An artificial intelligence based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot

Alistair Martin, DPhil[a], Jama Nateqi, MD[b,c,*], Stefanie Gruarin, MD[b], Nicolas Munsch, MSc[a], Isselmou Abdarahmane, BSc[a], Bernhard Knapp, PhD[a]

[a]*Symptoma, Data Science Department, Vienna, Austria*
[b]*Symptoma, Medical Department, Attersee, Austria*
[c]*Paracelsus Medical University, Salzburg, Austria*

Abstract

Background As the pandemic of the novel coronavirus disease 2019 (COVID-19) progresses worldwide, many governments have established phone hotlines to pre-screen potential COVID-19 cases. These hotlines face a deluge of callers which far exceeds their capabilities, thus leading to waiting times of hours or, in many cases, a complete inability to get into contact with health authorities.

Methods Symptoma is a symptom-to-disease digital health assistant that can differentiate more than 20,000 different diseases with an accuracy of more than 90%. We tested the accuracy of Symptoma to identify COVID-19 both with regards to a diverse set of clinical cases and diseases similar in presentation to COVID-19.

Findings We showed that Symptoma can accurately distinguish COVID-19 from diseases with similar symptoms in 96.32% of clinical cases. When considering only COVID-19 symptoms and risk factors, Symptoma identified 100% of those infected when presented with only three symptoms. Lastly, we showed that Symptoma's accuracy exceeds that of simple "yes-no" questionnaires widely available online.

Interpretation Symptoma provides unparalleled accuracy in systematically identifying cases of COVID-19 while concurrently considering over

*Corresponding author

20,000 other diseases. Furthermore, Symptoma offers predefined questions alongside free text input in 36 languages. This makes Symptoma a key tool in taking pressure off from health authorities worldwide. The Symptoma predictor is freely available as a web application at `https://www.symptoma.com`.

Keywords: Science, More Science, Even More Science

1. Introduction

Currently, the world is facing an unprecedented health crisis caused by the novel coronavirus disease 2019 (COVID-19). In order to curb this crisis, among many other measures, large-scale COVID-19 laboratory testing is carried out. However, capacities are far from being able to test whole populations. Therefore many countries have established phone hotlines to pre-screen persons who are unsure about their COVID-19 infection status. Only after talking to an operator and being identified as a potential case will laboratory testing occur. However, these hotlines are severely overrun worldwide, leading to hour-long waiting periods and disconnected lines, which leads to many COVID-19 cases going undiagnosed.

Here computer-based approaches can step in. Approaches can be grouped into two categories. Firstly, a large number of simple yes/no online questionnaires are available. These questionnaires lead straight to the point but are limited in their informative value as they do not provide a deeper understanding of a patient's health situation, do not allow for the consideration of additional symptoms, do not allow the generation of additional data for analysis, and are often language- and country-specific. Secondly, there are several general-purpose symptom checkers available that have already been developed over several years (benchmarked in Nateqi et al. [1]). However, most of these symptom checkers are highly restricted in the number of diseases taken into account as building up the database is quite cost-intensive, slow, and language ambiguities are hard to overcome leading to small disease databases and users can only choose from a limited list of pre-defined symptoms which makes those tools not viable. We have also recently shown that our Symptoma engine outperforms other symptom checkers by a large margin [1]. This was also confirmed by another study [2]. In the following, we present COVID-19 specific Symptoma version that allows for predictions and analysis way beyond currently available methods.

2. Methods

2.1. Test cases

In order to show the performance of Symptoma for COVID-19 we analysed a total of 1,142 medical test cases. The different sets and sources of these cases are described below.

2.1.1. BMJ Cases

A total of 1,112 cases were sourced from the British Medical Journal (BMJ) [3, 4] and transcribed by a medical clinician into sets of symptoms, both negative and positive, alongside other risk factors, the patient's age and sex when available. The cases cover a hugely diverse range of causes, including but not limited to rib fractures, rabies, or metastatic cancer. The number of symptoms and keywords per case ranges from one to 33 (median eight) including terms like "right true vocal cord is immobile" and "metal buttons".

