ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.3, pp.191-207
DOI:10.33168/LISS.2022.0314

Feature Weighting Based Food Recognition System

Ahmed Kawther Hussein

Department of Computer Science, College of Education, Mustansiriyah University, Baghdad-Iraq

ahmedkawther@uomustansiriyah.edu.iq (corresponding author)

Abstract. Food recognition is an important pattern recognition problem. It involves identification various food types for certain applications such as food industry, calories counting or restaurant automation. Feature extraction is an essential stage in food recognition. It involves extracting various mathematical or handcrafted features for classification. The numerous types of extracted features create an issue of redundancy and over fitting. This article proposes feature weighting based on the concept of entropy for food recognition. The feature weighting indicates to selecting the top 25% of features after sorting them according to their importance. We evaluate Local Binary Pattern, Gabor features, Histogram of Gradient for features weighting based on entropy. The evaluation is done based on two types of neural networks, namely, extreme learning machine ELM and fast learning machine fast learning network for whole features extraction and for only 25% selection of most important features. The model has shown a superiority of ELM and fast learning network when top 25% of features were selected for both ELM and fast learning network with more superiority of fast learning network.

Keywords: food recognition, features extraction, feature selection, extreme learning machine, fast learning network.

1. Introduction

Mobile tablets and smartphones now have portable technology that has progressed and grown in strength. Consequently, it became possible to execute many kinds of applications. The applications range from those that assist communication, like social media platforms, to those that are geared toward health, including biometric analysis and illness detection. One of the most important and quickly evolving applications uses the camera and processing capacity of smartphones to detect food and calculate the number of calories in various meals. Such an application is helpful for facilitating the diet process that many people adhere to and for lowering the number of diabetes cases that are on the rise globally as a result of most people's lack of calorie knowledge. Since the food we consume at restaurants often has a high calorie and carbohydrate content, it is necessary to count them and determine whether the total falls within the doctor's recommended daily allowance. Therefore, developing a portable system that makes use of the camera and processing capacity of a mobile device to detect the type of food and its caloric content is a current area of research, especially given the rapid advancement of artificial intelligence algorithms.

In the past, computer vision algorithms could only detect a small subset of objects, necessitating the use of several heuristics. However, it has become much simpler to design a quick and accurate object identification algorithm as a result of the quick growth of artificial intelligence in general and machine learning in particular. The strategy is built on the traditional pattern recognition stack, which begins with feature extraction, progresses to training a neural network or classifier, and concludes with making a prediction using the taught classifier.The characteristics that must be extracted must be discriminative, able to identify the item and set it apart from other things of a similar kind. In addition, the classifier must be over-fitting-free and able to update its knowledge as necessary.

The automated food categorization is crucial in many industrial and application fields. First off, it may be utilized as a subsystem for estimating the calories in food to inform the user of such information before to eating. This is helpful for diabetics and others who follow a diet plan. Additionally, it helps restaurants set up automatic food counting of the prepared meals each day. Thirdly, before putting different goods in food boxes, it may be used in food manufacturers to categorize them.

Traditionally, researchers have extracted several kinds of features and used them to train extreme learning machines, which are shallow machine learning models. Directly employing features without giving them any weight, however, may result in poorer success with training or prediction. This article makes feature weighting based on entropy possible.

Here is a summary of what this article contributes. First, it assesses how texture traits contribute to the identification of food items. It also concentrates on the Arabic morning and its meals. Thirdly, it suggests using entropy as a criterion to group characteristics according to their significance and choose a subset of them for categorization.

The rest of the article is divided into the following sections. We offer the associated works in section 2. Section 3 then offers the technique. Following that, section 4 provides the evaluation, followed by section 5's conclusion and recommendations for further study.

