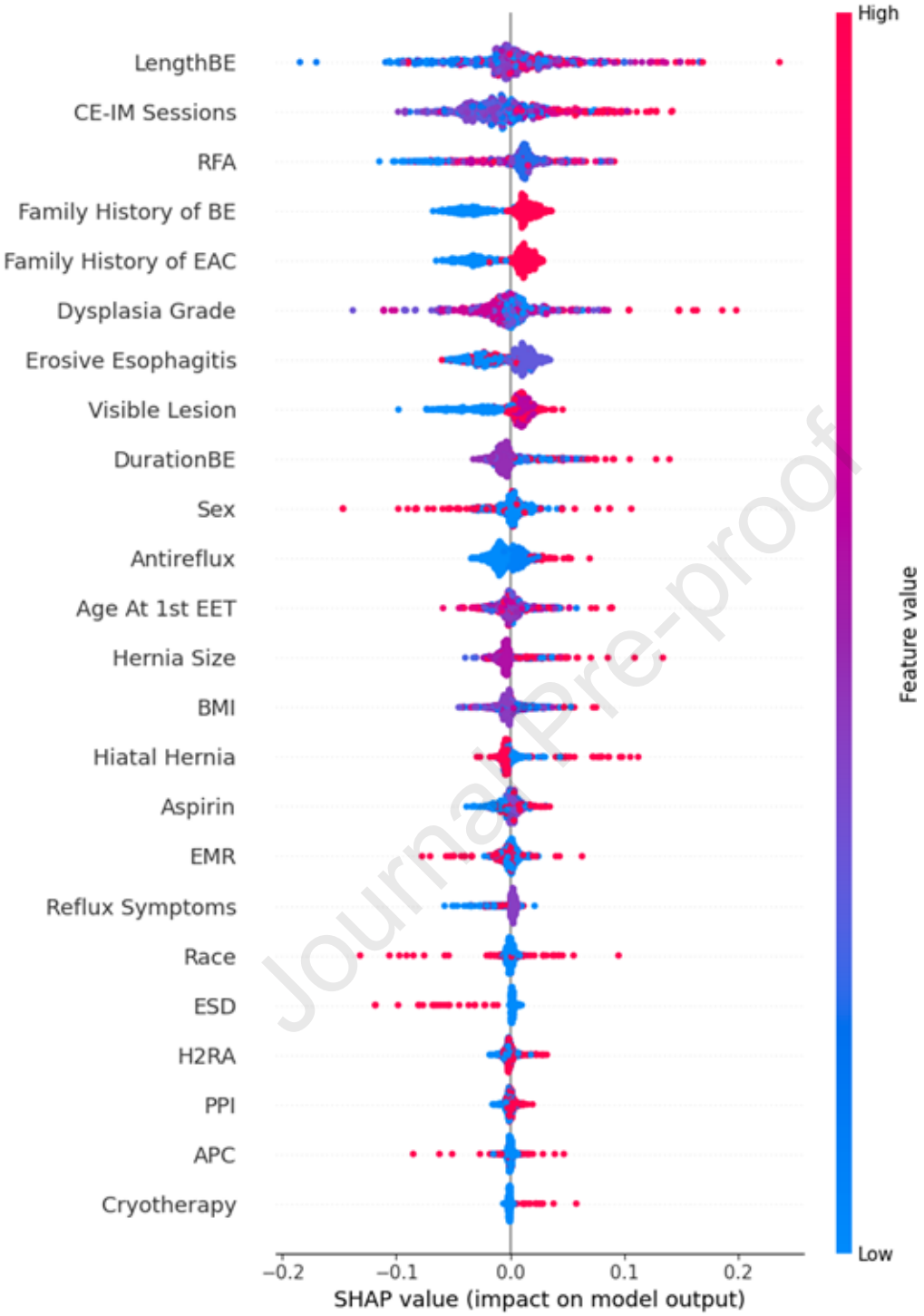
B. Distribution of time to recurrence of dysplastic Barrett's esophagus



Journal Pre-proof



Journal Pre-proof



Journal Pre-proof

Journal Pre-proof

WHAT YOU NEED TO KNOW

Background

Endoscopic eradication therapy represents an effective treatment option for dysplastic Barrett's esophagus, but data are lacking regarding optimal surveillance intervals, for which this study utilized machine-learning to develop an individual-specific risk prediction model.

Findings

Utilizing data from 2511 patients, our model had an area under the receiver operating characteristic curve of 0.91 (0.87-0.95) on external validation for determining the risk of recurrence of Barrett's esophagus after achieving complete eradication of intestinal metaplasia.

Implications for patient care

This risk-prediction model can be utilized by providers to develop a personalized surveillance program for each patient based on patient-, disease-, and treatment-related factors.

Journal Pre-proof

Journal Pre-proof

Supplementary Material

METHODS

Database and Registries

The following databases were utilized for the development of the model: (i) the Treatment with Resection and Endoscopic Ablation Techniques for Barrett's Esophagus Consortium (TREAT-BE): This is an ongoing prospective multicenter study reporting outcomes after EET at four tertiary care centers and was developed to establish quality indicators in EET and report clinical outcomes related to EET.[1-4] (ii) the University of North Carolina hospitals and (iii) the Mayo Clinic (Rochester, Arizona, and Florida) EET databases that included cohorts of patients with BE undergoing EET.[5-8] We ensured that there was no overlap of patients in these 3 databases with the US RFA registry, the external validation cohort.

