

1 **Development of an Interactive Web Dashboard to Facilitate the Reexamination of**
2 **Pathology Reports for Instances of Underbilling of CPT Codes**

3 Jack Greenburg¹, Yunrui Lu², Shuyang Lu², Uhuru Kamau², Robert Hamilton³, Jason
4 Pettus⁴, Sarah Preum⁵, Louis Vaickus⁴, Joshua Levy^{2,4,6,7,*}

- 5 1. Department of Computer Science, Middlebury College, Middlebury, VT
6 2. Program in Quantitative Biomedical Sciences, Dartmouth College Geisel School
7 of Medicine, Hanover, NH
8 3. Department of Pathology, Johns Hopkins Hospital, Baltimore, MD
9 4. Emerging Diagnostic and Investigative Technologies, Department of Pathology
10 and Laboratory Medicine, Dartmouth Health, Lebanon, NH
11 5. Department of Computer Science, Dartmouth College, Hanover, NH
12 6. Department of Epidemiology, Dartmouth College Geisel School of Medicine,
13 Hanover, NH
14 7. Department of Dermatology, Dartmouth Health, Lebanon, NH

15

16 * To whom correspondence should be addressed: joshua.j.levy@dartmouth.edu

17

18 **Corresponding Author Contact Information:**

19 Joshua J. Levy PhD

20 Assistant Professor of Pathology and Dermatology

21 Adjunct Assistant Professor of Epidemiology

22 Faculty, Quantitative Biomedical Sciences

23 Machine Learning Co-Director, Emerging Diagnostic and Investigative Technologies

24 Biostatistics and Bioinformatics Shared Resource, Dartmouth Cancer Center

25 Dartmouth-Hitchcock Medical Center

26 1 Medical Center Drive, Department of Pathology and Laboratory Medicine, Lebanon, NH

27 03756

28 Phone: (925) 457-5752 | Email: joshua.j.levy@dartmouth.edu

29

30

31

32 **Abstract**

33 Current Procedural Terminology Codes is a numerical coding system used to bill for medical procedures
34 and services and crucially, represents a major reimbursement pathway. Given that Pathology services
35 represent a consequential source of hospital revenue, understanding instances where codes may have
36 been misassigned or underbilled is critical. Several algorithms have been proposed that can identify
37 improperly billed CPT codes in existing datasets of pathology reports. Estimation of the fiscal impacts of
38 these reports requires a coder (i.e., billing staff) to review the original reports and manually code them
39 again. As the re-assignment of codes using machine learning algorithms can be done quickly, the
40 bottleneck in validating these reassignments is in this manual re-coding process, which can prove
41 cumbersome. This work documents the development of a rapidly deployable dashboard for examination
42 of reports that the original coder may have misbilled. Our dashboard features the following main
43 components: 1) a bar plot to show the predicted probabilities for each CPT code, 2) an interpretation plot
44 showing how each word in the report combines to form the overall prediction, 3) a place for the user to
45 input the CPT code they have chosen to assign. This dashboard utilizes the algorithms developed to
46 accurately identify CPT codes to highlight the codes missed by the original coders. In order to
47 demonstrate the function of this web application, we recruited pathologists to utilize it to highlight reports
48 that had codes incorrectly assigned. We expect this application to accelerate the validation of re-
49 assigned codes through facilitating rapid review of false positive pathology reports. In the future, we will
50 use this technology to review thousands of past cases in order to estimate the impact of underbilling has
51 on departmental revenue.

52
53 **Keywords:** Natural language processing, current procedural terminology, pathology reports, web
54 development, machine learning, misbilling

55
56

57 Introduction

58 The assignment of Current Procedural Terminology (CPT) Codes represents a critical compensatory
59 component of a hospital's financial system. As the pathology laboratory is responsible for a
60 consequential fraction of services billed, assignment of CPT codes (i.e., coding) via the pathology
61 reporting system is essential for generating hospital revenue. Misbilled codes (i.e., over/underbilling) can
62 cost a hospital significant revenue. The value of identifying misbilled codes has been recognized across
63 many medical subspecialties¹⁻⁷. Pathology reports contain auditable information describing diagnostic
64 case information and additional background information pertaining to services rendered, including what
65 tests/services had been run for the patient and subjective assignment of case complexity. In order for the
66 hospital to receive compensation for these tests, hospitals employ billing staff (i.e., coders) to read,
67 identify, and assign the CPT codes that dictate what tests/services were performed. Accuracy in this
68 coding process is very important since a failure to report any code represents revenue lost by the
69 hospital. Any system that could serve as a "second check" or "suggestion" would have clear and
70 immediate benefits⁸.

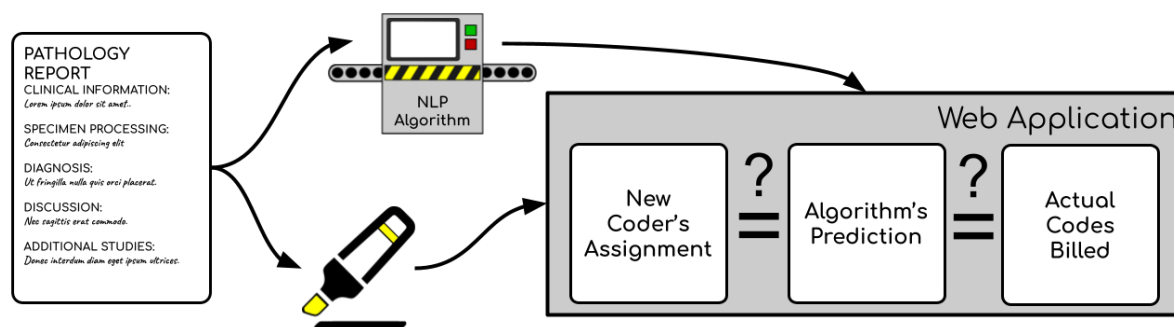
71
72 There are several factors that complicate assignment of CPT codes. Pathology reports do not follow a
73 strict format (i.e., reporting standardized to such a degree that a series of clicks are all that are necessary
74 to both provide a case description and assign codes). In fact, reports may vary drastically from each other
75 (i.e., subjective component) based on the personal style of the signing pathologist, their primary language
76 and where they were trained. Not only does each patient present a unique set of diagnostic/prognostic
77 criteria (which may require ordering many different procedures or stains to hone in on a clear diagnosis),
78 but pathologists may describe similar cases with widely varying lexicon (e.g., certain pathologists may be
79 more verbose than their colleagues or vice versa). An added challenge is that the assignment of primary
80 CPT codes (i.e., CPT 88300-88309), is partially based on the perceived complexity of the case at hand
81 (there do exist guiding principles), meaning that for some codes the coding process relies on a subjective
82 interpretation of report text as it aligns to these criteria.

83
84 Difficulties of assigning CPT codes may be ameliorated through the use of natural language processing
85 (NLP), which provides an automated and objective means to parse through reporting text to arrive at a
86 key finding. While many NLP methods require the use of hard-programmed rules (e.g., regular
87 expressions to identify specific words), machine learning methods provide the level of adaptiveness that
88 a complex search task (i.e., identifying the CPT code) requires by learning, so to speak, patterns of text
89 that signify assignment of specific codes⁹. Many current NLP projects exist to convert free text from
90 pathology reports into structured electronic health record (EHR) information for synoptic case reporting
91 and experiment planning (i.e., a structured search tool for identifying cases to form a study cohort)¹⁰⁻¹⁵.

