






Explaining Socio-Demographic and Behavioral Patterns of Vaccination Against the Swine Flu (H1N1) Pandemic

Clara Punzi^{1,2,3} , Aleksandra Maslennikova^{2,3} , Gizem Gezici^{1,2} ,
Roberto Pellungrini^{1,2}, and Fosca Giannotti^{1,2}

¹ Scuola Normale Superiore, Pisa, Italy
`clara.punzi@sns.it`

² KDD Lab, ISTI-CNR, Pisa, Italy

³ Department of Computer Science, University of Pisa, Pisa, Italy

Abstract. Pandemic vaccination campaigns must account for vaccine skepticism as an obstacle to overcome. Using machine learning to identify behavioral and psychological patterns in public survey datasets can provide valuable insights and inform vaccination campaigns based on empirical evidence. However, we argue that the adoption of local and global explanation methodologies can provide additional support to health practitioners by suggesting personalized communication strategies and revealing potential demographic, social, or structural barriers to vaccination requiring systemic changes. In this paper, we first implement a chain classification model for the adoption of the vaccine during the H1N1 influenza outbreak taking seasonal vaccination information into account, and then compare it with a binary classifier for vaccination to better understand the overall patterns in the data. Following that, we derive and compare global explanations using post-hoc methodologies and interpretable-by-design models. Our findings indicate that socio-demographic factors play a distinct role in the H1N1 vaccination as compared to the general vaccination. Nevertheless, medical recommendation and health insurance remain significant factors for both vaccinations. Then, we concentrated on the subpopulation of individuals who did not receive an H1N1 vaccination despite being at risk of developing severe symptoms. In an effort to assist practitioners in providing effective recommendations to patients, we present rules and counterfactuals for the selected instances based on local explanations. Finally, we raise concerns regarding gender and racial disparities in healthcare access by analysing the interaction effects of sensitive attributes on the model's output.

Keywords: Explainable AI · Chain classification · Vaccine hesitancy · Vaccination Patterns · Protected Groups

1 Introduction

In recent years, the Covid-19 outbreak has considerably raised global awareness about pandemics. While the long-term effects of the strategies employed

The original version of this chapter was previously published non-open access. A Correction to this chapter is available at https://doi.org/10.1007/978-3-031-44067-0_33

to defeat Covid-19 have yet to be determined, studies about other pandemics, such as the 2009 pandemic caused by the A(H1N1)pdm09¹ virus (abbreviated as H1N1 or “swine flu” which is responsible for between 150.000 and 575.000 deaths globally in 2009²), revealed that vaccination is a crucial tool whose effectiveness extends beyond single-person immunisation by protecting entire communities through a phenomenon known as “herd immunity” [13,29]. Therefore, national governments must allocate the necessary resources and prepare the population, beginning with informational and awareness-raising campaigns, so that the highest possible vaccination rates can always be achieved. Notably, understanding local contexts and health-related behaviors is essential to the success of a vaccination campaign [18,41]. Vaccine-related concerns in particular pose a major threat to adequate coverage [26]. Indeed, *vaccine hesitancy*, which the World Health Organization (WHO) defines as “the delay in acceptance or refusal of vaccination despite the availability of vaccination services” [28], is listed as one of the top 10 threats to global health³.

Within the broader context of *vaccine hesitancy*, we simulate a real case scenario of H1N1 flu vaccine prediction and further examine the factors that examine *vaccine hesitancy* with Explainable AI (XAI) techniques. We foresee that explanations corresponding to the outcomes of the predictions will lead to insightful observations. Health officers and practitioners could elicit pivotal communication strategies to adopt based on the objectives of the vaccination campaign (e.g., by refuting or supporting specific opinions or behaviors). Moreover, explanations can reveal demographic or social barriers to immunisation that health officers primarily responsible for planning should address in order to implement the required systemic changes (such as the elimination of administration fees). Additionally, within the EU, i.e., if the proposed model is implemented in the EU, or its decisions affect EU citizens, explicability is required by law for high-risk AI applications such as the ones pertaining to health⁴. In the scope of this work, distinct explainable methods enable us to investigate the most influential features in the overall decision-making process of the presented AI-based models as well as case-specific justifications, i.e., *local* explanations. We also provide *counterfactual* explanations for *what-if* inquiries, as research shows that, in everyday life, individuals often rely on counterfactuals, i.e., what the model would predict if the input were marginally tweaked [8]. Specifically, we devote a substantial component of our analysis to the subsample of individuals that are not-vaccinated (H1N1) despite being at risk for developing severe symptoms. We also conduct an in-depth analysis of the correlation and impact of sensitive attributes, such as ethnicity and gender, on *vaccine hesitancy*.

¹ https://web.archive.org/web/20120505042135/http://www.who.int/influenza/gisrs_laboratory/terminology_ah1n1pdm09/en/.

² <https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/>.

³ <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>.