Data Collection

While information contained in the separate databases and registries varied, the following information was typically collected: (i) demographics, medical comorbidities, smoking history, medications [aspirin, non-steroidal anti-inflammatory drugs, proton pump inhibitors (PPIs), histamine-2 receptor blockers (H2RA)], surgical history, family history of BE and EAC, (ii) highest histologic grade of BE prior to, during and post-EET, (iii) endoscopic findings with BE length, presence of visible lesions at inception and during follow up, hiatal hernia, (iv) details of EET such as performance of EMR, ablative modality, number of treatment sessions, and (v) endoscopic and histologic data in patients undergoing surveillance endoscopy after achieving CE-IM. We did not apply normalization to center continuous variables to a mean of 0 and

standard deviation of 1. Instead, we examined the range and distribution of all continuous predictors and found them to be on comparable scales. As a result, we determined that standardization was not necessary, as the model's performance was not sensitive to the original variable scales.

Data Integrity and Relevance

The dataset utilized in this study underwent rigorous preprocessing to ensure data integrity and relevance. Records with missing outcome data, specifically those related to the recurrence of intestinal metaplasia, were excluded. Due to the missing data in some of the databases, several imputation strategies were explored to mitigate potential biases and improve model robustness. Initial approaches included basic statistical methods such as mean and median imputation, followed by more advanced techniques like Multivariate Imputation by Chained Equations (MICE) and deep generative models, including diffusion-based imputation. Deep generative models, which attempt to generate plausible values in the missing data by learning from the distribution of the observed data. Diffusion models, a more recent generative technique, were also evaluated for their potential in imputing missing values, particularly in complex data distributions The rate of missing entries is described below. In parallel, a range of ML algorithms were evaluated for predictive performance, including Random Forest, XGBoost, Support Vector Machines (SVM), and Decision Tree classifiers. Model selection was guided primarily by prediction accuracy and the area under the receiver operating characteristic curve (AUC).

Journal Pre-proof

Predictors and statistics analysis

Key demographic predictors included age at first endoscopic eradication therapy (EET), with only 4 missing values (0.16%), and Race, which had 112 missing values (4.46%). Sex was nearly complete, with just 2 missing entries (0.08%). Body Mass Index (BMI) exhibits substantial missingness, with 1,412 missing entries (56.23%).

Clinical predictors display a wide range of completeness. Anti-reflux therapy data was missing in 953 cases (37.95%), while family history of BE and esophageal adenocarcinoma (EAC), exhibited high missing rates, each with 2,284 missing entries (90.96%). Medication use, such as aspirin and proton pump inhibitors (PPI), shows moderate missingness, with 1,410 (56.15%) and 1,839 (73.24%) missing values, respectively. Similarly, data on H2 receptor antagonists (H2RA) was incomplete in 1,850 cases (73.68%). Key clinical factors associated with disease progression, such as duration of BE, hiatal hernia presence, hernia size, and dysplasia grade, showed variable missingness. Duration of BE was absent in 1,942 cases (77.34%), and hiatal hernia data was missing in 1,917 cases (76.34%). Hernia size had the highest rate of missingness among these (2,020 missing values; 80.45%), whereas dysplasia grade was complete.

Other important predictors, such as length of BE, visible lesion, and erosive esophagitis, had moderate missing rates, with 1,627 (64.79%), 1,400 (55.75%), and 1,849 (73.64%) missing values, respectively. Reflux symptoms data showed one of the highest levels of missingness, with 2,088 missing values (83.15%).

Procedural predictors, including APC, Cryotherapy, Endoscopic Submucosal Dissection (ESD), and Endoscopic Mucosal Resection (EMR), showed complete data. Radiofrequency Ablation (RFA) data has minimal missingness (8 missing values; 0.31%). Finally, CE-IM sessions exhibited a similarly low missing rate, with only 8 entries missing (0.36%). Of note, we included RFA in the modeling framework despite nearly all participants receiving it because treatment history is clinically relevant in real world practice and may influence recurrence risk in heterogeneous ways depending on patient context.

Model Development

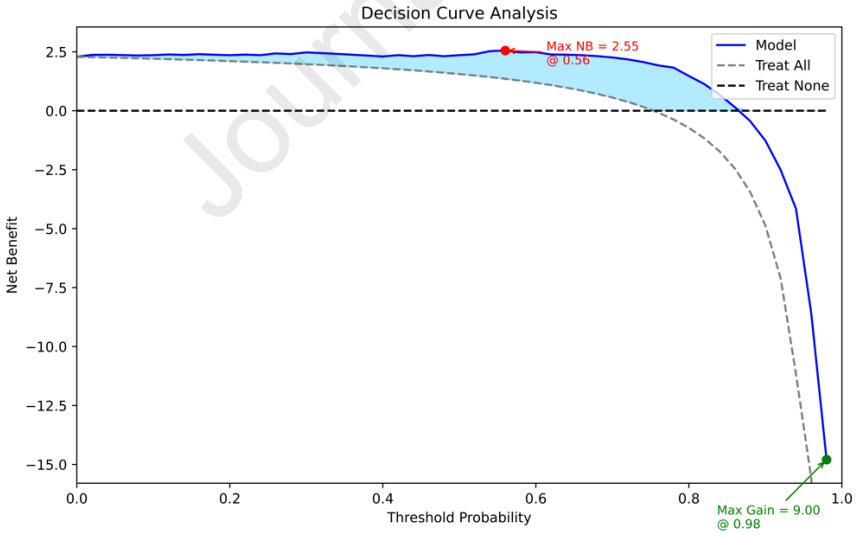
To predict the grade of dysplasia on recurrence, the dysplasia grade of recurrence (0-5) was categorized as follows: grade 2: low-grade dysplasia, grade 3: high-grade dysplasia, grade 4: esophageal adenocarcinoma, and grade 5: indefinite for dysplasia. Conversely, non-dysplastic Barrett's esophagus (grade 1) and no recurrence (grade 0) were grouped into a separate category.