92
93 The topic of CPT code assignment remains relatively less explored in comparison. A 2019 paper by JJ Ye
94 began to explore the concept of applying machine learning to assist in coding. Ye concluded that
95 augmenting existing coding pipelines with machine learning technologies may prove beneficial and could
96 potentially help avoid misbilling¹⁶.

97
98 Our research group is keen to understand the impact of underbilling on lost hospital revenue. We
99 hypothesize that codes predicted by the machine learning model that are higher than the manually
100 assigned codes have a high likelihood of representing erroneous underbilling. However, we are unable to
101 explore this possibility as no publicly available dashboards exist to help facilitate the process of
102 estimating the scope and impact of misbilling. While there do exist dashboards to assist in pathology
103 report analysis, they do not operate within our specific niche¹⁷. In this study, we document the
104 development of a web application that assists with interrogating the results from an NLP prediction

105 pipeline and comment on its potential application for identifying misbilling. We imagine this tool would
106 serve as a research resource for others to employ in similar assessments. In a future work, we hope to
107 employ such a dashboard to review thousands of past cases in order to provide information on how
108 much money is lost by the hospital due to mistakes in coding. Further in the future, **the dashboard will be**
109 **useful tool to easily test the viability of models, accelerating the automation timeline for the assignment**
110 **of CPT codes (Figure 1).**



111
112 **Figure 1:** Graphical depiction of key innovations: Pathology report is processed by coder and NLP
113 algorithm; coder reviews candidate underbilled codes to determine if code was actually underbilled
114

115 **Methods**

116 **Data Collection:**

117 We utilized the data from 93,039 pathology individual pathology reports, as previously documented in a
118 prior study (5 year period)¹⁸. Of the data that we had access to, we utilized only the report text and the
119 CPT codes that were identified by the coders. While many CPT codes were assigned to these cases, we
120 trained machine learning models to recognize the 38 most prevalent codes. Most of these are ancillary
121 codes (codes assigned with the click of a button for ordered procedures), while we denoted five such
122 codes as primary codes based on case complexity (CPT 88302, 88304, 88305, 88307, 88309). Data
123 collection and preprocessing follows a prior work. Diagnostic and other reporting subsections were
124 formatted using the spaCy and scikit-learn python packages to generate count matrices (i.e., rows denote
125 reports, columns denote words, elements represent term frequency).

126 **Model Training and Dataset Partition**

127 We train the models using XGBoost. XGBoost is gradient boosting decision tree library which generates
128 decision trees that are iteratively morphed based on the errors obtained from the previously fit trees¹⁹.
129 However, other machine learning models (e.g., BERT or another algorithm) could be substituted with
130 minor respecification of the preprocessing workflow²⁰. We chose XGBoost because it was faster to
131 implement, and was found to be slightly more accurate than BERT in a previous work¹⁸. Comparing the
132 accuracy of specific algorithms is outside of the scope of this work as our aim is to document the display
133 outputs that interface with these algorithms.
134

135
136 We trained two separate sets of XGBoost models which can interface with this dashboard. First, we
137 trained a model to differentiate primary CPT codes 88302, 88304, 88305, 88307, and 88309. These codes
138 pertain to varied degrees of complexity for assessment of different conditions. For each set of codes, we
139 trained two versions of each model, one with just the diagnosis section and another with the full report
140 text (i.e., adding other reporting subfields). Next, we trained a second set of models on the 38 most
141 common CPT codes as detailed in our previous work. These models can interrogate CPT codes that

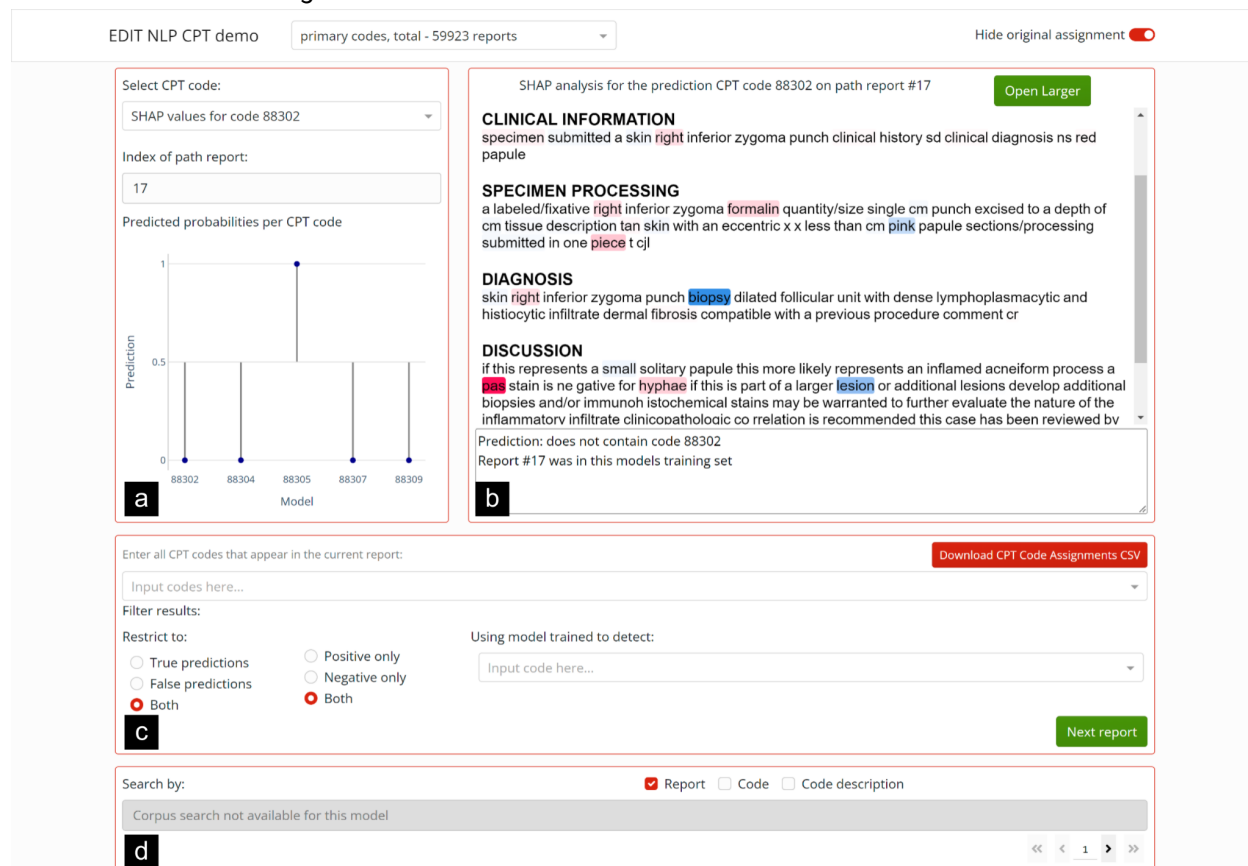
142 correspond to common ancillary tests (e.g., immunostaining), though this was not the primary focus of
143 this work.

144

145 Web Application Development

146

147 As potential ambiguity of free text within pathology reports could lead to the erroneous assignment of
148 primary codes on part of the human rater, we would expect these inaccuracies to be well accounted for
149 using our modeling approach. We developed an interactive web application to facilitate the assessment
150 of instances where our model disagreed from the original CPT code assignment. We have included
151 further detail on the design of the dashboard in this section.