⁴ <https://gdpr.eu/tag/gdpr/>.

To the best of our knowledge, this is the first work that presents an Explainable AI-based Clinical Decision Support System (CDSS)⁵ that uses a comprehensive, carefully curated national survey benchmark dataset regarding the 2009 H1N1 flu pandemic, jointly prepared by the United States (US) National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC). Our proposed Explainable CDSS predicts whether a certain individual will receive the H1N1 vaccine based on the given behavioral and socio-demographic features, including one related to the uptake of the seasonal vaccine. Additionally, we implement a baseline model consisting of a binary classifier that only predicts whether a particular individual will get vaccinated or not regardless of the type of vaccine (i.e., seasonal or H1N1) to disclose general vaccination patterns in the US. The most similar work to ours is a recent preprint that presents an AI-based CDSS for COVID-19 *vaccine hesitancy* [2]. Yet, in [2], researchers do not use a comprehensive benchmark dataset that has been prepared by an official agency, but rather they employ a small survey dataset that they collected using Qualtrics (a web-based survey tool), which includes only 2000 instances in total. In addition to this, the authors present a more coarse-grained study in which the XAI methods are only utilised to find the most significant factors that impact a person’s decision in the overall dataset and among different ethnic groups without using local explanations or counterfactuals.

Our main contributions can be summarised as follows:

1. We propose an AI-based CDSS to predict *vaccine hesitancy* in the US using a comprehensive benchmark dataset collected during the 2009 H1N1 flu pandemic by the US National Center for Health Statistics.
2. We leverage various XAI techniques to identify the most critical behavioral, socio-demographic, and external factors that have the greatest influence on *vaccine hesitancy*, primarily in the critical situation of the H1N1 flu outbreak, with the aim of providing evidence-based recommendations that could aid health officials and practitioners in developing effective vaccination campaigns.
3. Our findings demonstrate that doctor recommendations are essential for alleviating *vaccine hesitancy*, hence, we incorporate both *local* and *global* explanations to assist healthcare providers by providing sample tailored recommendations, particularly for the patients deemed at high risk of the H1N1 flu. These explanations can be used to select the optimal communication strategy based on a given patient, and if this patient is a non-vaccinated high-risk individual, then we further generate *counterfactuals* that can be exploited to persuade the patient.
4. As anticipated, our results from a real-world scenario also reveal *social injustice* issues in accessing healthcare services and report that the lack of health insurance is one of the most significant factors in *vaccine hesitancy*, which is typically associated with *sensitive attributes* such as belonging to particular gender and ethnic groups.

⁵ CDSS: An application that analyzes data to help healthcare providers make decisions and improve patient care.

The remainder of the paper is structured as follows. In Sect. 2 we first provide some related work, then in Sect. 3 we describe the technical details of our *vaccine hesitancy* prediction framework, which is composed of the classification models and the XAI methods we used. In Sect. 4 we detail the experimental setup, present the results and further discuss them. Finally, in Sect. 5 we mention the limitations with several potential future work directions and conclude the paper.

2 Background and Related Work

In recent times, XAI has drawn significant attention [1, 19–21, 27, 35–37, 39, 40] primarily due to the growing concern surrounding the lack of transparency in AI applications. Humans seem to be programmed to investigate the causes behind the action; hence, they are reluctant to adopt techniques that are not explicitly interpretable, tractable, and trustworthy [24], particularly in light of the growing demand for ethical AI [5]. Studies demonstrate that providing explanations can increase understanding, which can help improve trust in automated systems [1]. Thus, XAI methods provide justifications that enable users to comprehend the reason behind a system output in a specific context. These methods can be divided into post-hoc, i.e. explanations obtained by external methods, such as SHAP (*SHapley Additive exPlanations*) [27], LIME (*Local Interpretable Model-Agnostic Explanations*) [35]), and LORE (*LOcal Rule-based Explanations*) [19], and explainable-by-design (transparent) methods, i.e. built to be explainable, such as linear models, k -nearest neighbours, and decision trees. The post-hoc XAI methods can be classified as *model-specific* or *model-agnostic* based on the underlying model to be explained and if an explainer does not consider the black box internals and learning process, it is a *model-agnostic* approach. In addition to the aforementioned post-hoc methods, ANCHOR [36] which is a successor of LIME and outputs easy-to-understand if-then rules is a *model-agnostic* explainer, as well. Moreover, the state-of-the-art XAI methods can also be differentiated as *global*, or *local*. The global approaches explain the whole decision logic of a black box model, whereas the local approaches focus on a specific instance. Based on this categorisation, SHAP is a global explainer, whereas LIME, LORE, and ANCHOR are local explainers. INTGRAD [40], DEEPLIFT [39], and GRAD-CAM [37] are saliency mapping-based methods for neural networks that are model-specific, and local explainers.

