Developing Machine Learning and Statistical Tools to Evaluate the Accessibility of Public Health Advice on Infectious Diseases among Vulnerable People

Xie, Wenxiu; Ji, Meng; Zhao, Mengdan; Lam, Kam-Yiu; Chow, Chi-Yin; Hao, Tianyong

Published in:
Computational Intelligence and Neuroscience

Published: 01/01/2021

Document Version:
Final Published version, also known as Publisher’s PDF, Publisher’s Final version or Version of Record

License:
CC BY

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1155/2021/1916690

Publication details:
Xie, W., Ji, M., Zhao, M., Lam, K-Y., Chow, C-Y., & Hao, T. (2021). Developing Machine Learning and Statistical Tools to Evaluate the Accessibility of Public Health Advice on Infectious Diseases among Vulnerable People. Computational Intelligence and Neuroscience, 2021, Article 1916690. Advance online publication. https://doi.org/10.1155/2021/1916690

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher’s copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 13/04/2024
Developing Machine Learning and Statistical Tools to Evaluate the Accessibility of Public Health Advice on Infectious Diseases among Vulnerable People

Wenxiu Xie 1*, Meng Ji 2*, Mengdan Zhao 2*, Kam-Yiu Lam 1, Chi-Yin Chow 1, and Tianyong Hao 3

1Department of Computer Science, City University of Hong Kong, Hong Kong, China
2School of Languages and Cultures, University of Sydney, Sydney, Australia
3School of Computer Science, South China Normal University, Guangzhou, China

Correspondence should be addressed to Meng Ji; christine.ji@sydney.edu.au

Received 25 October 2021; Accepted 30 November 2021; Published 17 December 2021

Copyright © 2021 Wenxiu Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. From Ebola, Zika, to the latest COVID-19 pandemic, outbreaks of highly infectious diseases continue to reveal severe consequences of social and health inequalities. People from low socioeconomic and educational backgrounds as well as low health literacy tend to be affected by the uncertainty, complexity, volatility, and progressiveness of public health crises and emergencies. A key lesson that governments have taken from the ongoing coronavirus pandemic is the importance of developing and disseminating highly accessible, actionable, inclusive, coherent public health advice, which represent a critical tool to help people with diverse cultural, educational backgrounds and varying abilities to effectively implement health policies at the grassroots level.

Objective. We aimed to translate the best practices of accessible, inclusive public health advice (purposefully designed for people with low socioeconomic and educational background, health literacy levels, limited English proficiency, and cognitive/functional impairments) on COVID-19 from health authorities in English-speaking multicultural countries (USA, Australia, and UK) to adaptive tools for the evaluation of the accessibility of public health advice in other languages. Methods. We developed an optimised Bayesian classifier to produce probabilistic prediction of the accessibility of official health advice among vulnerable people including migrants and foreigners living in China. We developed an adaptive statistical formula for the rapid evaluation of the accessibility of health advice among vulnerable people in China. Results. Our study provides needed research tools to fill in a persistent gap in Chinese public health research on accessible, inclusive communication of infectious diseases’ prevention and management. For the probabilistic prediction, using the optimised Bayesian machine learning classifier (GNB), the largest positive likelihood ratio (LR+) 16.685 (95% confidence interval: 4.35, 64.04) was identified when the probability threshold was set at 0.2 (sensitivity: 0.98; specificity: 0.94). Conclusion. Effective communication of health risks through accessible, inclusive, actionable public advice represents a powerful tool to reduce health inequalities amidst health crises and emergencies. Our study translated the best-practice public health advice developed during the pandemic into intuitive machine learning classifiers for health authorities to develop evidence-based guidelines of accessible health advice. In addition, we developed adaptive statistical tools for frontline health professionals to assess accessibility of public health advice for people from non-English speaking backgrounds.

1. Introduction

From Ebola, Zika, to the novel coronavirus pandemic, outbreaks of highly infectious diseases continue to reveal severe consequences of social and health inequalities in both developing and developed countries [1, 2]. Vulnerable people are affected more by the uncertainty, complexity, volatility, and progressiveness of public health crises and emergencies [3, 4]. A key lesson that governments have taken from the ongoing coronavirus pandemic is the importance of developing disseminating highly accessible, inclusive, actionable, coherent public health advice [5–8],
which represents a critical tool to help people with diverse cultural, educational backgrounds and varying abilities to effectively implement health policies at the grassroots level. High variability in people’s socioeconomic background, education, and health literacy levels, varying intellectual, cognitive abilities, English proficiency, and religious beliefs can cause barriers to access and implement public health advice under health emergencies and crises [9–11]. Increasing inclusiveness and accessibility of health advice recommendations among diverse vulnerable populations has emerged as an important topic in both public health education and domestic and international policy making as a highly cost-effective measure to reduce health inequalities [12–15].

Currently, there is a lack of national or international guidelines around the development of accessible and inclusive public health advice, especially for health emergencies and crises. However, the recent outbreak of coronavirus has prompted national health authorities to develop highly accessible health resources on COVID-19. Most current public-oriented health resources on infectious diseases belong to regular health resources (RHR) requiring higher levels of education, English proficiency, health knowledge, and literacy [16–19]. Typical RHR are resources published by the World Health Organisation which are intended for both general and professional readerships [20–22]. However, the practical accessibility of WHO health resources among the public remains unknown, as the international health organisation recently embarked on surveys to establish evidence of the accessibility of its public health resources [20].

Vulnerable people-oriented (VPO) health resources are known for their significantly improved language understandability, information relevance (minimal distracting details irrelevant to the readers), social inclusiveness (applicability among diverse people), and information actionability [23–25]. They are developed by medical/health professionals with extensive experiences of working firstly with diverse vulnerable populations to ensure the practical usability of VPO. In English-speaking multicultural societies, intralingual health translations and simplified or accessibility-enhanced English resources provide main sources of public health advice and information for diverse vulnerable populations [8, 26–28], although evidence-based guidelines to inform the development of these materials and associated quality control measures are yet to be established and validated.

Increasing amounts of VPO health resources in English, known as easy-to-read or easy read health materials, provide valuable firsthand materials for the development of assessment instruments and techniques to support best practices in the clinic, as well as global health policy making around accessible health recommendation design and social dissemination. The development of language-adaptive evaluation instruments can facilitate evidence-based health policy making by international health authorities. For health policy makers, we developed supervised Bayesian machine learning classifiers, and for frontline health professionals we developed convenient, language-adaptable statistical analysis tools to evaluate the accessibility of public health advice on top infectious diseases, including COVID-19.

It should be noted that the machine learning classifiers and statistical tools that we developed using Chinese health resources can be conveniently and reliably adapted to other languages using low-cost, relatively easy-to-obtain natural language annotation tools. In doing so, our study provides practical and useful tools for clinical and health professionals working directly with vulnerable people. The intuitive Bayesian probabilistic evaluation tool that we developed will help advance research-based global accessible health advice design and international benchmarking for the pandemic which continues to spread in many developing countries and any future public health crises that require rapid, effective, and efficient response from governments and health authorities to address the practical needs from diverse vulnerable populations and help minimise the environmental and health impacts on them.

2. Methods

2.1. Collection and Translation of Regular and Vulnerable People-Oriented Health Resources. We collected two sets of stylistically distinct public health advice on the prevention and self-management of infectious diseases. Regular health resources (RHR) (202) were selected from the website of the World Health Organisation (WHO). Vulnerable People-Oriented (VPO) materials (91), especially public health advice on COVID-19, were collected from websites of health authorities, including Centers for Disease Control and Prevention (CDC), Australian Ministry of Health and Public Health England. These resources were easily recognisable due to being labelled clearly as easy-to-read public health advice and instructions on COVID-19. As an international health authority, WHO provides verified professional translations of original English resources. Since our study was to develop assessment tools to evaluate accessibility of public health advice in Chinese, Chinese translations of regular WHO health advice on infectious diseases were collected. VPO resources collected from sources of CDC, the Australian Ministry of Health and Public Health, England, were translated into Chinese using forward and backward translation recommended by the WHO [29–31].