Journal Pre-proof

Decision Curve Analysis (DCA)

Decision Curve Analysis (DCA) is a method used to assess the clinical utility of predictive models by evaluating net benefits across various threshold probabilities. The X-axis represents the threshold probability, which indicates the likelihood at which a physician would choose to intervene, while the Y-axis shows the net benefit, a measure that accounts for the benefits of true positives and subtracts the harms associated with false positives (as demonstrated below). In DCA, the "Treat All" scenario assumes that all patients receive treatment, and its net benefit is used as a reference for comparison. The "Treat None" scenario, conversely, assumes no

treatment, resulting in a net benefit of zero. The model's net benefit curve demonstrates the performance of the predictive model at different threshold probabilities, providing insights into its clinical utility, particularly in guiding decision-making processes. The results demonstrated that the Random Forest model provided greater clinical utility compared to both the "Treat All" and "Treat None" strategies across a wide range of threshold probabilities. Specifically, the model yielded a maximum net benefit of 2.55 true positives per 100 patients at a threshold of 0.56 (red point). Notably, the model outperformed both alternative strategies across a clinically relevant threshold range of 0.02 to 0.86. The maximum gain in net benefit over the "Treat All" strategy was observed at a threshold of 0.98, reaching a value of 9.00 per 100 patients (green point). Between thresholds of 0.1 and 0.8, the model achieved an average net benefit gain of 0.90, and the area between the model and "Treat All" curve within this range was 0.6233 (light blue area), further supporting the model's overall clinical advantage.



Journal Pre-proof

Cox Proportional Hazards Model

The goal of this analysis was to leverage survival models to estimate and compare the recurrence and survival probabilities of specific given patients. Four synthetic patients based on their clinical predictors were provided here as demo case. The Cox proportional hazards model is applied to quantify the effect of 24 predictors (e.g., age, BMI, BE duration, dysplasia grade) on patient outcomes. The model assumes proportional hazards, meaning the relative risk of an event (e.g., recurrence) between patients remains constant over time. By combining the model's baseline hazard with patient-specific risk scores derived from the predictors, we can estimate time-dependent survival probabilities for each individual.

In our study, the Cox model integrated recurrence and timing information into a combined variable represented as (T/F, Time), where "True" indicated recurrence with the associated time to recurrence (in months), and "False" denoted no recurrence, with the total follow-up time recorded. Patients with zero follow-up time were excluded.

The Cox model is trained on a larger cohort to establish the baseline hazard and predictor coefficients. Using the synthetic patient data, risk scores are calculated and used to predict survival probabilities over time. Kaplan-Meier analysis is then performed to provide a non-parametric estimation of survival probabilities and visualize them as survival curves. To assess recurrence, the complementary probabilities are calculated, showing the cumulative risk of recurrence over time. These recurrence probabilities are summarized in a heatmap, highlighting the differences in risk among patients at various time points.

Journal Pre-proof

Model Assessment

To evaluate the performance of the prediction model, we generated a calibration plot comparing predicted versus observed BE recurrence rates in the external validation set. Predicted risks were grouped into bins, and for each bin, the mean predicted recurrence risk was plotted against the corresponding observed recurrence rate of BE. The observed risk was estimated empirically as the average recurrence rate within each group, while the predicted risk was derived from the model outputs. This method provided a visual assessment of how closely predicted probabilities reflect actual outcomes. In addition to the calibration plot, we performed the Hosmer–Lemeshow goodness-of-fit test to statistically evaluate the calibration of the model. This test compared the observed and expected event frequencies across deciles of predicted risk, with a non-significant p-value indicating no evidence of poor fit. Together, the calibration plot and Hosmer–Lemeshow test provided a comprehensive assessment of the model's ability to generate reliable risk estimates in clinical practice.

Journal Pre-proof

Discriminative Ability (Gain Plot Analysis) and Model Interpretability – SHAP Analysis

To quantify the relative contribution of each predictor to the model's performance, we calculated feature importance scores using the gain metric. A gain plot was generated to visualize the importance of all 24 predictors used in the BE recurrence model. Predictors were ranked by their average gain, with higher values indicating greater influence on the model's predictive accuracy. This analysis identified the variables that contributed most significantly to risk stratification and offered interpretability into the model's underlying decision-making process.

To enhance the interpretability of the Random Forest model and understand the impact of individual predictors on BE recurrence risk, we performed a SHAP (Shapley Additive exPlanations) analysis. SHAP values provide a unified measure of feature contribution by quantifying the marginal effect of each variable on the model's output, based on cooperative game theory. Two visualizations were generated: a bar plot displaying the mean absolute SHAP value for each of the 24 predictors, and a beeswarm plot showing the distribution of SHAP values across all samples. The bar plot ranked predictors by their overall importance in the model, while the beeswarm plot captured both the magnitude and direction (positive or negative) of each feature's impact on individual predictions. Together, these plots offered a comprehensive view of global and local interpretability, highlighting key variables that drive the model's risk predictions and how they vary across patients.

SUPPLEMENTARY RESULTS

SHAP Analysis

The SHAP analysis (Supplementary Figure 4) ranked predictors by their average contribution to model output, identifying BE length, number of sessions needed to achieve CE-IM, and treatment with RFA as the most impactful predictors. These predictors exerted the largest influence on the predicted recurrence risk, aligning closely with the top-ranked predictors in the gain plot. Additional predictors such as family history of BE, family history of EAC and dysplasia grade also demonstrated strong predictive importance.

Given that both the gain plot and SHAP analysis identified BE length as the most influential predictor in the ML model for BE recurrence, we stratified the development cohort into two

subgroups based on BE segment length: <5 cm and ≥5 cm. The BE recurrence predictive model was retrained accordingly, and in the external validation cohort, it achieved an accuracy of 0.76 (95% CI: 0.71–0.82), a sensitivity of 0.73 (95% CI: 0.70–0.78), a specificity of 0.80 (95% CI: 0.76–0.84), and an AUC of 0.86 (95% CI: 0.83–0.91).