XAI in Healthcare. AI-based CDSSs are computer systems developed to assist in the delivery of healthcare and can be helpful as a *second set of eyes* for clinicians [3]. The trust issue is particularly obvious in CDSS where health professionals have to interpret the output of AI systems to decide on a specific patient's case. Therefore, it is vital that XAI applications to AI-based CDSS increase trust by allowing healthcare officials to investigate the reasons behind its suggestions. Cai et al. reveal that clinicians expressed a desire for preliminary information regarding fundamental, universal characteristics of a model, such as its inherent strengths and limitations, subjective perspective, and overarching design objective, rather than solely comprehending the localized, context-dependent rationale

behind each model decision. There have been many attempts to leverage XAI in healthcare [9–11, 17, 33, 38]. In [9], scholars investigate the expectation of pathologists from the AI-based CDSS assistant. This qualitative lab study reveals that the medical experts have a desire for preliminary information regarding fundamental, universal characteristics of a model, such as its inherent strengths and limitations, subjective perspective, and overarching design objective, rather than solely comprehending the localized, context-dependent rationale behind each model decision. In [17], researchers analyse an AI-based imaging CDSS designed to assist health practitioners in detecting COVID cases in the scope of examining the explanation needs of different stakeholders. In [10], scholars propose an AI-based CDSS that predicts COVID-19 diagnosis using clinical, demographic, and blood variables and employs XAI to extract the most essential markers. In [33], authors present the results of a user study on the impact of advices from a CDSS on healthcare practitioners' judgment. For detailed surveys, please refer to [11]. Finally, in [38], the authors propose instead a classification model on a social media dataset that first distinguishes misleading from non-misleading tweets pertaining to COVID-19 vaccination, then extract the principal topics of discussion in terms of *vaccine hesitancy* and finally apply SHAP to identify important features in model prediction.

Classification Models in Tabular Data. The state-of-the-art approaches for prediction tasks on tabular data suggest the employment of ensemble tree-based models. In general, boosting methods build models sequentially using the entire dataset, with each model reducing the error of the previous one. Differently from other gradient-boosting ensemble algorithms, such as XGBoost [14] and LightGBM [25], CatBoost (proposed by Yandex) [15] employs balanced trees that not only allow for quicker computations and evaluation but also prevent overfitting. For such a reason, together with the peculiar structure of our dataset, we decided to firstly rely on this model. Notably, Catboost includes a built-in function for feature selection that removes features recursively based on the weights of a trained model. Feature scores provide an estimate of how much the average prediction changes when a feature's value is altered⁶. Consequently, despite being classified as a black box, CatBoost retains some global interpretability. As a second classification model, we use TabNet (proposed by Google) [4], a deep neural network devised specifically for tabular data and classified as an explainable-by-design model. TabNet's architecture combines two important advantages of state-of-the-art classification approaches: the explainability of tree-based algorithms and the high performance of neural networks. In addition to global interpretability, Tabnet implements local interpretability for instance-wise feature selection, unlike CatBoost.

⁶ <https://catboost.ai/en/docs/concepts/fstr>.

3 The Explainable AI-Based CDSS of Vaccine Hesitancy

3.1 Dataset

We used the dataset from the *National 2009 H1N1 Flu Survey (NHFS)*, a questionnaire conducted in the US during the 2009 H1N1 flu outbreak⁷ to monitor vaccination coverage and produce timely estimates of vaccination coverage rates⁸. The survey contains questions about influenza-related behaviours, opinions regarding vaccine safety and effectiveness as well as disease history etc. (the full NHFS questionnaire can be found on the CDC website⁹). The dataset contains 26.707 instances, 36 categorical features (the first being the ID of each anonymized individual), all of which are binary, ordinal, or nominal, and two additional binary variables that can be used as targets, namely, the seasonal and H1N1 flu vaccination status. As anticipated, the features include demographic data (e.g. sex, race, geographic location), health-related behaviors (e.g., washing hands, wearing a face mask), and opinions about flu and vaccine risks. Note that a competition has been launched on this benchmark dataset¹⁰ hence, for a complete description of the dataset, please refer to the competition website.

Preprocessing. All features in the dataset are conceptually categorical, but most of them are reported as numerical rankings or binary variables, so we only applied transformation on the remaining 12 categorical features (4 ordinal, 3 binary, and 5 multinomial). We used manual ordinal encoding for the ordinal and binary, and one-hot encoding for the multinomial ones. Also, since the dataset contains missing values in most columns, we applied iterative imputation: a strategy that models each feature with NaNs as a function of other features in a round-robin fashion. We initialized it as the most frequent value of the given variable and we set the Random Forest Classifier as the base model for the iteration step. To avoid the imputation of missing values from other synthetic data, we substituted the imputed values only at the end of the process. Lastly, in the baseline model that does not consider vaccination type, to better interpret the explanations, we merged vaccine-specific features by computing the average of corresponding H1N1 and seasonal vaccine feature scores (for instance, instead of having two separate features representing opinions about seasonal and H1N1 vaccine effectiveness, we used their average as a proxy for overall opinion about vaccine effectiveness).