2.2. Morphological-Lexical-Structural (MLS) Features. The Chinese translations of RHR and VPO were annotated using Chinese Readability Index Explorer (CRIE) [32, 33]. CRIE annotation provided 26 morphological-lexical-structural (MLS) features and 46 part-of-speech (POS) features (72 in total) of the annotated Chinese texts for our machine learning classifier development. MLS included average sentence number per paragraph, type token ratio (TTR), low-stroke characters (1–10 strokes), middle-stroke characters (11–20 strokes), high-stroke characters (21 or above), average strokes per character, 2-character words, 3-character words, average words per sentences, ratio of noun phrases, normalised frequency of noun phrases, average number of...
idioms per sentence, content words (verbs, nouns, adverbs, and adjectives), adverbs of negation, sentences with complex semantic (polysemous) categories, density of content words, average logarithmic frequency of content words, pronouns, personal pronouns, conjunctions, positive conjunctions, negative conjunctions, and difficult words ratio. These MLS features were studied extensively in Chinese readability research.

2.3. Parts of Speech (POS) Features. A POS tagging system developed by Academia Sinica as one of the most comprehensive automatic analysers of Chinese was applied. The POS features collected included nonpredictive adjective (A), coordinate conjunction (Ca), conjunctions (Cb), conjunctions (Cba), correlative conjunctions (Cbb), adverbs (D), nominal/adverbial/complement markers (DE), adverbial noun-modifiers (Da), adverbs of degree before verbs (Dfa), adverbs of degree after verbs (Dbf), tense markers (Di), sentential adverbs (Dk), interjections (I), common nouns (Na), proper nouns (Nb), geographical names (Nc), location names (Ncd), time adverbs (Nd), attributive adjectives (Nep), modifiers of quantitative measures (Nes), quantities (Neu), measures of words (Nf), post-positions (Ng), pronouns (Nh), prepositions (P), Be verbs (Shi), auxiliary words (T), intransitive predicates (VA), causeative verbs (VAC), transitive verbs placed after objects (VB), transitive verbs (VC), verbs placed before locations (VCL), predicates used with both direct and indirect objects (VD), action verbs as sentence objects (VE), action verbs as predicate objects (VF), classification verbs (VG), modifiers of verbs and nouns (VH), causeative modifiers (VHC), adjectives or past particles placed after objects (VI), transitive verbs to describe states (VJ), mental states and processes (VK), causative predicates (VL), and have verbs (V_2).

2.4. Statistical Analysis. Table 1 shows that, among the 72 natural language features (26 MLS, 46 POS), statistically significant differences \( p < 0.05 \) of nonparametric Mann-Whitney U tests (26) between regular health resources (RHR) and vulnerable people-oriented (VPO) health resources were present in 88.9% of the entire feature set: 96.15% (25/26) of morphological-lexical-structural (MLS) features and 84.78% (39/46) of part-of-speech (POS) features. In addition to 2-sided \( p \) values, we computed Hedges' \( g \) [34] as corrected Cohen's \( d \) (1, 36, and 37) and 95% confidence interval of the effect size estimates. We also provided common language effect sizes (CLES), also known as probability of superiority [35–37]. In our study, CLES allowed an intuitive interpretation of the likelihood of the mean of a certain natural language feature randomly selected from VPO as higher than the mean of that feature in RHR on infectious diseases. We calculated Hedges' \( g \) and CLES alongside commonly reported \( p \) values for two purposes: first, this was to help us interpret the result of automatic feature selection using Bayesian machine learning classifiers (Gaussian Naive Bayes selected due to the presence of normally distributed continuous variables as features); second, this facilitated the determination of sample sizes for follow-up studies or the comparison of effects across studies.

The following formula shows the corrected effect size or Hedges' \( g \):

\[
Hedges'\ g = Cohen's\ d_i \times \left(1 - \frac{3}{4(n_i + n_j) - 9}\right)
\]

The following formula shows common language effect size (CLES):

\[
Z = \frac{|x_1 - x_2|}{\sqrt{SD_1^2 + SD_2^2/2}}.
\]

When computing effect sizes (Hedges' \( g \)) and associated probabilities of superiority (CLES), we used RER as reference class. As a result, positive values indicated that features were statistically higher in more difficult regular health resources; and negative values suggested that features were prevalent in highly accessible, vulnerable people-oriented public health advice. MLS features which had large (absolute value larger than 1), corrected effect sizes, Hedges' \( g \), and CLES included the following: average sentences per paragraph (VPO: M = 0.601, SD = 0.345; RHR: M = 3.316, SD = 1.163, \( p < 0.001 \), Hedges' \( g \) = 2.588, 95% CI [2.301, 2.874], and CLES = 0.966); type token ratio (TTR) (VPO: M = 0.461, SD = 0.119; RHR: M = 0.623, SD = 0.079, \( p < 0.001 \), Hedges' \( g \) = 1.828, 95% CI [1.568, 2.088], and CLES = 0.902); frequency of difficult words (VPO: M = 141.813, SD = 96.381; RHR: M = 70.872, SD = 35.986, \( p < 0.001 \), Hedges' \( g \) = 2.376, 95% CI [-2.655, -2.098], and CLES = 0.954); pronouns (VPO: M = 41.692, SD = 35.043; RHR: M = 1.469, SD = 1.652, \( p < 0.001 \), Hedges' \( g \) = -2.530, 95% CI [-2.814, -2.245], and CLES = 0.963); personal pronouns (VPO: M = 37.868, SD = 31.838; RHR: M = 0.705, SD = 1.139, \( p < 0.001 \), Hedges' \( g \) = -2.577, 95% CI [-2.864, -2.291], and CLES = 0.966); low-stroke characters (VPO: M = 641.593, SD = 444.264; RHR: M = 284.707, SD = 140.930, \( p < 0.001 \), Hedges' \( g \) = -1.507, 95% CI [-1.758, -1.256], and CLES = 0.823); single sentences (VPO: M = 0.883, SD = 0.114; RHR: M = 0.460, SD = 0.191, \( p < 0.001 \), Hedges' \( g \) = -2.225, 95% CI [-2.498, -1.952], and CLES = 0.942).

Part of speech (POS) features which had large (absolute value larger than 1), corrected effect sizes, Hedges' \( g \), and CLES included the following: Nn (pronouns) (VPO: M = 41.692, SD = 35.043; RHR: M = 1.469, SD = 1.652, \( p < 0.001 \), Hedges' \( g \) = -2.530, 95% CI [-2.814, -2.245], and CLES = 0.963); VF (action verbs as predicate objects) (VPO: M = 4.198, SD = 3.964; RHR: M = 0.222, SD = 0.629, \( p < 0.001 \), Hedges' \( g \) = -2.119, 95% CI [-2.388, -1.849], and CLES = 0.933); VA (intransitive predicates) (VPO: M = 11.659, SD = 10.140; RHR: M = 2.088, SD = 2.214, ...
## Table 1: Mann-Whitney U test.

