Food Recognition Continual Learning based on Online Clustering Method and Class Incremental Extreme

Dr Shashikala Parameshwarappa #1, Dr B.V.Dhandra*2,

Computer Sceince and Engineering Department
Government Engineering College ,Raichur -584135
State:Karnataka, Country: India

Dr BV Dhandra
Department of Master of Computer Applications
Gulbarga University Gulbarga 58102

***Abstract*—"Catastrophic interference" or "catastrophic forgetting" describes an artificial neural network's propensity to drastically and unexpectedly lose previously learnt information while receiving new input. This paper introduces a novel framework for continuous, open-ended learning by combining features selected using Relief F, extracted using transfer learning on deep models, and classified using an adaptive reduced class incremental kernel extreme learning machine (ARCIKELM). In order to reduce processing complexity, Relief F sorts the recovered features in a predetermined order. The novel ARCIKELM classifier makes real-time adjustments to the network topology in order to reduce catastrophic forgetting. Resolves domain adaption difficulties upon receipt of new samples from the current class. The findings portray as proposed framework is capable of learning new information. Item's name, country of origin, and ingredients may be shown when the system was enhanced. additional clues as to whether the meal is nourishing or fatty.**

***Keywords*— ARCIKELM, open-ended continual learning, Catastrophic, artificial neural network**

I. INTRODUCTION

For open-ended continuous learning, new courses are continually appearing, and fresh photographs of current classes are always coming in. Included in open-ended continuous learning are two essential kinds of incremental learning: 1) Incremental learning based on data 2) classroom learning via little steps. Data incremental learning adjusts domain changes with newly obtained pictures and constantly enhances identification efficacy of existing classes, whereas class incremental learning learns from novel classes. Learning new ideas and enhancing classification efficacy of current classes are both aided by these two types of incremental learning, which are similar to both aspects of human learning.

Because dataset is always changing and new ideas become relevant over time, open-ended continuing learning is crucial for many real-world identification tasks like food recognition.

Concerning food identification problem, individuals daily submit images of food from their cellphones, including both new and old classes. Since different users having varied notions about same class, it's important to assume that their perceptions of existing classes vary. Datasets utilized to test the suggested framework also show substantial levels of inter-class similarity and intra-class variance in food imagery .A "big vocabulary" means that the food dataset can include an unlimited number of additional tags. The proliferation of food-related media such as social media, television series, blogs, etc., has resulted in an explosion of new food tags, often called food porn [1]. The food databases are therefore both dynamic and open-ended. Regardless, most of the traditional approaches to food identification start with fixed datasets and a lot of class variability. While open-ended continuing learning may address food recognition in practical settings, it is important to address the issues that come with this approach.

II. RELATED WORK

Ghalib Ahmed Tahir, et.al,**[6]**, 2021, suggested an AI architecture that is user-centered and explainable in order to boost confidence by learning from users' requirements and profiles to draw conclusions and provide explanations that make sense. As far as nutritional assessment apps go, Framework covers all the bases by identifying Food/Non-Food, food categories, and substances. Authors Sanjeev T K and Merin Meleet, [7] Online data augmentation suggestivelyimproved model's capacity for classifying food photographs taken from many possible angles in 2022. Second, by performing feature rankings utilising Relief F method, this research endeavor reduces the dimensionality of the recovered features. It becomes increasingly difficult for the model to compute due to the duplicated properties. When the validation accuracy is 92% and the training accuracy is 98%, it is considered the best epoch.

C. Chen, et.al [2], developed a Support Vector Machine (SVM) based Hybrid Incremental Learning (HIL) approach to address the issue. This system can learn new item samples and ideas via human interaction, allowing it to gradually improve its identification capacity.

G. Csurka, et.al [3], Methods for large-scale picture categorization that can continually and cheaply include new classes and training pictures. In this regard, we implemented a novel metric learning strategy for closest NCM classifier and compared it to k-NN classifier, both of which are distance-based. Additionally, we provide an enhancement to the NCM classifier that enables more detailed representations of classes.

Researchers D. Maltoni and V. Lomonaco [4], Training deep models sequentially on several separate tasks without losing previously learned information is possible using architecture, regularization, and rehearsal procedures. Next, we provide a novel method, AR1, that integrates architectural and regularization techniques. AR1 is well-suited for online learning because to its minimal overhead, which includes both memory and compute. Compared to other regularization algorithms, AR1 fared much better on CORe50 and iCIFAR-100.