3.2 Classification Models

We implemented two binary classification models for predicting the uptake of the H1N1 vaccine and the vaccine in general (regardless of the vaccine type,

⁷ <https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html>.

⁸ <https://www.drivendata.org/competitions/66/flu-shot-learning/page/213/>.

⁹ https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NIS/nhfs/nhfsauf_DUG.PDF.

¹⁰ <https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/>;

seasonal or H1N1), with the latter serving as a baseline model. In both cases, we used two state-of-the-art machine learning algorithms for classification on categorical tabular data, namely, CatBoost [34] and TabNet [4]. For the main task of predicting the uptake of the H1N1 vaccine, we decided to rely on a multi-label classifier chain since we discovered, during the data exploration phase, a positive correlation between the two target variables of seasonal and H1N1 vaccination (moderate Pearson coefficient: $\rho = 0.38$). We performed an exhaustive grid search with cross-validation on the training dataset to determine the best hyperparameters, which were then used to train the classifiers. Furthermore, given the significant imbalance in the distribution of the dataset with respect to the joint combination of the seasonal and H1N1 vaccines, we compared the performance of the selected models on augmented training datasets derived through various upsampling strategies. These techniques included a naive random over-sampling approach, where new instances of the underrepresented class were generated by picking samples at random with replacement, as well as the Synthetic Minority Oversampling Technique (SMOTE, [12]) and the Adaptive Synthetic sampling method (ADASYN, [22]). Nevertheless, none of these methods led to a significant improvement in the F1 score (see Table 1), hence we opted to maintain the initial dataset for subsequent analyses. It should be noted that, in contrast to the H1N1 model, the baseline classification model did not exhibit an imbalanced class distribution. The best performance for both the baseline and H1N1 model was achieved by CatBoost classifier.

Table 1. Model performances.

Model	Upsampling	AUC (weighted)	F1-score (weighted)	AUC (macro)	F1-score (macro)
CatBoost Classifier Chain	-	0.75	0.87	0.77	0.77
	Random oversampling	0.79	0.83	0.79	0.75
	SMOTE	0.77	0.84	0.77	0.76
	ADASYN	0.77	0.84	0.77	0.76
TabNet Classifier Chain	-	0.73	0.82	0.73	0.73
	Random oversampling	0.75	0.80	0.75	0.72
	SMOTE	0.71	0.81	0.71	0.72
	ADASYN	0.76	0.80	0.76	0.73
CatBoost Baseline	-	0.77	0.77	0.77	0.77
TabNet Baseline	-	0.75	0.75	0.75	0.75

3.3 XAI Methods

We initially obtained the global feature importance scores from TabNet [4] and CatBoost’s [15] built-in functions, and compared them to SHAP-based feature rankings. This choice is based on the fact that SHAP [27] offers a wide range of analysis tools and its feature rankings have demonstrated greater stability compared to the built-in functions of tree-based ensemble models [42]. Then, we inspected the interaction effects between features; in particular, we examined the impact of *sensitive* attributes, such as ethnicity and gender, on the model prediction. After that, we *locally* explained specific test set instances: we computed local feature importance scores with SHAP [27] and LIME [35] and extracted *counterfactuals* from LORE [19]¹¹. The instances were chosen from the subpopulations of high-risk individuals declared by the US H1N1 recommendations¹², for further discussion please see Sect. 4.2.

The goodness, usefulness, and satisfaction of an explanation should be considered when assessing the validity and convenience of an explanation technique [6]. In the scope of this study, we conducted both quantitative and qualitative assessments. On the one hand, we ensured that our explainers had a high degree of fidelity, i.e., that they could accurately approximate the prediction of the black box model [30]. On the other hand, we discussed the actual usefulness of the explanations from the perspective of the end-user, i.e., a health official or practitioner.

4 Results and Discussion

4.1 H1N1 Vaccine Hesitancy Model vs Baseline

In this part, we compare the global explanations of the baseline and H1N1 vaccine hesitancy models. First of all, we retrieved feature importance rankings using CatBoost, which is a black-box model that enables a certain degree of global interpretability, and TabNet, which is an explainable-by-design method. Figure 1a displays the feature importance rankings of the baseline model. Both models significantly rely on whether a doctor recommended a vaccination, personal opinion regarding vaccine efficacy, and age. Notably, the CatBoost model prioritises personal judgment about the risks of getting sick without vaccination and the availability of health insurance, while TabNet disregards these features entirely. In the H1N1 model (See Fig. 1b), the feature importance ranking of CatBoost differed considerably from TabNet. Both models significantly rely on the doctor’s recommendation and opinion on vaccine efficacy, but age was not a determining factor. The features of opinions about the risk of getting sick and health insurance were only considered by CatBoost in the baseline model,

¹¹ We did not use the recent version of LORE [20] which is more stable and generates actionable features as claimed by the authors since we could not execute the code in their github repo.