152

153 **Figure 2. Dashboard overview:** a) The bar plot shows the predictions for each CPT code. b) Contains the
154 interpretation plot for a specific report and specific CPT code. c) Where the user selects which reports to
155 display; it is also where the user will input their assignment. d) Search box to search all reports for
156 specific language.

157

158 *Web application overview.* We designed an interactive display output– designed using the Plotly Dash
159 framework– that interfaces with trained machine learning models and allows domain experts to select
160 specific reports to interrogate (**Figure 2**)²¹. Coders operate this dashboard by first selecting a code or
161 report of interest through a text box at the bottom of the dashboard which can search for reports with
162 specific code assignments, text present (**Figure 2D**), or whether they were predicted to be false positives
163 (and particularly false positives of higher case complexity) as denoted by the model (**Figure 2C**). The
164 prediction plot on the left indicates the predicted probabilities of specific codes for the report in question
165 (**Figure 2A**). The interface also has a prediction box which allows coders utilizing the dashboard to

166 reassign CPT codes based on presented information (i.e., original code assignment, model prediction,
167 etc.). A key functionality here is the ability to apply filters to limit which reports the dashboard will display
168 to them, for example allowing the user to assess false negatives (**Figure 2C**). In order to facilitate
169 reassessment, this dashboard includes interpretation plots which highlight words associated with select
170 CPT codes to rapidly corroborate findings with established CPT reporting guidelines (linked references to
171 guidelines / code reference). The interpretation plots utilize SHAP (SHapley Additive exPlanations) values
172 to provide additional insight into how the algorithm made its decisions by demonstrating how each
173 specific input word contributes to the overall CPT code prediction²². SHAP values fall under the umbrella
174 of additive feature attribution methods (i.e. methods that assign a numerical importance value to each
175 feature such that all features sum to the prediction value). These values help to unpack the machine
176 learning algorithm black box, and, in our case, they will help to reveal what patterns the algorithms detect
177 in the text.

178
179 *Assigning and submitting new CPT codes.* The dashboard allows codes to be displayed in the following
180 ways: 1) originally assigned code, 2) reassigned code, and 3) hide code. Display of the original code can
181 be toggled using the *prediction mode* feature that reveals this prediction box and hides the CPT codes
182 previously chosen by the coders, so as not to bias the user inadvertently. Coders can submit new CPT
183 code reassignments, blinded by assignment of the original code but informed by the machine learning
184 derived prediction. The submitted prediction is added to an output file for later analysis (i.e., underbilling).
185 These reassignments can also be used to help train future iterations of these prediction models to reduce
186 further errors/edits.

187
188 *Methods to explore validity of CPT code assignment and underbilling.* We have included additional text on
189 how the SHAP interpretation plots can be configured for different CPT codes and how to filter reports to
190 focus on false positives to facilitate exploration of underbilling candidates in the supplementary material
191 (section "Supplementary Methods").

192
193 **Estimating the Potential Impact of Underbilling**
194 Using data tables downloaded from the Medicare Physician Fee Schedule (MPFS) Look-Up Tool (2022),
195 omitting the modifiers "26" and "TC" (incomplete examination and technical component respectively), we
196 recorded the average non-facility prices for CPT codes 88302-88309: 1) 88302– \$33.32, 2) 88304–
197 \$43.72, 3) 88305– \$74.39, 4) 88307– \$301.33, and 5) 88309– \$457.70²³. We assessed a subset of
198 pathology reports over a five-year span for an estimate of how these primary codes could potentially be
199 reassigned given reexamination. Multiplying the price difference between candidate underbilled reports
200 (e.g., 88307 vs 88305 indicates \$301.33-\$74.39=\$226.94 difference) by the number of reports yields an
201 upper estimate for the potential savings under the assumption that all candidate underbilled reports were
202 truly underbilled. Using our laboratory information system, we pulled and counted all reports over a ten-
203 year span, from January 2011 to December 2021 to extrapolate these findings. This database search
204 identified 749,136 reports containing diagnostic text and including addendums which will be the subject
205 of future inquiry through the developed dashboard.

206
207 **Code Availability**
208 The dashboard is freely available and can be adapted to fit a wide variety of models and purposes. We
209 have included the source code and a small set of examples at the following URL:
210 https://github.com/jackgreenburg/cpt_code_app. The software can be installed locally using Python
211 (*pip*) from this GitHub repository. Note that installing web application locally requires installation of
212 Python 3.8 or above. We have also included a small online demo at the following URL:
213 <https://edit.cpt.code.demo.levylab.host.dartmouth.edu/> (**user:** edit_user, **password:** qdp_2022).

214

215 **Results**

216 **Examples to Illustrate Operation and Usage of the Web App**

217 Someone capable of assigning CPT codes (i.e., a coder) presented with the dashboard will reassign CPT
218 codes via the following process: 1) Select which refined sequence of reports they want to view, 2) View
219 the interpretation plots for each code, 3) Submit the new CPT code assignment. In the following section,
220 we demonstrate several examples of reports analyzed using the web application to provide an idea on the
221 scope of its usage:

222

223 *Example of true positive cases correctly assigned, e.g., CPT 88305.* Using the dashboard with the primary
224 code model (i.e., case complexity of examination under a microscope—CPT codes 88302, 88304, 88305,
225 88307, and 88309) we evaluated whether reports were likely to be improperly billed. In Figure 2, we
226 demonstrate a simple example of how the SHAP interpretation plot can be employed to assess the
227 validity of the finding. In this example, the primary code model assigned CPT code 88305. The
228 interpretation plot identified words relating to punch biopsies for assignment of the code for a
229 Dermatopathology case. Punch biopsies are commensurate with assignment of code 88305 as
230 compared to more complex excisions in Dermatopathology²⁴. If this specific report was presented to a
231 coder through this dashboard prior to assignment of a CPT code, they would benefit from being able to
232 use that information to help rapidly determine the assignment. The report in question was assigned the
233 CPT code 88305 by the original coder, so we can be reasonably confident that this was an accurate
234 billing. Because the original coder and our algorithm agree, we are unlikely to reevaluate this case for
235 underbilling. As our models were highly accurate across this corpus, the majority of reports were true
236 positives. For this reason, we have developed methods to automatically display cases of interest (i.e.,
237 underbilling candidates).

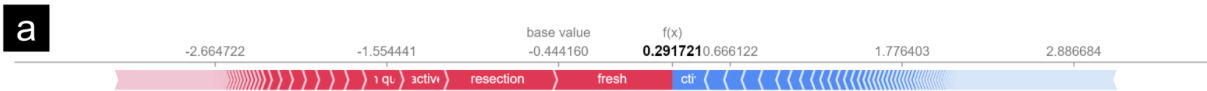
238

239 *Example of underbilled code identified by the dashboard (CPT 88305 to 88307).* If we restrict the dashboard
240 to only show all false positive predictions, then the dashboard displays to the coder a more focused
241 subset of the reports that all have heightened potential to contain missed CPT codes.

