Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Techniques for the Identification of Risk Factors Associated with Food Insecurity among Adults in Arab countries during the COVID-19 Pandemic

Radwan Qasrawi (radwan@staff.alquds.edu)
Al Quds University: Al-Quds University https://orcid.org/0000-0001-8671-7026
Maha Hoteit
Lebanese University: Universite Libanaise
Reema Tayyem
Qatar University College of Health Sciences
Khlood Bookari
Taibah University Faculty of Medicine
Haleama Al Sabbah
Zayed University Department of Natural Science and Public Health
Iman Kamel
Cairo Research Center
Somaia Dashti
Public Authority for Applied Education and Training
Sabika Allehdan
University of Bahrain College of Health Sciences
Hiba Bawadi
Qatar University College of Health Sciences
Mostafa Waly
Sultan Qaboos University
Mohammad Ibrahim
Mutah University
Stephanny Vicuna Polo
Al Quds University: Al-Quds University
Diala Abu Al-Halawa
Al Quds University School of Medicine: Al Quds University Faculty of Medicine

Research Article

Keywords: food insecurity1, COVID-192, food consumption score3, machine learning4, prediction5, Arab countries6

Posted Date: March 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2545270/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at BMC Public Health on September 16th, 2023. See the published version at https://doi.org/10.1186/s12889-023-16694-5.

Abstract

Background: A direct consequence of global warming, and strongly correlated with poor physical and mental health, food insecurity is a rising global concern associated with low dietary intake. The Coronavirus pandemic has further aggravated food insecurity among vulnerable communities, and thus has sparked the global conversation of equal food access, food distribution, and improvement of food support programs. This research was designed to identify the key features associated with food insecurity during the COVID-19 pandemic using Machine learning techniques. Seven machine learning algorithms were used in the model, which used a dataset of 32 features. The model was designed to predict food insecurity across ten Arab countries in the Gulf and Mediterranean regions. A total of 13,443 participants were extracted from the international Corona Cooking Survey conducted by 38 different countries during the COVID -19 pandemic.

Results: The findings indicate that Jordanian, Palestinian, Lebanese, and Saudi Arabian respondents reported the highest rates of food insecurity in the region (15.4%,13.7%,13.7% and 11.3% respectively). On the other hand, Oman and Bahrain reported the lowest rates (5.4% and 5.5% respectively). Our model obtained accuracy levels of 70%-82% in all algorithms. Gradient Boosting and Random Forest techniques had the highest performance levels in predicting food insecurity (82% and 80% respectively). Place of residence, age, financial instability, difficulties in accessing food, and depression were found to be the most relevant features associated with food insecurity.

Conclusions: Overall, ML algorithms seem to be an effective method in early detection and prediction of food insecurity. Future research would benefit from utilizing the proposed model in developing more complex and accurate models aiming to enhance granularity, with the ability to share data, to incorporate wide range of variables, and to make use of automation for effective prevention and intervention programs at the regional and individual levels.

Introduction

Food insecurity (FI) has been a major global health concern, particularly in low- and middle-income countries. FI is defined as a lack of access to sufficient and affordable food [1, 2]. Improper diet, unhealthy eating habits and lifestyles[3–5], limited food availability, lack of access to food, and proper food consumption to meet dietary needs have been evidenced as major factors that contribute to food insecurity [4, 6, 7] y. According to the State of Food Security and Nutrition in the World (SOFI) report, food insecurity can have a serious impact on people's health of all ages and may include several adverse health effects such as depression, diabetes, obesity, anxiety, and hypertension [8, 9]. Furthermore, the SOFI report indicated that the COVID-19 pandemic increased chronic hunger between 2019 and 2020[8]. The Global Humanitarian Overview report indicated that about 265 million people were in need of humanitarian assistance during the year 2021 due to the COVID-19 pandemic [8, 10]. Several studies reported that the COVID-19 pandemic significantly impacted global food insecurity and poverty, mainly in

low- and middle-income countries that have limited access to food, as they experienced high rates of unemployment, income inequality, and disruption of social safety programs[11–14].

In the Arab region, mainly in Jordan, Lebanon and Palestine, the pandemic has severely affected the countries' national economies, in particular the services sectors, the supply chain, the markets, and trade, all of which have directly impacted food supply, demand, and access [15–17]. Several studies reported that low- and middle-income countries were most affected by the COVID-19 pandemic, and thus presented an increase in food insecurity[8, 18]. A United Nations Economic and Social Commission for Western Asia (ESCWA) study reported that around 8.3 million people in the Arab region ran the risk of falling into poverty and income crises due to lockdowns, which increased the number of food-insecure and undernourished persons[19]. Furthermore, Kharroubi et al.'s study indicated that food insecurity among Lebanese increased from 27–36%-39% due to the pandemic[12]. Similarly, World Food Program (WFP) reports indicated that 53% and 34% of Jordanians and Palestinians are vulnerable to food insecurity respectively[20–22]. Although Gulf Cooperation Council (GCC) countries have been considered the most food-secure countries in the Arab world, the COVID-19 pandemic and the current global crises have affected food accessibility and caused food insecurity concerns in the GCC region[23, 24].

Food insecurity is an interdimensional concept that encompasses food availability, access to food, and food consumption diversity. Several indicators have been developed for the assessment and depiction of households' food insecurity. The Food Consumption Score (FCS) was used to measure the quantity and diversity of food consumption and has been considered a valid indicator for food insecurity[25–28]. Data mining and Machine Learning (ML) techniques have been used in several studies as efficient tools for predicting and identifying the risk factors associated with food insecurity [29–32]. K-Nearest Neighbor (KNN), Random Forest (RF), Logistic regression and Support Vector Machine (SVM) are among the machine learning models that have been assessed for the prediction of food insecurity[29, 30, 33].

Gao et al. examined the machine learning models used to identify vulnerable household characteristics and to predict food-insecure households. The models indicated that household size, income, and access to farm production and resources are risk factors associated with food insecurity[34]. In Georgiana et al.'s study, the machine learning models were used for identifying food insecurity in the food sharing network, the authors indicated that Random Forest and AdaBoost had higher prediction accuracies and produced a complex features structure that contributes to food insecurity[29]. In Doreswamy et al.'s study, the machine learning models were used in household food insecurity classifications, among them are the KNN, Logistic Regression (LR), RF, and SVM models; the RF reported the best accuracy for features classifications and importance with an accuracy rate of 99.98% [30].