¹² <https://www.cdc.gov/h1n1flu/vaccination/acip.htm>.

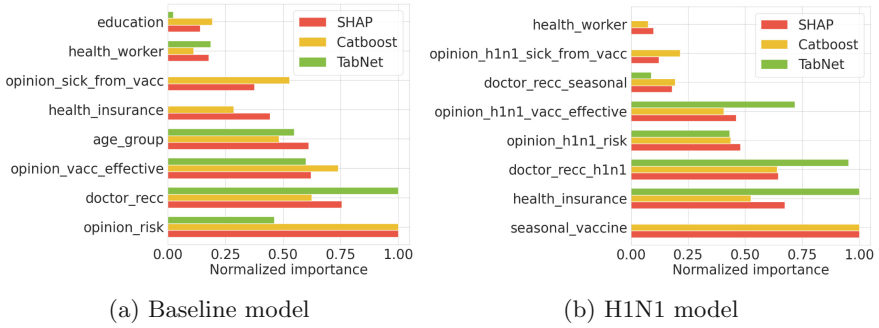


Fig. 1. Comparison of different feature importance rankings, sorted according to SHAP rankings.

while both models deem them significant for the H1N1 prediction. Interestingly, TabNet ignores the most crucial feature of CatBoost which is the seasonal vaccination status.

In addition, we computed post-hoc explanations by applying SHAP [27] to the model with the best classification performance, namely CatBoost [15]. It is noteworthy that SHAP achieved a significantly *high fidelity* score of 0.92, which is indicative of its capacity to accurately mimic the underlying black-box model. Using Tree SHAP as the algorithm to compute Shapley values, we discovered, as expected, that SHAP feature rankings were comparable to those provided by CatBoost for both the baseline and H1N1 models. In the following sections, we will refer primarily to SHAP when discussing about global explanations.

4.2 Vaccine Hesitancy in High-Risk Individuals

Due to the H1N1 vaccine’s limited availability during the campaign’s initial phase, health officials advised people at the highest risk for viral effects or those caring for them to receive the vaccine first. These target subpopulations were (1) adults who live with or care for children under 6 months, (2) healthcare workers, (3) adults aged 25 to 64 with certain chronic health conditions, (4) people aged 6 months to 24 years, and (5) pregnant women. In our work, however, we note that the target group (5) could not be analyzed since the dataset did not contain the related information, and condition (4) was slightly modified to (4’) 18-to-34-year-old, as this is the lowest age group reported in the dataset.

We used XAI techniques to understand why some high-risk individuals **do not** vaccinate in order to lay the basis for effective doctor recommendations. Indeed, the findings discussed in Sect. 4.1 indicate that doctor recommendations are crucial for promoting vaccination not only among the general population but also, and most importantly, among individuals at high risk of being severely affected by a pandemic influenza outbreak. In the following, we show how local explanations generated by SHAP [27], LIME [35], and LORE [19] can be leveraged by physicians to design effective, patient-specific communication strategies

for recommending vaccination. As a first example, consider the subject with the identifier $id = 24210$, a white woman who satisfies criteria (3) and (4'). In this instance, our model accurately predicted that she had declined the H1N1 vaccination against the doctor's recommendation. As depicted in Figs. 2a and 2b, the feature importance scores computed by SHAP and LIME concur that her belief that the vaccine was not very effective and her refusal to receive the seasonal vaccine had a substantial negative impact on the vaccination outcome. Based on LORE's counterfactual ($fidelity = 0.99$), we found that the doctor's recommendation was ineffective because she or he failed to raise the subject's opinion about the vaccine's efficacy and the swine flu's threat. Furthermore, LORE identified having health insurance and living in a particular geographical region as conditions for a positive vaccination outcome. Unfortunately, the actionability of these features is debatable, revealing the existence of social disparities in vaccination.

As a second example, we consider the subject with $id = 23241$, a black woman who meets criteria (1), (3), and (4'). Similar to the previous subject, the model accurately predicted that she had declined the H1N1 vaccination, but this time we know she did not receive a doctor's recommendation. SHAP and LIME ($fidelity = 1$) evaluate this fact to be extremely negative in terms of feature importance, along with other factors such as not having received the seasonal vaccine, having a very low opinion of the risk of becoming sick with H1N1 flu without vaccination, and not having health insurance. In addition, LIME scored unfavorably for its lack of employment in specific industries and professions. LORE ($fidelity = 0.99$) provided a coherent decision rule and a few counterfactual explanations that, first and foremost, required a doctor's recommendation and that, additionally, indicate that an effective recommendation would be one capable of increasing the subject's opinion regarding the effectiveness of the H1N1 vaccine, allowing her to obtain health insurance, and convincing her to also receive the seasonal vaccine. Interestingly, some counterfactuals also included conditions indicating non-belonging to the "black" or "other or multiple" ethnic group, as well as geographically-based criteria, which however are subject to the same limitations as those previously noted regarding the actionability of certain counterfactual.

