FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 2 | Pages 308-311 | e-ISSN: 2347-470X

A Comparison of Machine Learning and Deep Learning Models for Predicting Household Food Security Status

Mersha Nigus[1] and H.L Shashirekha[2]

[1,2]Department of Computer Science, Mangalore University, Mangalore, India

*****Correspondence:** Mersha Nigus**:** mbezadtu@gmail.com

ABSTRACT- ML and DL algorithms are becoming more popular to predict household food security status, which can be used by the governments and policymakers of the country to provide a food supply for the needy in case of emergency. ML models, namely: k-Nearest Neighbor (kNN), Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), Multi-Layer Perceptron (MLP) and DL models, namely: Artificial Neural Network (ANN) and Convolutional Neural network (CNN) are investigated to predict household food security status in Household Income, Consumption and Expenditure (HICE) survey data of Ethiopia. The standard evaluation measures such as accuracy, precision, recall, F1-score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are used to evaluate the models' predictive performance, and the experimental results reveal that ANN, a DL model surpassed the ML classifiers with an accuracy of 99.15%.

Keywords: Deep Learning, Machine Learning, Food insecurity

ARTICLE INFORMATION

Author(s): Mersha Nigus and H.L Shashirekha
Received: 12/04/2022; **Accepted**: 27/06/2022; **Published**: 30/06/2022;
e-ISSN: 2347-470X;
Paper Id: IJEER220413;
Citation: 10.37391/IJEER.100241
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-10/ijeer-100241.html

crossref member
CROSSREF.ORG
THE CITATION LINKING BACKBONE

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Food insecurity is a concern in any society that is growing and becoming a significant problem due to manufactured wars, climate change, and economic downturns. In any place around the globe, food insecurity problems are becoming more frequent and more severe. Predicting the state of food insecurity is essential to take early intervention action, such as food distribution and other resource distribution handled by humanitarian organizations [Westerveld et al., 2021].

A large number of individuals have been malnourished for at least the last 15 years. Despite significant reductions in the proportion of undernourished persons, from 69% in 1994/95 to 35% in 2013/14 (FAO-FSI, 2014), the number still remains high. In 2018, it was estimated that more than 821 million individuals throughout the globe were food insecure.

The population of Sub-Saharan Africa accounted for 239 million food-insecure people, and the statistics coming out of Ethiopia, Kenya, Somalia, and South Sudan are particularly alarming. For many years, food insecurity has been a significant source of worry in Ethiopia [Ayenew, 2017], surpassing all other countries in Sub-Saharan Africa regarding aid. Ethiopia's food insecurity is caused by a complex interplay of several causes, including climate change, civil strikes, natural catastrophes, and societal norms [Alelign et al., 2021].

Faster action in times of food crisis may save lives and resources. Food aid and humanitarian assistance can not be provided unless the extent and breadth of these emergencies are accurately and appropriately identified. On the other hand, policymakers often lack the knowledge of people for whom the food aid has to be distributed [Zhou and Baylis, 2019]. Household income Consumption and Expenditure Surveys (HICE) are complex surveys conducted on a nationally representative sample to describe important aspects of household socioeconomic conditions, including food acquisition and consumption. It primarily reflects the income aspect of the household and gives information on households' income, consumption, and expenditures. This study uses the HICE survey data of Ethiopia to predict household food security status using the ML and DL models.

The remainder of the paper is organized as follows: A literature review is presented in section 2, followed by methodology in section 3, experiments and findings in section 4, and finally, the conclusion and future research in section 5.

2. RELATED WORKS

Several researchers have previously undertaken a substantial study on predicting household food insecurity. [Meerza et al., 2021] used Bangladesh's Integrated Household Survey (BIHS), the most comprehensive nationally representative rural household data, to obtain household-level data on the dynamics of poverty, food security, and agricultural development in Bangladesh. This survey provides household consumption information of 321 food items received via own production, purchase, and other sources for seven days. The authors used a food consumption table to translate food intake into calories and measured the calorie intake per family member per day by dividing the household calorie intake per day by the household size. By considering the fact that the recommended dietary need varies by age and gender, they estimated an average requirement of 2,122 calories per household member per day. A household is identified as food

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 2 | Pages 308-311 | e-ISSN: 2347-470X

insecure when the per capita calorie intake falls below 2,122 per day per person. According to BIHS data, while 4,461 households are food insecure, the remaining 1,955 are food secure. Experiments were conducted using SVM, Random Forest (RF), and LR classifiers to predict household food insecurity, and RF, SVM, and LR classifiers obtained accuracies of 83.3%, 71.1%, and 69.6%, respectively.