Thus, this study aims to further investigate the performance of different machine learning techniques in predicting and identifying risk factors associated with food insecurity in low- and middle-income countries during the COVID-19 pandemic. Furthermore, we seek to identify the direct consequences of the COVID-19 pandemic on people's health and poverty indices in the region. The study utilized ML

algorithms to detect food insecurity and food secure groups based on the food consumption score, key features were extracted from the Regional Corona Cooking Survey dataset. Furthermore, the study determined the features most associated with food insecurity.

Methods

Aim

The study aimed to identify the key determinants of food insecurity within the Arab region during the COVID-19 pandemic through a novel approach, Machine Learning algorithms.

Setting

The “Corona Cooking Survey April 2020”[35] conducted by 38 different countries during the COVID – 19 pandemic was used for this study. The survey was distributed to over 300 participants per country in each of the following Arab countries: (Palestine, Lebanon, Jordan, Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates, and Bahrain. The survey data was collected by the research team between April 17th and June 25th, 2020.

Design

Data Source

The dataset was extracted from the international “Corona Cooking Survey April 2020”[35]. The survey was designed as a multi-language survey including the Arabic language to facilitate data collection among different countries. The data collection instrument assessed the effect of the COVID-19 lockdown on adults’ health and nutrition. The survey included several types of information, such as COVID-19 Lockdown Measures, COVID-19 Feelings (Kessler K6 scale), Food Literacy Scale, Shopping Experiences and Behavior, Cooking Behavior and Attitudes, Seeking Recipes & Food Content, Eating Behaviors (Food Frequency Questionnaire), Food Advice Sources, E-drinking and E-dining, anthropometric measurements, and lifestyle and eating habits. The survey was distributed through different social media platforms and through the partner universities’ networks. The final sample included over 300 participants per country. Countries that acquired a lesser number of samples were excluded from the analysis. The data relating to food insecurity in Arab countries were extracted from the International Survey. Overall, 13,443 participants aged over 18 years were included in the ML model development.

Study Features

The features were extracted from the main study variables, accessible via https://osf.io/nz9xf/files/ [35]. The study features contain respondents’ data from before and after the COVID-19 lockdown. The machine learning models used food insecurity as the main target variable for assessing the performance of machine learning in predicting and identifying associated risk factors. Food insecurity was defined in reference to the Food Consumption Score (FCS), which is an indicator used to assess dietary diversity

before and after the pandemic. The FCS was categorized into two groups: Low FCS if the FCS < 42 (Unacceptable), and high FCS score if the FCS scores > = 42 (Acceptable). Thus, food insecurity was determined based on the FCS classification. The low FCS group was considered a food insecure group, while the high FCS group was considered a normal group. Detailed information about the study variables' definition and calculation can be found in Hoteit et al.[36].

Data Preprocessing

Data underwent a preprocessing procedure prior to building the ML models. The preprocessing phase included data cleaning, formatting, missing data treatment, and data categorization. The data cleaning process included the null value data, the text-to-numeric conversion, and the missing data treatment.

In the data set, imbalanced data was encountered as 1529 participants were categorized as food insecure, and 11914 participants were categorized as food secure. An imbalanced data set might bias the ML model's estimation by providing more weight to the dominant class[26]. The simple and effective Min-Max Normalization technique was used to scale features to a common range to ensure that all the features were on the same scale, and to allow the model to make more accurate predictions. The minimum and maximum values for each feature were first calculated, followed by subtracting the minimum value from all feature values and dividing the resulting tables by the range (i.e., the difference between the maximum and minimum values) to ensure that all the features were scaled between 0 and 1.

Machine Learning Model

Seven machine learning models were used in this study to assess the performance of ML in predicting food insecurity among Arab countries following the COVID-19 pandemic. Logistic Regression (LR), Gradient Boosting (GB), Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), Naïve Bayes (NB), and k-nearest neighbors' algorithm (k-NN) were built and evaluated considering their performance measures. The models' features were categorized into two groups: 1) the dependent variable, represented by the food consumption score, and 2) the independent variables, including all associated features. The features are further represented in Table 1.

Table 1
The list of variables used in the machine learning model

Code	Name	Description	Value
1	Gender	Gender	1 = Male, 2 = Female
2	FCS_category	FCS Status	1 = Not Acceptable, 2 = Acceptable
3	FCS_variation	Variation of FCS during lockdown	1 = Decreased, 2 = The same, 3 = Increased
4	Region_of_living	Living region	1 = MENA, 2 = GULF
5	Country _Living	Arabic Countries	Bahrain, Egypt, Jordan, Palestine, Lebanon, Saudi Arabia, Emirate, Oman, Kuwait
6	Family_Income	Family Income Level	1 = Low, 2 = Average, 3 = High
7	Education	Education Level	>= Secondary, >Secondary
8	WatchTV	Number of Hours watching TV	1 = < One Hour, 2 = 1–2 Hours, 3 = 3–4 Hours, 5 = 5hours +
9	Computer use	Number of Hours using computer	1 = < One Hour, 2 = 1–2 Hours, 3 = 3–4 Hours, 5 = 5hours +
10	Employment	Employment status	1 = Employed, 2 = Unemployed
11	BMI	Body Mass Index (BMI)	1 = Normal, 2 = Overweight, 3 = Obese
12	AgeGroup	Age in years	1 = 18–23, 2 = 24–29, 3 = 30–39. 4 = 40+
13	Cooking_Money	Don't have Money for Cooking	1 = Yes, 2 = No
14	Access_Food	Don't have access to food for Cooking	1 = Yes, 2 = No
15	Cooking_Facilities	Don't have access to cooking tools for cooking	1 = Yes, 2 = No
16	Famiy_Size	Family Size	1= <= 5 persons, 2 = > 5 Person
17	Financial_Problems	Financial difficulties until the end of the month	> 5 per
18	Financial_Shopping	Financial difficulties in Shopping	1 = Yes, 2 = No
19	PHA	Physical Activity during lockdown	1 = Low Active, 2 = Moderate, 3 = Highly Active
20	Smoking	Smoking Before COVID-19	1 = Yes, 2 = No

