

NLP COVID-19, Schöning (2021)

1 **Automatic identification of risk factors for SARS-CoV-2**
2 **positivity and severe clinical outcomes of COVID-19 using Data**
3 **Mining and Natural Language Processing**

4

5 **Authors:** Verena Schöning¹, Evangelia Liakoni¹, Jürgen Drewe², Felix
6 Hammann^{1,*}

7

8 (1) Clinical Pharmacology and Toxicology, Department of General Internal
9 Medicine, Inselspital, Bern University Hospital, University of Bern, Bern,
10 Switzerland

11 (2) Department of Clinical Pharmacology, University Hospital Basel, Basel,
12 Switzerland

13

14 * Corresponding Author:

15 Felix Hammann, MD, PhD

16 Clinical Pharmacology and Toxicology, Department of General Internal
17 Medicine, Inselspital, Bern University Hospital, Freiburgstrasse, 3010 Bern,
18 Switzerland

19 Felix.Hammann@insel.ch

20

21 **Keywords**

22 SARS-CoV-2, COVID-19, risk factors, electronic health records, natural
23 language processing

24

25

26 **ABSTRACT**

27 **Objectives:** Several risk factors have been identified for severe clinical
28 outcomes of COVID-19 caused by SARS-CoV-2. Some can be found in
29 structured data of patients' Electronic Health Records. Others are included as
30 unstructured free-text, and thus cannot be easily detected automatically. We
31 propose an automated real-time detection of risk factors using a combination
32 of data mining and Natural Language Processing (NLP).

33 **Material and methods:** Patients were categorized as negative or positive for
34 SARS-CoV-2, and according to disease severity (severe or non-severe COVID-
35 19). Comorbidities were identified in the unstructured free-text using NLP.
36 Further risk factors were taken from the structured data.

37 **Results:** 6250 patients were analysed (5664 negative and 586 positive; 461
38 non-severe and 125 severe). Using NLP, comorbidities, i.e. cardiovascular and
39 pulmonary conditions, diabetes, dementia and cancer, were automatically
40 detected (error rate $\leq 2\%$). Old age, male sex, higher BMI, arterial hypertension,
41 chronic heart failure, coronary heart disease, COPD, diabetes, insulin only
42 treatment of diabetic patients, reduced kidney and liver function were risk
43 factors for severe COVID-19. Interestingly, the proportion of diabetic patients
44 using metformin but not insulin was significantly higher in the non-severe
45 COVID-19 cohort ($p < 0.05$).

46 **Discussion and conclusion:** Our findings were in line with previously reported
47 risk factors for severe COVID-19. NLP in combination with other data mining
48 approaches appears to be a suitable tool for the automated real-time detection
49 of risk factors, which can be a time saving support for risk assessment and

NLP COVID-19, Schöning (2021)

50 triage, especially in patients with long medical histories and multiple
51 comorbidities.

NLP COVID-19, Schöning (2021)

52 **1 BACKGROUND**

53 Coronavirus disease 19 (COVID-19) is caused by the severe acute respiratory
54 syndrome coronavirus 2 (SARS-CoV-2). It was first identified in Wuhan, China,
55 in December 2019,(1) and spread globally with rising numbers of cases and
56 deaths.(2)

57 Several risk factors have been identified for severe COVID-19 disease and
58 need to be taken into consideration when assessing patients in the early stages
59 since they can affect the clinical outcome and are therefore important for patient
60 triage and management decisions. Some of the reported risk factors, such as
61 age,(3, 4) sex,(5, 6) or obesity,(7-10) are included in tabulated form in the
62 electronic health records (EHRs). However, information on other relevant
63 comorbidities such as diabetes,(11-13) cardiac(14-16) and pulmonary
64 diseases,(17, 18) cancer,(19) or dementia(20) is usually included as free-text
65 in the medical history of the EHR and can be time consuming to retrieve. This
66 holds in particular in patients with a long medical history and conditions that
67 might not be included separately in the list of diagnoses (e.g. pre-diabetes).
68 Encoding into ICD (International Classification of Diseases) identifiers – which
69 would be highly suitable for machine processing - is often performed only after
70 discharge or death of the patient and is thus not readily available for the risk
71 evaluation on admission, requires human intervention, and might also only
72 include the main diagnoses relevant to billing purposes.

73 Here we propose a real-time support for patient triage by automatically
74 identifying relevant risk factors for infection with SARS-CoV-2 and for the
75 development of severe COVID-19 (i.e. need for critical care, mechanical

NLP COVID-19, Schöning (2021)

76 ventilation support, or death of any cause). A combination of data mining and
77 natural language processing (NLP) was used to analyse different parts of the
78 EHRs. Structured (tabulated) administrative, demographic and laboratory data
79 as well as unstructured (free-text) medical history information were processed
80 to identify specific risk factors (age, sex, high BMI, diabetes, cardiovascular
81 disease, pulmonary disorders, dementia, cancer, and decreased kidney and
82 liver function) for SARS-CoV-2 infection or severe courses of COVID-19.(15-
83 17, 21, 22) Furthermore, for diabetic patients, the treatment regimen (untreated,
84 insulin, oral anti-diabetics (OAD) or combination of insulin and OAD) on
85 admission was taken into consideration to classify disease severity.

86 After development and internal validation, these methods were used in a
87 retrospective analysis of patients seen at the Insel Hospital Group (IHG) in
88 Switzerland, a tertiary hospital network and the biggest health care provider in
89 Switzerland with six locations and about 860'000 patients treated per year.
90 Patients who received testing for SARS-CoV-2 at any point during the 'first' and
91 'second' waves of the pandemic in 2020 were included.

92 **2 METHODS**

93 **2.1 Study population**

94 The study was carried out at the IHG. The protocol was approved by the
95 Cantonal Ethics Committee of Bern (Project-ID 2020-00973). We considered
96 all individuals tested for SARS-CoV-2 at the IHG between February 1st through
97 November 16th 2020– covering the 'first wave' and part of the 'second wave' of
98 COVID-19 in the country, and who did not reject the IHG general research

NLP COVID-19, Schöning (2021)

99 consent. For patients with no registered general research consent status, a
100 waiver of consent was granted by the ethics committee. Patients who objected
101 to the general research consent of the IHG were excluded from the study.
102 Participation in other trials (including COVID-19 related treatment studies) was
103 not an exclusion criterion and was not recorded separately. A reverse-
104 transcriptase polymerase chain reaction (RT-PCR) assay on nasopharyngeal
105 swabs was in use throughout the entire observational period as diagnostic test
106 for SARS-CoV-2 detection. Detailed information on the selection of the study
107 population is provided in **Figure 1**. All patients had been discharged or died by
108 the time of the analysis.