2.1.2. Covid-19 Cases

A set of 30 case reports were derived from the current literature (e.g. [5, 6, 7, 8]). For each case, a list of symptoms and risk factors the patient is presented with is given alongside age and sex.

2.1.3. COVID-19 - computer generated

We make use of the World Health Organisation (WHO) COVID-19 symptom list to construct example queries from those infected with COVID-19 [9]. The ten most frequent symptoms are combined with both "contact with someone infected with COVID-19" and "visiting/living in a COVID-19 risk area", to give 12 possible symptoms and/or risk factors. All possible combinations of these are then taken as potential COVID-19 cases yielding a total number of 4096 artificial cases.

2.2. Accuracy Measurements

For any given set of symptoms, many possible causes could give rise to that specific presentation. We count a prediction as true positive if the true cause is listed within the top 30 results returned by Symptoma. With respect to the possible 20,000 causes within Symptoma this is the top 0.15%.

Focussing on COVID-19, we can generate the following classification:

- True positive: C19 case and C19 returned in top 30 results

- False positive: Non-C19 case and C19 returned in top 30 results
- True negative: Non-C19 case and C19 not returned in top 30 results
- False negative: C19 case and C19 not returned in top 30 results

3. Results

3.1. Sensitivity and specificity

Symptoma classifies nearly all 30 COVID-19 case descriptions correctly as COVID-19 cases (96.6% sensitivity), failing only when presented with a case containing no defining symptoms of COVID-19 (Fever, Fatigue, Dizziness, Constipation, Rhonchi , Tachypnea, and Bilateral pneumonia) . Achieving 100% sensitivity is however easy e.g. by constructing a test that simply classifies every case as COVID-19. To address this issue we also tested how well Symptoma performs on cases of non-COVID-19 patients. For this purpose we use the above described 1,112 BMJ cases that stretch over 84 fields of medicine. Of these 1,112 cases, only 41 are classified as potential COVID-19 cases by Symptoma, with only seven of these ranking COVID-19 higher than the correct diagnoses. These seven cases relate to diseases that present similarly to COVID-19, however, have far lower incidence rates and, therefore, are deemed less likely, e.g., Severe Acute Respiratory Syndrome (SARS-CoV) or the Avian influenza A (H5N1) virus infection (bird flu). The results are summarized in Table 1.

	n cases	**Flagged as COVID-19 Risk**	**Not flagged as COVID-19 Risk**
COVID-19 cases	30	29 (TP)	1 (FN)
BMJ cases	1112	41 (FP)	1071 (TN)
Sensitivity	96.66% (29 of 30 detected)		
Specificity	96.31% (1071 of 1112 not wrongly detected)		
Accuracy	96.32% (1158 of 1166 predictions correct)		

Table 1: Sensitivity and specificity of Symptoma in COVID-19 cases and BMJ negative controls.

3.2. Discovery speed and sensitivity

Identifying patients presenting with COVID-19 both quickly and efficiently is of utmost importance to digital diagnoses. However, achieving both speed and accuracy simultaneously is remarkably difficult. Short, and therefore quick, questionnaires will typically have low specificity, while conversely, long questionnaires lack efficiency and speed, often containing many questions not pertinent to any given patient. Symptoma's free text search allows quick, efficient, and complex queries of symptom's without constraint to a predefined list of symptoms. To highlight this with regards to COVID-19, we show in Figure 1, the search rank of queries containing various numbers of symptoms known to be present in those infected with COVID-19 (see Methods). Key symptoms, such as suffering a fever or living in an area with a high incidence of COVID-19, leads to COVID-19 suggested within the top 30 search results immediately. This threshold is passed by 75%, 98.5%, and 100% of one, two, and three symptom queries respectively. At three symptoms, 99.1% of the possible combinations are returned within the top 10 results, and with four symptoms, all queries return COVID-19 within the top 10. These results highlight the speed with which a correct diagnosis can be observed, even when minimal symptoms are entered into the query.