2. Related Works

Because deep learning models have a strong potential for learning several samples of the same food type, the research on food identification focuses on them. A multi-food categorization system was developed in the work of (Pouladzadeh, P., & Shirmohammadi, S. 2017) .The classification system employs the following techniques: region proposal algorithm to generate candidate regions, coevolutionary neural network, maximum cover using submodular optimization algorithm to select positive regions for each food category, and particle cloud computing to speed up the procedure. Their approach has a restriction in that it is not entirely automated due to the user being required to specify the region of interest.Multi-food categorization based on multi-scale multi-view feature aggregation is done in the work of (S Jiang, S., Min, W., Liu, L., & Luo, Z. 2019). High-level semantic features, mid-level attribute features, and deep visual features are all subjected to multi-granularity aggregation. Unified representation is the outcome of the aggregation. They have employed ingredient-supervised CNN to broaden understanding of the cuisine with the highest likelihood. Therefore, the aggregated characteristics have the best likelihood of capturing the semantics of food photos. They have used multi-scale CNN activations for each type of feature to increase the discriminative power of the features in order to deal with the assumption that food armament does not display a distinctive spatial layout. Multi-level and multi-view fusion are features of the work. This piece has been improved upon. In the paper(jiang, S., Min, W., Lyu, Y., & Liu, L. 2020, a system known as multi view few shot learning was introduced. In order to fill the gap between the disparate training categories and test categories, it also incorporates category-oriented deep visual features and ingredient supervised neural network. A multi-view relation network with the integration of combined feature maps and convolution allows the learning of fine-grained features. Additionally, two types of neural networks—Siamese and matching—have been extended in this work. The work's computational complexity hasn't been addressed, though. In addition to the issue of food identification accuracy, several studies have also looked at the mobile device's energy usage and time delay. Convolutional neural networks have been used in the work of(Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Yunsheng, M., ... & Hou, P.2017), to enable edge computing and a communication layer between the front and back ends, which has allowed for the achievement of the three performance metrics of accuracy, time delay, and energy consumption. Their back-end computing architecture made use of an inception module and CNN work from leNet-5, AlexNet, and GoogleLeNet. To identify the food, a cloud or edge service is needed, which is not possible with mobile devices without an Internet connection. Building a smaller network that learns from a larger network in the cloud is another approach to enabling food identification on mobile devices. The Joint-learning Distilled Network (JDNet), developed by H. Zhao, K.-H. Yap, A. C. Kot, and L. Duan. 2020), learns from a vast instructor network to achieve great accuracy in food identification. With simultaneous training, they have used a framework for cooperative learning. Utilizing various intermediate layer characteristics in the student and instructor network, this is accomplished. With a four times lower network size, the accuracy that was attained was about 92%. Overall, the research focused on utilizing deep learning architecture, which has the benefit of achieving high identification accuracies of food even with varying instances of the same food kinds. The vast food databases have also been used to understand this. In (Lee, J. M., Jung, I. H., & Hwang, K. 2022),a Deep learning image classification was employed to categorize the beef , and Open CV technique was utilized to determine the freshness of the beef.

For the purpose of building a dataset for food recognition, many models were successful. A dataset comprised of 251 fine-grained food categories was created using online photos in the research of (Kaur, P., Sikka, K., Wang, W., Belongie, S., & Divakaran, A. 2019) There are 158k photos in the data. The work of (H. Kagaya and K. Aizawa. 2015) Convolutional neural network was proven to be accurate models for food recognition in another dataset referred to as food 101.

3. Methodology

The established approach for classifying food products is presented in this section. It is divided into five pieces. firstly, problem formulation in sub-section 3.1. Pre-processing is covered in sub-section 3.2, segmentation is covered in sub-section 3.3, and features extraction is covered in sub-section 3.4. Fourth, sub-section 3.5 has feature weighting.

3.1. Problem formulation

Suppose we have picture I, which shows a collection of food items $Items = \{item_1, item_2, \dots item_N\} \subseteq X$ where X denotes the set of all food items that are used for training. Without prior information of the quantity of food items as depicted in this image, we must construct a classifier, C, that takes the input image I and does some processing to predict the food items that are present in Fig. 1. Hence, the decision of the classifier $y = C(I) = \{item'_1, item'_2, \dots item'_k\}$. The recognition performance of the classifier is calculated based on the Equation (1):

$$recognition\ performance = \frac{\|Items \cap C(I)\| - \|C(I)/Items\|}{\|Items\|} \quad (1)$$

Using a training dataset, we want to choose the picture characteristics that may be utilized to improve the classifier.4

Fig. 1: Breakfast image.

3.2. Pre-processing

One role of pre-processing is to filter-out the noise from the image. Typically, various types of noises exist: high-frequency noise is removed by using median filter and low frequency noise is removed by using averaging filter. Another role of the pre-processing is to convert all images to the same size because. This is done by using the resize operation. Consequently, the feature that will be extracted from the image will have fixed dimension which makes the classification simpler.

3.3. Segmentation

To separate the ROI's region of interest from the backdrop is the function of segmentation. The food item on the dish will be represented by the ROI. The tablecloth, the fork and spoon, etc., as well as any additional patterns that could be present in the image, are unimportant. Therefore, it is not necessary to separate them from the picture.

3.4. Features extraction

There are many different kinds of characteristics that may be employed for feature extraction. We separate them into handmade and mathematical aspects. We split up mathematical features into global features and local features. They include the following.