<table>
<thead>
<tr>
<th>MLS features (1–26) POS (27–72)</th>
<th>EHR mean</th>
<th>EHR std.</th>
<th>RHR mean</th>
<th>RHR std.</th>
<th>Asymp. sig. (2-tailed)</th>
<th>Effect size $d_{Cohen}$ (corrected effect size or Hedges’ $g$)</th>
<th>Common language effect size CLES</th>
<th>95% CI</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average sentences per paragraph</strong></td>
<td>0.601</td>
<td>0.345</td>
<td>3.316</td>
<td>1.163</td>
<td>0.000</td>
<td>2.588</td>
<td>0.966</td>
<td>2.301 to 2.874</td>
<td>52.00</td>
<td>4238.00</td>
<td>−14.669</td>
</tr>
<tr>
<td><strong>TTR</strong></td>
<td>0.461</td>
<td>0.119</td>
<td>0.623</td>
<td>0.079</td>
<td>0.000</td>
<td>1.828</td>
<td>0.902</td>
<td>1.568 to 2.088</td>
<td>4389.00</td>
<td>8575.00</td>
<td>−10.680</td>
</tr>
<tr>
<td><strong>Difficult words</strong></td>
<td>141.813</td>
<td>96.381</td>
<td>70.872</td>
<td>35.986</td>
<td>0.000</td>
<td>−1.311</td>
<td>0.823</td>
<td>−1.558 to −1.065</td>
<td>8769.50</td>
<td>70897.50</td>
<td>−6.657</td>
</tr>
<tr>
<td><strong>Low-stroke characters</strong></td>
<td>641.593</td>
<td>444.264</td>
<td>284.707</td>
<td>140.930</td>
<td>0.000</td>
<td>−1.507</td>
<td>0.857</td>
<td>−1.758 to −1.256</td>
<td>8108.00</td>
<td>70236.00</td>
<td>−7.264</td>
</tr>
<tr>
<td><strong>Middle-stroke characters</strong></td>
<td>120.396</td>
<td>80.348</td>
<td>51.810</td>
<td>27.695</td>
<td>0.000</td>
<td>−1.562</td>
<td>0.865</td>
<td>−1.814 to −1.31</td>
<td>7357.00</td>
<td>69485.00</td>
<td>−7.955</td>
</tr>
<tr>
<td><strong>High-stroke characters</strong></td>
<td>0.055</td>
<td>0.229</td>
<td>0.452</td>
<td>0.164</td>
<td>0.005</td>
<td>0.277</td>
<td>0.578</td>
<td>−2.593 to 0.127</td>
<td>14144.00</td>
<td>18330.00</td>
<td>−2.815</td>
</tr>
<tr>
<td><strong>Average strokes per character</strong></td>
<td>7.677</td>
<td>2.747</td>
<td>7.711</td>
<td>0.342</td>
<td>0.195</td>
<td>0.103</td>
<td>0.529</td>
<td>−2.593 to 0.127</td>
<td>14144.00</td>
<td>18330.00</td>
<td>−2.815</td>
</tr>
<tr>
<td><strong>2-character words</strong></td>
<td>256.637</td>
<td>176.783</td>
<td>114.341</td>
<td>56.262</td>
<td>0.000</td>
<td>−1.509</td>
<td>0.857</td>
<td>−1.763 to −1.258</td>
<td>8100.50</td>
<td>70228.50</td>
<td>−7.271</td>
</tr>
<tr>
<td><strong>3-character words</strong></td>
<td>13.000</td>
<td>11.824</td>
<td>8.298</td>
<td>6.372</td>
<td>0.008</td>
<td>−0.603</td>
<td>0.665</td>
<td>−0.837 to −0.369</td>
<td>13139.50</td>
<td>75267.50</td>
<td>−2.647</td>
</tr>
<tr>
<td><strong>Average words per sentences</strong></td>
<td>11.592</td>
<td>4.963</td>
<td>11.893</td>
<td>1.962</td>
<td>0.018</td>
<td>0.106</td>
<td>0.530</td>
<td>−0.125 to 0.336</td>
<td>13438.00</td>
<td>17624.00</td>
<td>−2.368</td>
</tr>
<tr>
<td><strong>Single sentences</strong></td>
<td>0.883</td>
<td>0.114</td>
<td>0.623</td>
<td>0.079</td>
<td>0.000</td>
<td>−2.376</td>
<td>0.954</td>
<td>−2.655 to −2.098</td>
<td>949.00</td>
<td>63077.00</td>
<td>−13.851</td>
</tr>
<tr>
<td><strong>Ratio of noun phrases</strong></td>
<td>0.267</td>
<td>0.104</td>
<td>0.415</td>
<td>0.198</td>
<td>0.000</td>
<td>0.810</td>
<td>0.717</td>
<td>−0.573 to 1.046</td>
<td>8396.50</td>
<td>12582.50</td>
<td>−6.999</td>
</tr>
<tr>
<td><strong>Frequency of noun phrases per 10K</strong></td>
<td>322.369</td>
<td>37.240</td>
<td>314.664</td>
<td>37.002</td>
<td>0.028</td>
<td>−0.208</td>
<td>0.558</td>
<td>−0.439 to 0.023</td>
<td>13620.50</td>
<td>75748.50</td>
<td>−2.200</td>
</tr>
<tr>
<td><strong>Average idioms per sentences</strong></td>
<td>0.001</td>
<td>0.012</td>
<td>0.029</td>
<td>0.004</td>
<td>0.004</td>
<td>0.424</td>
<td>0.618</td>
<td>0.192 to 0.656</td>
<td>13969.00</td>
<td>18155.00</td>
<td>−2.894</td>
</tr>
<tr>
<td><strong>Content words</strong></td>
<td>392.582</td>
<td>265.421</td>
<td>163.074</td>
<td>80.538</td>
<td>0.000</td>
<td>−1.642</td>
<td>0.877</td>
<td>−1.896 to −1.387</td>
<td>6988.50</td>
<td>69116.50</td>
<td>−8.293</td>
</tr>
<tr>
<td><strong>Adverbs of negation</strong></td>
<td>2.758</td>
<td>3.067</td>
<td>0.935</td>
<td>1.228</td>
<td>0.000</td>
<td>−1.032</td>
<td>0.767</td>
<td>−1.272 to −0.729</td>
<td>9522.50</td>
<td>71650.50</td>
<td>−6.283</td>
</tr>
<tr>
<td><strong>Sentences with complex semantic categories</strong></td>
<td>25.396</td>
<td>22.039</td>
<td>7.614</td>
<td>4.758</td>
<td>0.000</td>
<td>−1.643</td>
<td>0.877</td>
<td>−1.896 to −1.388</td>
<td>6696.50</td>
<td>68824.50</td>
<td>−8.578</td>
</tr>
<tr>
<td><strong>Density of content words</strong></td>
<td>0.828</td>
<td>0.026</td>
<td>0.815</td>
<td>0.031</td>
<td>0.000</td>
<td>−0.433</td>
<td>0.620</td>
<td>−0.665 to −0.2</td>
<td>11857.50</td>
<td>73985.50</td>
<td>−3.820</td>
</tr>
<tr>
<td><strong>Average logarithmic frequency of content words</strong></td>
<td>1.738</td>
<td>0.169</td>
<td>1.337</td>
<td>0.183</td>
<td>0.000</td>
<td>−2.225</td>
<td>0.942</td>
<td>−2.498 to −1.952</td>
<td>1854.00</td>
<td>63982.00</td>
<td>−13.009</td>
</tr>
<tr>
<td><strong>Idioms</strong></td>
<td>0.077</td>
<td>0.268</td>
<td>0.224</td>
<td>0.510</td>
<td>0.009</td>
<td>0.312</td>
<td>0.587</td>
<td>0.081 to 0.544</td>
<td>14168.00</td>
<td>18354.00</td>
<td>−2.623</td>
</tr>
<tr>
<td><strong>Pronouns</strong></td>
<td>41.692</td>
<td>35.043</td>
<td>1.469</td>
<td>1.652</td>
<td>0.000</td>
<td>−2.530</td>
<td>0.963</td>
<td>−2.245 to −2.864</td>
<td>620.50</td>
<td>62748.50</td>
<td>−14.398</td>
</tr>
<tr>
<td><strong>Personal pronouns</strong></td>
<td>37.868</td>
<td>31.838</td>
<td>0.705</td>
<td>1.139</td>
<td>0.000</td>
<td>−2.577</td>
<td>0.966</td>
<td>−2.291 to −2.291</td>
<td>326.00</td>
<td>62454.00</td>
<td>−15.334</td>
</tr>
</tbody>
</table>
Table 1: Continued.