Model calibration assessment

Model calibration was assessed by constructing a calibration curve comparing predicted versus observed probabilities of BE recurrence in the external validation cohort. As shown in the calibration plot (**Figure 4**), the points closely follow the 45-degree diagonal reference line, indicating strong agreement between predicted and observed recurrence probabilities across the entire risk spectrum (Hosmer-Lemeshow using 20 bins of 5% of observations in each bin [P = 0.12] and using 10 bins of 10% in each bin [P = 0.16]). This alignment suggests that the model is well-calibrated, providing reliable probability estimates for BE recurrence.

Journal Pre-proof

Supplementary Tables

Supplementary Table 1: Baseline patient and disease characteristics and endoscopic eradication therapy details of development cohort (n=1114)

Patient and Disease Characteristics	Mean and Standard deviation or n (%)
Age at index EET	66±11 years
Race	White: 1003 (90.0%), Others: 35 (3.1%)
Sex	Male: 867 (77.9%), Female: 245 (22.0%)
Body Mass Index	29.8 ±5.9 kg/m^2
Anti-reflux Surgery	86 (7.7%)
Family history of Barrett's esophagus	11 (1.0 %)
Family history of esophageal adenocarcinoma	15 (1.3%)
Barrett's esophagus disease duration	18.7±24.4 months
Reflux symptoms	170 (3.8%)
Endoscopy Findings, Treatment and Recurrence Details	
Hiatal Hernia	506 (20.2%)
Hiatal Hernia length	3.5±1.6 cm
Grade of Dysplasia	Low-grade dysplasia: 420 (37.7%)
	High-grade dysplasia: 532 (47.8%)
	Intramucosal adenocarcinoma: 162 (14.5%)
Length of Barrett's esophagus	5.1±3.6 cm
Visible Lesion	661 (26.3%)
Erosive Esophagitis	258(10.3%)
EET Treatments (may be >1 therapy/patient)	
APC	191 (17.1%)
Cryotherapy	51 (4.6%)

Journal Pre-proof

RFA	814 (73.1%)
EMR	631 (56.6%)
ESD	70 (6.3%)
Treatment sessions to achieve CE-IM	Median: 3.7 [Q1 2.0, Q3 5.0]
Recurrence of BE (with or without neoplasia)	288 (20.6%)
Recurrence of neoplastic BE	105 (9.4%)
	IND 2, LGD 60, HGD 27, IMC 16

EET: endoscopic eradication therapy; APC: argon plasma coagulation; RFA: radiofrequency ablation; EMR: endoscopic mucosal resection; ESD: endoscopic submucosal dissection; CE-IM: complete eradication of intestinal metaplasia; IND: indefinite for dysplasia; LGD: low-grade dysplasia; HGD: high-grade dysplasia; IMC: intramucosal cancer

Journal Pre-proof

Supplementary Table 2: Baseline patient and disease characteristics and endoscopic eradication therapy details of the external validation cohort (n=1397)

Patient and Disease Characteristics	**Mean and Standard deviation or n (%)**
Age at index EET	64±10 years
Race	White: 1319 (94.4%), Others: 42 (3.0%)
Sex	Male: 1153 (82.5%), Female: 244 (17.5%)
Body Mass Index	NA
Anti-reflux Surgery	38 (2.7%)
Family history of Barrett's esophagus	NA
Family history of esophageal adenocarcinoma	NA
Barrett's esophagus disease duration	NA
Reflux symptoms	NA
Endoscopy Findings, Treatment and Recurrence Details	
Hiatal Hernia	NA
Hiatal Hernia length	NA
Grade of Dysplasia	Low-grade dysplasia: 648 (46.4%)
	High-grade dysplasia: 620 (44.4%)
	Intramucosal adenocarcinoma: 129 (9.2%)
Length of Barrett's esophagus	NA
Visible Lesion	NA
Erosive Esophagitis	NA
EET Treatments (may be >1 therapy/patient)	
APC	16 (1.1%)
Cryotherapy	3 (0.2%)
RFA	310 (22.3%)