3.4.1. Local binary pattern LBP

An invariant measure for texture in grayscale format is the local binary pattern(hang, B., Gao, Y., Zhao, S., & Liu, J. 2009) The fundamental LBP operates by employing the 3×3 neighborhood surrounding each pixel(Zhao, Y., Huang, D. S., & Jia, W. 2012). Based on Equation (2), the mask is utilized to create a binary pattern by comparing each of the surrounding pixels with the center one: There are many different kinds of characteristics that may be employed for feature extraction. We separate them into handmade and mathematical aspects. We split up mathematical features into global features and local features. They include the following:

Where

$$f\big(I(z_0), I(z_i)\big) = \begin{cases} 1, if\ I(z_i) - I(z_0) > threshold \\ 0, if\ I(z_i) - I(z_0) \le threshold \end{cases} \quad (2)$$

$i = 1, \dots 8$ denotes the index of the surrounding pixels

z_0 indicates to the central pixel

Equation (3) gives the generic formulation of LBP with regard to radius R and neighbor P.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\big(g_p - g_c\big)2^p \quad (3)$$

Where

$$s(x) = \begin{cases} 1, x \ge 0 \\ 0, x < 0 \end{cases}$$

g_p denotes the grayscale of the pixel

g_c denotes the grayscale of the center

3.4.2. Gabor features

Gabor features are provided for an image $I(x, y)$ as a vector $f^{(R)}$ that contains rotation-invariant features as provided in Equations(4)-(7) (Han, J., & Ma, K. K. 2007)

$$f^{(R)} = [\mu_0^{(R)}, \sigma_0^{(R)}, \mu_1^{(R)}, \sigma_1^{(R)}, \dots, \mu_{S-1}^{(R)}, \sigma_{S-1}^{(R)}] \quad (4)$$

$$\mu_m^{(R)} = \frac{1}{N} \sum_x \sum_y |J_m^{(R)}(x, y)| \quad (5)$$

$$\sigma_m^{(R)} = \sqrt{\frac{1}{N} \sum_x \sum_y (|J_m^{(R)}(x, y)| - \mu_m^{(R)})^2} \quad (6)$$

$$J_{m,n}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) g_{m,n}(x - x_1, y - y_1) \quad (7)$$

Where

$m = 0, 1, \dots, S - 1$

3.4.3. Histogram of gradient HoG

The gradient histogram procedure is performed in three phases, starting with determining the gradient orientation at each pixel. Second, create a tiny rectangular section for each orientation's histogram. Concatenating the histogram of all tiny sections is the third step (N. Dalal and B. Triggs. 2005). Figure 2 shows the three actions.

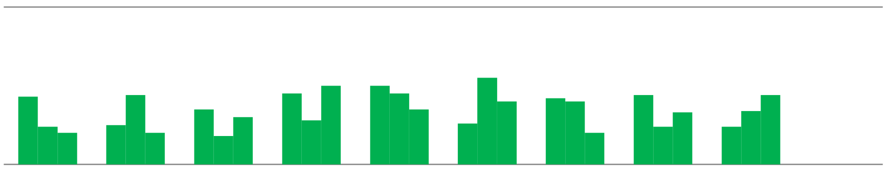

Fig. 2: Histogram of gradient HoG steps.

3.5. Feature weighting

According to the feature weighting, the top 25% of features will be chosen after being sorted by significance. Entropy, which is determined using Equation (8), was used to gauge the significance of the trait.

$$E = \sum_i^N -p(x_i)\log(p(x_i)) \tag{8}$$

where

i denotes the index of the value of the feature

$p(x_i)$ denotes the probability of the appearance of certain value

The higher the entropy is, the more importance of its corresponding feature

3.6. Classification

Extreme learning machine ELM will be utilized for categorization. Equation (9) gives the weights of the hidden output matrix assuming that Y is the ground truth of the training data.

$$\beta^\star = \Phi^\top(\frac{\mathrm{I}}{C} + \Phi\Phi^\top)^{-1}\mathrm{Y}^\top \tag{9}$$

Additionally, the effectiveness of ELM is contrasted with that of a fast learning machine fast learning network, which has parallel weights between the input and hidden layer and the input and output layer and is similar to ELM.

3.7. Dataset

Various photographs of breakfast dishes were taken for examination. They are displayed in Fig. 3. Seventy photos were included in the sample of photographs that were acquired. To make the issue more difficult, the photographs were taken with no restrictions.

Fig.2: Examples of breakfast meals captured by the camera of the smartphone.

3.8. Evaluation metrics

For evaluation metrics, the system will be evaluated based on the accuracy of classification. The formula of the classification accuracy is given in Equation (10)

$$ACC = \frac{N_m}{N} \tag{10}$$

Where

N_m denotes the total number of misclassifications

N denotes the total number of tested records

Each entry is a member of a certain food item class C. If it is successfully anticipated, the outcome is TP; if it is wrongly predicted to not exist in the food picture, the outcome is FN. Similar to this, if the outcome is projected to exist mistakenly, it can be FP, whereas if it is properly predicted to not exist, it can be TN. The equations for precision, recall, F-measure, and G-mean may be determined using the set of TP, TN, FP, and FN.

$$Precision = \frac{TP}{FP + TP} \tag{11}$$

$$Recall = \frac{TP}{FN + TP} \tag{12}$$

$$F - measure = 2\frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

$$G - mean = \sqrt{Precision \times Recall} \tag{14}$$

4. Experimental Works and Results

The experimental assessment is presented in this part. ELM and rapid learning network were the two classifiers employed. All or just 25% of the most crucial features were utilized. The part is divided into three smaller portions. The assessment of LBP-based feature recognition is presented first in sub-section 4.1. The assessment of Gabor-based feature recognition is then presented in sub-section 42. The assessment of HoG-based feature recognition is provided in subsection 4.3. By employing the entropy measure to include just the top 25% of features, the case of feature weighting is labeled as (25%-the classifier name), whereas the case of non-feature weighting is referred to as (All-the classifier name).