4.3 Social Injustice in Healthcare

The US healthcare system has been widely acknowledged and recorded to exhibit structural inequalities that are often linked to particular ethnic and gender categories [16, 23]. The same holds true specifically in the campaigns for H1N1 [7], COVID-19 [31], and seasonal vaccine [32]. Therefore, socio-demographic factors like gender and ethnicity, as well as social injustice in healthcare access, should be taken into account when interpreting studies about vaccine hesitancy, as the refusal to be vaccinated may be due to structural barriers, such as a lack of health insurance in a country where public health is not guaranteed. Indeed, our results confirm that health insurance coverage is one of the most important predictive factors, especially in the H1N1 model, as shown in Sect. 4.1, and the counterfactual explanations in Sect. 4.2 consistently identified health insurance

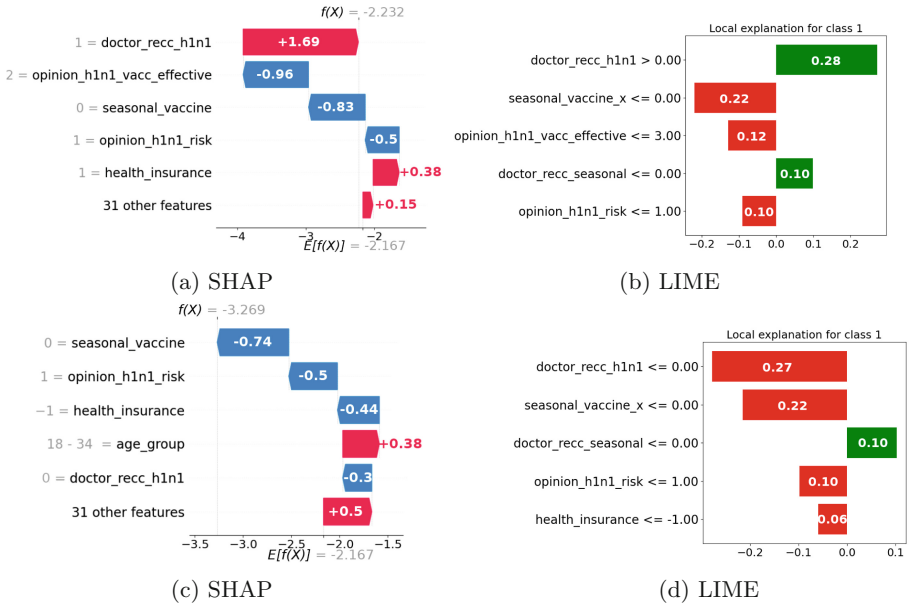


Fig. 2. Local explanations for $id = 24210$ (top row, true class = 0, predicted class = 0), and $id = 23241$ (bottom row, true class = 0, predicted class = 0).

as a key driver in promoting vaccination in the subpopulation at high-risk with respect to H1N1.

The impact of *health insurance*, *ethnicity*, and *sex* on the model’s predictions is illustrated in the dependence scatter plots in Fig. 3. In these three plots, points are displayed based on their coordinates (x, y) as feature value (x) and Shapley value (y), where each point refers to an observation. For instance, Fig. 2a displays that the perceived threat posed by H1N1 has the greatest interactive effect with health insurance in predicting vaccine uptake, while in Fig. 2b and Fig. 2c, for the sensitive attributes: ethnicity and gender, health insurance is the most interactive feature. In Fig. 2b, ethnicity does not significantly impact the model’s decision among the *white* subpopulation, since the corresponding data points are not dispersed, whereas other three subpopulations exhibit a greater degree of variation which might point to racial disparities in access to vaccination campaigns. In terms of gender, the plot in Fig. 2c reveals that men are more likely to be vaccinated irrespective of their health insurance, as most Shapley values are positive. This observed bias of the H1N1 classifier towards men conveys that there may have been real-world factors that favored men’s access to the vaccine. Interestingly, women with health insurance are less likely to be vaccinated, whereas men are more likely. The aforementioned trend in the decision rules of SHAP [27], LIME [35], and LORE [19] is corroborated by the plot in Fig. 2a, as only a minimal fraction of points without health insurance (or with no information provided) are associated with positive Shapley values. For repro-

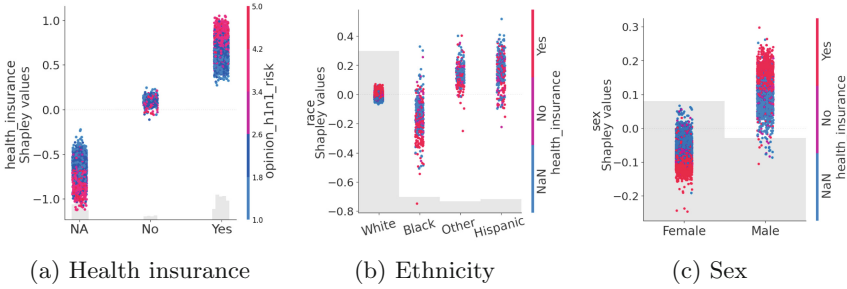


Fig. 3. Dependence scatter plots for the H1N1 model – the x -axis denotes the feature values, the y -axis refers to Shapley values, coloring is based on the values of the feature in the secondary y -axis (most interactive feature chosen by SHAP).

ducibility purposes, our code is publicly available at <https://github.com/gizem-gg/H1N1-VaccineHesitancy-CDSS>.