242

SHAP analysis for the prediction CPT code 88307 on path report #1438

✕



CLINICAL INFORMATION

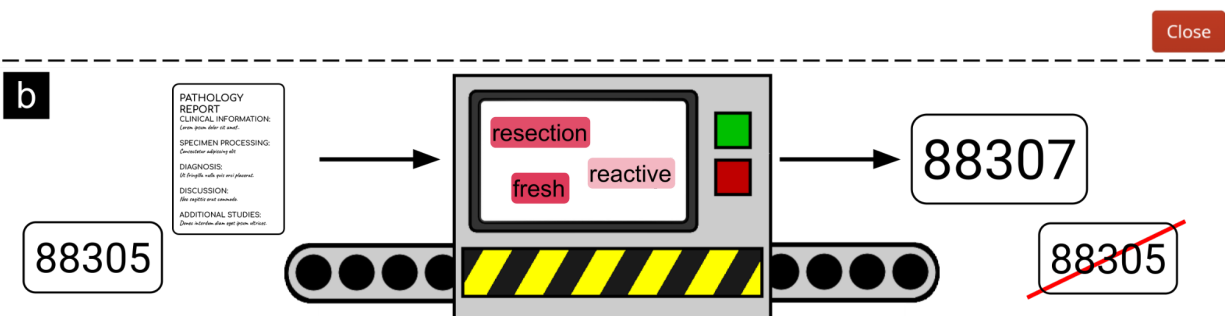
specimen submitted a distal colon b proximal descending colon ostomy clinical history trisomy history of av canal status post repair imperforate anus status post colostomy and psarp now ready for colo stomy closure clinical diagnosis same

SPECIMEN PROCESSING

a labeled/fixative distal colon **fresh** quantity/size single x x cm tissue description annular portion of colonic mucosa sections/processing a representative section is submitted r b labeled/fixative proximal descending colon ostomy **fresh** quantity/size single x x cm tissue description annular portion of colonic mucosa sections/processing a representative section is submitted r sns

DIAGNOSIS

a distal colon partial **resection** normally ganglionated colon with mild chronic and focal **active** neutrophilic mucosal inflammation and **reactive** surface epithelial changes b proximal descending colon **resection** at ostomy closure partial normally ganglionated colon with occasional hypertrophic and aganglionic submucosal nerves mild chronic and focal **active** neutrophilic and eosinophilic mucosal inflammation and **reactive** surface epithelial changes



243
244 **Figure 3.** Example interpretation plot for primary CPT code 88307. This report was originally assigned
245 88305, but our model assigned 88307. The interpretation plot highlighted the words “fresh” and
246 “resection” as important.

247
248 In this example, the case was a distal colon partial resection which was deemed to have been inflamed
249 with reactive epithelial changes. The code originally assigned was CPT code 88305, but the model had
250 revised this to CPT 88307. The example shown in **Figure 3** contains the interpretation plots for
251 assignment of the predicted CPT code 88307. Interpretation plots identified “colon” and “resection” as
252 strong indicators that the report contained the CPT code 88307. While colon biopsies are typically 88305,
253 resections should be 88307 or greater, given that they involve much more pre and post analytic work²⁵.
254 These interpretation plots inform the user of the dashboard on the model’s reasons for assigning 88307,
255 helping them with their assignment. It is also worth pointing out the negative contribution of the word
256 “mucosa.” The algorithm picked up the fact that mucosa examination is not typically associated with
257 codes higher than 88305, and it provides that information to the user as well. In this case, it was not
258 significant enough to change the overall prediction, so the model predicted 88307 despite it.

259
260 If the user wanted more information about why the model chose not to assign 88305, then they could
261 examine the interpretation plots for that code as well (**supplementary material contains the**

262 **interpretation plots for the class indicating the presence of the code 88305; Supplementary Figures 1-**
263 **4).** This interpretation plot shows the new coder exactly why our model disagrees with the original
264 assignment (particularly important for re-assessing underbilled codes, see **Supplementary Figure 5**). In
265 this particular example (**Figure 3**), the model took the presence of “representative” and “resection” to
266 mean that this code was unlikely to represent 88305. The algorithm picked up on the fact that the
267 majority of resections are coded as 88307 or 88309, and it predicted accordingly. The user can use these
268 indicators to rapidly reassess the case, as salient portions of text are highlighted for them. The savings
269 for the hospital associated with this single example would have been in excess of two hundred dollars
270 (**see Methods, “Estimating the Potential Impact of Underbilling”**).

271
272 *Additional examples.* In the supplementary material, we have included two additional examples of true
273 positives and two additional examples of false positives / candidate underbilled reports (**Supplementary**
274 **Figures 6-9**). For instance, **Supplementary Figure 6** displays a report where our model’s assignment and
275 the original coder both assigned the CPT code 88305 and using the SHAP plot correctly identifies
276 esophagus as one of the most informative words (esophagus biopsy falls under criteria for CPT 88305).
277 Similarly, **Supplementary Figure 7** displays a report which demonstrates the true positive assignment of
278 CPT 88307. While examination of other placenta specimen is considered to be a CPT 88305, third
279 trimester pregnancies should be assigned a CPT 88307. This is corroborated using the SHAP
280 functionality in our web application. In **Supplementary Figure 8**, the machine learning model reassigned a
281 CPT 88302 to a CPT 88304. This candidate underbilled case featured an appendectomy as elucidated
282 using SHAP. CPT 88304 is the correct determination because CPT code 88304 does account for several
283 common specimens in which malignancy may rarely present, such as the gallbladder, the appendix and
284 the tonsil. Finally, **Supplementary Figure 9** features inspection of a mass excised from the left palmer
285 hand. A soft tissue mass excised typically warrants assignment of an 88307, save for Lipoma, which is an
286 88304²⁴.

287
288 **Extrapolating Maximal Underbilling Savings to Future Cohort Assuming Reports were Truly Underbilled**
289 Calculating the price differences for candidate underbilled reports in the study cohort yielded maximal
290 costs savings of \$97,084.13 under the assumption that underbill candidates were truly underbilled.
291 Extrapolating these findings to all reports from all pathology subspecialties identified between January
292 2011 to December 2021, the maximum potential savings could total up to \$110,337.38 per year. Of
293 course, while this upper bar sets an estimate of what could be saved if all candidate underbilled reports
294 were in fact underbilled, we expect only a fraction of these reports to be underbilled which will be the
295 subject of future study.

296 297 **Discussion**

298
299 The semi-autonomous assignment of CPT codes will play an important role in billing and revenue for
300 pathology departments which aim to augment their clinical reporting infrastructure. To this end, we have
301 developed a dashboard which can help facilitate the evaluation of these semi-autonomous billing
302 augmentation tools by providing helpful display graphics that inform reassessment. Utilizing interactive
303 digital display dashboards can help further usher in the adoption and embedding of such applications in
304 the electronic health record system pending further validation and user-centered design.

305
306 The dashboard featured in this work will allow our team to efficiently review thousands of past cases to
307 provide information on potential revenue lost by the hospital due to underbilling of CPT codes, but it could
308 be used for any task where machine learning is used to interpret free text reports. Overall, model

309 predictions and interpretations from select example cases correlated well with known CPT coding
310 guidelines. This confirmed that the interpretation plots featured in the dashboard could be meaningfully
311 utilized to determine whether the code predicted by the model should replace the original assignment.
312 Every accurate adjustment equates to revenue that our model could have helped save.