[Barbosa and Nelson, 2016] used a two-stage stratified random sampling to collect household survey data from 80 households belonging to six distinct municipalities, identifying 480 households. Out of 75 features, they used 14 features to train an SVM classifier to classify household data into food secure and insecure categories and obtained 77% accuracy and 84% recall. [Gao et al., 2020] identified household indicators that separate food-insecure families from food-secure ones, allowing for more accurate targeted aid to the latter. 2008, Afghanistan National Risk and Vulnerability Assessment (NRVA) survey data included 20,511 randomly selected households nationwide. They trained Decision Tree (DT) and RF models to predict food secure and food-insecure families, and the RF model obtained an accuracy of 80%. The results demonstrate that the six variables: (i) income and expenditure items, (ii) household size, (iii) farm-related measures, (iv) access to particular resources, (v) short-term shocks, and (vi) dwelling wall composition are the critical determinants of food security level. [Alsharkawi et al., 2021] employed ML algorithms to predict household food poverty in Jordan, using the dataset that contained the details of 63,211 families with 47 unique features. Out of 16 ML approaches used to predict household food poverty, the Light Gradient Boosting method (LightGBM) and Bagged Decision Trees outperformed the others with an F1 score of 81%.

Research conducted by [Okori and Obua, 2011] for the prediction of food security has used Uganda Bureau of Statistics secondary data for two agricultural seasons on agricultural households in four regions of Uganda. The dataset comprises 3,030 instances with 24 features, though the authors used only 15 of them in their experiments. Nine out of 24 were not included in the classification because they were either transformed or reconstructed into other variables. The authors applied SVM, kNN, NB, and DT to predict food insecurity, and SVM and kNN outperformed the other classifiers. [Van der Heijden et al., 2018] proposed a method for predicting food insecurity in Ethiopia at the sub-national level using open data linked to food insecurity drivers, and RF model for predicting food insecurity obtained an accuracy of 90%.

3. METHODOLOGY

This section outlines the proposed study of comparing ML and DL models to predict household food security status, and the framework for ML and DL models are shown in Figures 1 and 2, respectively. ML classifiers, namely: kNN, SVM, LR, NB, MLP, and DL Models, namely: ANN and CNN, have been used to predict the household food security status.

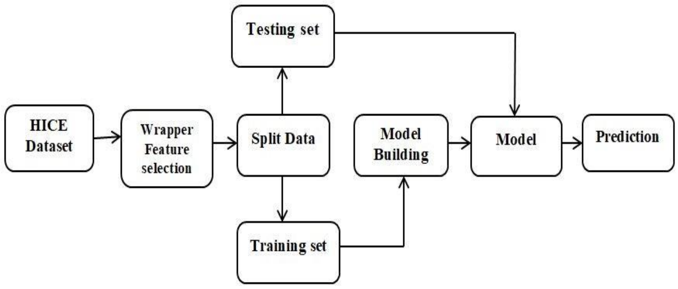

Figure 1: ML framework to predict household food security status

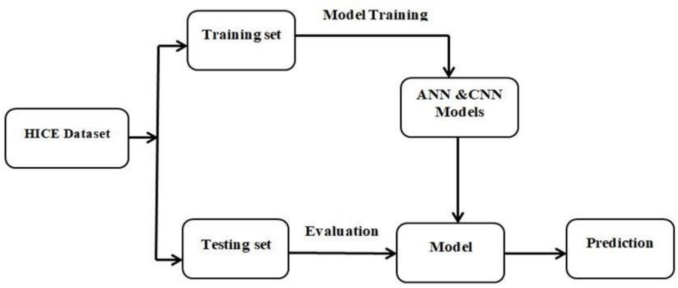

Figure 2: DL framework to predict household food security status

3.1 Feature Selection

Feature selection is an approach for dimensionality reduction that aims to choose a subset of the better features from the original features by removing unnecessary and redundant features [Dorsewamy and Nigus, 2020]. A backward feature selection technique chooses the most important features from the given dataset. Backward feature selection method starts with all the features and eliminates the most unnecessary and redundant features in each iteration. The features selected using backward feature selection are Region: Zone, Household serial number, Household size, Ecology, Household-head age, Household-head marital status, Weight, Year, and Net-calorie.