5 Conclusion and Future Work

In this work, we proposed an AI-based Explainable CDSS for predicting and assessing hesitancy towards the swine flu vaccination uptake. XAI methodologies assist us in identifying doctor recommendation, health insurance, seasonal vaccine adoption, and personal opinion regarding vaccine efficacy as the most influential factors in H1N1 vaccination. On the basis of counterfactual explanations, we provided physicians with suggestions for effectively conveying to their patients the need to receive the H1N1 vaccine, with a focus on those at high risk for severe symptoms. In particular, we discovered that communication strategies that can improve the subject’s opinion of the effectiveness of the H1N1 vaccine and the threat posed by the swine flu are more likely to function as catalysts for change. Moreover, our analysis highlights the crucial role of health insurance, which reflects actual disparities in healthcare access in the US, and illustrates how vaccination campaigns can be hampered not only by vaccine reluctance but also by economic constraints. Likewise, it has been found that membership in marginalized groups based on gender, ethnicity, or geography can result in individuals with a higher risk profile opting out of vaccination. A major limitation of our analysis is the large number of missing values regarding health insurance, which is one of the most important features for our model. Second, our algorithm of choice for counterfactual explanation is based on a genetic algorithm for neighborhood generation. It could be interesting to compare different algorithms for neighborhood generation. Moreover, the choice of the attribute to consider in counterfactual generation should be guided by the principle of actionability, to focus on feature that healthcare professional can act upon. As future work, we plan to address these limitations and evaluate the efficacy of the proposed Explainable AI-based CDSS framework by conducting a comprehensive user case study with health officials and physicians.

Acknowledgements. We sincerely thank Dr. Andrea Beretta (CNR-ISTI) for his invaluable guidance and insightful discussions.

This work has been supported by the European Union under ERC-2018-ADG GA 834756 (XAI), by HumanE-AI-Net GA 952026, and by the Partnership Extended PE00000013 - “FAIR - Future Artificial Intelligence Research” - Spoke 1 “Human-centered AI”.

It has been realised also thanks to computational (data and algorithms) resources of “SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics” (<http://www.sobigdata.eu>), G.A.No.871042 and by NextGenerationEU - National Recovery and Resilience Plan Resilienza, PNRR) - Project: “SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics” - Prot. IR000001 3 - Notice n. 3264 of 12/28/2021.

References

1. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* **6**, 52138–52160 (2018)
2. Alharbi, R., Chan-Olmsted, S., Chen, H., Thai, M.T.: Cultural-aware machine learning based analysis of covid-19 vaccine hesitancy. *arXiv preprint arXiv:2304.06953* (2023)
3. Antoniadi, A.M., et al.: Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. *Appl. Sci.* **11**(11), 5088 (2021)
4. Arik, S.O., Pfister, T.: TabNet: attentive interpretable tabular learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8 (2021)
5. Birhane, A.: Algorithmic injustice: a relational ethics approach. *Patterns* **2**(2), 100205 (2021)
6. Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D., Rinzivillo, S.: Benchmarking and survey of explanation methods for black box models. *CoRR abs/2102.13076* (2021)
7. Burger, A.E., Reither, E.N., Mamelund, S.E., Lim, S.: Black-white disparities in 2009 H1N1 vaccination among adults in the United States: a cautionary tale for the COVID-19 pandemic. *Vaccine* **39**(6), 943–951 (2021)
8. Byrne, R.M.: *The Rational Imagination: How People Create Alternatives to Reality*. MIT Press, Cambridge (2007)
9. Cai, C.J., Winter, S., Steiner, D., Wilcox, L., Terry, M.: “Hello AI”: uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proc. ACM Hum.-Comput. Interact.* **3**(CSCW), 1–24 (2019)
10. Chadaga, K., Prabhu, S., Bhat, V., Sampathila, N., Umakanth, S., Chadaga, R.: A decision support system for diagnosis of covid-19 from non-covid-19 influenza-like illness using explainable artificial intelligence. *Bioengineering* **10**(4), 439 (2023)
11. Chaddad, A., Peng, J., Xu, J., Bouridane, A.: Survey of explainable AI techniques in healthcare. *Sensors* **23**(2), 634 (2023)
12. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *J. Artif. Int. Res.* **16**(1), 321–357 (2002)
13. Chen, D.S.: Hepatitis b vaccination: the key towards elimination and eradication of hepatitis B. *J. Hepatol.* **50**(4), 805–816 (2009)
14. Chen, T., et al.: Xgboost: extreme gradient boosting. *R Package Version 0.4-2*, vol. 1, no. 4, pp. 1–4 (2015)