313

314 **Limitations**

315 There are a few limitations to our study. The model presented in this paper was developed using data at a
316 single hospital and may not be representative of pathology reporting at large. Individuals may write with
317 specific lexical patterns that evolve because of training, peer influence and experience which can
318 complicate analyses and warrant adjustment. XGBoost models stochastically sample predictors, which
319 may not cover the entire vocabulary across the corpus, causing failures in select edge cases. We did not
320 prospectively evaluate the usability of this tool (e.g., execution speed, ergonomics, etc) and plan to do so
321 in a subsequent study. We also plan to evaluate the potential fiscal impact of underbilling on an expanded
322 cohort. There is also considerable clinical information outside of the Pathology report that may prove
323 relevant for case assessment that is unavailable for these approaches. We acknowledge that additional
324 algorithmic finetuning and annotation may improve modeling results. For instance, we assumed that CPT
325 codes used to train the model were taken as ground truth. As it is unlikely our cohort was significantly
326 biased by a few sporadic errors, it should be noted that original codes were likely assigned by a single
327 individual and not re-verified for the purpose of training. We expect the introduced application to rapidly
328 facilitate re-review of mismatches for model training in future iterations of this approach. Furthermore,
329 there may be additional model explanation tools more appropriate than SHAP for medical decision-
330 making tools as SHAP has been critiqued for limited causal interpretability and ethical challenges
331 associated with interpretation in the real-world context (e.g., fairness and stakeholder viewpoints)²⁶⁻³⁰.
332 This approach may benefit from methods to incorporate expert interpretation. Utilization of these model
333 explanation techniques also requires appropriate education to ensure these tools are appropriately
334 utilized.

335

336 **Future Direction**

337 Our next immediate goal is to utilize this application to quantify lost hospital revenue based on model
338 prediction of codes that are higher in complexity than the original assignment (i.e., underbilling). We will
339 make this estimate by comparing the originally assigned codes to the codes that would be assigned with
340 the help of the algorithm when they are flagged as potential false positives in order to quantify how many
341 reports were underbilled. From there we could extrapolate the fiscal burden that missed codes are
342 placing on the hospital³¹⁻³⁵.

343

344 Additionally, if we find that these models highlight missed codes with high accuracy, that will further
345 enforce the utility of machine learning models for pathology code prediction. As one corrected code
346 could mean the difference of several hundred dollars (e.g., 88309 is priced \$450 on average as compared
347 to 88302, which is priced \$30 on average), it would not take many to make this correction procedure
348 economically viable. For instance, we found in this study that if underbilling candidates were truly
349 underbilled, this could potentially equate to over \$100,000 per year in savings for our institution as
350 extrapolated across an 11-year span, including addendums. This number will be amended/reduced
351 pending review of mismatched cases from local coders to understand the true specificity of this
352 approach though is out of scope for the current work. Should findings prove significant and affirm a
353 consequential estimate, we would therefore want to identify mechanisms to integrate into the coders'
354 workflow as soon as possible¹⁶. Nationally, the average maximum salary for medical coders is \$35 per
355 hour³⁶⁻³⁹. Given that around 1.5% of reports were deemed to be candidate underbilling reports,
356 reexamination of these reports is a worthwhile endeavor versus the amount of time committed (i.e., only

357 one out of every hundred reports would require reexamination). Regardless of the immediate research
358 findings, we will continue to utilize the dashboard to evaluate future, more accurate models (e.g., ones
359 that could potentially improve the coding efficiency of the billing).

360
361 We also acknowledge that such a tool could also be configured to evaluate instances of overbilling (e.g.,
362 potential fraud), which is especially relevant if there was a systematic finding in a specific specimen type
363 or subspecialty. There are many potential impacts associated with developing algorithms to detect
364 overbilling– for instance, whether detection of such cases leads to loss of revenue or requires additional
365 incentives or additional educational trainings for such a search⁴⁰. While this is not the specific focus of
366 this research project, utilization of this application can identify general cases of misbilling to improve the
367 ability to “audit” emerging reporting systems and improve their accuracy.

368
369 We plan to submit a full research article upon gathering our findings from the analysis of our first models.
370 However, regardless of the algorithm chosen, the dashboard largely serves the same function (i.e.,
371 reexamine codes). The basic functions of the dashboard are interoperable to other research domains,
372 and significant changes to improve the functionality are likely not required (while usability improvements
373 will be pursued). The strength of the dashboard is in its ease of use and relative simplicity, allowing
374 researchers to focus more of their time on model development and improvement.

375
376

377 **Conclusion**

378 We have developed an interactive dashboard that can be used to interrogate the assignment of CPT
379 codes via a dashboard agnostic NLP model. The dashboard is additionally model-agnostic, allowing for
380 incorporation of future state-of-the-art approaches. We plan to use the dashboard described in this report
381 to determine the viability of detecting underbilling with our current classification models.

382
383

384 **References**

385

386 1. Cossio M. Digital pathology incorporation in CPT codes: a gateway for expansion into clinical
387 practice? Published online August 16, 2022. doi:10.33774/coe-2022-87dcw

388 2. Wilcock AD, Schwamm LH, Zubizarreta JR, et al. Legislation Increased Medicare Telestroke Billing,
389 But Underbilling And Erroneous Billing Remain Common. *Health Aff (Millwood)*. 2022;41(3):350-359.
390 doi:10.1377/hlthaff.2021.00791

391 3. A. Scherer M. Osteoporotic Fractures Often Undercoded or Miscoded. *Dtsch Arztebl Int*.
392 2022;119(13):232. doi:10.3238/arztebl.m2022.0081

393 4. Joo H, Burns M, Lakshmanan SSK, Hu Y, Vydiswaran VGV. Neural Machine Translation–Based
394 Automated Current Procedural Terminology Classification System Using Procedure Text:
395 Development and Validation Study. *JMIR Form Res*. 2021;5(5):e22461. doi:10.2196/22461

396 5. Shayan O, Tassavor M, Han J, Alexandra G, Cula S, Ungar J. Impact of Coding Curriculum on
397 Dermatology Resident Billing. *Cureus*. 2022;14(4). doi:10.7759/cureus.24148

398 6. López-Úbeda P, Martín-Noguerol T, Aneiros-Fernández J, Luna A. Natural Language Processing in
399 Pathology: Current Trends and Future Insights. *Am J Pathol*. 2022;192(11):1486-1495.
400 doi:10.1016/j.ajpath.2022.07.012