109 Patients were classified according to their test result and disease severity, with
110 the worst outcome at any point determining the class:

- 111 - *Negative*: Patients who always tested negative for SARS-CoV-2
- 112 - *Positive*: Patients who tested positive for SARS-CoV-2 at any point
 - 113 ○ *Non-severe*: Patients who tested positive for SARS-CoV-2, but
 - 114 were neither admitted to the intensive care unit (ICU) nor died of
 - 115 any cause during their hospital stay.
 - 116 ○ *Severe*: Composite outcome for patients who tested positive for
 - 117 SARS-CoV-2 and required ICU admission at any stage during the
 - 118 disease or died of any cause during their hospital stay.

119 Given that the study was retrospective and observational, sample sizes were
120 dictated by the dynamics of the pandemic in the greater Bern region.
121 Consequently, no formal power calculations were performed *a priori*.

NLP COVID-19, Schöning (2021)

122 The NLP validation dataset consisted of 138 patients, who were tested positive
123 for SARS-CoV-2 during the first wave in Switzerland (February - August 2020)
124 and whose medical history were available. The EHR was screened manually
125 by two of the authors for the following comorbidities of interest: arterial
126 hypertension, chronic heart failure, atrial fibrillation or flutter, coronary heart
127 disease, asthma, chronic obstructive pulmonary disease (COPD), diabetes
128 (including type I, type II and pre-diabetes), dementia (including mild cognitive
129 impairment), and cancer.(15-17, 21, 22)

130 **2.1.1 Data mining and Natural Language Processing**

131 Age and sex were taken directly from the structured parts of the EHR. Although
132 BMI was also included in the structured parts, some values were not correctly
133 calculated. Therefore, BMI were recalculated using weight and height, which
134 are also documented in the EHR, after removing highly unrealistic values (e.g.
135 weight of 80 kg for a <1 year old child or height of 16.3 cm for an adult), which
136 were probably due to typing errors. Laboratory values indicative for kidney
137 failure (estimated glomerular filtration rate (GFR) according to the Chronic
138 Kidney Disease Epidemiology Collaboration (CKD-EPI) equation) or hepatic
139 dysfunction (alanine aminotransferase (ALAT) and alkaline phosphatase (AP))
140 were also taken from the structured parts of the EHR. We used the lowest and
141 highest values measured, respectively, in the time frame between three days
142 before and one day after the Sars-CoV-2 test as described in a previous
143 study.(23) Kidney function was categorized into five stages: normal
144 (GFR \geq 90 mL/min), mild impairment (GFR 60-89 mL/min), moderate
145 impairment (GFR 30-59 mL/min), severe impairment (GFR 15-29 mL/min) and
146 kidney failure (GFR < 15 mL/min).(24) Liver function was considered as

NLP COVID-19, Schöning (2021)

147 impaired if ALAT > 90 U/L (~3x ULN (upper limit of normal)) and/or
148 AP > 200 U/L (~2x ULN).

149 The comorbidities of interest were identified in the medical history, which is the
150 unstructured, free-text part of the EHR. In a first step, we merged the
151 information of different time points into a single text file per patient and removed
152 duplicate sentences. We then read the text files into Python as data corpus,
153 where we performed the actual NLP (**Figure 2**). General data cleaning of these
154 files consisted of converting special characters (e.g. umlauts such as 'ä' to 'ae'),
155 setting the whole text to lower case, and removing dates. Then sentences were
156 tokenized, i.e. the text was separated into single (half-) sentences (tokens).
157 Tokens with less than two characters or information related to family history
158 were removed. As the medical history consisted mainly of catch phrases and
159 half-sentences, we decided against a removal of stop words. For each disease,
160 a specific list of key terms was generated, which consisted of common terms
161 (one or more words) and abbreviations used to describe the diseases in
162 German EHR systems. As there are many different terms for the diverse types
163 of cancer, another approach was chosen for this element. We first scanned all
164 EHRs for words containing cancer-specific terms such as 'cancer', 'carcinoma',
165 or 'neoplasia', and created a key terms list. This list was then appended with
166 other common terms and abbreviations for specific types of cancer, for instance
167 'ALL' for acute lymphoblastic leukaemia. For correction of spelling errors and
168 lemmatization, the key terms list was transformed into a dictionary, where each
169 word only appears once. Words within the medical history were replaced with
170 the corresponding word of the dictionary when the similarity threshold of the
171 Levenshtein distance (minimum number of characters required to transform

NLP COVID-19, Schöning (2021)

172 one word to another) was equal or bigger than 90%. Then, for each token within
173 the medical history, we checked for the presence of the defined key terms. If
174 the key term consisted of more than one word (e.g. 'arterial hypertension'), the
175 algorithm was set to allow up to three words between the single words of the
176 key term. This further reduced the need for removing stop words.

177 If the key term was present, we determined the context of the key term
178 (affirmation or negation) using a German implementation of *NegEx*.⁽²⁵⁾ The
179 output of the negation detection was screened, and the negation trigger list
180 appended accordingly where necessary (e.g. 'non-insulin'). In case of an equal
181 or bigger number of affirmed than negated statements, the patient was tagged
182 as 'positive' for the disease. The disease status of each patient as determined
183 by NLP was compared to the manually tagged label. For cases where the
184 automatically detected disease status varied from the manual one, the medical
185 history was screened again, and, if possible, the key term list was corrected
186 accordingly. The NLP detection was considered suitable if the error rate per
187 disease was $\leq 2\%$.

188 As the structured medication on admission table was not always filled in, we
189 decided to use a combined approach to identify anti-diabetic treatment: we
190 selected individuals which were identified as diabetes patients by NLP and
191 searched the EHR for key terms of treatment with insulin or OADs. We then
192 combined these NLP results with the medication at admission table. If either
193 the NLP or the medication at admission table showed use of insulin and/or
194 OAD, we tagged the patient as receiving that treatment. If the patient received
195 neither, we categorized the patient as 'untreated'. For the subgroup analysis of
196 the OAD only cohort, we were only able to use the recorded medication on

NLP COVID-19, Schöning (2021)

197 admission to identify metformin, the only biguanide marketed in Switzerland, as
198 medical histories and diagnosis lists often do not mention the specific OAD
199 used.

200 Age, sex, BMI, disease status, anti-diabetic treatment and kidney and liver
201 function of all patients with available information were then analysed and
202 stratified according to the cohorts described above.

203 **2.1.2 Software and statistical tests**

204 NLP was performed in Python (version 3.8.5). The text file corpus was
205 generated using the Natural Language Toolkit (*nltk*) package (version 3.5).(26)
206 Steps requiring regular expressions (e.g. text cleaning, sentence tokenization,
207 word replacement) were performed using the *re* package (version 2.2.1). For
208 spelling correction and lemmatization, we used the *FuzzyWuzzy* package
209 (version 0.18.0) and specific, manually generated keyword lists. Detection of
210 negation was performed using *NegEx*(27) in combination with its German
211 implementation.(25)

212 Data wrangling, analysis and visualization was performed in GNU R (version
213 4.0.2, R Foundation for Statistical Computing, <http://www.R-project.org>,
214 Vienna, Austria). Standard statistics, e.g. Shapiro-Wilk test, Chi square,
215 Fisher's exact test, Wilcoxon rank sum test, were conducted using the *stats*
216 package (version 4.0.2).