4.1. LBP evaluation

Fig. 4 to 7 show the classification metrics for categorizing foods using quick learning networks and ELM with and without feature weighting. The accuracy of ELM was the highest after doing 25% of feature weighting, as seen in the hardlim figure. Similar to ELM without feature weighting, this accuracy was. However, for the two scenarios of features weighting and non-feature weighting, rapid learning networks have offered poorer accuracy. Overall, we see that the accuracy of the rapid learning network was higher when the activation function was changed from hardlim to other types of activation functions. This demonstrates how the findings are affected by the type of activation function.

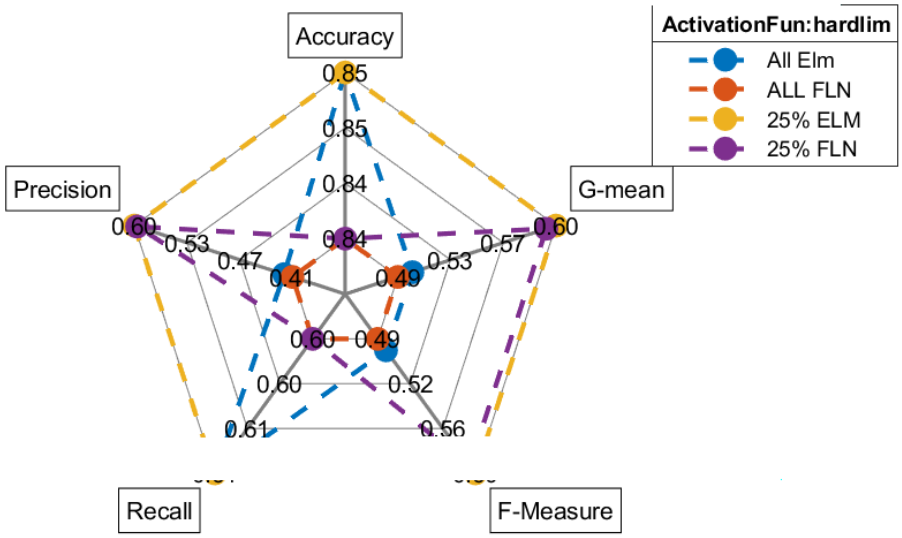

Fig.3: Classification metrics of food classification for hardlim activation function based on LBP.

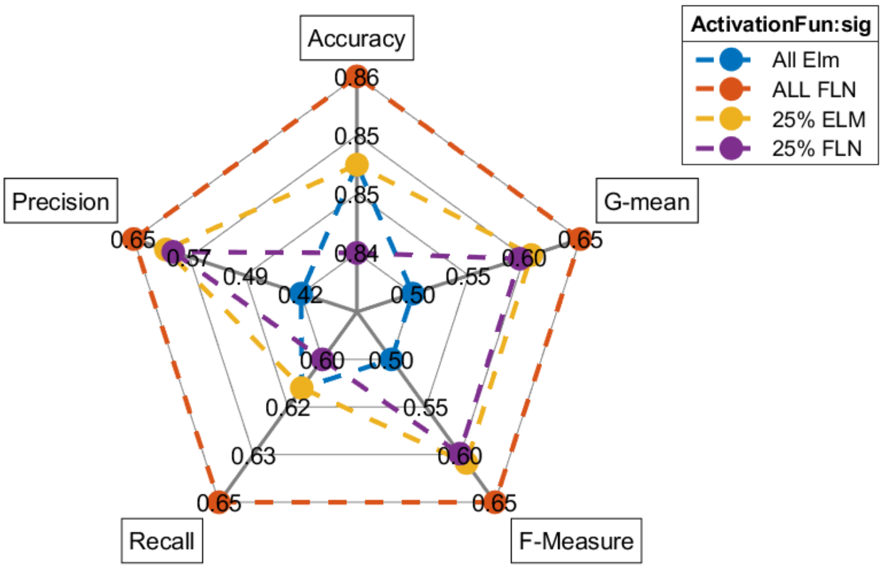

Fig.4: Classification metrics of food classification for sigmoid activation function based on LBP.