- 401 7. Wright JR, Chan S, Morgen EK, et al. Workload Measurement in Subspecialty Placental Pathology in
402 Canada. *Pediatr Dev Pathol*. Published online September 8, 2022:10935266221118150.
403 doi:10.1177/10935266221118150
- 404 8. Morris WC, Heinze DT, Warner Jr HR, et al. Assessing the accuracy of an automated coding system in
405 emergency medicine. *Proc AMIA Symp*. Published online 2000:595-599.
- 406 9. Chowdhary KR. *Fundamentals of Artificial Intelligence*.; 2020. doi:10.1007/978-81-322-3972-7_19
- 407 10. Hammami L, Paglialonga A, Pruneri G, et al. Automated classification of cancer morphology from
408 Italian pathology reports using Natural Language Processing techniques: A rule-based approach. *J*
409 *Biomed Inform*. 2021;116:103712. doi:10.1016/j.jbi.2021.103712
- 410 11. Zaccaria GM, Colella V, Colucci S, et al. Electronic case report forms generation from pathology
411 reports by ARGO, automatic record generator for onco-hematology. *Sci Rep*. 2021;11(1):23823.
412 doi:10.1038/s41598-021-03204-z
- 413 12. Borjali A, Magneli M, Shin D, Malchau H, Muratoglu OK, Varadarajan KM. Natural Language
414 Processing with Deep Learning for Medical Adverse Event Detection from Free-Text Medical
415 Narratives: A Case Study of Detecting Total Hip Replacement Dislocation. *Comput Biol Med*.
416 2021;129:104140. doi:10.1016/j.combiomed.2020.104140
- 417 13. Glaser AP, Jordan BJ, Cohen J, Desai A, Silberman P, Meeks JJ. Automated Extraction of Grade,
418 Stage, and Quality Information From Transurethral Resection of Bladder Tumor Pathology Reports
419 Using Natural Language Processing. *JCO Clin Cancer Inform*. 2018;(2):1-8. doi:10.1200/CCI.17.00128
- 420 14. Santos T, Tariq A, Das S, et al. PathologyBERT – Pre-trained Vs. A New Transformer Language Model
421 for Pathology Domain. Published online May 13, 2022. doi:10.48550/arXiv.2205.06885
- 422 15. Marchesin S, Giachelle F, Marini N, et al. Empowering digital pathology applications through
423 explainable knowledge extraction tools. *J Pathol Inform*. 2022;13:100139.
424 doi:10.1016/j.jpi.2022.100139
- 425 16. Ye JJ. Construction and Utilization of a Neural Network Model to Predict Current Procedural
426 Terminology Codes from Pathology Report Texts. *J Pathol Inform*. 2019;10. doi:10.4103/jpi.jpi_3_19
- 427 17. Arvisais-Anhalt S, Lehmann CU, Bishop JA, et al. Searching Full-Text Anatomic Pathology Reports
428 Using Business Intelligence Software. *J Pathol Inform*. 2022;13:100014.
429 doi:10.1016/j.jpi.2022.100014
- 430 18. Levy J, Vattikonda N, Haudenschild C, Christensen B, Vaickus L. Comparison of machine-learning
431 algorithms for the prediction of current procedural terminology (CPT) codes from pathology reports.
432 *J Pathol Inform*. 2022;13(1):3. doi:10.4103/jpi.jpi_52_21
- 433 19. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22Nd ACM*
434 *SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM; 2016:785-
435 794. doi:10.1145/2939672.2939785
- 436 20. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for
437 Language Understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of*
438 *the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and*
439 *Short Papers)*. Association for Computational Linguistics; 2019:4171-4186. doi:10.18653/v1/N19-
440 1423

- 441 21. Modern Analytic Apps for the Enterprise. Plotly. Accessed April 11, 2019. <https://plot.ly>
- 442 22. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with
443 explainable AI for trees. *Nat Mach Intell.* 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9
- 444 23. Wynn BO, Burgette LF, Mulcahy AW, et al. Development of a Model for the Validation of Work Relative
445 Value Units for the Medicare Physician Fee Schedule. *Rand Health Q.* 2015;5(1):5.
- 446 24. Dimenstein IB. CPT Coding in Dermatopathology. *Lab Med.* 2009;40(3):151-156.
447 doi:10.1309/LMPB2BIWVHB9RPNK
- 448 25. Dimenstein IB. Principles and Controversies in CPT Coding in Surgical Pathology. *Lab Med.*
449 2011;42(4):242-249. doi:10.1309/LM6TN3ECGMRQUGRV
- 450 26. Barocas S, Selbst AD, Raghavan M. The hidden assumptions behind counterfactual explanations and
451 principal reasons. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and*
452 *Transparency.* ; 2020:80-89.
- 453 27. Yildirim N, Zimmerman J, Preum S. Technical Feasibility, Financial Viability, and Clinician Acceptance:
454 On the Many Challenges to AI in Clinical Practice. In: *HUMAN@AAAI Fall Symposium.* ; 2021.
- 455 28. Hancox-Li L, Kumar IE. Epistemic values in feature importance methods: Lessons from feminist
456 epistemology. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and*
457 *Transparency.* ; 2021:817-826.
- 458 29. McCradden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions
459 in health care machine learning. *Lancet Digit Health.* 2020;2(5):e221-e223.
- 460 30. Jesus S, Belém C, Balayan V, et al. How can I choose an explainer? An Application-grounded
461 Evaluation of Post-hoc Explanations. In: *Proceedings of the 2021 ACM Conference on Fairness,*
462 *Accountability, and Transparency.* FAccT '21. Association for Computing Machinery; 2021:805-815.
463 doi:10.1145/3442188.3445941
- 464 31. Arispe Angulo KR, Farooq A, Samra H, Weldemichael W, Anderson J, Jorns JM. Data-Driven
465 Development of an Institutional "Gross-Only" Policy for the Examination of Select Surgical Pathology
466 Specimens. *Am J Clin Pathol.* 2020;154(4):486-493.
- 467 32. Deeken-Draisey A, Ritchie A, Yang GY, et al. Current Procedural Terminology Coding for Surgical
468 Pathology: A Review and One Academic Center's Experience With Pathologist-Verified Coding. *Arch*
469 *Pathol Lab Med.* 2018;142(12):1524-1532. doi:10.5858/arpa.2017-0190-RA
- 470 33. Harris CK, Kwong JY, Cohen MH, et al. Changes in surgical pathology case volume and amendment
471 rates during the COVID-19 pandemic. *Am J Clin Pathol.* Published online 2022.
- 472 34. Harris CK, Pyden A, Onken AM, et al. Prioritizing Patient Safety and Minimizing Waste: Institutional
473 Review of Cases and a Proposed Process for Designing a Surgical Pathology Gross-Only Examination
474 Policy. *Am J Clin Pathol.* 2022;158(5):598-603.
- 475 35. Johnson SM, Samulski TD, O'Connor SM, et al. Clinical and Financial Implications of Second-Opinion
476 Surgical Pathology Review: Experience of a Comprehensive Cancer Center With Regional Affiliates.
477 *Am J Clin Pathol.* 2021;156(4):559-568.
- 478 36. DAVENPORT TH, MILLER SM. The future of work now: Medical coding with AI. *Forbes.* Published
479 online January 3, 2020:1-3.

- 480 37. Buck C.J. *Step-by-Step Medical Coding, 2017 Edition - E-Book*. Elsevier Health Sciences; 2016.
- 481 38. Gray K. Professional and Industry Certifications for the Health Information Workforce. In: *The Health*
482 *Information Workforce*. Springer; 2021:97-114.
- 483 39. Paolone D. Coding and Billing. In: *Navigating Organized Urology*. Springer; 2022:123-132.
- 484 40. Hedgepeth J, Randall C, Massey T, Bookhout C. Pathologist-Verified Billing: Correction Rates of
485 Miscoded Frozen Section Cases. *Arch Pathol Lab Med*. Published online October 26, 2022.
486 doi:10.5858/arpa.2022-0158-OA
- 487

488 **Supplementary:**

489

490 **Supplementary Methods**

491

492 *Toggling different SHAP interpretations for CPT code models.* This dashboard works primarily with multi-
493 class or multi-target classification of clinical text. Users can toggle between displays that predict the
494 output of *ancillary codes* (i.e., separate models predicting binary endpoints to identify 38 CPT codes, each
495 given a probability 0-1, multiple codes predicted simultaneously) and *primary codes* (i.e., one model to
496 predict presence of one CPT code from a set made up of codes 88302, 88304, 88305, 88307, and 88309).
497 For the *ancillary code* model, the application can only display the SHAP values for one CPT code at a time,
498 so we included a selectable plot of each model's predictions across the codes in order to quickly switch
499 between the CPT code-specific models. The *primary code* model that we developed also has unique
500 SHAP values for each potential code, so the scatter plot is also used to switch between codes to display
501 correspondent text identified by the model.