217 Comparisons were performed between a) patients tested negative and positive
218 for SARS-CoV-2 and b) patients with a non-severe and severe clinical
219 manifestation of COVID-19. Statistical significance levels were determined
220 using Wilcoxon rank sum test for non-normally distributed parameters, as

NLP COVID-19, Schöning (2021)

221 confirmed by Shapiro-Wilk test, and the Chi square test or the Fisher's exact
222 test (sample size ≤ 5) for categorical parameters. A p value of <0.05 was
223 considered statistically significant.

224 **3 RESULTS**

225 A total of 6'250 patients tested for SARS-CoV-2 at the IHG during the study
226 period were included and categorized as either SARS-CoV-2 negative
227 (n=5'664) or positive (n=586), and – among the latter- as having a non-severe
228 (n=461) or severe (n=125) clinical outcome. The comorbidities of interest, i.e.
229 arterial hypertension, chronic heart failure, atrial fibrillation or flutter, coronary
230 heart disease, asthma, COPD, diabetes, dementia and cancer, could be
231 automatically detected in the German EHR of the IHG using NLP with an error
232 rate of $\leq 2\%$ compared to the manually tagged EHR. The results of the
233 validation and the key terms used for each disease are shown in the
234 supplementary material (**Table S1** and **Table S2**).

235 Stratification of patients to the outcome showed that the proportions of patients
236 with old age, male sex, higher BMI, arterial hypertension, or diabetes were
237 significantly higher in the positive SARS-CoV-2 and in the severe COVID-19
238 cohorts (**Table 1**). Patients with atrial fibrillation or flutter were more likely to be
239 tested negative for SARS-CoV-2, but despite a tendency for more severe
240 clinical manifestations of COVID-19, these results were not significant. The
241 same proportion of patients with coronary heart disease was tested negative
242 and positive for SARS-CoV-2, but with a higher proportion of patients with a
243 severe course.

NLP COVID-19, Schöning (2021)

244 **Table 1: Distribution of risk factors as determined by automated analysis**
 245 **of the Electronic Health Records (EHRs)**

	<i>Test for SARS-CoV-2</i>			<i>COVID-19 clinical manifestation</i>		
	<i>negative</i> (N = 5'664)	<i>positive</i> (N = 586)	<i>p-value</i>	<i>non-severe</i> (N = 461)	<i>severe</i> (N = 125)	<i>p-value</i>
Age (years)						
<i>Median (IQR)</i>	63 (39, 77)	65 (48, 77)	0.013[§]	62 (44, 75)	69 (59, 81)	<0.002[§]
Sex						
<i>Female (%)</i>	2'650 (46.8%)	246 (42.0%)	0.029*	211 (45.8%)	35 (28.0%)	<0.002*
BMI						
<i>Median (IQR)</i>	24.6 (21.1, 28.7)	26.5 (23.5, 29.7)	<0.002[§]	25.9 (23.2, 29.4)	27.9 (24.8, 30.9)	0.003[§]
Diseases						
<i>aHT (%)</i>	2,270 (40.1%)	270 (46.1%)	0.006*	192 (41.7%)	78 (62.4%)	<0.002*
<i>Chronic heart failure (%)</i>	1,591 (28.1%)	140 (23.9%)	0.035*	97 (21.0%)	43 (34.4%)	0.003*
<i>Atrial fibrillation or flutter (%)</i>	1,025 (18.1%)	84 (14.3%)	0.027*	59 (12.8%)	25 (20.0%)	0.058*
<i>Coronary heart disease (%)</i>	1,052 (18.6%)	92 (15.7%)	0.098*	56 (12.2%)	36 (28.8%)	<0.002*
<i>Asthma (%)</i>	350 (6.2%)	52 (8.9%)	0.015*	45 (9.8%)	7 (5.6%)	0.203*
<i>COPD (%)</i>	612 (10.8%)	45 (7.7%)	0.023*	28 (6.1%)	17 (13.6%)	0.009*
<i>Diabetes (%)</i>	1,129 (19.9%)	155 (26.5%)	<0.002*	111 (24.1%)	44 (35.2%)	0.017*
<i>Dementia (%)</i>	516 (9.1%)	55 (9.4%)	0.885*	40 (8.7%)	15 (12.0%)	0.339*
<i>Cancer (%)</i>	1,407 (24.8%)	96 (16.4%)	<0.002*	73 (15.8%)	23 (18.4%)	0.582*
Kidney function						
<i>Normal, GFR >90 mL/min (%)</i>	2,131 (36.16%)	196 (32.13%)	0.054	163 (36.22%)	33 (20.62%)	< 0.002
<i>Mild impairment, GFR 60-89 mL/min (%)</i>	1,871 (31.74%)	239 (39.18%)	<0.002*	187 (41.56%)	52 (32.50%)	0.055
<i>Moderate impairment, GFR 30-59 mL/min (%)</i>	1,299 (22.04%)	116 (19.02%)	0.095	77 (17.11%)	39 (24.38%)	0.058

NLP COVID-19, Schöning (2021)

	<i>Test for SARS-CoV-2</i>			<i>COVID-19 clinical manifestation</i>		
	<i>negative</i> (N = 5'664)	<i>positive</i> (N = 586)	<i>p-value</i>	<i>non-severe</i> (N = 461)	<i>severe</i> (N = 125)	<i>p-value</i>
Severe impairment, GFR 15-29 mL/min (%)	385 (6.53%)	37 (6.07%)	0.720	13 (2.89%)	24 (15.00%)	<0.002*
Kidney failure, GFR <15 mL/min (%)	208 (3.53%)	22 (3.61%)	1	10 (2.22%)	12 (7.50%)	0.005*
Liver function Elevated liver enzymes (%)	837 (16.89%)	74 (13.99%)	0.101	30 (9.62%)	26 (26.26%)	<0.002*
Diabetes management						
<i>Untreated (%)</i>	336 (29.8%)	38 (24.5%)	0.210*	27 (24.3%)	11 (25.0%)	1*
<i>Insulin and OADs (%)</i>	258 (22.9%)	43 (27.7%)	0.213*	30 (27.0%)	13 (29.6%)	0.907*
<i>Insulin only (%)</i>	231 (20.5%)	23 (14.8%)	0.124*	10 (9.0%)	13 (29.6%)	0.003*
<i>OAD only (%)</i>	304 (26.9%)	51 (32.9%)	0.143*	44 (39.6%)	7 (15.9%)	0.008*
<i>Incl. metformin (%)</i>	-	-	-	21 (18.9%)	2 (4.6%)	0.024**
<i>Excl. metformin (%)</i>	-	-	-	23 (20.7%)	5 (11.46%)	0.247**

246 Bold numbers indicate significant differences (p <0.05) between compared groups. aHT: arterial
 247 hypertension, BMI: body mass index, COPD: chronic obstructive pulmonary disease, GFR:
 248 glomerular filtration rate, IQR: interquartile range, OAD: oral anti-diabetics, \$ Wilcoxon rank
 249 sum test, * Chi Square test **Fisher test

250

251 Patients suffering from chronic heart failure or COPD were less likely to be
 252 tested positive for SARS-CoV-2, but more likely to develop a severe COVID-19
 253 course. In contract, significantly more asthmatic patients were tested positive
 254 for SARS-CoV-2 than negative. However, with regards to clinical manifestation,
 255 no difference could be observed between the two groups. Cancer patients were
 256 less likely to be tested positive for SARS-CoV-2, but no significant difference in
 257 severity was observed.