<table>
<thead>
<tr>
<th>MLS features (1–26) POS (27–72)</th>
<th>EHR mean</th>
<th>EHR std.</th>
<th>RHR mean</th>
<th>RHR std.</th>
<th>Asymp. sig. (2-tailed)</th>
<th>Effect size $d_{Cohen}$ (corrected effect size or Hedges’ g)</th>
<th>Common language effect size CLES</th>
<th>95% CI Mann-Whitney U</th>
<th>95% CI Wilcoxon W</th>
<th>95% CI Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjunctions</td>
<td>18.967</td>
<td>16.440</td>
<td>11.520</td>
<td>6.407</td>
<td>0.001</td>
<td>-0.795</td>
<td>0.713</td>
<td>-1.031 to -0.558</td>
<td>12507.50</td>
<td>74635.50</td>
</tr>
<tr>
<td>Positive conjunctions</td>
<td>16.824</td>
<td>14.159</td>
<td>9.000</td>
<td>5.193</td>
<td>0.000</td>
<td>-0.991</td>
<td>0.758</td>
<td>-1.23 to -0.751</td>
<td>10808.00</td>
<td>72936.00</td>
</tr>
<tr>
<td>Negative conjunctions</td>
<td>0.846</td>
<td>1.584</td>
<td>1.440</td>
<td>1.393</td>
<td>0.000</td>
<td>0.414</td>
<td>0.615</td>
<td>0.182 to 0.646</td>
<td>10710.50</td>
<td>14896.50</td>
</tr>
<tr>
<td>Difficult words ratio</td>
<td>30.384</td>
<td>5.724</td>
<td>35.847</td>
<td>8.175</td>
<td>0.000</td>
<td>0.706</td>
<td>0.691</td>
<td>0.471 to 0.941</td>
<td>9400.50</td>
<td>13586.50</td>
</tr>
<tr>
<td>A</td>
<td>3.099</td>
<td>3.774</td>
<td>3.026</td>
<td>2.969</td>
<td>0.236</td>
<td>-0.023</td>
<td>0.507</td>
<td>-0.254 to 0.207</td>
<td>14740.50</td>
<td>18926.50</td>
</tr>
<tr>
<td>VI</td>
<td>0.143</td>
<td>0.485</td>
<td>0.159</td>
<td>0.424</td>
<td>0.348</td>
<td>0.037</td>
<td>0.510</td>
<td>-0.194 to 0.267</td>
<td>15418.50</td>
<td>19604.50</td>
</tr>
<tr>
<td>Dk</td>
<td>0.011</td>
<td>0.105</td>
<td>0.017</td>
<td>0.130</td>
<td>0.048</td>
<td>0.514</td>
<td></td>
<td>0.028 to 0.278</td>
<td>15919.00</td>
<td>20105.00</td>
</tr>
<tr>
<td>VG</td>
<td>2.154</td>
<td>2.454</td>
<td>2.054</td>
<td>1.919</td>
<td>0.505</td>
<td>-0.049</td>
<td>0.514</td>
<td>-0.32 to 0.142</td>
<td>14132.00</td>
<td>18318.00</td>
</tr>
<tr>
<td>Nv</td>
<td>9.967</td>
<td>10.390</td>
<td>9.290</td>
<td>6.711</td>
<td>0.083</td>
<td>-0.089</td>
<td>0.525</td>
<td>-0.138 to 0.323</td>
<td>15810.00</td>
<td>19996.00</td>
</tr>
<tr>
<td>Neqb</td>
<td>0.033</td>
<td>0.180</td>
<td>0.057</td>
<td>0.277</td>
<td>0.092</td>
<td>0.526</td>
<td></td>
<td>0.145 to 0.609</td>
<td>13151.00</td>
<td>17337.00</td>
</tr>
<tr>
<td>Cab</td>
<td>0.121</td>
<td>0.390</td>
<td>0.398</td>
<td>0.799</td>
<td>0.000</td>
<td>0.377</td>
<td>0.605</td>
<td>0.638 to 0.174</td>
<td>15488.00</td>
<td>77616.00</td>
</tr>
<tr>
<td>I</td>
<td>0.033</td>
<td>0.180</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>-0.406</td>
<td>0.613</td>
<td>-0.638 to -0.174</td>
<td>14493.00</td>
<td>76621.00</td>
</tr>
<tr>
<td>VAC</td>
<td>0.154</td>
<td>0.392</td>
<td>0.048</td>
<td>0.215</td>
<td>0.001</td>
<td>-0.406</td>
<td>0.613</td>
<td>-0.638 to -0.174</td>
<td>14377.50</td>
<td>77021.50</td>
</tr>
<tr>
<td>Nd</td>
<td>4.253</td>
<td>7.872</td>
<td>2.298</td>
<td>3.408</td>
<td>0.002</td>
<td>-0.418</td>
<td>0.616</td>
<td>-0.65 to -0.186</td>
<td>12764.50</td>
<td>74892.50</td>
</tr>
<tr>
<td>Nb</td>
<td>1.451</td>
<td>2.423</td>
<td>0.739</td>
<td>1.420</td>
<td>0.000</td>
<td>-0.425</td>
<td>0.618</td>
<td>-0.657 to -0.193</td>
<td>12369.50</td>
<td>74497.50</td>
</tr>
<tr>
<td>Dfb</td>
<td>0.066</td>
<td>0.291</td>
<td>0.003</td>
<td>0.053</td>
<td>0.000</td>
<td>-0.451</td>
<td>0.625</td>
<td>-0.683 to -0.219</td>
<td>15181.00</td>
<td>77309.00</td>
</tr>
<tr>
<td>Neu</td>
<td>4.505</td>
<td>6.339</td>
<td>2.395</td>
<td>3.209</td>
<td>0.001</td>
<td>-0.521</td>
<td>0.644</td>
<td>-0.82 to -0.288</td>
<td>12438.00</td>
<td>74566.00</td>
</tr>
<tr>
<td>VJ</td>
<td>10.011</td>
<td>8.836</td>
<td>6.455</td>
<td>5.125</td>
<td>0.000</td>
<td>-0.586</td>
<td>0.661</td>
<td>-0.81 to -0.352</td>
<td>11896.00</td>
<td>74024.00</td>
</tr>
<tr>
<td>VL</td>
<td>2.912</td>
<td>2.946</td>
<td>1.710</td>
<td>1.741</td>
<td>0.001</td>
<td>-0.588</td>
<td>0.661</td>
<td>-0.81 to -0.354</td>
<td>12457.00</td>
<td>74585.00</td>
</tr>
<tr>
<td>Cba</td>
<td>0.110</td>
<td>0.379</td>
<td>0.003</td>
<td>0.053</td>
<td>0.000</td>
<td>-0.602</td>
<td>0.665</td>
<td>-0.836 to -0.369</td>
<td>14652.50</td>
<td>76780.50</td>
</tr>
<tr>
<td>Caa</td>
<td>12.846</td>
<td>11.117</td>
<td>8.705</td>
<td>5.150</td>
<td>0.103</td>
<td>-0.608</td>
<td>0.666</td>
<td>-0.921 to -0.401</td>
<td>14244.50</td>
<td>76372.50</td>
</tr>
<tr>
<td>VB</td>
<td>0.912</td>
<td>1.488</td>
<td>0.321</td>
<td>0.722</td>
<td>0.000</td>
<td>-0.635</td>
<td>0.673</td>
<td>-0.921 to -0.401</td>
<td>12839.50</td>
<td>74967.50</td>
</tr>
<tr>
<td>VHC</td>
<td>0.527</td>
<td>1.089</td>
<td>1.705</td>
<td>1.958</td>
<td>0.000</td>
<td>0.649</td>
<td>0.677</td>
<td>0.145 to 0.884</td>
<td>9100.00</td>
<td>13286.00</td>
</tr>
<tr>
<td>Da</td>
<td>0.516</td>
<td>1.058</td>
<td>0.125</td>
<td>0.348</td>
<td>0.000</td>
<td>-0.686</td>
<td>0.686</td>
<td>-0.921 to -0.451</td>