502

503 *Filtering documents to focus on false positive reports as candidate underbilling instances.* In prediction
504 mode, certain filters can then be applied to select the type of report that the dashboard will present. It is
505 through these filters that we can restrict the dashboard to display false positives only; this is done by
506 selecting the "False predictions" and "Positive only" filters (**Figure 2C**). This filter will show the reports
507 that the model predicts to feature a certain code, but that the original coder did not assign. The coder can
508 then select which codes it identifies from a dropdown menu. Pressing "Next report" updates a CSV file
509 containing the coder's predictions. From there we can compare their predictions to both the original
510 predictions and our model predictions.

511

512 *Focused search of keywords.* Even when not in the hands of coders, the dashboard has a search feature
513 that makes it advantageous to work with. For example, if we believed that our model was particularly
514 good (or perhaps particularly weak) at assigning codes pertaining to examination under a microscope to
515 reports containing the word cytogenetics, then we easily analyze those further. To do this we would
516 select the search by "Report" and search by "Code description" options, and search for "cytogenetics
517 microscope" (**see Figure 2**). Clicking on any of the reports would then display that report on the main
518 dash.

519

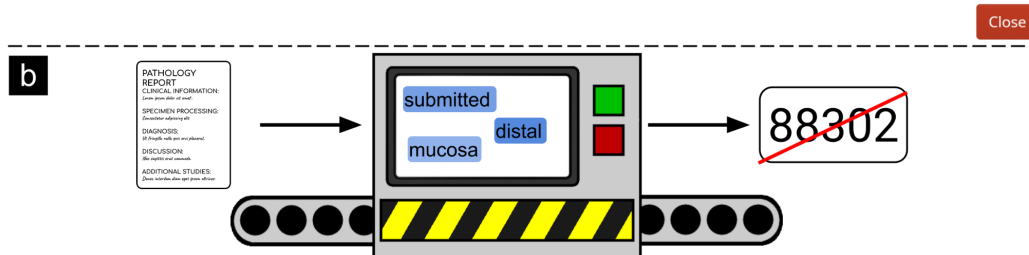
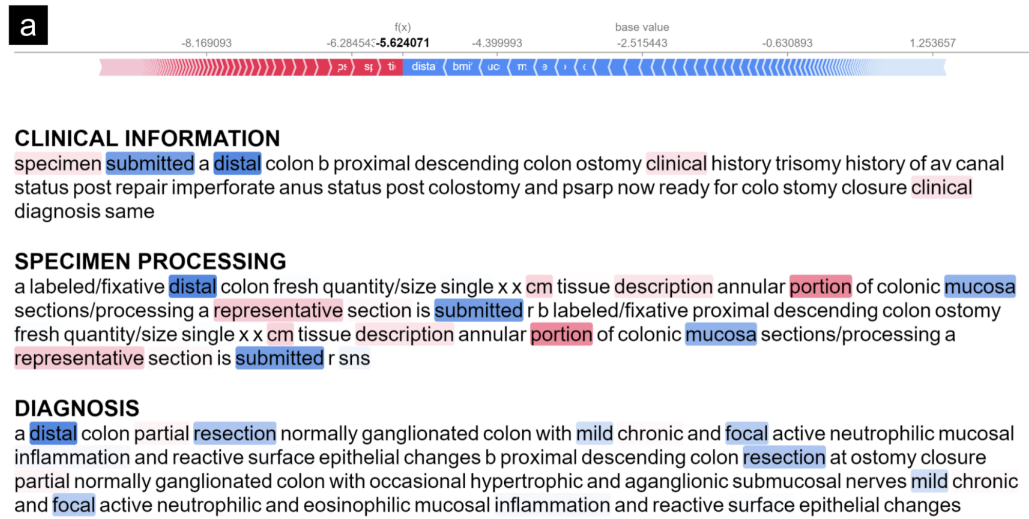
520

521

522 **Supplementary Results**

523

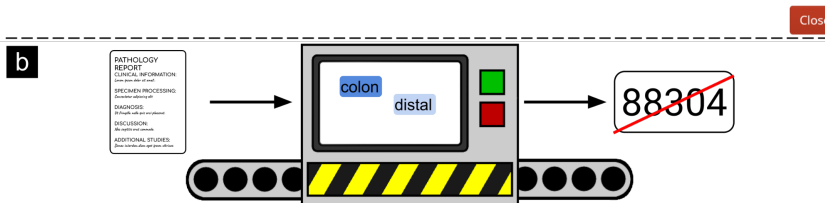
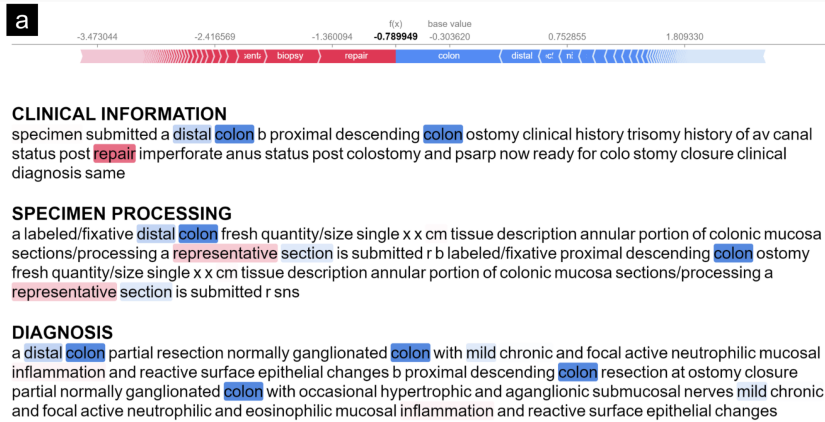
SHAP analysis for the prediction CPT code 88302 on path report #1438



524
525
526
527

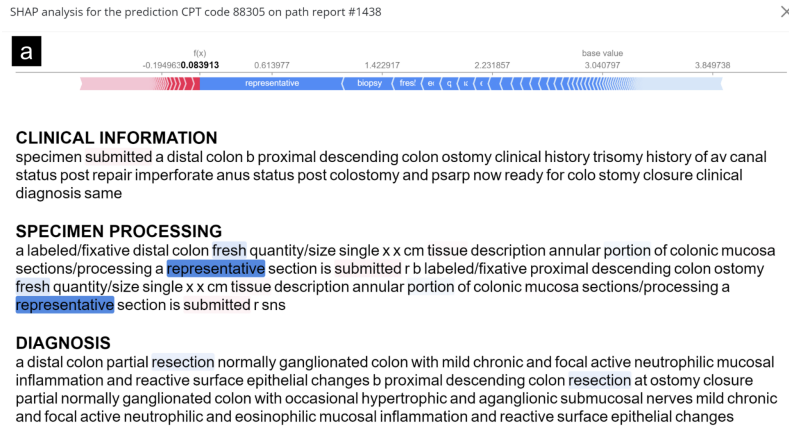
Supplementary Figure 1: Evaluation of same case from Figure 3, depicting words related to CPT 88302 using SHAP on #1438