NLP COVID-19, Schöning (2021)

258 Diabetes treatment was not associated with the risk for tested positive for
259 SARS-CoV-2, but with its clinical consequences. Treatment with OAD only was
260 significantly associated with less cases of a severe outcome, whereas
261 treatment with insulin only showed the opposite effect. A subgroup analysis of
262 the OAD only group showed that this effect might be linked to metformin. No
263 differences were seen for untreated and insulin/OAD co-medicated patients.

264 Normal kidney function around the time of SARS-CoV-2 testing was associated
265 with benign courses, whereas a severely impaired kidney function of kidney
266 failure was more often seen in severe cases. Additionally, elevated liver
267 enzymes were also seen more often in cases with severe COVID-19.

268 **4 DISCUSSION**

269 During the ongoing pandemic, the swift identification of patients at risk for
270 infection and severe courses of COVID-19 is highly important. Manual
271 identification of risk factors is a time-consuming task, especially for patients with
272 co-morbidities, as the respective medical histories might be very long. Using
273 NLP on EHRs poses several challenges, as written clinical text contains
274 abbreviations, acronyms, spelling errors, and nested negations.(28)
275 Furthermore, most available software solutions focus on English language text.
276 We were able to develop a suitable NLP algorithm to automatically detect
277 relevant risk factors for COVID-19 in the medical history of patients at time of
278 admission. The selected key terms allowed for a detection in German EHRs
279 with an error rate of $\leq 2\%$ on the validation dataset and thus can be used for
280 automated real-time detection to support risk assessment and triage at point-
281 of-care.

NLP COVID-19, Schöning (2021)

282 Old age, male sex, higher BMI, arterial hypertension, and diabetes were
283 identified as risk factors for a positive COVID-19 test result as well as for a more
284 severe clinical outcome. This is in line with observations of other studies and
285 reviews.(29) Old age is generally accepted as a risk factor for severe clinical
286 manifestations, probably due to coexisting age-dependent risk factors (e.g.
287 diabetes, arterial hypertension, coronary heart disease, malnutrition) and
288 remodelling of the immune system (immunosenescence).(3, 4, 30) Sex-specific
289 differences in immune response during COVID-19 might be responsible for the
290 higher vulnerability of males,(5, 31) but also life style (e.g. personal hygiene
291 and smoking)(32) might contribute to this observation. Obesity is not only
292 associated with other comorbidities such as diabetes and arterial hypertension,
293 but also several physiological changes, e.g. chronic inflammation, endocrinal
294 dysfunction, impaired pulmonary perfusion, and immune dysregulation, which
295 may contribute to severe clinical manifestations.(7-10) Furthermore, several
296 comorbidities can often also coexist and a number of two or more comorbidities
297 were a significant factor for a more severe COVID-19 clinical outcome in
298 previous studies.(33)

299 Diabetes was seen in approximately one third of patients with a severe disease
300 course. This proportion was within the range of other studies.(34, 35)
301 Monotherapy of diabetes before admission in COVID-19 patients with insulin
302 was associated with a higher percentage of severe outcomes (29.6% vs. 9.0%).
303 By contrast, OAD treatment that included metformin was associated with a
304 higher percentage of non-severe disease progression (18.9% vs. 4.6%). This
305 decrease in severity was significant for the metformin group ($p < 0.02$). OAD
306 treatment that did not include metformin was not significantly associated with

NLP COVID-19, Schöning (2021)

307 disease severity. This is in line with recent clinical reports that metformin
308 treatment of diabetic COVID-19 patients resulted in a reduced mortality
309 rate.(36-38) Metformin acts by activating the AMP-activated protein kinase
310 AMPK,(39) which then phosphorylates angiotensin-converting enzyme 2
311 (ACE2) causing increase in ACE2 expression.(40) Even though ACE2 serves
312 as molecular target for SARS-CoV-2, it also exhibits protective effects.(41-45)
313 Furthermore, AMPK also inhibits the protein kinase B (AKT)/mammalian target
314 of rapamycin (mTOR) pathway that is essential for viral translation.(46) Finally,
315 metformin, inhibits inflammatory pathways and elevated inflammatory immune
316 responses (cytokine storms).(47)

317 Possible reasons for the observation that COPD was associated with lower
318 proportion of a positive test for COVID-19, which is also seen in other published
319 data,(48) include a potential effect of smoking, the most predominant risk factor
320 for the development of COPD. Smokers have been reported to be less
321 susceptible to COVID-19.(49, 50) Furthermore, given the common symptoms
322 of COPD such as cough, increased phlegm, shortness of breath, and
323 decreased oxygen levels, it is also possible that due to the similarity in clinical
324 presentation with a SARS-CoV-2 infection might lead to a higher test rate in
325 COPD patients and thus more negative SARS-CoV-2 test results, while
326 respiratory diseases including COPD might be under-diagnosed in some
327 positively tested cases.(33) A possible protective role of regular treatment with
328 inhaled corticosteroids in patients with COPD or asthma was not confirmed by
329 Schultze et al.(51) Despite the often similar symptoms, and the presence of
330 chronic airway inflammation in both COPD and asthma, the generally better
331 outcomes in cases with asthma might be due to the usually younger age of

NLP COVID-19, Schöning (2021)

332 these patients or to the non-reversible structural changes of the airways in case
333 of COPD.

334 Our study has several limitations. As the IHG is a major hospital centre in the
335 region, patients admitted to the hospital for other reasons were also tested for
336 SARS-CoV-2 if they displayed any symptoms indicative for COVID-19. The
337 patients in the SARS-CoV-2 negative cohort probably have more health-related
338 problems than the general population. This effect is further corroborated as we
339 only analysed patients with available EHRs, and, due to the retrospective
340 nature of the study, had no information on patients tested at the ambulant
341 COVID-19 test centre without being admitted to the IHG as in- or out-patients.
342 This was partially mitigated by including records from the three months
343 preceding diagnosis. Therefore the higher incidence of specific comorbidities in
344 the SARS-CoV-2 negative cohort might represent a selection bias. Additionally,
345 we did not perform a case-controlled study or adjusted for confounding factors
346 such as smoking or age. Furthermore, with the exception of diabetes and renal
347 function, we did not differentiate between the different stages and severities of
348 the diseases. This issue can be addressed by refining the key terms list in a
349 subsequent study.