<td>12813.00</td>
<td>74941.00</td>
</tr>
<tr>
<td>MLS features (1–26) POS (27–72)</td>
<td>EHR mean</td>
<td>EHR std.</td>
<td>RHR mean</td>
<td>RHR std.</td>
<td>Asymp. sig. (2-tailed)</td>
<td>Effect size corrected effect size or Hedges’ g</td>
<td>Common language effect size</td>
<td>95% CI</td>
<td>Mann-Whitney U</td>
<td>Wilcoxon W</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------</td>
<td>-------</td>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Nes</td>
<td>2.033</td>
<td>2.100</td>
<td>0.960</td>
<td>1.278</td>
<td>0.000</td>
<td>−0.723</td>
<td>0.696</td>
<td>−0.959 to −0.488</td>
<td>10783.50</td>
<td>72911.50</td>
</tr>
<tr>
<td>Dfa</td>
<td>3.473</td>
<td>3.854</td>
<td>1.608</td>
<td>1.752</td>
<td>0.000</td>
<td>−0.797</td>
<td>0.713</td>
<td>−1.033 to −0.561</td>
<td>10952.00</td>
<td>73080.00</td>
</tr>
<tr>
<td>VH</td>
<td>22.154</td>
<td>15.913</td>
<td>13.767</td>
<td>8.222</td>
<td>0.000</td>
<td>−0.817</td>
<td>0.718</td>
<td>−1.053 to −0.58</td>
<td>11401.00</td>
<td>73529.00</td>
</tr>
<tr>
<td>Nf</td>
<td>7.418</td>
<td>8.071</td>
<td>3.094</td>
<td>3.956</td>
<td>0.000</td>
<td>−0.852</td>
<td>0.727</td>
<td>−1.089 to −0.615</td>
<td>9110.50</td>
<td>71238.50</td>
</tr>
<tr>
<td>Nc</td>
<td>7.780</td>
<td>7.781</td>
<td>2.688</td>
<td>5.303</td>
<td>0.000</td>
<td>−0.864</td>
<td>0.729</td>
<td>−1.110 to −0.627</td>
<td>8121.00</td>
<td>70249.00</td>
</tr>
<tr>
<td>Nep</td>
<td>3.527</td>
<td>4.388</td>
<td>1.327</td>
<td>1.653</td>
<td>0.000</td>
<td>−0.890</td>
<td>0.736</td>
<td>−1.227 to −0.653</td>
<td>12279.50</td>
<td>74407.50</td>
</tr>
<tr>
<td>T</td>
<td>1.165</td>
<td>1.662</td>
<td>0.261</td>
<td>0.649</td>
<td>0.000</td>
<td>−0.953</td>
<td>0.750</td>
<td>−1.032 to −0.715</td>
<td>10192.00</td>
<td>72320.00</td>
</tr>
<tr>
<td>Di</td>
<td>2.901</td>
<td>3.169</td>
<td>0.912</td>
<td>1.538</td>
<td>0.000</td>
<td>−1.003</td>
<td>0.761</td>
<td>−1.243 to −0.763</td>
<td>9200.50</td>
<td>71328.50</td>
</tr>
<tr>
<td>SHI</td>
<td>4.396</td>
<td>4.942</td>
<td>1.676</td>
<td>1.709</td>
<td>0.000</td>
<td>−1.006</td>
<td>0.762</td>
<td>−1.246 to −0.766</td>
<td>10776.00</td>
<td>72904.00</td>
</tr>
<tr>
<td>VD</td>
<td>1.703</td>
<td>2.355</td>
<td>0.375</td>
<td>0.861</td>
<td>0.000</td>
<td>−1.012</td>
<td>0.763</td>
<td>−1.252 to −0.772</td>
<td>9587.50</td>
<td>71715.50</td>
</tr>
<tr>
<td>Cbb</td>
<td>5.890</td>
<td>6.402</td>
<td>2.415</td>
<td>2.001</td>
<td>0.000</td>
<td>−1.022</td>
<td>0.765</td>
<td>−1.263 to −0.782</td>
<td>10325.50</td>
<td>72453.50</td>
</tr>
<tr>
<td>Neqa</td>
<td>6.011</td>
<td>5.997</td>
<td>2.537</td>
<td>2.227</td>
<td>0.000</td>
<td>−1.034</td>
<td>0.768</td>
<td>−1.306 to −0.824</td>
<td>10036.50</td>
<td>72164.50</td>
</tr>
<tr>
<td>Na</td>
<td>113.209</td>
<td>77.721</td>
<td>64.639</td>
<td>32.606</td>
<td>0.000</td>
<td>−1.065</td>
<td>0.774</td>
<td>−1.331 to −0.848</td>
<td>11326.50</td>
<td>73454.50</td>
</tr>
<tr>
<td>Ng</td>
<td>6.473</td>
<td>8.048</td>
<td>2.068</td>
<td>1.980</td>
<td>0.000</td>
<td>−1.090</td>
<td>0.780</td>
<td>−1.444 to −0.953</td>
<td>8874.50</td>
<td>71002.50</td>
</tr>
<tr>
<td>V_2</td>
<td>4.560</td>
<td>5.879</td>
<td>1.128</td>
<td>1.214</td>
<td>0.000</td>
<td>−1.197</td>
<td>0.801</td>
<td>−1.482 to −0.993</td>
<td>9003.50</td>
<td>71131.50</td>
</tr>
<tr>
<td>VE</td>
<td>7.824</td>
<td>7.207</td>
<td>2.670</td>
<td>2.913</td>
<td>0.000</td>
<td>−1.237</td>
<td>0.809</td>
<td>−1.583 to −1.09</td>
<td>8111.50</td>
<td>70239.50</td>
</tr>
<tr>
<td>DE</td>
<td>27.396</td>
<td>21.300</td>
<td>11.653</td>
<td>7.618</td>
<td>0.000</td>
<td>−1.336</td>
<td>0.828</td>
<td>−1.672 to −1.15</td>
<td>8606.50</td>
<td>70734.50</td>
</tr>
<tr>
<td>P</td>
<td>22.571</td>
<td>16.570</td>
<td>9.330</td>
<td>6.188</td>
<td>0.000</td>
<td>−1.424</td>
<td>0.843</td>
<td>−1.777 to −1.175</td>
<td>8358.00</td>
<td>70486.00</td>
</tr>
<tr>
<td>Ncd</td>
<td>6.747</td>
<td>5.960</td>
<td>1.869</td>
<td>1.933</td>
<td>0.000</td>
<td>−1.526</td>
<td>0.86</td>
<td>−1.911 to −1.401</td>
<td>7167.50</td>
<td>69295.50</td>
</tr>
<tr>
<td>VCL</td>
<td>5.242</td>
<td>5.730</td>
<td>0.693</td>
<td>1.033</td>
<td>0.000</td>
<td>−1.656</td>
<td>0.879</td>
<td>−2.031 to −1.401</td>
<td>5248.50</td>
<td>67376.50</td>
</tr>
</tbody>
</table>
2.6. Training and Testing Machine Learning Classifiers. To train machine learning classifiers for automatic information accessibility evaluation, the total number of RHR used was 202 and the total number of VPO used was 91. Next, 73.3% (148) of RHR and 62.6% (57) of VPO were used as the training data and the remaining texts (54 RHR and 34 VPO) were used as testing data. We applied 5-fold cross-validation with the training data to produce the mean and standard deviation of area of curve (AUC) of the GNB classifier. Review of the model performance was on the remaining 30% test data in terms of AUC, accuracy, sensitivity, specificity, and macro F1 (Table 3).