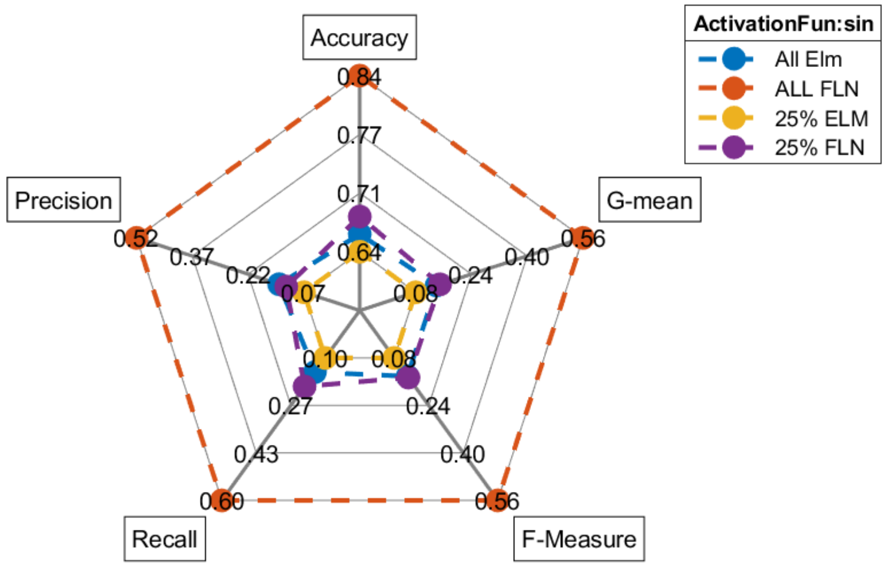

Fig.5:Classification metrics of food classification for sin activation function based on LBP.

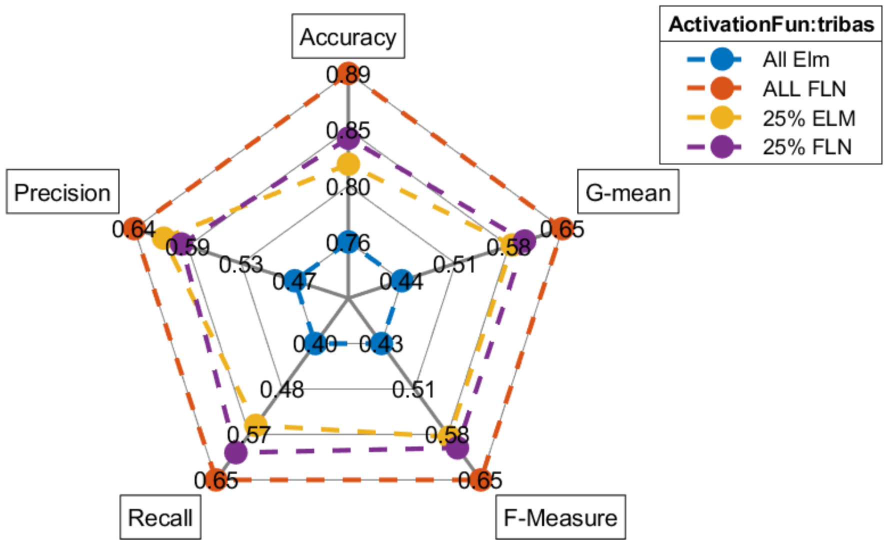

Fig.7: Classification metrics of food classification for tribes activation function based on LBP.

The results also demonstrate that FLN outperformed ELM in terms of utilizing all characteristics when the tribes activation function was applied. The input hidden output weights, which are included in FLN and were helpful when all features were used, interpret this.

4.2. Gabor evaluation

Similar to LBP, Figures 8 to 11 show the accuracy achieved by ELM and rapid learning networks when all features were chosen and when just the top 25% of features were chosen. It has been noted that ELM has demonstrated advantage when the top 25% of features were chosen for hardlim characteristics.

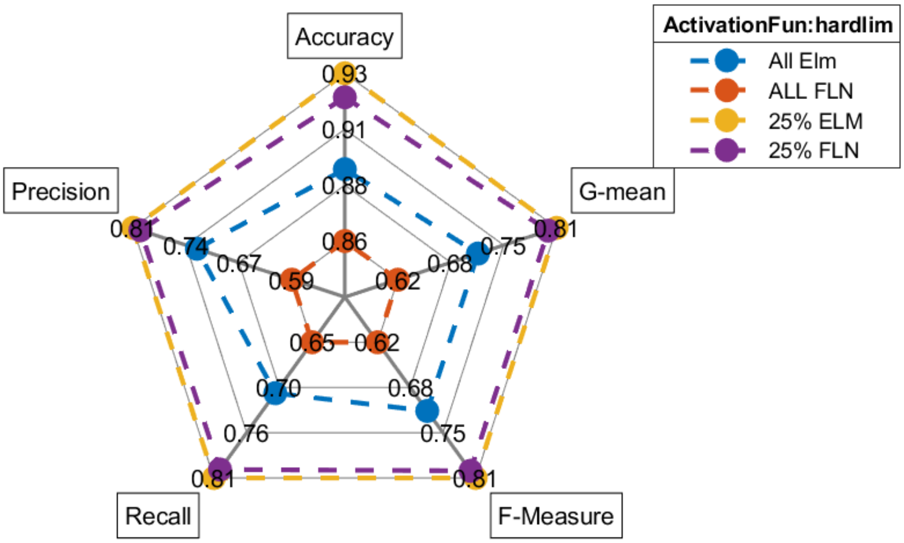

Fig. 6: Classification metrics of food classification for hardlim activation function based on Gabor.