350 The validated error rate for the NLP detection of the different disease is around
351 2%. Due to this low error rate, the pronounced contrast between significant
352 features, and the large amount of EHRs analysed, it is not expected to affect
353 statistical conclusions. The main focus of the analysis were patients tested
354 positive for SARS-CoV-2, and approximately 25% of these patients were used
355 for the validation of NLP and thus manually screened for the diseases of
356 interest.

NLP COVID-19, Schöning (2021)

357 **5 CONCLUSION**

358 In conclusion, we were able to develop a suitable NLP algorithm to
359 automatically detect risk factors for COVID-19 in German EHRs at time of
360 admission with an error rate of $\leq 2\%$. Our technique should be relative easily
361 transferable to other languages by amending the key terms list and the negation
362 detection accordingly. Use of NLP and data mining can provide a timesaving
363 support for risk assessment and triage at point-of-care, especially in patients
364 with long medical histories and multiple comorbidities. In the future, the same
365 technique could also be used for vaccine prioritization for the automated and
366 time efficient identification of persons at risk.

367 **6 DECLARATIONS**

368 **6.1 Authors' contributions**

369 FH and VS conceptualized this study. VS and FH performed the data analysis.
370 FH and EL contributed to data extraction. All authors critically revised and
371 approved the final manuscript.

372 **6.2 Acknowledgements**

373 We thank Noel Frey, Myoori Wijayasingham, and the Insel Data Science Center
374 for database and infrastructure support.

375 **6.3 Competing interest**

376 None.

NLP COVID-19, Schöning (2021)

377 **6.4 Funding**

378 None.

379 **6.5 Ethics approval and consent to participate**

380 The study was approved by the Cantonal Ethics Committee of Bern (Project-ID
381 2020-00973). Participants either agreed to a general research consent or, for
382 participants with no registered general research consent status (neither
383 agreement nor rejection), a waiver of consent was granted by the ethics
384 committee.

385 **6.6 Availability of data and materials**

386 The datasets used and/or analyses during the current study are available from
387 the corresponding author on reasonable request.

388 **7 REFERENCES**

- 389 1 Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human
390 respiratory disease in China. *Nature*. 2020 2020/03/01;**579**(7798):265-9.
391 2 Johns Hopkins University. COVID-19 Dashboard by the Center for Systems
392 Science and Engineering (CSSE) at Johns Hopkins University (JHU). 2020
393 [cited 2020 17.11.2020]; Available from: <https://coronavirus.jhu.edu/map.html>
394 3 Bencivenga L, Rengo G, Varricchi G. Elderly at time of COroNaVirus disease
395 2019 (COVID-19): possible role of immunosenescence and malnutrition.
396 *GeroScience*. 2020 2020/08/01;**42**(4):1089-92.
397 4 Cunha LL, Perazzio SF, Azzi J, Cravedi P, Riella LV. Remodeling of the
398 Immune Response With Aging: Immunosenescence and Its Potential Impact
399 on COVID-19 Immune Response. *Frontiers in immunology*. 2020;**11**:1748-
400 5 Peckham H, de Gruijter NM, Raine C, et al. Male sex identified by global
401 COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nature*
402 *communications*. 2020 2020/12/09;**11**(1):6317.
403 6 Klein SL, Dhakal S, Ursin RL, Deshpande S, Sandberg K, Mauvais-Jarvis F.
404 Biological sex impacts COVID-19 outcomes. *PLOS Pathogens*.
405 2020;**16**(6):e1008570.
406 7 Kang Z, Luo S, Gui Y, et al. Obesity is a potential risk factor contributing to
407 clinical manifestations of COVID-19. *International journal of obesity*. 2020
408 Dec;**44**(12):2479-85.

NLP COVID-19, Schöning (2021)

- 409 8 Kwok S, Adam S, Ho JH, et al. Obesity: A critical risk factor in the COVID-19
410 pandemic. *Clinical obesity*. 2020 Dec;**10**(6):e12403.
- 411 9 Foldi M, Farkas N, Kiss S, et al. Obesity is a risk factor for developing critical
412 condition in COVID-19 patients: A systematic review and meta-analysis.
413 *Obesity reviews : an official journal of the International Association for the Study*
414 *of Obesity*. 2020 Oct;**21**(10):e13095.
- 415 10 Sattar N, McInnes IB, McMurray JJV. Obesity Is a Risk Factor for Severe
416 COVID-19 Infection: Multiple Potential Mechanisms. *Circulation*. 2020 Jul
417 7;**142**(1):4-6.
- 418 11 Zhu L, She Z-G, Cheng X, et al. Association of Blood Glucose Control and
419 Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes. *Cell*
420 *metabolism*. 2020 2020/06/02/;**31**(6):1068-77.e3.
- 421 12 Tadic M, Cuspidi C. The influence of diabetes and hypertension on outcome
422 in COVID-19 patients: Do we mix apples and oranges? *The Journal of Clinical*
423 *Hypertension*. 2020 2020/12/25;**n/a**(n/a).
- 424 13 Lim S, Bae JH, Kwon H-S, Nauck MA. COVID-19 and diabetes mellitus: from
425 pathophysiology to clinical management. *Nature Reviews Endocrinology*. 2021
426 2021/01/01/;**17**(1):11-30.
- 427 14 Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular Disease,
428 Drug Therapy, and Mortality in Covid-19. *N Engl J Med*. 2020 Jun
429 18;**382**(25):e102.
- 430 15 Ssentongo P, Ssentongo AE, Heilbrunn ES, Ba DM, Chinchilli VM.
431 Association of cardiovascular disease and 10 other pre-existing comorbidities
432 with COVID-19 mortality: A systematic review and meta-analysis. *PLOS ONE*.
433 2020;**15**(8):e0238215.
- 434 16 Inciardi RM, Adamo M, Lupi L, et al. Characteristics and outcomes of
435 patients hospitalized for COVID-19 and cardiac disease in Northern Italy.
436 *European Heart Journal*. 2020;**41**(19):1821-9.
- 437 17 Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities and its effects
438 in patients infected with SARS-CoV-2: a systematic review and meta-analysis.
439 *International Journal of Infectious Diseases*. 2020 2020/05/01/;**94**:91-5.
- 440 18 Olloquequi J. COVID-19 Susceptibility in chronic obstructive pulmonary
441 disease. *European Journal of Clinical Investigation*. 2020
442 2020/10/01/;**50**(10):e13382.
- 443 19 Desai A, Khaki AR, Kuderer NM. Use of Real-World Electronic Health
444 Records to Estimate Risk, Risk Factors, and Disparities for COVID-19 in
445 Patients With Cancer. *JAMA Oncology*. 2021;**7**(2):227-9.
- 446 20 Wang Q, Davis PB, Gurney ME, Xu R. COVID-19 and dementia: Analyses
447 of risk, disparity, and outcomes from electronic health records in the US.
448 *Alzheimer's & Dementia*. 2021 2021/02/09/;**n/a**(n/a).
- 449 21 Nandy K, Salunke A, Pathak SK, et al. Coronavirus disease (COVID-19): A
450 systematic review and meta-analysis to evaluate the impact of various
451 comorbidities on serious events. *Diabetes & Metabolic Syndrome: Clinical*
452 *Research & Reviews*. 2020 2020/09/01/;**14**(5):1017-25.
- 453 22 Liu N, Sun J, Wang X, Zhao M, Huang Q, Li H. The Impact of Dementia on
454 the Clinical Outcome of COVID-19: A Systematic Review and Meta-Analysis.
455 *Journal of Alzheimer's Disease*. 2020;**78**:1775-82.
- 456 23 Schöning V, Liakoni E, Baumgartner C, et al. Development and validation of
457 a prognostic COVID-19 severity assessment (COSA) score and machine