Journal Pre-proof

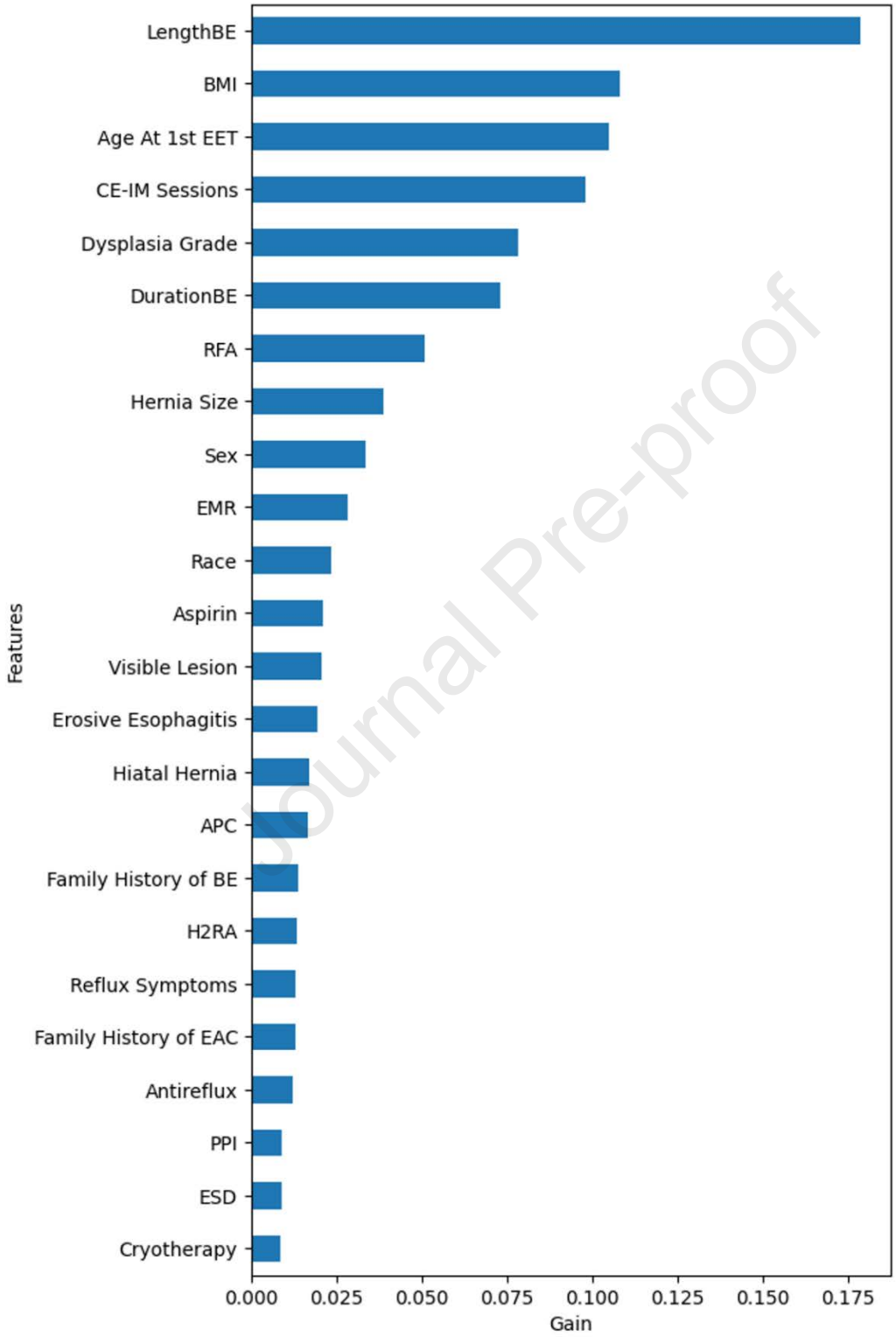
EMR	62 (4.4%)
ESD	0 (0.0%)
Treatment sessions to achieve CE-IM	Median: 2.9 [Q1 2.0, Q3 4.0]
Recurrence of BE (with or without neoplasia)	446 (32.0%)
Recurrence of neoplastic BE	160 (11.5%)
	IND 38, LGD 52, HGD 42, IMC 28

EET: endoscopic eradication therapy; APC: argon plasma coagulation; RFA: radiofrequency ablation; EMR: endoscopic mucosal resection; ESD: endoscopic submucosal dissection; CE-IM: complete eradication of intestinal metaplasia; IND: indefinite for dysplasia; LGD: low-grade dysplasia; HGD: high-grade dysplasia; IMC: intramucosal cancer

Journal Pre-proof

Supplementary Figures

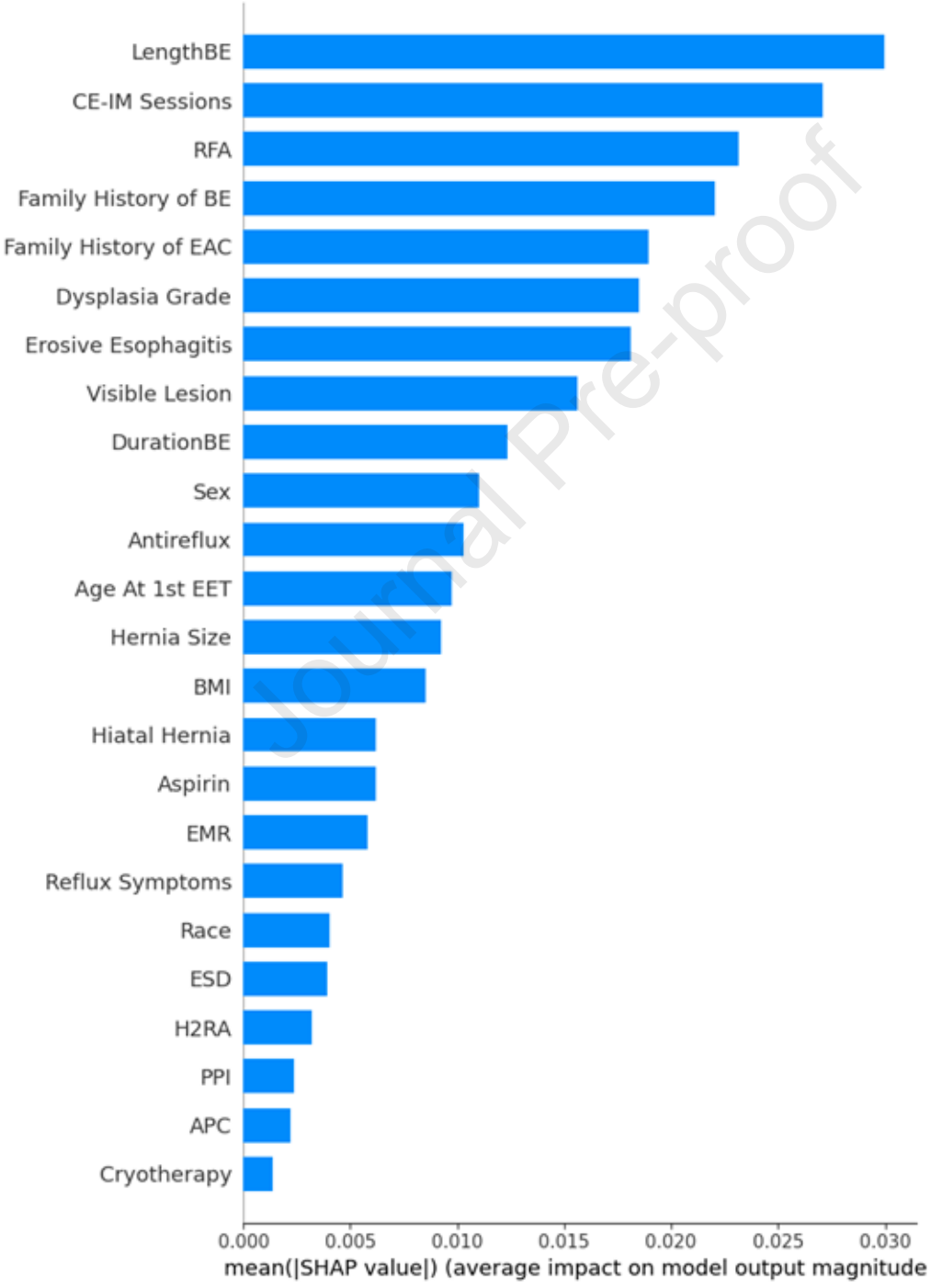
Supplementary Figure 1: Gain plot of predictors for BE recurrence prediction model