Please note that the Symptoma web application gives immediate feedback to the user after each added symptom and/or answered question thereby making use of a slight gamification approach. Therefore quick convergence is important and this is shown by the above analysis.

3.3. Symptoma performs better than simple approaches

Next we show how well Symptoma performs in comparison to relatively simple COVID-19 symptom checkers. These symptom checkers aim to determine (given a limited set of symptoms as input) the likelihood of suffering from COVID-19 in comparison to influenza, common cold or hay fever. These symptom checkers are based on literature derived symptom frequencies (see Table S1) for each disease. For this purpose we have implemented four different methods: vector calculus based distance in space between case presentation and symptom frequency (SF-DIST), distance normalised by the standard deviation (=z-score) (SF-SD), distance based on principal component analysis (PCA)(SF-PCA) and cosine similarity (SF-COS) (these methods are described in more detail in the SI).

To evaluate the performance of these approaches in comparison to Symptoma, we classified the combined COVID-19 and BMJ cases of Table 1

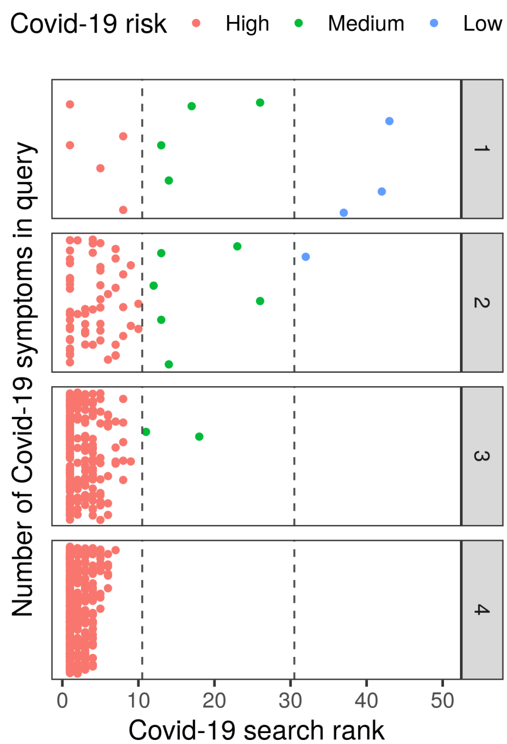

Figure 1: Identification of COVID-19 cases with regards to number of query terms entered. On the x-axis the search rank of the query in Symptoma against the y-axis, where each panel considers a different number of symptoms in the query. Only reported COVID-19 symptoms are considered. Points are jittered vertically for clarity only.

that have at least one COVID-19 symptom (n=394) with all four simple approaches. A case is classified as COVID-19-positiv if the probability of COVID-19 is at least 5% higher than the probability for influenza, common cold or hay fever. As Symptoma weights COVID-19 against more than 20.000 diseases we use the definition as stated within the Methods.

The results are summarised in Figure 2. It can be seen that Symptoma performs considerably better than any of the more simplistic approaches. This is surprising as the simple approaches just take COVID-19, common cold, flu, and hay fever into account which gives them the significant advantage of a 25% chance of a random guess to be correct.