R. Polikar, et.al [5], Massive amounts of data are becoming accessible in a streaming way due to the proliferation of mobile phones, IOT technologies, and sensor networks [1]-[5]. It is often believed, either tacitly or in an outright way, that the mechanism producing this data stream is stationary, meaning that data are taken through fixed but unknowable distribution of probabilities. In 2021[6], Ghalib et.al, explained Due to its capacity to acquire objective measures for nutritional intake, food recognition systems have lately attracted a lot of interest from researchers in the relevant area. With the ability to identify Food/Non-Food, food categories, and ingredients, our framework is all-encompassing for a nutritional evaluation app. Experiments using the newly added Malaysian food dataset for component identification and the conventional food benchmarks showed that the combined set of metrics outperformed the alternatives.

III. Proposed System

Clustering methods, such as self-organizing incremental neural networks, may help choose closest nodes during live classification and identify which mapping nodes best represent class for ARCIKELM. This has potential to remain noise-invariant even when categorization input is far from current neurons, which could lessen catastrophic forgetting.Inverse is true in a cloud setting with regard to the autoscaling of computing resources. We extended the recommended structure to indicate the name, country, and ingredients of the cuisine. also shows if the meal is fatty or healthy.

.

IV. Methodology

There are two aspects of flexible learning that the proposed framework takes into account. 1) Classwork that builds upon itself 2) Leveraging data for incremental learning.

The 3 modules are:

A. Module for feature extraction;

B. Selection of features; and

C. Classification.

During training, specified deep feature extractor module receives all incoming images and uses them to extract features. After that, we rank them using the Relief F approach and then utilize our recommended process to choose the greatest characteristics.

Characteristics selected using Relief F method constitute final picture presentation. However, there is great generalizability in the recovered features from the deep model. Fixed class architecture softmax doesn’t make use of this capacity for final classification. Using these representations as a learning tool, this study applies a novel ARCIKELM that meets the requirements of both continuous and open-ended learning. In case of new class visual representations, it augments output and hidden neurons.

It employs our proposed strategy to incrementally update the model, adding hidden neurons only when necessary, if they are members of preexisting classes. In the classification step, we utilize similar deep feature extractor that we used for extracting features from testing picture. We choose most accurate representations according to how the Relief F method ranks the attributes. The final decision is with ARCIKELM. To train the ELM base classifier, we use the publicly available ELM_1 starting classes. In order to make room for new class, ELM_1 is incrementally updated to become ELM_2 utilising labelled data from both the new and old classes. As the network grows, it acquires additional output neurons and changes its ELM structure. Here is the mathematical reasoning behind it. First, Equation (1) is used to determine the weight output, in line with ELM method, once training stage of database of d is finished.

$$\beta 0 = k_0^{-1} H^T{}_0 T_0 \text{-----}(1)$$

where,

where,

$$H_0 = \begin{bmatrix} G(a_1, b_1, x_d) & \cdots & G(a_1, b_1, x_1) \\ \vdots & \cdots & \vdots \\ G(a_1, b_1, x_n) & \cdots & G(a_n, b_n, x_n) \end{bmatrix}$$

$$and\ T_0 = \begin{bmatrix} t_1^T \\ \vdots \\ t_n^T \end{bmatrix}\ and\ K_0 = H_0^T H_0$$

Equation (3) is utilised for deliberating H1 from new class dataset.

$$\boldsymbol{H_1} = \begin{bmatrix} \boldsymbol{G(a_1, b_1, x_s^1)} & \cdots & \boldsymbol{G(a_1, b_1, x_s^1)} \\ \vdots & \cdots & \vdots \\ \boldsymbol{G(a_1, b_1, x_s^n)} & \cdots & \boldsymbol{G(a_n, b_n, x_s^n)} \end{bmatrix} \quad (3)$$

Here, m is number of columns in T0, and Equation (4) represents number of samples in T1 with m+1 dimensions.

$$\mathbf{T_1} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1} \end{bmatrix}_{\mathbf{N_1 * (m+1)}} \quad (4)$$

One uses Equation (5) to determine β1 after merging d and s datasets.

$$\boldsymbol{\beta_1} = \boldsymbol{K_1^{-1}} \begin{bmatrix} \boldsymbol{H_0} \\ \boldsymbol{H_1} \end{bmatrix}^{\boldsymbol{T}} \begin{bmatrix} \boldsymbol{T_0 \cdot M} \\ \boldsymbol{T_1} \end{bmatrix} \quad (5)$$

where Equation (6) denotes M, a transformation matrix.

$$\boldsymbol{M} = \begin{bmatrix} \mathbf{1} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \ddots & \vdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{1} & \mathbf{0} \end{bmatrix} \quad (6)$$

It is clear that β1 is determined via incremental learning without the need to provide a prior dataset.