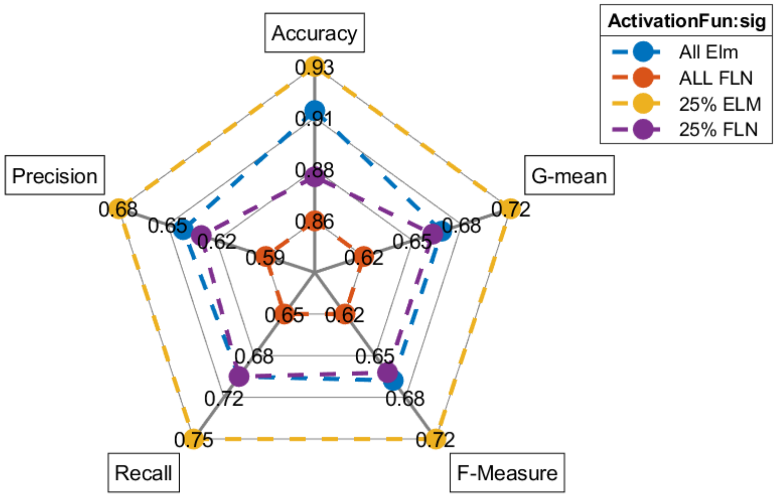

Fig. 7: Classification metrics of food classification for sig activation function based on Gabor.

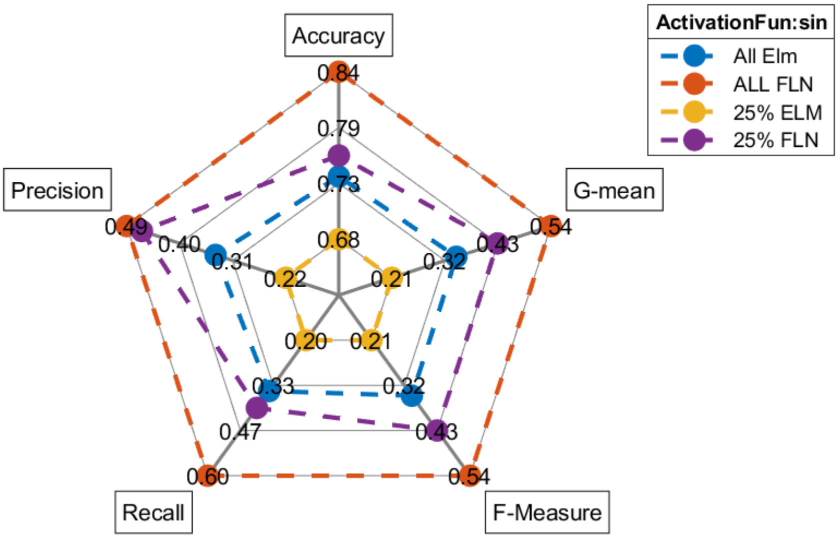

Fig.8: Classification metrics of food classification for sin activation function based on Gabor.

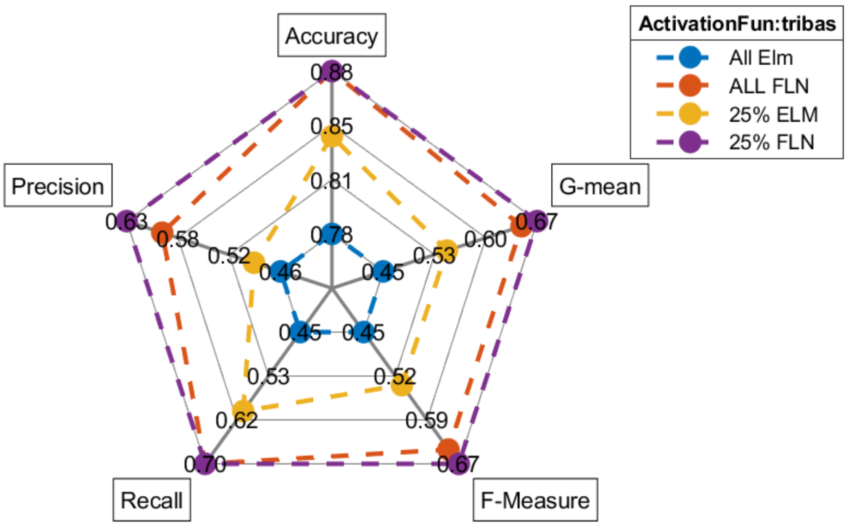

Fig. 9:Classification metrics of food classification for tribas activation function based on Gabor.

Additionally, we note that FLN produced superior outcomes when the top 25% of features were employed as opposed to utilizing all features with the same classifier.