4. EXPERIMENT AND RESULTS

The HICE dataset of nine regional states and two city administrations: Addis Ababa and Dire Dawa of Ethiopia, is used to predict the household food security status. This dataset consists of 32,378 instances labeled as food secure and 25,686 instances labeled as food insecure with 21 features, and the description of 21 features is shown in Table 1. Only the selected features are used for training ML classifiers, and all the features are used to train DL models. The parameters and values used for both ANN and CNN models are shown in Table 2. The experiments are conducted on Microsoft Windows 10 with Intel® Core™ i7- 9700 CPU running at 3.00 GHz, 8 processors, 16 GB RAM, and a 1 TB hard disc, and the ML and DL models are implemented in Python 3.6 tool using Keras [Chollet, 2021] and Tenser-flow.

Review Article | Volume 10, Issue 2 | Pages 308-311 | e-ISSN: 2347-470X

Table 1: Description of the features of the HICE dataset

No	Features	Data Type	Description
1	Region	Categorical	The largest administrative unit next to the state
2	Zone	Categorical	An administrative unit under Region
3	Woreda	Categorical	The administrative unit under Zone
4	Town	Categorical	Capital city or Regional cities
5	K/ketema	Categorical	Sub-city with big town
6	Kebele	Numeric	The smallest administrative unit of the woreda
7	HH-Sno	Numeric	Household Selection Number
8	HH-Size	Numeric	Total number of families living together
9	Adquiv	Numeric	Food Consumption Distribution based on age and sex
10	Place of Residence	Categorical	The place where the household resides
11	REP	Categorical	Enumeration area where household survey conducted
12	Ecology	Categorical	The climate Zone where households live
13	HH-age-sex	Categorical	Gender of Household head
14	HH-head-Age	Numeric	Age of the household head
13	HH-head-Martial status	Categorical	Marital status of household head
16	HH head-Education	Categorical	Education level of household head
17	Weight	Numeric	A sample representative of the household
18	Annual-expenditure	Numeric	The annual expenditure of the household
19	Survey-Year	Numeric	Year of survey data collected
20	Net Calorie	Numeric	Average calorie intake per day per person
21	Security status	Categorical	Food security status of household

Table 2: ANN and CNN models Parameters and their values

Parameter	Values
Unit	32,16,8
Layers	3
Input Activation function	ReLU
Output Activation function	Softmax
Learning Rate	0.001
Loss	binary_crossentropy
Epochs	300
Optimizer	Adam
Decay	0.00001
Momentum	0.3,0.5

Figure 3 displays the accuracy, precision, recall, F1-score, MAE, and RMSE of ML and DL models. Among all ML and DL models, ANN scores better than other classifiers, with an accuracy of 99.15%.

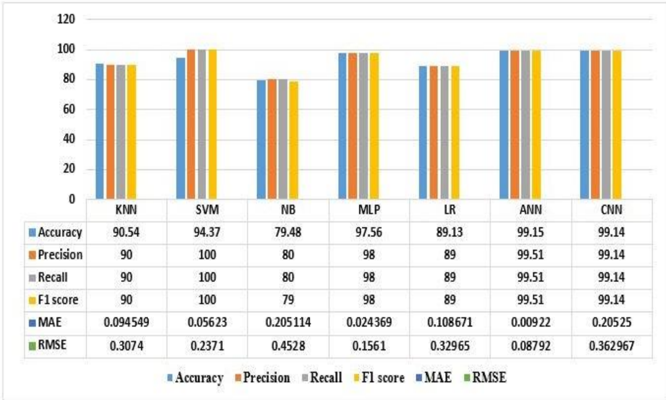

Figure 3: Comparison of the performance of ML and DL models

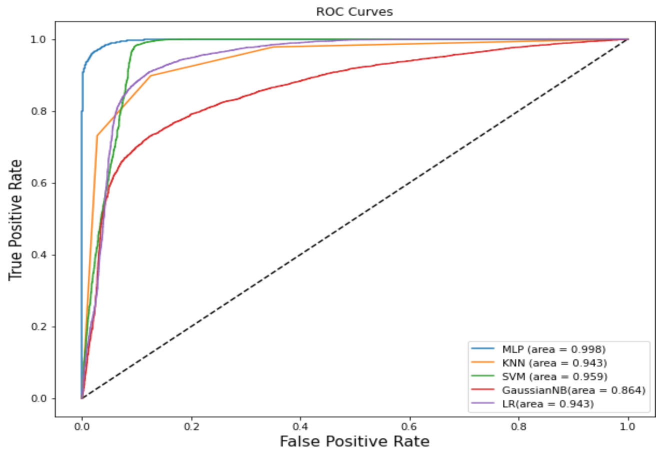

Figure 4: ROC curves of ML classifiers

The ROC curve of ML classifiers and DL models are shown in Figures 4 and 5, respectively, and the results illustrate that the ANN model has a larger ROC with a score of 100%, while the CNN model has a score of 95%.