Code	Name	Description	Value
21	Mother_SHOP_3	Mother usually did food shopping	1 = Yes, 2 = No
22	Father_SHOP_4	Father usually did food shopping	1 = Yes, 2 = No
23	Hopeless	I feel hopeless	1 = Yes, 2 = No
24	restless	I feel restless or fidgety	1 = Yes, 2 = No
25	require_efforts	I feel that everything requires effort	1 = Yes, 2 = No
26	worthless	I feel worthless	1 = Yes, 2 = No
27	nervous	I feel nervous	1 = Yes, 2 = No
28	depressed	I feel so depressed that nothing could cheer me up	1 = Yes, 2 = No
29	Moretime_activity	I feel I have more time than usual in doing activities	1 = Yes, 2 = No
30	struggle_financially	I feel I struggle financially	1 = Yes, 2 = No
31	connected	I feel more connected than usual	1 = Yes, 2 = No
32	lockdown_duration	Lockdown duration (Weeks)	1 = < 12 weeks, 2 > = 12 weeks

The ML models were built based on the data ratio of 70:20:10 for training, testing, and validation. The grid search method and cross-validation with 10-folds were used for parameters' optimizations. The following parameters were set for the ANN, RF, and SVM:

- In Artificial Neural Networks, the hidden layer had 1000 neurons, with a 600-maximum number of iterations in reference to the logistic activation function.
- The Random Forest trees were set to 1000 with 5 maximum depth trees, and the leaf node minimum number was set to 1, while the maximum number of samples to split the internal nodes was set to 2.
- The SVM regularization parameter was set to 10, the RFB kernel was set to 0.001, and the bias error control factor was set to 1.

Based on the parameters optimization results, the optimized algorithms (ANN, SVM, RF) were used in identifying and predicting food insecurity.

Data Analysis

Data cleaning, transformation, and normalization processes were conducted prior to building the ML data analysis. The final dataset consisted of 13,446 participants. The seven ML models were built and performed using the python orange data mining software[37], which was then used for testing and validating the machine learning models.

Different performance measures were used to evaluate whether the ML models are able to predict food insecurity levels and the associated risk factors, such as accuracy, specificity, precision, recall, and F-measure. The calculating equations for performance measures are as follows:

$$Specificity = \frac{TN}{FP + TN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$Precision = \frac{TP}{TP + FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$Recall = \frac{TP}{TP + FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$F - Measure = \frac{2 * precision * recall}{precision + recall} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Results

Statistical Analysis

The food consumption score (FCS) was used to identify the food insecure (FI) and food secure (FS) participants. The borderline and not acceptable FCS group were considered food insecure, while those with acceptable FCS were considered food secure. Overall, 9.3% out of 13,443 participants reported low and borderline FI rates among Arab countries. Results in Fig. 1 show the distribution of FI and FS by country. The findings indicate that the Jordanian, Palestinian, Lebanese, and Saudi Arabian respondents reported the highest FI rates in the region (15.4%,13.7%,13.7% and 11.3% respectively). On the other hand, Oman and Bahrain reported the lowest FI rates (5.4% and 5.5% respectively).

The ML models used a balanced data set of 4,259 participants extracted from the general data set of 13,443 participants. The connectivity between the associated features and food insecurity was described by the correlation matrix as illustrated in Table 2. The correlation matrix compares the target variable (FCS) with the study features and identifies which features are most correlated with the outcome variable. The results indicated that most of the study features are significantly correlated with the outcome feature except for five features: “Family Size”, “Smoking”, “I feel hopeless”, “I feel that everything requires effort”, “I feel worthless”, “I feel so depressed”, “Food shopping by mother”, “Food shopping by father”, and the

"Lockdown period". The highest correlations were found between the region and Body Mass Index (BMI) (correlation values = 0.202 and 0.143 respectively).

Table 2
Correlation between study features and food consumption score

Features	Correlation-Value	Features	Correlation-Value
Gender	− .090**	Use Computer	.149**
BMI	.143**	Education Level	− .057**
Region	.202**	I feel hopeless	0.017
Don't have money for cooking	.090**	I feel restless or fidgety	.046**
Don't have access to food for cooking facilities	.103**	I feel that everything requires effort	0.023
Don't have access to cooking tools for cooking	.110**	I feel worthless	-0.017
Family Size	-0.029	I feel nervous	.037*
Facing Financial difficulties to the end of the month	.035*	I feel so depressed	0.012
Financial difficulties in Shopping	0.014*	I feel I have more time than usual	.045**
Physical Activity during lockdown	.075**	I feel I struggle financially	0.04*
Smoking During COVID-19	0.004	I feel more connected than usual	.063**
Food shopping by mother	0.003	Lockdown duration (Weeks)	0.015
Food shopping by father	0.005	Age (Year)	− .375**
Watch TV	.071**	Country of living during COVID19	− .139**
*P< 0•05, **P< 0•01			

Machine Learning Models Performance Analysis

Table 3 depicts the performance measures in predicting participants' FI, their accuracy, sensitivity, specificity, F1-score, and Receiver Operating Characteristic (ROC) classification curve. The accuracy score presents the crossover over the accuracy of the model, sensitivity measures the segment of FI participants correctly predicted, while specificity shows the identified segment of FS participants. According to the models' results, the overall accuracy in predicting food insecurity ranged from 70–82%. Sensitivity ranged from 72–81.6%, while the f1-score ranged from 70.3–80.5%. Comparing the different

ML algorithms, results indicated that GB and RF reported the highest accuracy rates (81.6% and 81.4% respectively), while the KNN algorithm had the lowest accuracy rate.