NLP COVID-19, Schöning (2021)

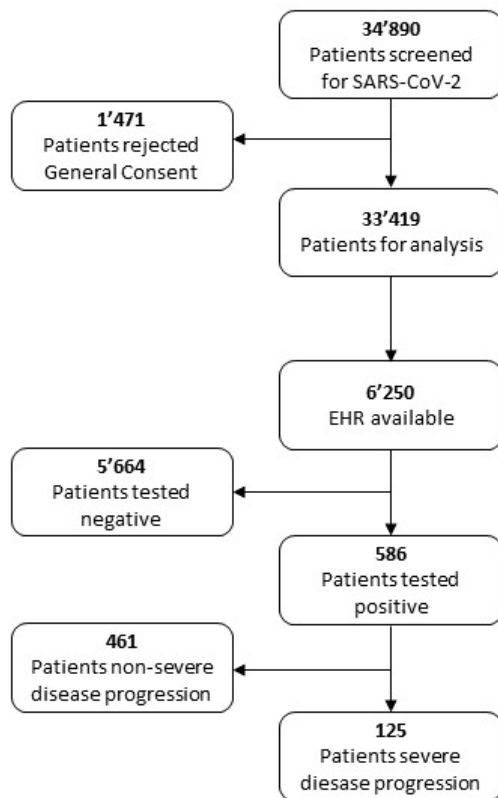
- 458 learning models for patient triage at a tertiary hospital. *Journal of Translational*
459 *Medicine*. 2021 2021/02/05;**19**(1):56.
- 460 24 National Kidney Foundation. K/DOQI clinical practice guidelines for chronic
461 kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis*
462 2002;**39**(2 Suppl 1):S1-266.
- 463 25 Cotik V, Roller R, Xu F, Uszkoreit H, Budde K, Schmidt D. *Negation*
464 *Detection in Clinical Reports Written in German*; 2016.
- 465 26 Bird S, Klein E, Loper E. *Natural Language Processing with Python*: O'Reilly
466 Media, Inc.; 2009.
- 467 27 Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A
468 Simple Algorithm for Identifying Negated Findings and Diseases in Discharge
469 Summaries. *Journal of Biomedical Informatics*. 2001 2001/10/01/;**34**(5):301-10.
- 470 28 Dalianis H. *Clinical Text Mining: Secondary Use of Electronic Patient*
471 *Records*: Springer Nature; 2018.
- 472 29 Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity
473 and fatality: a structured literature review. *Infection*. 2020 2020/08/28.
- 474 30 Romero Starke K, Petereit-Haack G, Schubert M, et al. The Age-Related
475 Risk of Severe Outcomes Due to COVID-19 Infection: A Rapid Review, Meta-
476 Analysis, and Meta-Regression. *International Journal of Environmental*
477 *Research and Public Health*. 2020;**17**(16):5974.
- 478 31 Takahashi T, Ellingson MK, Wong P, et al. Sex differences in immune
479 responses that underlie COVID-19 disease outcomes. *Nature*. 2020
480 2020/12/01;**588**(7837):315-20.
- 481 32 Anca PS, Toth PP, Kempler P, Rizzo M. Gender differences in the battle
482 against COVID-19: Impact of genetics, comorbidities, inflammation and lifestyle
483 on differences in outcomes. *International Journal of Clinical Practice*. 2021
484 2021/02/01;**75**(2):e13666.
- 485 33 Guan W-j, Liang W-h, Zhao Y, et al. Comorbidity and its impact on 1590
486 patients with Covid-19 in China: A Nationwide Analysis. *European Respiratory*
487 *Journal*. 2020:2000547.
- 488 34 Bhatraju PK, Ghassemieh BJ, Nichols M, et al. Covid-19 in Critically Ill
489 Patients in the Seattle Region - Case Series. *N Engl J Med*. 2020 May
490 21;**382**(21):2012-22.
- 491 35 Arentz M, Yim E, Klaff L, et al. Characteristics and Outcomes of 21 Critically
492 Ill Patients With COVID-19 in Washington State. *Jama*. 2020;**323**(16):1612-4.
- 493 36 Nafakhi H, Alareedh M, Al-Buthabhak K, Shaghee F, Nafakhi A, Kasim S.
494 Predictors of adverse in-hospital outcome and recovery in patients with
495 diabetes mellitus and COVID-19 pneumonia in Iraq. *Diabetes Metab Syndr*.
496 2020 Dec 5;**15**(1):33-8.
- 497 37 Lally MA, Tsoukas P, Halladay CW, O'Neill E, Gravenstein S, Rudolph JL.
498 Metformin is Associated with Decreased 30-Day Mortality Among Nursing
499 Home Residents Infected with SARS-CoV2. *J Am Med Dir Assoc*. 2021
500 Jan;**22**(1):193-8.
- 501 38 Kow CS, Hasan SS. Mortality risk with preadmission metformin use in
502 patients with COVID-19 and diabetes: A meta-analysis. *J Med Virol*. 2021
503 Feb;**93**(2):695-7.
- 504 39 Kahn BB, Alquier T, Carling D, Hardie DG. AMP-activated protein kinase:
505 ancient energy gauge provides clues to modern understanding of metabolism.
506 *Cell Metab*. 2005 Jan;**1**(1):15-25.