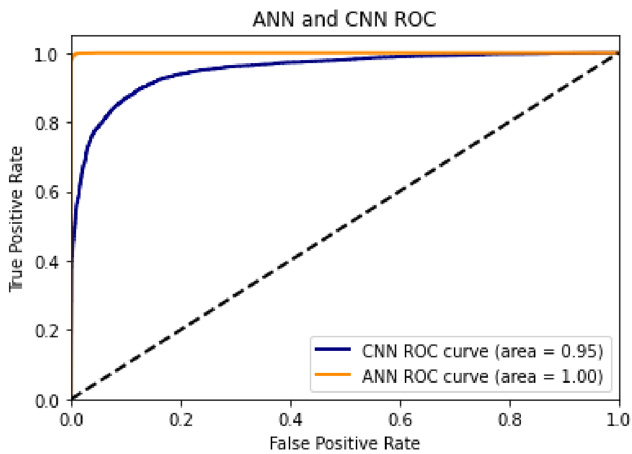

Figure 5: ROC curves of ANN and CNN models

International Journal of
Electrical and Electronics Research (IJEER)
Review Article | Volume 10, Issue 2 | Pages 308-311 | e-ISSN: 2347-470X

4. CONCLUSION AND FUTURE WORK

In this study, ML and DL models are trained on the HICE dataset to predict the food security status of households. The ML methodologies, namely: NB, LR, MLP, kNN, and SVM, and DL models, namely: ANN and CNN, are investigated, and the results illustrate that the ANN model surpassed the CNN model and ML classifiers with an accuracy of 99.15%. The future work includes conducting experiments using a larger dataset and state-of-the-art DL models. Further, additional data such as land information, weather data, price data, conflict-related data, and so on will also be considered to predict household food security status.

5. REFERENCE

[1] J. J. Westerveld, M. J. van den Homberg, G. G. Nobre, D. L. van den Berg, A. D. Teklesadik, and S. M. Stuit, "Forecasting transitions in the state of food security with machine learning using transferable features," Science of the Total Environment, vol. 786, p. 147366, 2021.
[2] M. M. Ayenew, "The dynamics of food insecurity in Ethiopia," pp. 1177–1195, 2017.
[3] M. Alelign, T. M. Abuhay, A. Letta, and T. Dereje, "Identifying risk factors and predicting food security status using supervised machine learning techniques," in 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 2021, pp. 12–17.
[4] Y. Zhou and K. Baylis, "Predict food security with machine learning: Application in eastern Africa," 2019.
[5] S. I. A. Meerza, and A. Ahamed, "Food insecurity through machine learning lens: Identifying vulnerable households," 2021.
[6] R. M. Barbosa and D. R. Nelson, "The use of support vector machine to analyze food security in a region of Brazil," Applied Artificial Intelligence, vol. 30, no. 4, pp. 318–330, 2016.
[7] C. Gao, C. J. Fei, B. A. McCarl, and D. J. Leatham, "Identifying vulnerable households using machine learning," Sustainability, vol. 12, no. 15, p. 6002, 2020.
[8] A. Alsharkawi, M. Al-Fetyani, M. Dawas , H. Saadeh, and M. Alyaman, "Poverty classification using machine learning: The case of Jordan," Sustainability, vol. 13, no. 3, p. 1412, 2021.
[9] W. Okori and J. Obua, "Machine learning classification technique for famine prediction," in Proceedings of the world congress on engineering, vol. 2, no. 1. Citeseer, 2011, pp. 4–9.
[10] W. van der Heijden, M. van den Homberg, M. Marijnis, M. de Graaff, and H. Daniels, "Combining open data and machine learning to predict food security in Ethiopia," in 2018 International Tech4Dev Conference: UNESCO Chair in Technologies for Development: Voices of the Global South, 2018.
[11] D. Dorsewamy and M. Nigus, "Feature selection methods for household food insecurity classification," in 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE, 2020, pp. 1–7.
[12] F. Chollet, Keras, 2018 (accessed June 10, 2021). [Online]. Available: https://keras.io/.