Table 3
FI ML prediction performance measures

#	Model	AUC	CA	F1	Precision	Recall
1	Gradient Boosting	85	82	83	83	82
2	Random Forest	84	82	80	83	82
3	Logistic Regression	84	80	82	81	81
4	Support Vector Machine	83	80	83	81	81
5	Neural Network	84	80	81	81	81
6	Naive Bayes	80	76	77	77	77
7	K-nearest neighbor	72	70	71	72	72

Figures' Legends

The results in Table 3 illustrate that the algorithms' performance is determined according to the performance measure. Thus, GB had the highest accuracy, F1-score, sensitivity, and AUC scores. However, the other algorithms reported acceptable performance measures and can be used in predicting food insecurity, except KNN and NB, which had the lowest performance levels and therefore are not recommended. The models' performance was further evaluated using the Area Under the Curve-Receiver Operating Characteristics curve (AUC- ROC). The food consumption scores were classified into two categories: FI and FS. The ROC was obtained for the FI as indicated in Fig. 2. Results in Fig. 2 illustrate that the ROC characteristics are in the upper left side of the curve for the gradient boosting and logistic regression models, thus the two models reported the highest accuracy rates (AUC of 82%, and 80% respectively).

The SHapley Additive exPlanations (SHAP) values analysis was used to identify the importance of study features in predicting the outcome variable and associated features. The SHAP value analysis was **conducted** on the GB model as it presented the best FI prediction accuracy levels. Figure 3 illustrates the correlation plot of FI participants with model predictors. The y-axis shows the FI predictors, while the x-axis shows the SHAP value. Based on the results in Fig. 3, the most important features that positively affect FI show to be (1) country of residence, (2) age, (3) financial difficulties in food shopping, (4) depression, (5) having financial problems, (6) don't have access to food. The place of residence (living country) had the highest positive significant impact on the outcomes of FI participants. On the other hand, BMI, physical activity, smoking, food shopping by the father, and food shopping by the mother had a negative impact on the outcome of FI participants. Notably, age, depression, and feeling nervous were found to be relevant factors that play a significant role in predicting respondents' food insecurity.

Discussion

Food insecurity and food inequity have been increasingly relevant during and following the COVID-19 pandemic [11, 13, 27]. Millions of households around the world relied on food support programs during COVID-19[38]. The lockdown-imposed work and movement limitations that significantly increased the risk of food insecurity at the global level[38][14, 39, 40]. In this paper, we used machine learning models to predict food insecurity using a dataset containing relevant data from 10 Arab countries during the COVID-19 pandemic. The models utilized the food consumption score (FCS) as a key variable in identifying food insecure (FI) and food secure (FS) participants as it has been used in several studies as an indicator for the prediction and identification of household FI [26, 41, 42].

This study has found that some of the features used are associated to food insecurity. Nonetheless, limited studies have used ML algorithms in identifying and predicting food insecurity, particularly in the Arab region and during the COVID-19 pandemic. The results showed significant FI levels in Arab countries, mainly in Jordan, Palestine, and Lebanon. Similarly, significantly high FI levels were evidenced in Saudi Arabia and Kuwait in the Gulf states. Furthermore, the study showed a significant relationship between the country of residence, age, and participants' mental health status with FI levels. These findings are consistent with other similar studies that indicated a significant association between household FI and age, place of residence, and other sociodemographic variables [12, 43–47]. Interestingly, we found that BMI had a strong correlation and association with FI. The negative association between BMI and FCS indicated that food-insecure people had higher BMI scores. This finding was consistent with other similar studies[6, 48, 49].

The implementation of the ML models demonstrates the power of ML in predicting FI from the associated factors. The performance analysis included the accuracy, recall, precision, F1-score, and ROC measures. The results showed that GB and RF are among the highest-performing ML models in predicting FI. However, the other employed models showed an acceptable performance rate, while KNN reported the lowest accuracy rate of all. Our results were found to be consistent with other studies that used ML in predicting FI and confirmed the feasibility of ML use in identifying FI [41, 50–52]. The models in this study reported accuracy rates ranging from 70–82%, while other models reviewed from previous research reported accuracy rates ranging from 55–85%. Other studies found that RF and SVM had high performance rates in predicting FI [34]. Furthermore, the ML models identified a significant association between mental health factors and FI outcomes, including depression, stress, hopelessness, and negative feelings. The findings were consistent with other studies that showed a significant association between FI outcomes and depression, anxiety, despair, and hopelessness [38, 53].

In this study, the importance of each feature on FI prediction was examined using the SHAP values analysis[51]. The country of residence, age, financial difficulties in shopping, depression, inaccessibility to food, and financial problems were found to be the most important features affecting the model outcomes. Our findings were consistent with other studies indicating that food prices, poor access to

food and market, poverty, living place, and lack of education were associated with food insecurity. Notably, depression, nervous feelings and age were among the most important factors associated with FI outcomes [41, 51, 52].

The pandemic encouraged the consideration of food access equity at the global level, mainly with regards to ensuring the availability of basic food, housing, health, and education. Our approach introduced an important contribution to improving the currently available methods of predicting and early warning of food insecurity as it has proven effective in identifying and predicting food insecure people from food consumption levels. The development of this model was based on international data from Arab countries to ensure its replicability at the global level, among marginalized and conflict-prone areas.

Strengths and Limitations

The pandemic calls for a shift in thinking and consideration of access to food and ensuring food security among low- and middle-income countries. This model was designed to respond to the need of early detection of food insecurity to ensure rapid and accessible humanitarian response. The model provides a precise and accurate FI identification and prediction tool that is less dependent on traditional assessment and analysis. Thus, the model improves automatic and early detection of food insecurity, which in turn can enhance rapid intervention and policy-making programs for combatting food insecurity. Thus, this research study not only introduces ML techniques in predicting FI during and after pandemic conditions, but also enhances data driven decision making and early intervention.

Nonetheless, the study was limited by the number of features used in predicting FI. In some countries, the model data set presented a lower number of participants, which limited the ability of building ML models per country. The research findings encourage future research by building more complex ML models that might improve prediction and classifications accuracy.

Conclusions

The study assessed the performance measures of seven ML algorithms in predicting the risk factors associated with food insecurity during the COVID-19 pandemic. The research found that Gradient Boosting and Random Forest are among the highest-performing Machine Learning algorithms for the accurate prediction of food insecurity. The study developed an ML model that can enhance the early detection of FI, and that can be replicated in other regions. The study contributes to the literature on food insecurity based on FCS by utilizing the ML methods in identifying the key characteristics of food insecurity in the Arab region, to determine the relationship between the food consumption score and the associated factors, and to support policy makers with advanced food insecurity identification and prediction tools. Furthermore, the use of ML models is a valuable tool in improving food insecurity prediction and detection over time, with enhanced granularity, with the ability to share data, to incorporate wide range of variables, and to make use of automation for effective prevention and intervention programs at the regional and individual levels.