2.7. Classifier Optimisation. High dimensional features can reduce the performance of machine learning classifiers due to the forced inclusion of irrelevant parameters in the model. To counter the issue of classifier underperformance caused by the presence of redundant features, we applied different classifier optimisation techniques to reduce the original features collected. First, we applied integral optimisation by selecting the optimised feature set from the combined MLS and POS features (72 in total). This led to a combinedly optimised feature set of 6 features (around 8% of the original total features): average sentences per paragraph (ASPP) \( p < 0.001 \), Hedges’ \( g_s = -2.577 \), 95% CI \([-2.864, -2.291]\), and CLES = 0.966), personal pronouns \( p < 0.001 \), Hedges’ \( g_s = -1.003 \), 95% CI \([-1.243, -0.763]\), and CLES = 0.966), Di (tense markers) \( p < 0.001 \), Hedges’ \( g_s = -0.418 \), 95% CI \([-0.65, -0.186]\), and CLES = 0.616), VF (action verbs as predicate objects) \( p < 0.001 \), Hedges’ \( g_s = -2.119 \), 95% CI \([-2.388, -1.849]\),

### Table 1: Continued.

<table>
<thead>
<tr>
<th>MLS features (1–26) POS (27–72)</th>
<th>EHR mean</th>
<th>EHR std.</th>
<th>RHR mean</th>
<th>RHR std.</th>
<th>Asymp. sig. (2-tailed)</th>
<th>Effect size ( d_{Cohen} ) (corrected effect size or Hedges’ ( g ))</th>
<th>Common language effect size CLES</th>
<th>Mann-Whitney U</th>
<th>Wilcoxon W</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>VC</td>
<td>48.791</td>
<td>35.985</td>
<td>14.759</td>
<td>10.551</td>
<td>0.000</td>
<td>-1.812</td>
<td>0.900</td>
<td>-2.071</td>
<td>5672.50</td>
<td>67800.50</td>
</tr>
<tr>
<td>D</td>
<td>44.110</td>
<td>31.491</td>
<td>14.489</td>
<td>8.247</td>
<td>0.000</td>
<td>-1.849</td>
<td>0.905</td>
<td>-1.552</td>
<td>5603.00</td>
<td>67731.00</td>
</tr>
<tr>
<td>VK</td>
<td>9.736</td>
<td>7.688</td>
<td>2.293</td>
<td>2.086</td>
<td>0.000</td>
<td>-1.889</td>
<td>0.909</td>
<td>-1.562</td>
<td>5158.50</td>
<td>67286.50</td>
</tr>
<tr>
<td>VA</td>
<td>11.659</td>
<td>10.140</td>
<td>2.088</td>
<td>2.214</td>
<td>0.000</td>
<td>-1.919</td>
<td>0.913</td>
<td>-2.181</td>
<td>4616.50</td>
<td>66744.50</td>
</tr>
<tr>
<td>VF</td>
<td>4.198</td>
<td>3.964</td>
<td>0.222</td>
<td>0.629</td>
<td>0.000</td>
<td>-2.119</td>
<td>0.933</td>
<td>-2.388</td>
<td>4026.50</td>
<td>66154.50</td>
</tr>
<tr>
<td>Nh</td>
<td>41.692</td>
<td>35.043</td>
<td>1.469</td>
<td>1.652</td>
<td>0.000</td>
<td>-2.530</td>
<td>0.963</td>
<td>-2.245</td>
<td>620.50</td>
<td>62748.50</td>
</tr>
</tbody>
</table>

\( p < 0.001, \) Hedges’ \( g_s = -1.919, 95\% CI \([-2.181, -1.656]\), and CLES = 0.913); \( p < 0.001, \) Hedges’ \( g_s = -1.889, 95\% CI \([-2.151, -1.627]\), and CLES = 0.909); D (adverbs) \( p < 0.001, \) Hedges’ \( g_s = -1.849, 95\% CI \([-2.11, -1.589]\), and CLES = 0.905); VC (transitive verbs) \( p < 0.001, \) Hedges’ \( g_s = -1.812, 95\% CI \([-2.071, -1.552]\), and CLES = 0.900); Ncd (location names) \( p < 0.001, \) Hedges’ \( g_s = -1.526, 95\% CI \([-1.777, -1.274]\), and CLES = 0.860); P (prepositions) \( p < 0.001, \) Hedges’ \( g_s = -1.424, 95\% CI \([-1.672, -1.175]\), and CLES = 0.843); and V_2 (have) \( p < 0.001, \) Hedges’ \( g_s = -1.197, 95\% CI \([-1.44, -0.953]\), and CLES = 0.801).
These optimised features were also those with the most significant statistical differences (indicated by p values, corrected effect size Hedges’ g, common language effect sizes CLES) between regular and vulnerable people-oriented health resources (Table 1). Next, we applied feature optimisation in the two sets of MLS (26) and POS (46) features separately. This led to an optimised MLS feature set of 2 features only (7.7% of the total MLS features): ASPP and ALFCW (p < 0.001, Hedges’ g = −2.225, 95% CI [−2.498, −1.952], and CLES= 0.942). The optimised POS feature set of 8 features (17.4% of the total POS features): A (nonpredicate adjectives) (p < 0.000, Hedges’ g = −0.023, 95% CI [−0.254, 0.207], and CLES= 0.507), Cbb (correlative conjunctions) (p < 0.000, Hedges’ g = −1.022, 95% CI [−1.263, −0.782], and CLES= 0.765), Dfa (adverbs of degree

and CLES = 0.933), and V_2 (have verbs) (p < 0.001, Hedges’ g = −1.197, 95% CI [−1.44, −0.953], and CLES= 0.801).

These optimised features were also those with the most significant statistical differences (indicated by p values, corrected effect size Hedges’ g, common language effect sizes CLES) between regular and vulnerable people-oriented health resources (Table 1). Next, we applied feature optimisation in the two sets of MLS (26) and POS (46) features separately. This led to an optimised MLS feature set of 2 features only (7.7% of the total MLS features): ASPP and ALFCW (p < 0.001, Hedges’ g = −2.225, 95% CI [−2.498, −1.952], and CLES= 0.942). The optimised POS feature set of 8 features (17.4% of the total POS features): A (nonpredicate adjectives) (p < 0.000, Hedges’ g = −0.023, 95% CI [−0.254, 0.207], and CLES= 0.507), Cbb (correlative conjunctions) (p < 0.000, Hedges’ g = −1.022, 95% CI [−1.263, −0.782], and CLES= 0.765), Dfa (adverbs of degree
before verbs) \( (p < 0.001, \text{Hedges' } g_s = -0.797, 95\% \text{ CI } [-1.033, -0.561], \) and CLES = 0.713), Nd (time adverbs) (see above), Ng (postpositions) \( (p < 0.001, \text{Hedges' } g_s = -1.090, 95\% \text{ CI } [-1.331, -0.848], \) and CLES = 0.780), Nh (pronomns) \( (p < 0.001, \text{Hedges' } g_s = -2.530, 95\% \text{ CI } [-2.814, -2.245], \) and CLES = 0.963), VCL (verbs placed before locations) \( (p < 0.001, \text{Hedges' } g_s = -1.656, 95\% \text{ CI } [-1.911, -1.401], \) and CLES = 0.879), and VHIC (causative modifiers) \( (p < 0.001, \text{Hedges' } g_s = 0.649, 95\% \text{ CI } [0.415, 0.884], \) and CLES = 0.677). For both integral and parallel classifier optimisations, we used backward feature elimination known as recursive feature elimination (RFE) with support vector machine as base estimator. Maximal validation accuracy/minimal classification errors were used as the feature optimisation criteria (Table 4 and Figure 2).