Journal Pre-proof

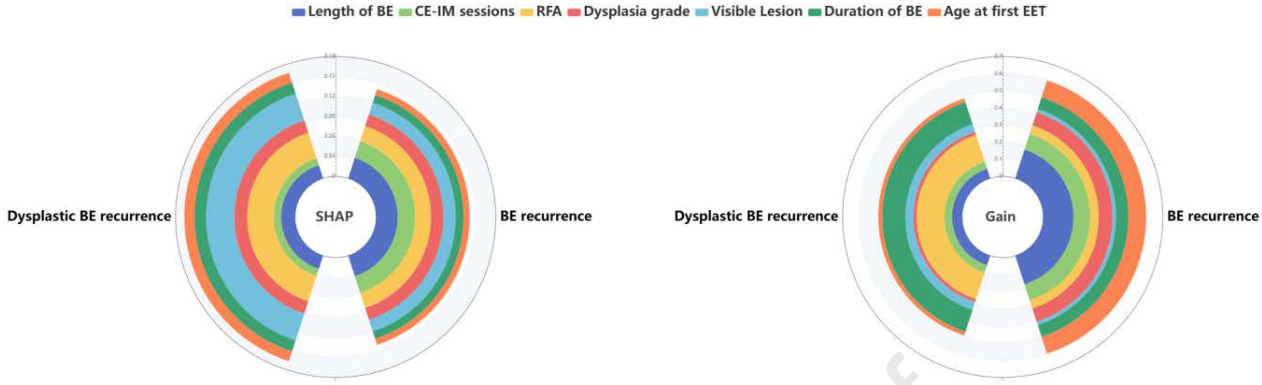
BE: Barrett's esophagus, BMI: body mass index; EET: endoscopic eradiction therapy; CE-IM: complete eradication of intestinal metaplasia; RFA: radiofrequency ablation; EMR: endoscopic mucosal resection; APC: argon plasma coagulation; H2RA: histamine 2 receptor antagonist; EAC: esophageal adenocarcinoma; PPI: proton pump inhibitor; ESD: endoscopic submucosal dissection

Supplementary Figure 2: SHAP analysis results for BE recurrence prediction model



Journal Pre-proof

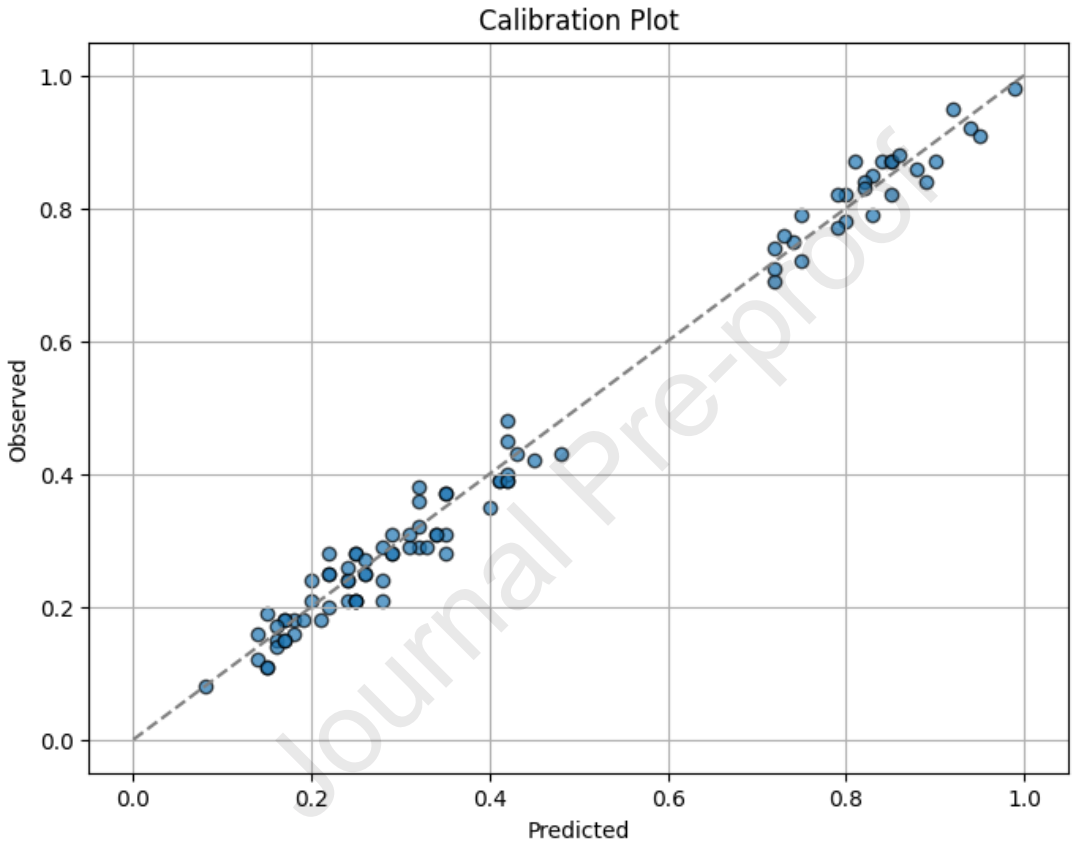
Supplementary Figure 3: Average absolute SHAP values and gain values of seven predictors for recurrence across different models.