Abbreviations

FI: Food insecurity

SOFI: State of Food Insecurity and Nutrition in the World

COVID-19: Coronavirus disease 2019

ESCWA: Social Commission for Western Asia

WFP: World Food Program

GCC: Gulf Cooperation Council

FCS: Food Consumption Score

ML: Machine Learning

KNN: k-Nearest Neighbor

RF: Random Forest

SVM: Support Vector Machine

LR: Logistic Regression

NB: Naïve Bayes

ANN: Artificial Neural Network

FS: Food Secure

BMI: Body Mass Index

ROC: Receiver Operating Characteristics

AUC: Area under the curve

Declarations

Ethics approval and consent to participate: This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving research study participants were approved by the University of Antwerp (SHW_46). Written consent was obtained from all subjects/patients at the start of the study.

Consent for publication: Not applicable

Availability of data and materials: The datasets generated and/or analysed during the current study are available in the Corona Survey repository, https://osf.io/nz9xf/files/.

Competing interests: The authors declare that they have no competing interests

Funding: This project did not receive any external funds.

Authors' contributions: CB, LT, IC, PD, SP, and KR: conceptualization, software, validation, supervision, project administration, and funding acquisition. CB, LT, IC, PD, SP, KR, MH, HM, AJ-J, RM, MA, KB, RT, HA, LC, RQ, RA, IK, SD, SA, MA-M, HB, MW: methodology. RQ, DA, and SV: Design the ML model, formal analysis, data curation and writing—original draft preparation. RQ, RT, HA, and MH: investigation. All authors and the research group contributed to the article and approved the submitted version.

Acknowledgements: The authors wish to thank the research participants for sharing their invaluable information. Furthermore, this paper and the research behind it would not have been possible without the support of the Corona Cooking Survey Regional Study Group and all its members, as listed below:

Belgium: Charlotte De Backer, Lauranna Teunissen, Kathleen Van Royen, Isabelle Cuykx, Paulien Decorte, Gaëlle Ouvrein, Karolien Poels, Heidi Vandebosch, Katrien Maldoy (University of Antwerp); Sara Pabian (Tilburg University), Christophe Matthys, Tim Smits, Jules Vrinten (KULeuven); Ann DeSmet (University of Antwerp, Université Libre de Bruxelles); Nelleke Teughels, Maggie Geuens, Iris Vermeir, Viktor Proesmans, Liselot Hudders (Ghent University);

Bahrain: Mariam Al-Mannai, Tariq Alalwan (University of Bahrain);

Lebanon: Elissa Naim (Lebanese University), Rania Mansour (Lebanese

University), Nour Yazbeck (Lebanese University);

Palestine: Hazem Agha, Rania Abu Seir (Al Quds University);

Saudi Arabia: Jamila Arrish, Ghadir Fallata, Omar Alhumaidan, Shihana Alakeel, Norah AlBuayjan, Sarah Alkhunein, Budur Binobaydan, and Aeshah Alshaya (National Nutrition Committee (NNC) at Saudi Food and Drug Authority (Saudi FDA);

United Arab Emirates: Ayesha Aldhaheri (United Arab Emirates University).

References

1. Hawkins M & Panzera A (2021) Food insecurity: A key determinant of health. *Arch Psychiatr Nurs* **35**, 113–117. Elsevier Inc.
2. Schroeder K & Smaldone A (2015) Food Insecurity: A Concept Analysis. *Nurs Forum (Auckl)* **50**, 274–284.

3. Flores HL & Amiri A (2019) CE: Addressing Food Insecurity in Vulnerable Populations. *American Journal of Nursing* **119**, 38–45.
4. Brown AGM, Esposito LE, Fisher RA, et al. (2019) Food insecurity and obesity: Research gaps, opportunities, and challenges. *Transl Behav Med* **9**, 980–987.
5. Frongillo EA & Bernal J (2014) Understanding the Coexistence of Food Insecurity and Obesity. *Curr Pediatr Rep* **2**, 284–290.
6. Frongillo EA & Bernal J (2014) Understanding the Coexistence of Food Insecurity and Obesity. *Curr Pediatr Rep* **2**, 284–290.
7. Jomaa L, Naja F, Cheaib R, et al. (2017) Household food insecurity is associated with a higher burden of obesity and risk of dietary inadequacies among mothers in Beirut, Lebanon. *BMC Public Health* **17**, 1–14. BMC Public Health.
8. Hoteit M, Al-Atat Y, Joumaa H, et al. (2021) Exploring the impact of crises on food security in lebanon: Results from a national cross-sectional study. *Sustainability (Switzerland)* **13**, 1–24.
9. Haskett ME, Kotter-Grühn D & Majumder S (2020) Prevalence and correlates of food insecurity and homelessness among university students. *J Coll Stud Dev* **61**, 109–114.
10. Doreswamy D & Nigus M (2020) Feature Selection Methods for Household Food Insecurity Classification. *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*, 1–7.
11. Niles MT, Bertmann F, Belarmino EH, et al. (2020) The early food insecurity impacts of covid-19. *Nutrients* **12**, 1–23.
12. Kharroubi S, Naja F, Diab-El-harake M, et al. (2021) Food insecurity pre-and post the covid-19 pandemic and economic crisis in lebanon: Prevalence and projections. *Nutrients* **13**, 1–15.
13. Kharroubi S, Naja F, Diab-El-harake M, et al. (2021) Food insecurity pre-and post the covid-19 pandemic and economic crisis in lebanon: Prevalence and projections. *Nutrients* **13**, 1–15.
14. Niles M, Bertmann F, Belarmino E, et al. The early food insecurity impacts of covid-19. Nutrients [revista en Internet] 2020 [acceso 2 de noviembre de 2021]; 12(7): 1-23. .
15. Economic UN & for Western Asia (ESCWA) SC (2020) Mitigating the impact of Covid19_Poverty and food insecurity in the Arab region. 1–5.
16. Economic UN & for Western Asia (ESCWA) SC (2020) Regional Emergency Response to Mitigate the Impact of COVID-19. *Covid Mptf*.
17. World Bank, IFAD, FAO, et al. (2020) Jordan Food Security Update-Implications of COVID-19. https://www.wfp.org/publications/jordan-food-security-update-implications-covid-19-july-aug-2020.
18. Smith MD & Floro MS (2020) Food insecurity, gender, and international migration in low- and middle-income countries. *Food Policy* **91**, 101837. Elsevier.
19. ESCWA (2020) to Mitigate the Impact of Introduction. *Covid Mptf*.
20. Of THES (2020) FAO, IFAD, UNICEF, WFP and WHO. In Brief to The State of Food Security and Nutrition in the World 2020. *Transforming food systems for affordable healthy diets*, 320.