3. Results

Table 3 shows the performance of GNB classifiers using optimised features outperformed those using original, larger feature sets. For example, on the testing data, the integrally optimised GNB with 2 MLS and 6 POS features achieved a higher AUC (0.993), sensitivity (0.963), specificity (0.9118), and accuracy (0.940) than the classifier using the full feature set which had AUC (0.940), sensitivity (0.944), specificity (0.8824), and accuracy (0.921). The separately optimised POS feature set (8 POS features) achieved a higher AUC (0.968), sensitivity (1.0), specificity (0.8824), and accuracy (0.955) than the original POS feature set (46 POS features) (AUC = 0.852, sensitivity = 0.889, specificity = 0.7941, and accuracy = 0.852). After comparing the 3 optimised feature sets: MLS/POS jointly optimised (6), MLS optimised (2), and POS optimised (8), we added the 2 sets of separately optimised features of 2 MLS features (ASPP, ALFCW) and 8 POS features (A, Cbb, Dfa, Nd, Ng, Nh, VCL, and VHIC) together and further refined the feature set to 2 features only: ALFCW and Nh (pronouns) using the same backward elimination RFE \_ SVM procedure. This resulted in a highly simplified model (model 8) which achieved largely comparable performance to the best-performing model, the one which integrated the 2 separately optimised features (model 7). Optimised model 8 achieved higher specificity (1.0) than the model 7 (0.8824), which indicated better detection of public health resources and advice suitable for use under health emergencies and crises for maximal social accessibility.

3.1. Bayesian Probabilistic Outputs. The outputs of Bayesian machine learning classifiers are in the form of a probability of belonging to the regular public health advice training data. In our study, for the best-performing classifier (model 8: refined MLS and POS separately optimised: 2 features: ALFCW and Nh), average mean output (probability) was 0.9557 \( (SD = 0.152; \text{range: } 0.02, 1; 95\% \text{ CI: } 0.915, 0.996) \) for regular public health resources and 0.022 \( (SD = 0.069; \text{range: } 0, 0.28; 95\% \text{ CI: } -0.00118, 0.04518) \) for vulnerable people-oriented public health advice. The differences among public health advice on infectious diseases (translated to Chinese) in terms of public accessibility (high: vulnerable people-oriented resources; low: regular resources) were statistically significant \( (p < 0.001, \text{Hedges' } g_s = 7.367, 95\% \text{ CI: } 6.197, 8.536, \) and CLES = 1). Figure 3 is a histogram which shows the number of regular (restricted accessibility) and vulnerable people-oriented (high accessibility) health pieces of advice that fell into each 10% probability bin based on the GNB outputs. One hundred percent of vulnerable people-oriented health advice was assigned a probability of highly accessible health resource equal to or smaller than 50% (specificity = 1); and 98.15% of regular public resources were assigned a probability of public advice of restricted/low accessibility larger than 50% (specificity = 98.15%). Around 2% of regular public health resources were misclassified as highly accessible information and advice for the public.

3.2. Thresholds and Positive/Negative Likelihood Ratios (LR+). Although it is intuitive to use 0.5 as the probability threshold (Figure 3), this is not the case in real life scenarios, because the criterion for a meaningful cut-off depends on the desired pair of sensitivity and specificity or the diagnostic utility of the research instrument. In our study, higher classifier sensitivity is indicative of higher precision with the prediction of regular public health advice (from WHO resources) which have restricted public accessibility; and higher classifier specificity implies increased accuracy with the detection vulnerable people-oriented health advice (from 3 national health authorities) which were designed by experienced health professionals to maximally enhance the language, cognitive accessibility, informational actionability, and communicative effectiveness of these emergency pieces of advice among diverse vulnerable populations with limited education, health literary, socioeconomic abilities, and varying intellectual/cognitive capabilities.

Table 5 shows various probability cut-offs and their associated sensitivity and specificity pairs using the best-performing GNB classifier that we developed using 2 features only (ALFCW: average logarithmic frequency of content words; Nh: pronouns). It shows that when setting the cut-offs lower than 0.1, sensitivity was the highest (0.9815) and specificity was 0.9118. This means that if this machine learning system were used to assist with public health advice design and dissemination, less than 2% of public health advice with restricted accessibility would be misclassified as accessible information and less than 10% of highly accessible materials would be misclassified as ineffective public health advice, potentially increasing the budgetary burden to hire experts to review health resources or extend the timeframe of releasing information to the public. When increasing the threshold from 0.1 to 0.2, sensitivity remained unchanged, and specificity increased. From 0.2 to 0.5, specificity (cost related) remained the same, but sensitivity decreased from 0.98 to 0.96, suggesting that 2% of health public with limited accessibility to vulnerable populations would be misclassified as suitable health advice and information. In high-risk scenarios such as the outbreaks of highly
infectious diseases among some of the most deprived communities, this misjudgement in health advice planning can be very costly, as vulnerable people would be given less effective, accessible, and protective health advice and information.

Sensitivity continued to decrease when thresholds raised higher than 0.5, despite the fact that specificity reached 1. Positive likelihood ratio (LR+) (the ratio between sensitivity and false positivity) is another measurement of diagnostic utility. A LR+ larger than 10 indicates a very large effect on posttest probability of disease or in the case of our study lack of accessibility of public health advice among some of our most vulnerable communities and people, who need accessible, actionable public health advice most. In our study, using the best-performing Bayesian classifier, the highest LR+ 16.685 (95% CI: 4.35, 64.04) was reached when setting

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Cross-validation accuracy (CVA)</th>
<th>Minimal classification error (MCE) = 1-CVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint optimisation (6)</td>
<td>0.9853</td>
<td>0.0147</td>
</tr>
<tr>
<td>Optimised MLS features (2)</td>
<td>0.9902</td>
<td>0.0098</td>
</tr>
<tr>
<td>Optimised POS features (8)</td>
<td>0.9854</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

Table 4: Criteria of classifier optimisation.

Figure 2: Recursive feature elimination using SVM as base estimator.

Figure 3: Percentage of vulnerable people-oriented (VPO) materials and regular health resources (RHR) assigned by GNB to 10 probability bins.
the threshold of probability at 0.2 (sensitivity = 0.982, 95% CI: 0.95, 1; specificity = 0.941, 95% CI: 0.86, 1). This represents the safest (highest sensitivity) and the most cost-effective (lowest budget investment in expert hire) model of machine learning-assisted predictive design of accessible public health advice and information for vulnerable people and populations. We hope that this promising result based on machine learning development using low-cost, relatively easy-to-obtain natural language tools would provide more support for evidence-based, user-oriented policy making around inclusive, accessible public health advice design and communication under both regular and emergency circumstances.