BE: Barrett's esophagus, CE-IM: complete eradication of intestinal metaplasia, EET: endoscopic eradication therapy.

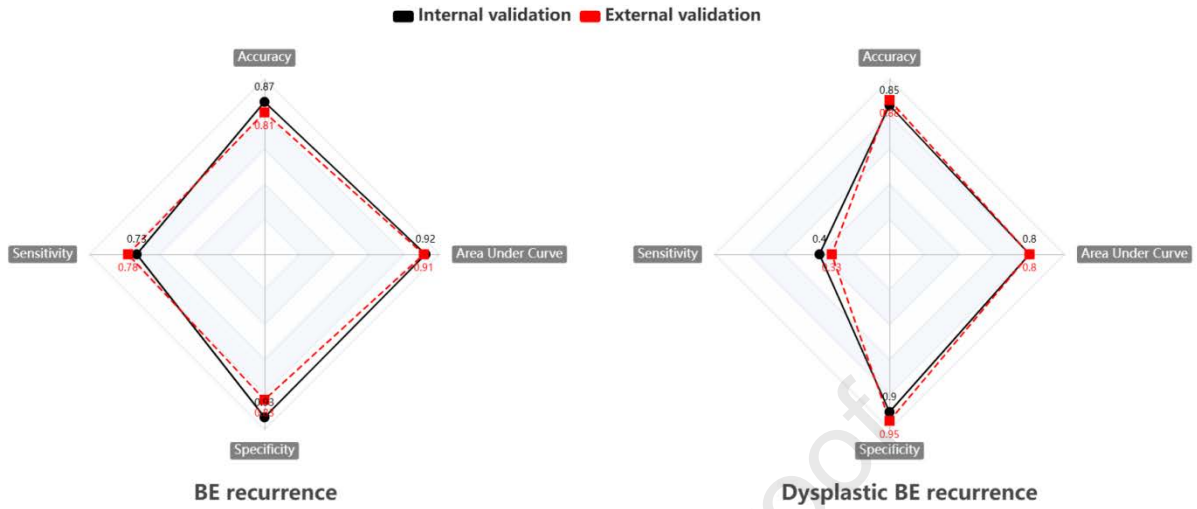
Journal Pre-proof

Supplementary Figure 4: Calibration plot of predicting recurrence of Barrett's esophagus



Journal Pre-proof

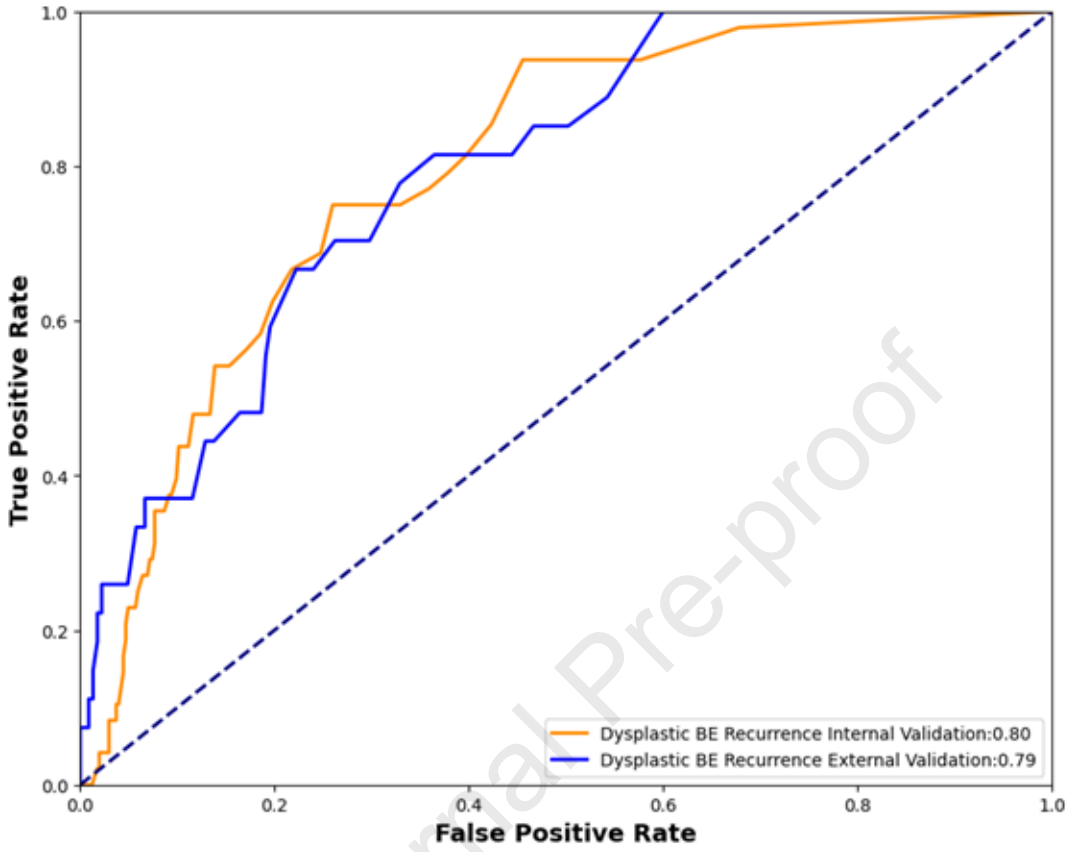
Supplementary Figure 5: **Radar plot of overall performance for recurrence models**



BE: Barrett's esophagus.