4.3. HoG

Similar to HoG, we show the accuracy achieved for ELM and rapid learning network when all features were picked and when the top 25% of features were selected in figures 12 to 15. It has been noted that when all characteristics were chosen, the rapid learning network demonstrated superiority. After a quick learning network, it is seen that the ELM with the top 25% outperforms the other models in sigmoid.

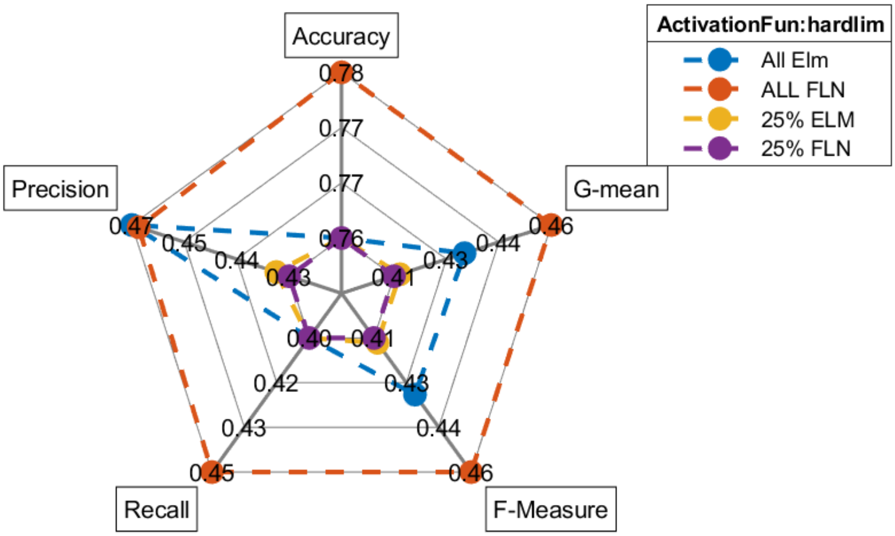

Fig.10:Classification metrics of food classification for hardlim activation function based on HoG.

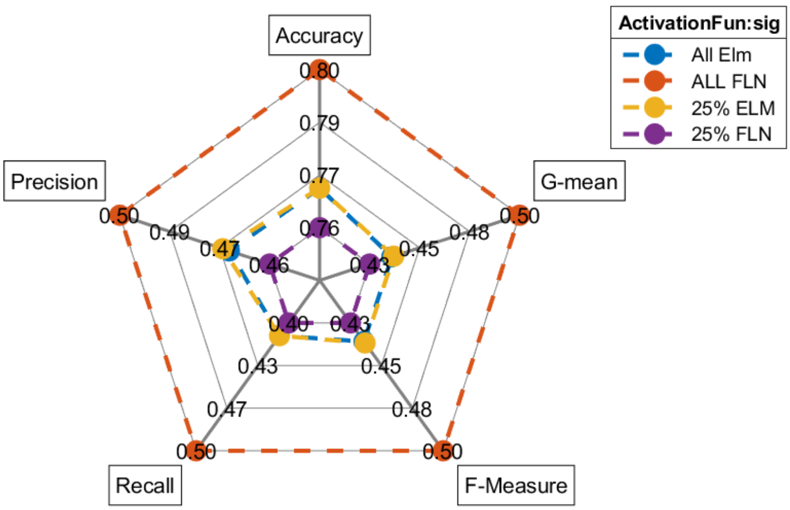

Fig.113: Classification metrics of food classification for sig activation function based on HoG.

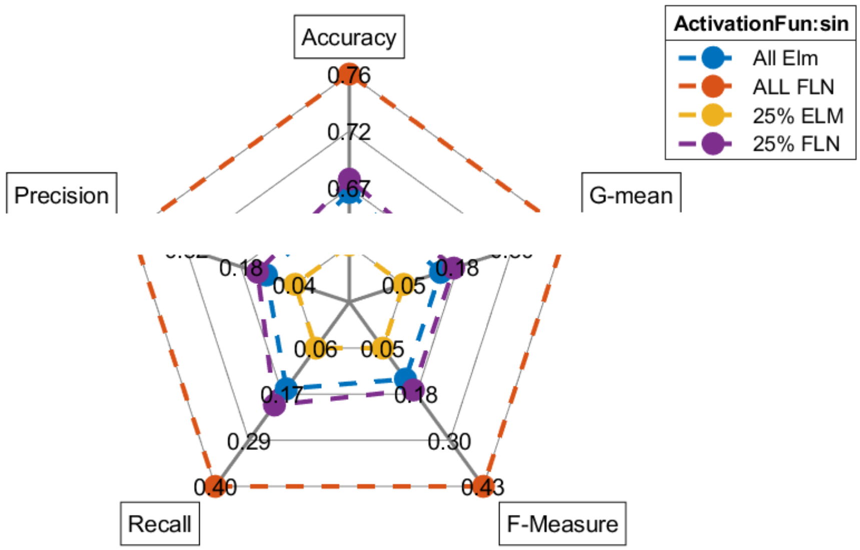

Fig. 12: Classification metrics of food classification for sin activation function based on HoG.