4. Discussion

4.1. Retrospective Assessment of Accessibility of Public Health Advice in Other Languages. Another major strength of our study was that we translated the best practices of accessible health advice on infectious diseases developed by English health authorities during the pandemic into adaptive analytical instruments (low-cost, fast-to-build machine learning algorithms) that can be used for the retrospective assessment of existing health advice in other languages. This was achieved by translating the English materials that we collected to the Chinese language using the forward and backward translation method recommended by WHO and comprehensively annotated the translations using Chinese natural language processing tools to allow automatic feature engineering and machine learning classifier development. Table 5 and Figure 4 illustrate how probabilistic outputs might assist with decision making among policy makers and health/medical professionals when designing public health advice for vulnerable people and communities speaking Chinese or migrants, foreigners living in China or greater China regions. Next, we adapted the machine learning classifier to a convenient statistical tool which can be effectively used in the clinic by health/medical professionals and educators to assess whether a certain piece of health advice on infectious diseases is accessible to vulnerable people (both Chinese and migrants) with limited education and health literacy or Chinese proficiency while living in Chinese speaking regions. We fitted a binary Logistic Regression (LR) model with the 70% training data collected. The model contained two independent variables which were borrowed from the best-performing classifier: ALFCW (average logarithmic frequency of content words) and Nh (number of pronouns). The fitted model with the regression coefficients was shown as follows.

The following is the binary regression formula:

$$\text{score}_{LR} = -2.67 \times \text{ALFCW} - 0.663 \times \text{Nh} + 7.536.$$  \hspace{1cm} (3)

We then used the Sigmoid Function to scale the score$\text{LR}$ to the region of $[0, 1]$:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}}.$$  \hspace{1cm} (4)

The following formula shows Chinese health advice accessibility assessment tool (CHAAAT):

$$\text{score} = 100 \times \text{Sigmoid}(-2.67 \times \text{ALFCW} - 0.663 \times \text{Nh} + 7.536).$$  \hspace{1cm} (5)

We examined the performance of the CHAAAT formula on the remaining 30% test data which contained 34 highly accessible COVID-19 prevention resources translated from 3 national health authorities in the USA, Australia, and UK and 54 regular health resources on infectious diseases developed by the WHO for the public. Average mean output (transformed score using the Sigmoid Function) was 5.8958 (SD = 18.128, range: 1.51908E-32, 73.5, and 95% CI: -0.194, 12) for highly accessible, vulnerable people-oriented public health advice and resources and 88.363 (SD = 2.34, range: 12.119, 99.466, and 95% CI: 83.665, 93.062) for regular WHO public health resources with limited accessibility among vulnerable people with low education, limited health literacy, and limited Chinese proficiency such as foreigner migrants or people with cognitive impairments.

Differences between the two sets of test data measured by CHAAAT formula were statistically very significant: $p < 0.001$, Hedges’ $g_s = 7.248$, 95% CI: 6.094, 8.401, and CLES = 1. These effect sizes were comparable to those ($p < 0.000$, Hedges’ $g_s = 7.367$, 95% CI: 6.197, 8.536, and CLES = 1) of probabilistic outputs of the best-performing GNB classifier using the same feature pair: ALFCW and Nh. Figure 5 shows the number of regular (restricted accessibility) and vulnerable people-oriented (high accessibility) health pieces of advice that fell into each 10-score bin based on the transformed scores computed using the CHAAAT regression formula. Ninety-four percent of vulnerable people-oriented health advice were assigned a transformed

<table>
<thead>
<tr>
<th>Probability thresholds</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive likelihood ratio (LR+) (95% CI)</th>
<th>Negative likelihood ratio (LR-) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.982 (0.95, 1.0)</td>
<td>0.9118 (0.82, 1.0)</td>
<td>11.12 (3.77, 32.79)</td>
<td>0.02031 (0.003, 0.142)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.982 (0.95, 1.0)</td>
<td>0.941 (0.86, 1.0)</td>
<td>16.69 (4.35, 64.04)</td>
<td>0.01968 (0.003, 0.137)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.963 (0.91, 1.0)</td>
<td>0.941 (0.86, 1.0)</td>
<td>16.37 (4.26, 62.87)</td>
<td>0.039 (0.010, 0.154)</td>
</tr>
<tr>
<td>0.6</td>
<td>0.963 (0.91, 1.0)</td>
<td>1.0 (1.0, 1.0)</td>
<td>Infinity</td>
<td>0.037 (0.010, 0.144)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.9444 (0.883, 1.0)</td>
<td>1.0 (1.0, 1.0)</td>
<td>Infinity</td>
<td>0.0556 (0.019, 0.167)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9074 (0.830, 0.985)</td>
<td>1.0 (1.0, 1.0)</td>
<td>Infinity</td>
<td>0.09259 (0.0402, 0.213)</td>
</tr>
</tbody>
</table>

Machine learning classifier: refined GNB classifier (with ALFCW and Nh features)
score of highly accessible health resource smaller than 50 (specificity = 94.12%) and 96.3% of regular public resources were assigned an indicative score of public advice of restricted/low accessibility equal to or larger than 50 (sensitivity = 96.3%).

Around 4% regular (requiring higher literacy and Chinese proficiency) resources were misclassified as highly accessible information and advice for vulnerable communities and people living in Chinese speaking regions. Since the 4% errors can still result in a costly overestimate of the accessibility or usability of public health advice among vulnerable populations who require highly accessible (simple, actionable, and implementable) public health information and advice, like with machine learning classifiers, we could adjust the thresholds of the CHAAAT formula to obtain the desired sensitivity and specificity pair according to the practical needs in the clinic. Table 6 shows that, like the threshold of the best-performing GNB classifier, setting the threshold of the transformed score (Formula (5)) to 29.906 will allow the regression calculator to achieve the same sensitivity and specificity as the GNB classifier. According to practical needs for accessible public health advice under different circumstances, further decreasing the threshold will lead to increased sensitivity of the assessment tool, which might be needed when the health risks being communicated are complex, and the vulnerability of the target populations is high.

The design and assessment of population-oriented public health advice are high-stakes activities which require highly precise, reliable research tools to support informed, evidence-based public health policy planning and delivery. Our study showed that valuable experiences gained in the design of accessible public advice as triggered by the coronavirus pandemic can be translated to useful, much-needed research instruments and tools to support the design of public health advice for any future health events or crises. Furthermore, we developed assessment tools adaptable to other languages to facilitate international benchmarking and support global public health policy planning around accessible health risk communication and public engagement. The limitation of our study is that we used Chinese as an illustrating example which is a distinct language from English. However, the underlying techniques and methods that we discussed and demonstrated can be conveniently modified for other languages or writing systems, especially underresourced languages such as African languages and minority languages, as we developed high performing
Bayesian machine learning classifiers based on small datasets. For illiterate populations, our methods need further adaptation for the assessment of oral public advice and information.

**Data Availability**

The data that support the findings of this study are available upon request from the corresponding author.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**


**Table 6: Thresholds of transformed scores of CHAAAT regression formula.**

<table>
<thead>
<tr>
<th>Score thresholds</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Score thresholds</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.828</td>
<td>1.000</td>
<td>0.853</td>
<td>64.118</td>
<td>0.926</td>
<td>0.941</td>
</tr>
<tr>
<td>7.685</td>
<td>1.000</td>
<td>0.882</td>
<td>66.222</td>
<td>0.907</td>
<td>0.941</td>
</tr>
<tr>
<td>16.462</td>
<td>0.981</td>
<td>0.882</td>
<td>68.542</td>
<td>0.889</td>
<td>0.941</td>
</tr>
<tr>
<td>25.144</td>
<td>0.981</td>
<td>0.912</td>
<td>70.964</td>
<td>0.870</td>
<td>0.941</td>
</tr>
<tr>
<td>29.906</td>
<td>0.981</td>
<td>0.941</td>
<td>72.994</td>
<td>0.870</td>
<td>0.971</td>
</tr>
<tr>
<td>43.160</td>
<td>0.963</td>
<td>0.941</td>
<td>73.327</td>
<td>0.852</td>
<td>0.971</td>
</tr>
<tr>
<td>59.899</td>
<td>0.944</td>
<td>0.941</td>
<td>76.001</td>
<td>0.852</td>
<td>1.000</td>
</tr>
</tbody>
</table>