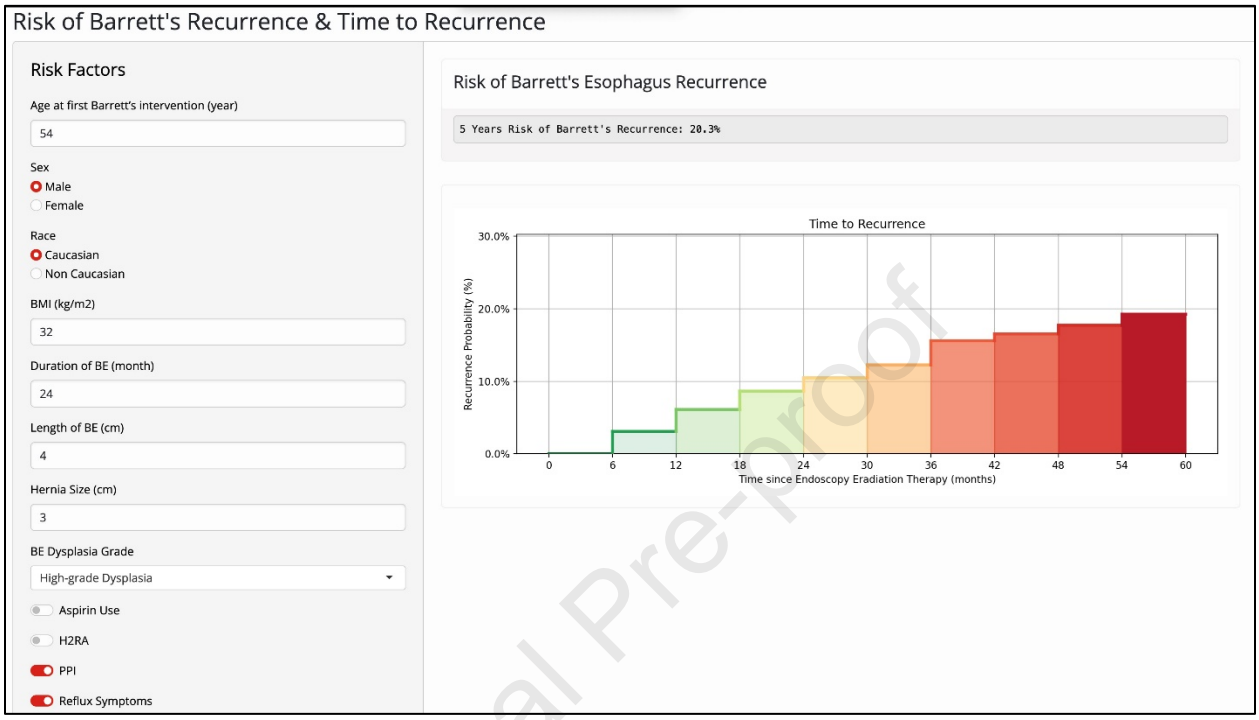
Journal Pre-proof

Supplementary Figure 6: Predictive performance of model in predicting recurrence of Barrett's-neoplasia



Journal Pre-proof

Supplementary Figure 7: Calculator for predicting the risk of recurrence and time to recurrence after endoscopic eradication therapy



Journal Pre-proof

Supplement References

1. Wani S, Muthusamy VR, Shaheen NJ, et al. Development of Quality Indicators for Endoscopic Eradication Therapies in Barrett's Esophagus: The TREAT-BE (Treatment With Resection and Endoscopic Ablation Techniques for Barrett's Esophagus) Consortium. Am J Gastroenterol 2017;112:1032-1048.
2. Wani S, Han S, Kushnir V, et al. Recurrence Is Rare Following Complete Eradication of Intestinal Metaplasia in Patients With Barrett's Esophagus and Peaks at 18 Months. Clin Gastroenterol Hepatol 2020;18:2609-2617 e2.
3. Omar M, Thaker AM, Wani S, et al. Anatomic location of Barrett's esophagus recurrence after endoscopic eradication therapy: development of a simplified surveillance biopsy strategy. Gastrointest Endosc 2019;90:395-403.
4. Mittal C, Muthusamy VR, Simon VC, et al. Threshold evaluation for optimal number of endoscopic treatment sessions to achieve complete eradication of Barrett's metaplasia. Endoscopy 2022;54:927-933.
5. Guthikonda A, Cotton CC, Madanick RD, et al. Clinical Outcomes Following Recurrence of Intestinal Metaplasia After Successful Treatment of Barrett's Esophagus With Radiofrequency Ablation. Am J Gastroenterol 2017;112:87-94.
6. Orman ES, Kim HP, Bulsiewicz WJ, et al. Intestinal metaplasia recurs infrequently in patients successfully treated for Barrett's esophagus with radiofrequency ablation. Am J Gastroenterol 2013;108:187-95; quiz 196.
7. Cotton CC, Wolf WA, Pasricha S, et al. Recurrent intestinal metaplasia after radiofrequency ablation for Barrett's esophagus: endoscopic findings and anatomic location. Gastrointest Endosc 2015;81:1362-9.
8. Sami SS, Ravindran A, Kahn A, et al. Timeline and location of recurrence following successful ablation in Barrett's oesophagus: an international multicentre study. Gut 2019;68:1379-1385.

Journal Pre-proof