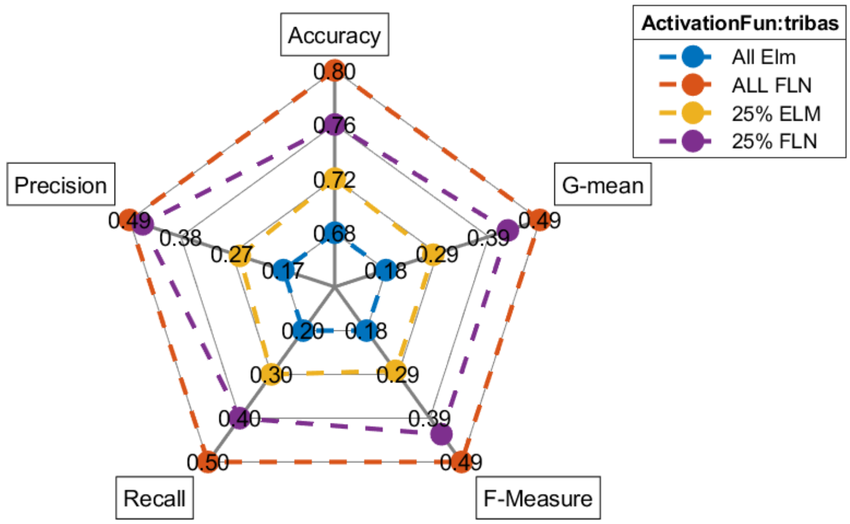

Fig.135: Classification metrics of food classification for tribas activation function based on HoG.

It was discovered that utilizing the whole collection of features with the ELM classifier resulted in the worst performance, as shown in figure 14, but using only the top 25% improved all metrics and increased accuracy from 28% to 0.72%.

5. Conclusion and Future work

A system for segmenting and categorizing calories has been shown in this article. The system's foundation is the extraction of several kinds of textural characteristics, specifically LBP, HoG, and Gabor. The examination was done using a unique picture of an Arabic breakfast. Extreme learning machine (ELM) and fast learning machine (FLM) fast learning network versions of neural networks were utilized. Each feature type is subject to five different forms of activation functions in the evaluation: sin, sigmoid, tribas, and hardlim. The majority of analyses have shown superiority in terms of the most important components for both ELM and fast learning networks. The study's limitations, nevertheless, stem from the fact that not all activation processes were successfully activated. The goal of future work is to develop an automated method for choosing the optimum balance of crucial selection criteria.

Acknowledgements

The author would like to thank the Mustansiriyah university (www.uomustansiriyah.edu.iq) Baghdad - Iraq for its support in the present work.

References

Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 1, 886-893.

Han, J. & Ma, K. K. (2007). Rotation-invariant and scale-invariant Gabor features for texture image retrieval. *Image and vision computing*, 25(9), 1474-1481.

Jiang, S., Min, W., Liu, L., & Luo, Z. (2019). Multi-scale multi-view deep feature aggregation for food recognition. *IEEE Transactions on Image Processing*, 29, 265-276.

Jiang, S., Min, W., Lyu, Y., & Liu, L. (2020). Few-shot food recognition via multi-view representation learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(3), 1-20.

Kagaya, H. & Aizawa, K., (2015). Highly accurate food/non-food image classification based on a deep convolutional neural network. In *International conference on image analysis and processing*, Springer, 350-357.

Kaur, P., Sikka, K., Wang, W., Belongie, S., & Divakaran, A. (2019). Foodx-251: A dataset for fine-grained food classification. *arXiv preprint arXiv:1907.06167*.

Lee, J. M., Jung, I. H., & Hwang, K. (2022). Classification of beef by using artificial intelligence. *Journal of Logistics, Informatics and Service Science*, 9(1), 1-10.

Liu, C. (2017). A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. 11(2), 249-261.

Muthevi, A. & Uppu, R. B. (2017). Leaf classification using completed local binary pattern of textures. In *IEEE 7th International Advance Computing Conference (IACC)*, 870-874.

Pouladzadeh, P. & Shirmohammadi, S. (2017). Mobile multi-food recognition using deep learning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 13(3), 1-21.

Zhang, B., Gao, Y., Zhao, S., & Liu, J. (2009). Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *IEEE transactions on image processing*, 19(2), 533-544.

Zhao, Y., Huang, D. S., & Jia, W. (2012). Completed local binary count for rotation invariant texture classification *IEEE transactions on image processing*, 21(10), 4492-4497.

Zhao, H., Yap, K.-H., Kot, A. C., & Duan. L. (2020), JDNet: A joint-learning distilled network for mobile visual food recognition, *IEEE Journal of Selected Topics in Signal Processing,* 14(4), 665-675