NLP COVID-19, Schöning (2021)

- 507 40 Saheb Sharif-Askari N, Saheb Sharif-Askari F, Mdkhana B, et al. Effect of
508 common medications on the expression of SARS-CoV-2 entry receptors in liver
509 tissue. *Arch Toxicol.* 2020 Dec;**94**(12):4037-41.
- 510 41 Behl T, Kaur I, Bungau S, et al. The dual impact of ACE2 in COVID-19 and
511 ironical actions in geriatrics and pediatrics with possible therapeutic solutions.
512 *Life Sci.* 2020;**257**:118075-.
- 513 42 Li Y, Zhou W, Yang L, You R. Physiological and pathological regulation of
514 ACE2, the SARS-CoV-2 receptor. *Pharmacol Res.* 2020;**157**:104833-.
- 515 43 Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. Trilogy of ACE2: a
516 peptidase in the renin-angiotensin system, a SARS receptor, and a partner for
517 amino acid transporters. *Pharmacol Ther.* 2010;**128**(1):119-28.
- 518 44 Imai Y, Kuba K, Rao S, et al. Angiotensin-converting enzyme 2 protects from
519 severe acute lung failure. *Nature.* 2005 2005/07/01;**436**(7047):112-6.
- 520 45 Ni W, Yang X, Yang D, et al. Role of angiotensin-converting enzyme 2
521 (ACE2) in COVID-19. *Critical Care.* 2020 2020/07/13;**24**(1):422.
- 522 46 Karam BS, Morris RS, Bramante CT, et al. mTOR inhibition in COVID-19: A
523 commentary and review of efficacy in RNA viruses. *J Med Virol.* 2021
524 Apr;**93**(4):1843-6.
- 525 47 Chen X, Guo H, Qiu L, Zhang C, Deng Q, Leng Q. Immunomodulatory and
526 Antiviral Activity of Metformin and Its Potential Implications in Treating
527 Coronavirus Disease 2019 and Lung Injury. *Front Immunol.* 2020;**11**:2056.
- 528 48 Deslée G, Zysman M, Burgel PR, et al. Chronic obstructive pulmonary
529 disease and the COVID-19 pandemic: Reciprocal challenges. *Respir Med Res.*
530 2020;**78**:100764-.
- 531 49 Tajlil A, Ghaffari S, Pourafkari L, Mashayekhi S, Roshanravan N. Nicotine
532 and smoking in the COVID-19 era. *J Cardiovasc Thorac Res.* 2020;**12**(2):136-
533 9.
- 534 50 Farsalinos K, Barbouni A, Niaura R. Systematic review of the prevalence of
535 current smoking among hospitalized COVID-19 patients in China: could
536 nicotine be a therapeutic option? *Internal and emergency medicine.* 2020
537 Aug;**15**(5):845-52.
- 538 51 Schultze A, Walker AJ, MacKenna B, et al. Risk of COVID-19-related death
539 among patients with chronic obstructive pulmonary disease or asthma
540 prescribed inhaled corticosteroids: an observational cohort study using the
541 OpenSAFELY platform. *The Lancet Respiratory Medicine.* 2020;**8**(11):1106-20.

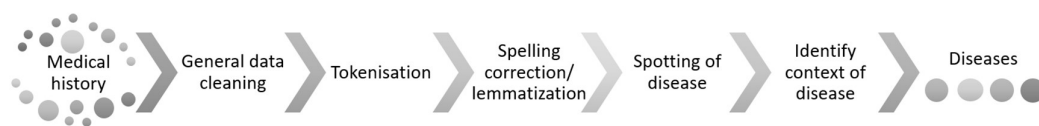
NLP COVID-19, Schöning (2021)

542 **8 FIGURES**



543

544 Figure 1: Flowchart of the selection of the patient population included in the
545 study



546

547 Figure 2: Data preparation steps for the detection of diseases in a free-text
548 medical history

549

NLP COVID-19, Schöning (2021)

550 **9 SUPPLEMENTARY MATERIAL**

551 **Table S1:** Overview on error rate of Natural Language Processing (NLP)
 552 detection of disease status on the validation dataset and the final key terms
 553 used for NLP.

Disease	Error rate on validation dataset (138 cases)	Key terms
Diabetes	0.7% (1/138)	diabetes, diabetisch, praediabetisch, praediabetes, mellitus, zuckerkrankheit, gestationsdiabetes
aHT	0.7% (1/138)	aht, arterielle hypertonie, hypertensive kardiopathie, hypertensive herzkrankheit, arterielle hypertonus, arterielle hypertension, hypertensive entgleisung, bluthochdruck, hypertensive notfall
Asthma	1.4% (2/138)	asthma, asthma bronchiale, bronchialasthma, belastungsasthma, belastungsasthma bronchiale
COPD	0.7% (1/138)	chronisch obstruktive pneumopathie, chronisch obstruktive atemwegserkrankung, chronisch obstruktive lungenerkrankung, copd, chronisch obstruktive bronchitis, chronisch bronchitis emphysem, chronisch obstruktive lungenkrankheit, obstruktive ventilationsstoerung, chronisch bronchitis lungenemphysem
Chronic heart failure	0.7% (1/138)	herzinsuffizienz, kardiomyopathie, kardiopathie, nyha, lfef, linksherzinsuffizienz, rechtsherzinsuffizienz, hypertensive herzkrankheit, hfref, hfmref, hfpef
Atrial fibrillation	0.7% (1/138)	vorhofflimmern, vhf, vorhofflattern, afib, vfli, vofli, vhfli, vfla, vhflla, vofla, ehra
Coronary heart disease	1.4% (2/138)	koronare herzerkrankung, khk, koronare herzkrankheit, stemi, nstemi, akute koronarsyndrom, acs, myokardiale minderperfusion, chronische koronarsyndrom, koronarstenose, ccs ii, ccs 2, ccs iii, ccs 3, ccs iv, ccs 4, instabile angina pectoris, ischaemische herzkrankheit, ischaemische herzerkrankung, ihk, koronaratheromatose, koronare gefaesserkrankung, koronare kardiopathie, myokardinfarkt
Dementia	1.4% (2/138)	demenz, dementielle, demenzielle, mild cognitive impairment, kognitive stoerung, mci, alzheimer, dement, kognitive beeintraechtigung, lewy

NLP COVID-19, Schöning (2021)

Disease	Error rate on validation dataset (138 cases)	Key terms
Cancer	2.1% (3/138)	Key terms according to Supplementary Material Table S2
Insulin	-	insulin, insulinpflichtig, insulinabhaengig, basisinsulintherapie, insulintherapie, insulinpumpe, insulinperfusor, insulinbasistherapie, levemir, novorapid, lantus, tresiba, toujeo, iddm]
OAD	-	oad, metformin, orale anti diabetikum, orale antidiabetikum, biguanid, glitazone, sulfonylharnstoffe, glinide, inkretine, alpha glucosidasehemmer, dpp 4 inhibitor, sgl 2 inhibitor, orale anti diabetika, orale antidiabetika, sitagliptin, janumet, kombiglyze