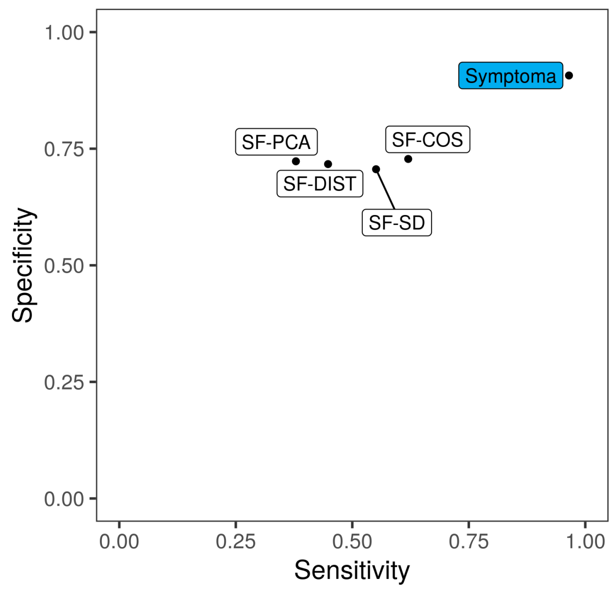

Figure 2: Comparison between four simple approaches and Symptoma using a scatter plot of sensitivity and specificity.

3.4. Free text input

A major limitation of other COVID-19 questionnaires and symptom checkers is that patients can only select from a predefined list of symptoms and answer fixed questions. Symptoma allows the patient to enter any type of symptom and the input is semantically understood. For example the symptom "Tiramisu" leads to the diagnosis "Salmonella Food Poisoning", "Pizza" and "Spaghetti" lead to "Overeating", "Donald Trump" to "Brachydactyly of Fingers" (short fingers) and "Fever" and "Italy" to "COVID-19". Please note that these are not hard-coded within Symptoma but a result of the Symptoma AI that automatically associates the meaning of symptoms with more than 20,000 diseases.

This free text input allows us to analyse if persons with a high risk of COVID-19 also have additional symptoms like the recently discovered anosmia [10, 11].

3.5. Availability in 36 languages

Symptoma is currently available in 36 languages aiming for ¿100 by the end of 2020. Due to standardization we overcome the language barrier and symptom-to-disease-predictions established from e.g. English scientific publications are also available in languages with fewer scientific publications. Vice versa the multi-language approach allows us to collect and analyse data from many countries around the world and provide a global view on entered symptoms and disease distributions.

3.6. Discussion

We present the application of the symptom-to-disease search engine Symptoma to COVID-19 cases and BMJ-derived decoy cases. Our methodology is superior to alternative approaches in multiple aspects. First, to the best of our knowledge there is no symptom-to-disease predictor that allows free text input that is semantically understood. Second, we are able to weigh COVID-19 not only against a few diseases but against more than 20,000 differential diagnoses which is far beyond the largest number of differential diagnoses by the second largest tool Isabel Healthcare with about 6,000 differential diagnoses [12]. Third, our predictive method is far beyond simplistic and pre-defined "if-then" or tree-like approaches. By constantly mining the current literature our system is up-to-date with the latest knowledge in almost real-time. Fourth, in contrast to other solutions Symptoma is available in 36 languages allowing a centralised approach on disease predictions and allowing standardised triage efforts. On these grounds we believe that Symptoma is a highly valuable tool in the global COVID-19 crisis.

Contributors

Study design: BK, AM, JN.
Data compilation: SG, NM, IA.
Data analysis: AM, BK, NM, IA.
Writing the manuscript: BK, AM.
Revising the manuscript critically: AM, BK, JN, SG.

Declaration of interests

All authors are employees of Symptoma GmbH. JN holds shares of Symptoma.

Data sharing

Results can be freely reproduced using the web interface https://www.symptoma.com/

Acknowledgments

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830017.

References

[1] J. Nateqi, S. Lin, H. Krobath, S. Gruarin, T. Lutz, T. Dvorak, A. Gruschina, R. Ortner, Vom symptom zur diagnose–tauglichkeit von symptom-checkern, HNO 67 (2019) 334–342.

[2] Pfizer, Comparison of symptom checkers, In preperation.

[3] Bmj best practice, `https://bestpractice.bmj.com/info/`, 2020.

[4] Bmj case reports, `https://casereports.bmj.com`, 2020.

[5] Z. Wang, X. Chen, Y. Lu, F. Chen, W. Zhang, Clinical characteristics and therapeutic procedure for four cases with 2019 novel coronavirus pneumonia receiving combined chinese and western medicine treatment, BioScience Trends (2020).