21. Nica-Avram G, Harvey J, Smith G, et al. (2021) Identifying food insecurity in food sharing networks via machine learning. *J Bus Res* **131**, 469–484. Elsevier Inc.
22. Haskett ME, Kotter-Grühn D & Majumder S (2020) Prevalence and correlates of food insecurity and homelessness among university students. *J Coll Stud Dev* **61**, 109–114.
23. Hassen T Ben & Bilali H El (2019) Food Security in the Gulf Cooperation Council Countries : Challenges and Food Security in the Gulf Cooperation Council Countries : Challenges and Prospects. .
24. Jensen SD The COVID-19 Coronavirus Pandemic and its Impact in the Gulf1. In *Security Dynamics in the Gulf and the Arabian Peninsula*, pp. 214–232. Routledge.
25. Lentz EC, Michelson H, Baylis K, et al. (2019) A data-driven approach improves food insecurity crisis prediction. *World Dev* **122**, 399–409. Elsevier Ltd.
26. Hossain M, Mullally C & Asadullah MN (2019) Alternatives to calorie-based indicators of food security: An application of machine learning methods. *Food Policy* **84**, 77–91. Elsevier.
27. Hoteit M, Mortada H, Al-Jawaldeh A, et al. (2022) Dietary Diversity in the Eastern Mediterranean Region Before and During the COVID-19 Pandemic: Disparities, Challenges, and Mitigation Measures. *Front Nutr* **9**.
28. Martini G, Bracci A, Riches L, et al. (2022) Machine learning can guide food security efforts when primary data are not available. *Nat Food* **3**, 716–728. Springer US.
29. Nica-Avram G, Harvey J, Smith G, et al. (2021) Identifying food insecurity in food sharing networks via machine learning. *J Bus Res* **131**, 469–484. Elsevier Inc.
30. Doreswamy D & Nigus M (2020) Feature Selection Methods for Household Food Insecurity Classification. *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*.
31. Doreswamy D & Nigus M (2020) Feature Selection Methods for Household Food Insecurity Classification. *2020 International Conference on Computer Science, Engineering and Applications, ICCSEA 2020*.
32. Balestier M, Cui M, Nandakumar R, et al. (2021) Regression Based Machine Learning to Generate and Validate a Metric for Food Insecurity. *2021 IEEE MIT Undergraduate Research Technology Conference, URTC 2021*, 1–5. IEEE.
33. Okori W & Obua J (2011) Machine learning classification technique for famine prediction. *Proceedings of the World Congress on Engineering 2011, WCE 2011* **2**, 991–996.
34. Gao C, Fei CJ, McCarl BA, et al. (2020) Identifying vulnerable households using machine-learning. *Sustainability (Switzerland)* **12**, 1–18.
35. Teunissen L, de Backer C & Dobson D et. al. (2020) Supplementary materials CoronaCookingSurvey. https://osf.io/nz9xf (accessed November 2022).
36. Hoteit M, Mortada H, Al-Jawaldeh A, et al. (2022) Dietary Diversity in the Eastern Mediterranean Region Before and During the COVID-19 Pandemic: Disparities, Challenges, and Mitigation Measures. *Front Nutr* **9**.

37. Demšar J, Curk T, Erjavec A, et al. (2013) Orange: data mining toolbox in Python. *the Journal of machine Learning research* **14**, 2349–2353. JMLR. org.
38. Hawkins M & Panzera A (2021) Food insecurity: A key determinant of health. *Arch Psychiatr Nurs* **35**, 113–117. Elsevier Inc.
39. Preis H, Mahaffey B, Heiselman C, et al. (2020) Vulnerability and resilience to pandemic-related stress among U.S. women pregnant at the start of the COVID-19 pandemic. *Soc Sci Med* **266**, 113348. Elsevier Ltd.
40. Preis H, Mahaffey B, Heiselman C, et al. (2020) Pandemic-related pregnancy stress and anxiety among women pregnant during the coronavirus disease 2019 pandemic. *Am J Obstet Gynecol MFM* **2**, 100155. Elsevier Inc.
41. Zhou Y, Lentz E, Michelson H, et al. (2022) Machine learning for food security: Principles for transparency and usability. *Appl Econ Perspect Policy* **44**, 893–910.
42. Hoteit M, Mortada H, Al-Jawaldeh A, et al. (2022) Dietary Diversity in the Eastern Mediterranean Region Before and During the COVID-19 Pandemic: Disparities, Challenges, and Mitigation Measures. *Front Nutr* **9**.
43. Hamad HJ & Khashroum A (2016) Household food insecurity (HFIS): Definitions, measurements, socio-demographic and economic aspects. *Journal of Natural Sciences Research* **6**, 63-75–75.
44. Bawadi HA, Tayyem RF, Dwairy AN, et al. (2012) Prevalence of food insecurity among women in northern Jordan. *J Health Popul Nutr* **30**, 49–55.
45. Long MA, Gonçalves L, Stretesky PB, et al. (2020) Food insecurity in advanced capitalist nations: A review. *Sustainability (Switzerland)* **12**, 1–19.
46. Broussard NH (2019) What explains gender differences in food insecurity? *Food Policy* **83**, 180–194. Elsevier.
47. Long MA, Gonçalves L, Stretesky PB, et al. (2020) Food insecurity in advanced capitalist nations: A review. *Sustainability (Switzerland)* **12**, 1–19.
48. Franklin B, Jones A, Love D, et al. (2012) Exploring mediators of food insecurity and obesity: A review of recent literature. *J Community Health* **37**, 253–264.
49. Jomaa L, Naja F, Cheaib R, et al. (2017) Household food insecurity is associated with a higher burden of obesity and risk of dietary inadequacies among mothers in Beirut, Lebanon. *BMC Public Health* **17**, 1–14. BMC Public Health.
50. Nica-Avram G, Harvey J, Smith G, et al. (2021) Identifying food insecurity in food sharing networks via machine learning. *J Bus Res* **131**, 469–484.
51. Ali Meerza SI & Ahamed A (2021) Food Insecurity Through Machine Learning Lens: Identifying Vulnerable Households. *Selected paper for presentation at the 2021 Agricultural and Applied Economics Association Annual Meeting*, 1–11.
52. Westerveld J, Stuit S, Heemskerk S, et al. (2019) Modelling Food Insecurity in Ethiopia Towards a machine learning model that predicts the transitions in food security using scalable features . 1–70.