554

555 **Table S2:** Key terms for the detection of cancer

Key terms
adenokarzinom
adenokarzinomatoes
adenokarzinominfiltrate
adenokarzinomzellen
adenomkarzinom
adenomkarzinominfiltrate
aderhautmelanom
adnexkarzinom
aktinische keratose
akustikusneurinom
all
alveolarzellkarzinom
aml
analkarzinom
analkrebs
angiosarkom
antrumkarzinom
astrozytom
augentumor
basaliom
basalkarzinom

NLP COVID-19, Schöning (2021)

Key terms
basalzellkarzinom
bauchspeicheldruesenkrebs
bauchspeichelkrebs
blasenkarzinom
blasenkarzinom, harnblasenkarzinom
bronchialkarzinom
bronchuskarzinom
brustkrebs
burkitt lymphom
cervixkarzinom
cholangiokarzinom
chondrosarkom
chondrosarkom
chondrosarkomanteilen
ciliarkoerpermelanom
cll
cml
coecumkarzinom
colonkarzinom
corpuskarzinom
darmkrebs
denokarzinom
dermatofibrosarkoms
dickdarmkarzinom
dickdarmkarzinom
dickdarmkrebs
doppelkarzinom
dünndarmkarzinom
duodenalkarzinom
eierstockkrebs
endometrioidesadenokarzinom
endometriumkarzinom
endometriumkarzinomrezidiv
endometriumskarzinomrezidiv
ependymome
ewing sarkom
ewingsarkom

NLP COVID-19, Schöning (2021)

Key terms
fibrosarkom
gallenblasenkarzinom
gallengangkarzinom
gastrointestinale stromakarzinom
gebaermutterhalskrebs
gebaermutterhalskrebs
gebaermutterkoerperkarzinom
gebaermutterkrebs
gehirntumor
genkarzinom
gist
glioblastom
gliome
haarzelleukaemie
haemangiosarkom
harnblasenkarzinom
harnroehrenkarzinom
hautkarzinomen
hautkrebs
hautkrebsnekrosen
hautkrebssetomelanomm
hemangiosarkom
hirntumor
hodenkarzinom
hodgkin
hypopharynxkarzinom
hypopharynxkarzinoms
hypophysentumor
kardiakarzinom
karzinoid
karzinom
karzinoma
karzinomablegers
karzinomanteile
karzinomatoese
karzinomatosa
karzinomatose

NLP COVID-19, Schöning (2021)

Key terms
karzinomdiagose
karzinome
karzinomgewebe
karzinomherd
karzinominfiltrate
karzinominfiltration
karzinomlaenge
karzinomnachweis
karzinomresektion
karzinomrezidiv
karzinomsuspekte
karzinomtypisch
karzinomvolumens
karzinomzellen
karzinomzellverbaende
karzinosarkom
karzinosarkoms
kehlkopfkarzinom
keimzellneoplasie
knochenkrebs
kolonkarzinom
kolorektalkarzinom
korpuskarzinom
krebs
krebsbehandlung
krebsdiagnose
krebserkrankung
krebserkrankungen
krebskrank
larynxkarzinom
leberkrebs
leberzellkrebs
leiomyosarkom
leukaemie
leukaemiediagnose
leukaemieinfiltrate
leukaemieklonspezifischen

NLP COVID-19, Schöning (2021)

Key terms
leukaemiepersistenz
leukaemierezidiv
leukaemiezellen
leyomyosarkom
lidtumor
liposarkom
liposarkomanteile
lippentumor
lungenadenokarzinom
lungenkarzinom
lungenkrebs
lymphangiokarzinomatose
lymphdruesenkrebs
lymphgranulomatose
lymphomlymphknotenmetastasenprogression
magenkarzinom
magenkrebs
malignes myelom
mammakarzinom
mastdarmkarzinom
medulloblastome
melanom
melanoma
melanomentfernung
melanomexzision
melanomnachweis
melanomzellen
meningeome
mikrokarzinom
mischgliom
morbus kahlermundhoehlenkarzinom
multiple myelom
mycosis fungoides
myxofibrosarkom
nasennebenhoehlentumor
nasentumor
nasopharynxkarzinom

NLP COVID-19, Schöning (2021)

Key terms
nebennierenkarzinom
nebennierenneoplasie
nebenschilddrüeskarzinom
neoplasie
neoplasieassoziiert
neoplasiebedingt
neoplasiein
neoplasien
nephroblastom
netzhauttumor
neurinom
neuroblastom
nierenkarzinom
nierenzellkarzinom
nstmammakarzinom
oesophaguskarzinom
oligoastrozytome
oligodendrogliome
organmetastasen
oropharynxkarzinom
osteosarkom
osteosarkom
osteosarkomresektion
ovarialkarzinom
pankeasneoplasie
pankreaskarzinom
pankreaskopfkarzinom
pankreaskopfneoplasie
pankreaskorpuskarzinom
pankreaskorpusneoplasie
pankreasneoplasie
pankreasschwanzkarzinom
papillenkarzinom
paraneoplasie
peniskarzinom
peritonealkarzinom
peritonealkarzinomatose

NLP COVID-19, Schöning (2021)

Key terms
pharynxkarzinom
plasmazellenneoplasie
plasmazellmyelom
plasmazellneoplasie
plasmozytom
platteneithelkarzinom
platteneithelkarzinominfiltrate
platteneithelungenkarzinom
platteneithelzellkarzinom
plexustumor
primaerkarzinom
primaertumor
promyelozytenleukaemie
prostatakarzinom
prostatakrebs
rachentumor
rektumkarzinom
retinoblastom
rezidivurothelkarzinom
rhabdomyosarkom
rhabdomyosarkomatoese
samenblasenkarzinom
sarkom
sarkomatoes
sarkomatoide
sarkomgewebe
sarkomresektion
sarkomrezidiv
sarkomrezidivresektion
sarkomzellen
schadelbasismetastase
schaedelmetastase
scheidenkarzinom
schildddruesenkarzinom
schildddruesenneoplasie
siegelringzellkarzinom
sigmakarzinom

NLP COVID-19, Schöning (2021)

Key terms
speicheldrüesenzkarzinom
speiseröhrenkarzinom
spinaliom
spinalzellkarzinom
stachelzellkarzinomstachelzellkarzinom
stimmlippenkarzinom
synovialsarkom
thymon
thymuskarzinom
tonsillenzkarzinom
tritonskarzinoms
tubenzkarzinom
unterleibskrebs
urachuskarzinom
urogenitalkarzinom
urothelkarzinomrezidiv
urothelkarzinomzellen
uteruskarzinom
uterussarkom
urothelkarzinom
vaginalkarzinom
vulvakarzinom
weichteilsarkom
wilmstumor
zervixkarzinom
zervixkarzinomnachsorge
zoekumkarzinom
zungengrundkarzinom
zungenzkarzinom
zweitkarzinom
zystadenokarzinom

556