SHAP analysis for the prediction CPT code 88304 on path report #1438

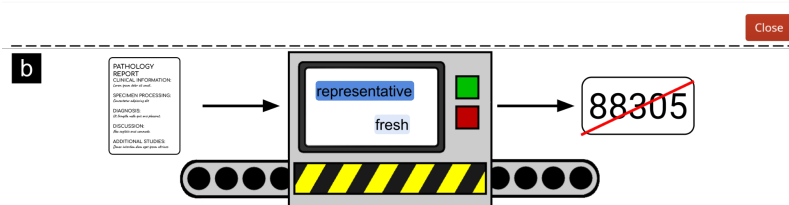


528

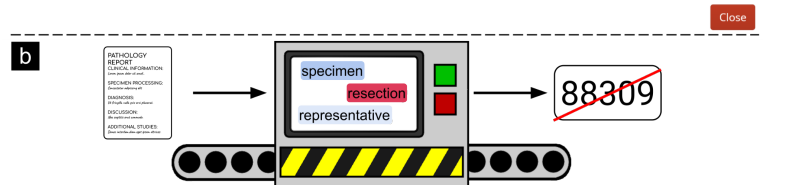
529 **Supplementary Figure 2:** Evaluation of same case from **Figure 3**, depicting words related to CPT 88304
 530 using SHAP on #1438
 531



532 **Supplementary Figure 3:** Evaluation of same case from **Figure 3**, depicting words related to CPT 88305
 533 using SHAP on #1438
 534
 535



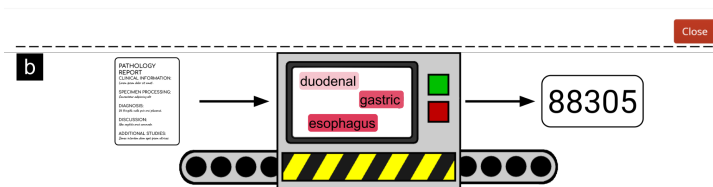
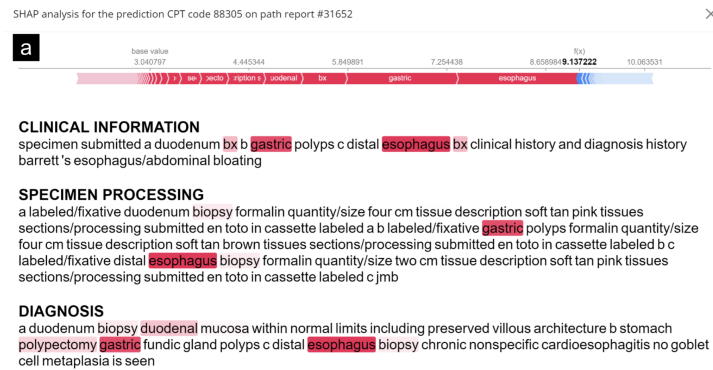
536 **Supplementary Figure 4:** Evaluation of same case from **Figure 3**, depicting words related to CPT 88309
 537 using SHAP on #1438
 538
 539



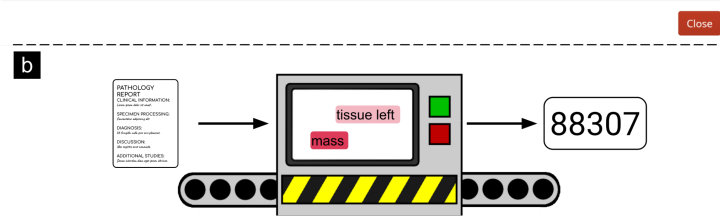
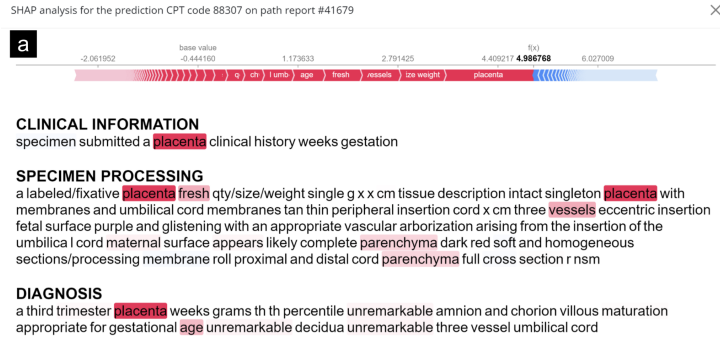
Predicted

	88302	88304	88305	88307	88309	
Billed	88302	334	33	44	1	0
88304	12	3418	515	82	2	
88305	8	376	50921	221	5	
88307	0	116	366	3178	31	
88309	0	5	38	97	120	

540
541 **Supplementary Figure 5:** Breakdown of all codes assigned a single primary code in our corpus. Candidate
542 underbilled codes are highlighted in green
543
544
545
546
547

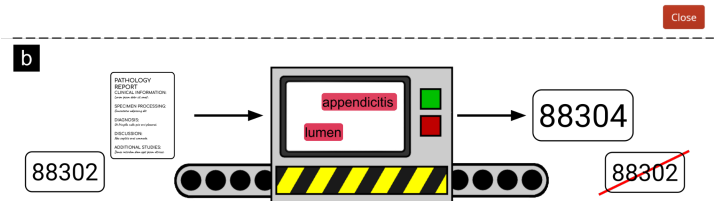
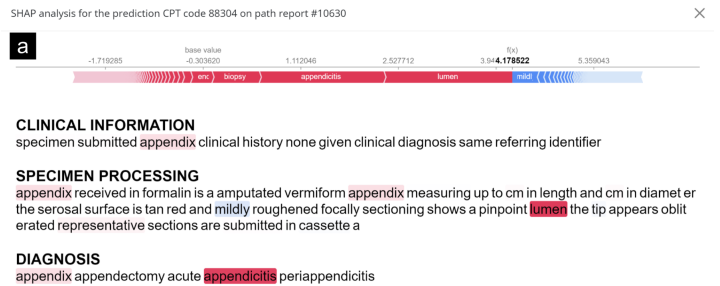


548
549 **Supplementary Figure 6:** Example interpretation plot for primary CPT code 88305. This report was
550 correctly assigned CPT 88305 by both the coder and model.
551



552
553
554
555

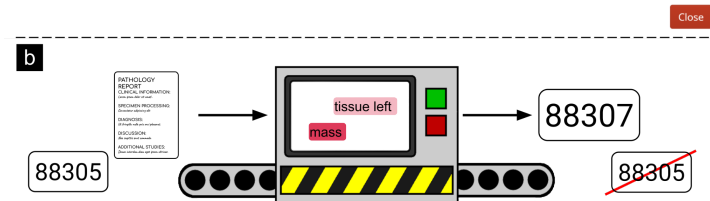
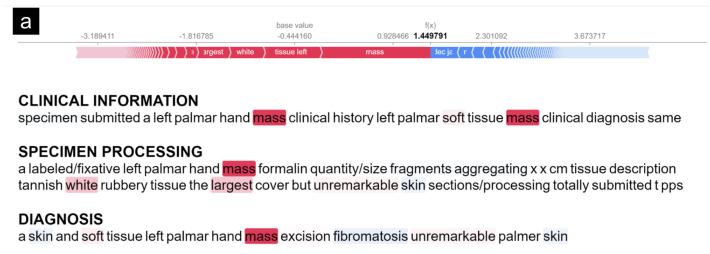
Supplementary Figure 7: Example interpretation plot for primary CPT code 88307. This report was correctly assigned CPT 88307 by both the coder and model.



556
557
558
559

Supplementary Figure 8: Example interpretation plot for primary CPT code 88304. This report was assigned CPT 88302 by the coder, while the model predicted CPT 88304.

SHAP analysis for the prediction CPT code 88307 on path report #46149



560
561
562

Supplementary Figure 9: Example interpretation plot for primary CPT code 88307. This report was assigned CPT 88305 by the coder, while the model predicted CPT 88307.