15. Dorogush, A.V., Ershov, V., Gulin, A.: Catboost: gradient boosting with categorical features support. arXiv preprint [arXiv:1810.11363](https://arxiv.org/abs/1810.11363) (2018)
16. Garfield, R., Majerol, M., Damico, A., Foutz, J.: The uninsured: a primer. Key facts about health insurance and the uninsured in America. The Henry James Kaiser Family Foundation, Menlo Park (2016)
17. Gerlings, J., Jensen, M.S., Shollo, A.: Explainable AI, but explainable to whom? An exploratory case study of XAI in healthcare. In: Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects (2022)
18. Glanz, K., Bishop, D.B.: The role of behavioral science theory in development and implementation of public health interventions. *Annu. Rev. Public Health* **31**(1), 399–418 (2010)
19. Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., Turini, F.: Factual and counterfactual explanations for black box decision making. *IEEE Intell. Syst.* **34**(6), 14–23 (2019)
20. Guidotti, R., et al.: Stable and actionable explanations of black-box models through factual and counterfactual rules. *Data Min. Knowl. Discov.* 1–38 (2022)
21. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. *ACM Comput. Surv.* **51**(5), 1–42 (2018)
22. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328 (2008). <https://doi.org/10.1109/IJCNN.2008.4633969>
23. Hoffman, C., Paradise, J.: Health insurance and access to health care in the united states. *Ann. N. Y. Acad. Sci.* **1136**(1), 149–160 (2008)
24. Holzinger, A., Biemann, C., Pattichis, C.S., Kell, D.B.: What do we need to build explainable AI systems for the medical domain? arXiv preprint [arXiv:1712.09923](https://arxiv.org/abs/1712.09923) (2017)
25. Ke, G., et al.: Lightgbm: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
26. Li, L., Wood, C.E., Kostkova, P.: Vaccine hesitancy and behavior change theory-based social media interventions: a systematic review. *Transl. Behav. Med.* **12**(2), 243–272 (2021)
27. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
28. MacDonald, N.E.: Vaccine hesitancy: definition, scope and determinants. *Vaccine* **33**(34), 4161–4164 (2015)
29. Macedo, C.G.D.: Director’s letter: the defeat of polio. *Bull. Pan Am. Health Organ. (PAHO)* **27**(1), 1993 (1993)
30. Molnar, C.: Interpretable Machine Learning. A Guide for Making Black Box Models Explainable, chap. Properties of Explanations (2022)
31. Njoku, A., Joseph, M., Felix, R.: Changing the narrative: structural barriers and racial and ethnic inequities in COVID-19 vaccination. *Int. J. Environ. Res. Public Health* **18**(18), 9904 (2021)
32. Okoli, G.N., Abou-Setta, A.M., Neilson, C.J., Chit, A., Thommes, E., Mahmud, S.M.: Determinants of seasonal influenza vaccine uptake among the elderly in the united states: a systematic review and meta-analysis. *Gerontol. Geriatr. Med.* **5**, 233372141987034 (2019)
33. Panigutti, C., Beretta, A., Giannotti, F., Pedreschi, D.: Understanding the impact of explanations on advice-taking: a user study for AI-based clinical decision sup-

- port systems. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–9 (2022)
34. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: Catboost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS 2018, pp. 6639–6649. Curran Associates Inc., Red Hook (2018)
 35. Ribeiro, M.T., Singh, S., Guestrin, C.: “Why should i trust you?” explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2016)
 36. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: high-precision model-agnostic explanations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
 37. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
 38. Sharma, S., Sharma, R., Datta, A.: (Mis) leading the COVID-19 vaccination discourse on twitter: an exploratory study of infodemic around the pandemic. *IEEE Trans. Comput. Soc. Syst.* (2022)
 39. Shrikumar, A., Greenside, P., Kundaje, A.: Learning important features through propagating activation differences. In: International Conference on Machine Learning, pp. 3145–3153. PMLR (2017)
 40. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: International Conference on Machine Learning, pp. 3319–3328. PMLR (2017)
 41. Weston, D., Ip, A., Amlôt, R.: Examining the application of behaviour change theories in the context of infectious disease outbreaks and emergency response: a review of reviews. *BMC Public Health* **20**(1) (2020)
 42. Zacharias, J., von Zahn, M., Chen, J., Hinz, O.: Designing a feature selection method based on explainable artificial intelligence. *Electron. Mark.* **32**(4), 2159–2184 (2022)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