[6] S. B. Stoecklin, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon, M. Mechain, L. Meurice, M. Nguyen, C. Bassi, et al., First cases of coronavirus disease 2019 (covid-19) in france: surveillance, investigations and control measures, january 2020, Eurosurveillance 25 (2020).

[7] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. Zhu, et al., Pathological findings of covid-19 associated with acute respiratory distress syndrome, The Lancet respiratory medicine (2020).

[8] J. Y. Kim, J.-H. Ko, Y. Kim, Y.-J. Kim, J.-M. Kim, Y.-S. Chung, H. M. Kim, M.-G. Han, S. Y. Kim, B. S. Chin, Viral load kinetics of sars-cov-2 infection in first two patients in korea, Journal of Korean Medical Science 35 (2019).

[9] W.-C. J. Mission, Report of the who-china joint mission on coronavirus disease 2019 (covid-19). geneva 2020, 2020.

[10] Aao-hns: Anosmia, hyposmia, and dysgeusia symptoms of coronavirus disease, https://www.entnet.org/content/aao-hns-anosmia-hyposmia-and-dysgeusia-symptoms-coronavirus-disease, 2020.

[11] G. Iacobucci, Sixty seconds on . . . anosmia, BMJ 368 (2020).

[12] Isabel - the symptom checker, https://symptomchecker.isabelhealthcare.com/the-symptom-checker/how-it-works, 2020.

Appendix A. Detailed methods of the simple predictors

We developed four different methods to weight the probability of a patient having COVID-19, influenza, common cold or hay fever. For this purpose we collected symptom frequencies for these four diseases from the literature (Table A.2). To determine the probability of each of the four diseases we represent each patient case as a 10-dimensional point where each dimension is either 1 (has the symptom), 0 (does not have the symptom), or 0.5 (does not know / unknown).

In the most simplistic approach (SF-DIST) we just calculate the distance in space between the patient and each of the four diseases (that can also be seen as 10-dimensional points). Normalisation yields the respective probabilities. In the second approach the same procedure is used but the distance in space is normalised by the standard deviation (=z-score) (SF-SD). In the third approach the influence of each symptom frequency is weighted by a PCA on all frequencies (SF-PCA). In the fourth approach we interpret the points as vectors and use the cosine similarity between them instead of the distance (SF-COS).

	COVID-19	Common cold	Influenza	Hay fever
Fever	87.9 [1]	15 [3]	68 [6]	NR
Fatigue	38.1 [1]	42 [4]	94 [6]	NR
Dry cough	67.7 [1]	80 [3]	93 [6]	22 [10]
Sneezing	NR	74 [4]	58 [7]	96 [11]
Malaise	14.8 [1]	30 [4]	94 [6]	NR
Rhinorrhea	4 [2]	95 [3]	91 [6]	62.1 [12]
Sore throat	13.9 [1]	70 [3]	84 [6]	30 [10]
Diarrhea	3.7 [1]	11 [4]	14.4 [8]	NR
Headache	13.6 [1]	80 [5]	91 [6]	50 [13]
Dyspnea	18.6 [1]	21 [4]	63 [9]	NR

Table A.2: Symptom frequencies as extracted from the literature.

[1] https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[2] https://www.thelancet.com/action/showPdf?pii=S0140-6736\%2820\%2930211-7

[3] https://www.sciencedirect.com/science/article/pii/S0095454305703559?via\%3Dihub

[4] https://www.ncbi.nlm.nih.gov/pubmed/3057962

[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4347877/pdf/nihms658637.pdf

[6] https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/485554

[7] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4915903/

[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4676820/

[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3650195/

[10] https://www.ncbi.nlm.nih.gov/pubmed/10971479

[11] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5806744/

[12] https://www.researchgate.net/publication/307953143_Inverse_correlation_of_soluble_programmed_cell_death-1_ligand-1_sPD-L1_with_eosinophil_count_and_clinical_severity_in_allergic_rhinitis_patients

[13] https://www.ncbi.nlm.nih.gov/pubmed/17300360