53. Carter MA nn, Dubois L & Tremblay MS (2014) Place and food insecurity: a critical review and synthesis of the literature. *Public Health Nutr***17**, 94–112.

Figures

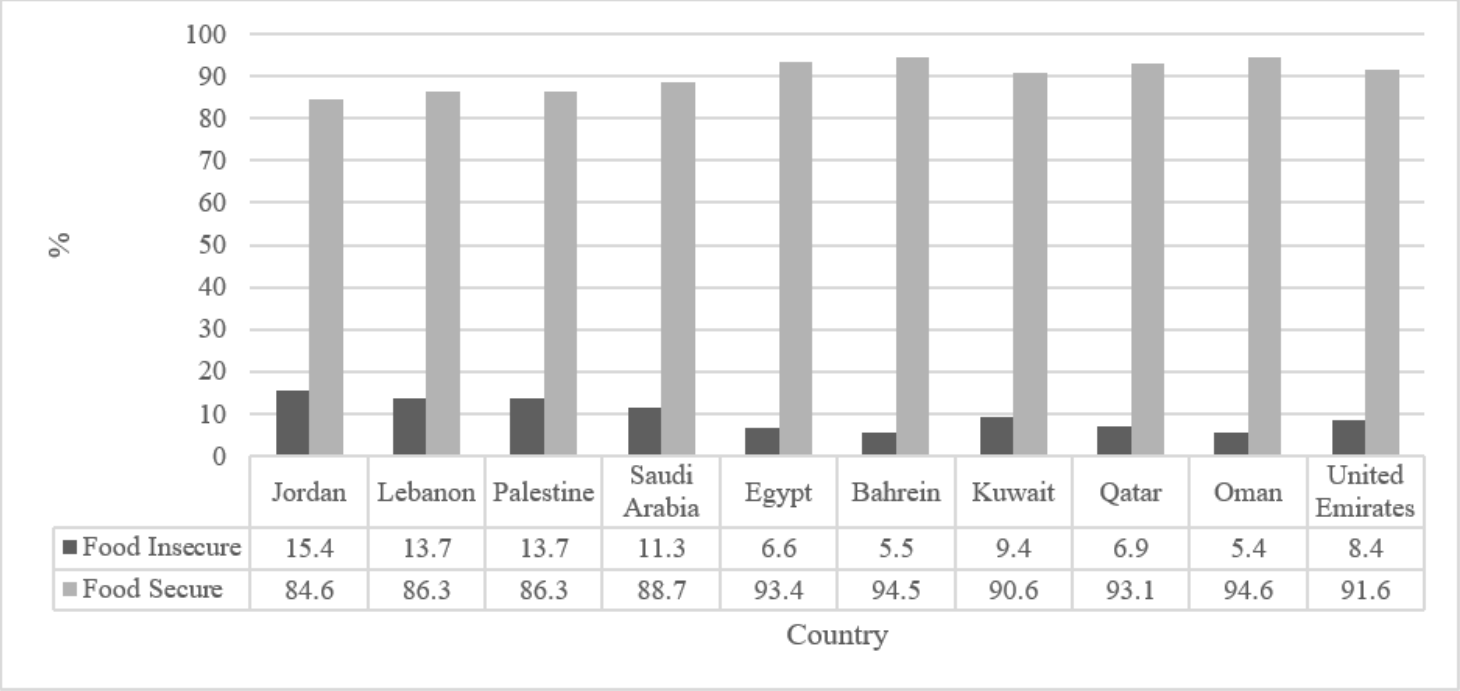


Figure 1

Food Consumption Score distribution by country

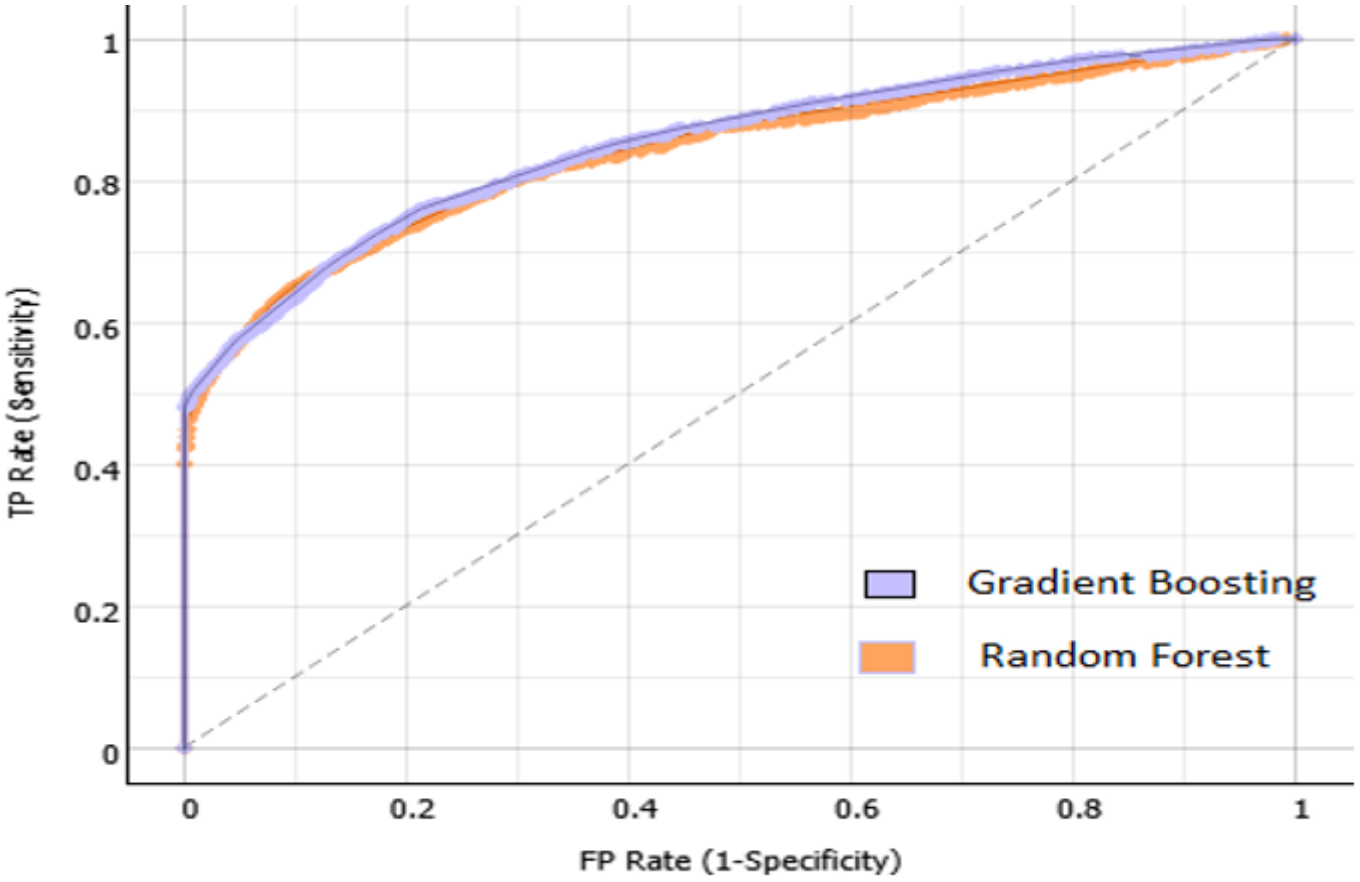


Figure 2

Gradient Boosting and Logistic Regression ROC curve for below average cognitive scores (TP rate of sensitivity against FP rate of specificity)

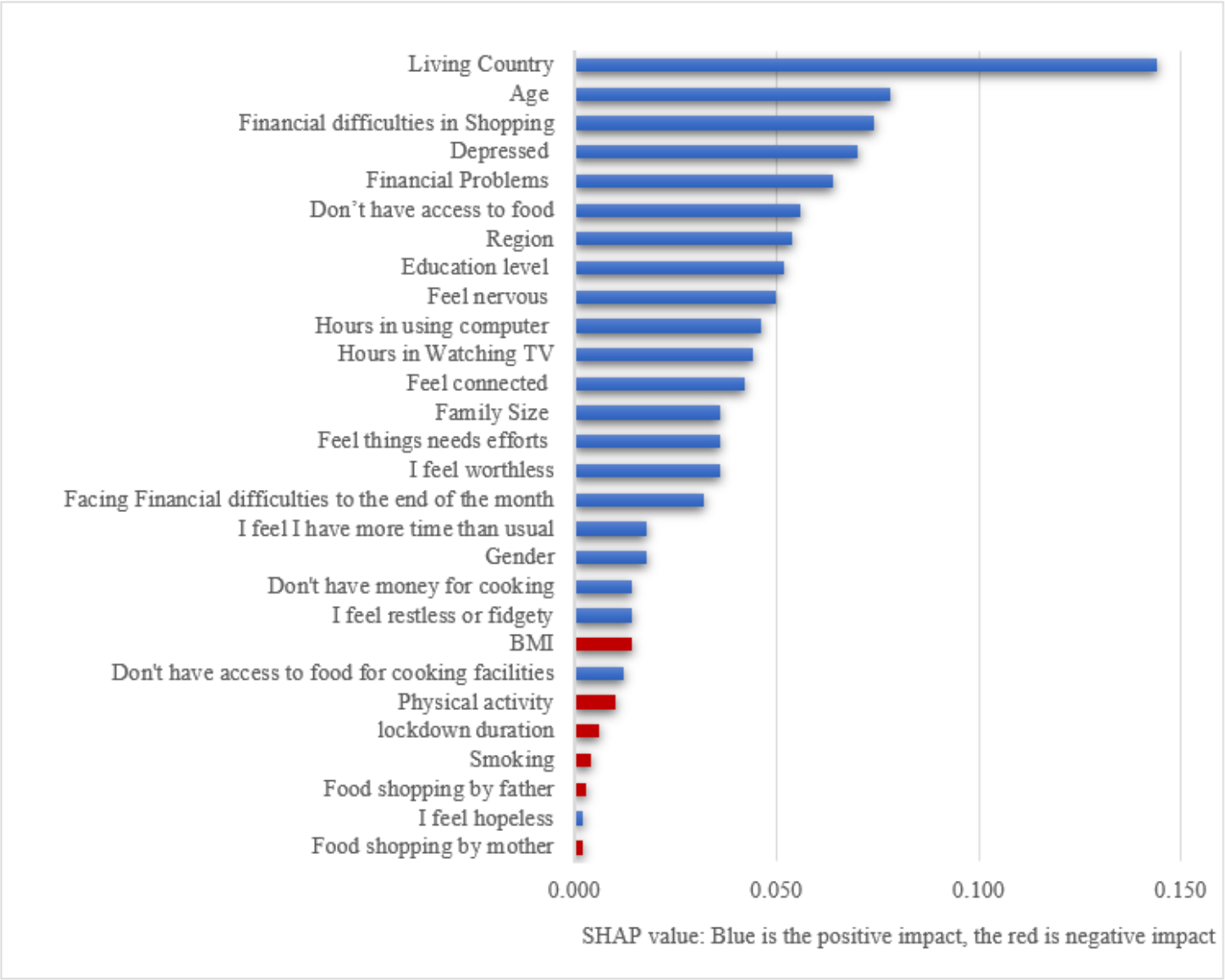


Figure 3

Correlation of Food Insecurity with the study features