With this method, extreme learning computers may now progressively learn new classes. However, there are several drawbacks to it. One big drawback is that there is a fixed amount of buried neurons. By increasing neuronal plasticity, this causes catastrophic forgetfulness. An increase in the number of neurons needed to process newly incoming courses is inevitable. Underfitting occurs, however, when the number of neurons remains constant. Another cause of overfitting is the initialization of a high number of hidden neurons. We have solved this issue using adaptive CIELM. Nevertheless, the robust experiments conducted by this research demonstrate that their methodology lacks stability. Our best guess is that this is because of the unpredictable input weights, which cause the generalization capacity to be weak.

Phase 1: Gathering information via the use of freely available software is known as data collection. Transformer model training data is crucial. Prior to doing any kind of analysis, it is necessary to collect the necessary data, which involves sorting and collecting structured quantitative information. Individuals with the same condition were the subjects of the data collection. Building high-performance models begins with selecting accurate and complete data to use as inputs. The models' final results will only be as good as the input data. You can get the dataset on the Kaggale website, and you can also find datasets that people in nations like India, Pakistan, China, Japan, and Korea have developed themselves.

1) Phase 2: Data Loading Training image data sometimes contains various mistakes or garbage values; these may be eliminated by checking for missing values and ensuring the value falls inside a certain range. It is necessary to eliminate variables that have a high number of missing values. Although selecting a model will not improve the model's accuracy, cleaning the picture data will lessen the impact of any negative outcomes.
2) Phase 3: Predictive modeling training based on outcomes from classification tasks. Classification performance is a common statistic for evaluating the model's accuracy using expected class labels. Classification precision isn't perfect, but it's good enough for most classification tasks. Since there is no universally accepted theory for applying algorithms to different types of problems, experts recommend that practitioners do controlled experiments to find out which combination of algorithms and methods yields the best results for certain categorization tasks. Using the training dataset, we will find the optimal way to assign input data samples to predefined class labels. There should be several examples of each classifier in the training food picture collection so that it accurately represents the problem.
3) Phase 4: Since it is a machine learning technique, fine-tuning can only work with large datasets of food images; training it on a small dataset will lead to overfitting since parameter estimates might be hundreds of thousands to millions. Therefore, it is often best to begin with a pre-trained model and then refine it using a small subset of the domain-relevant food picture collection.
4) Phase 5: After training a model, it is common practice to validate it. By comparing the trained model to a test dataset, it is possible to ascertain if the assignment was effective on the given dataset. In addition to evaluating the model's outputs, it examines the system's outputs.

V. System Architecture

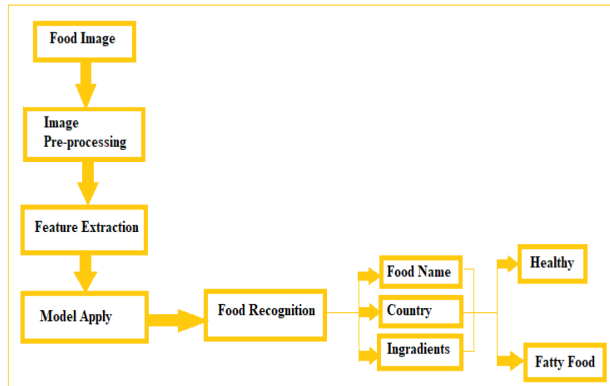

Fig-1: System Architecture

VI. Implementation

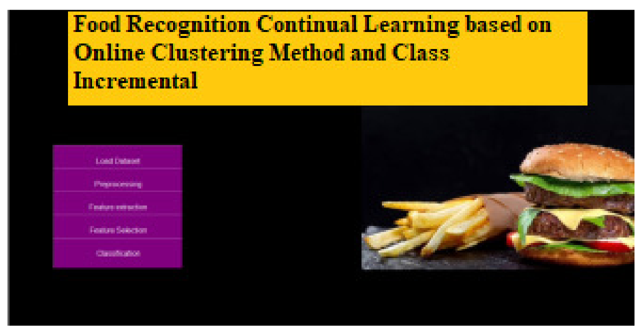

Figure 2: Menu
Menu consists of Input Data, Pre-processing, Feature Extraction, Recognition

Figure 3: Read Image

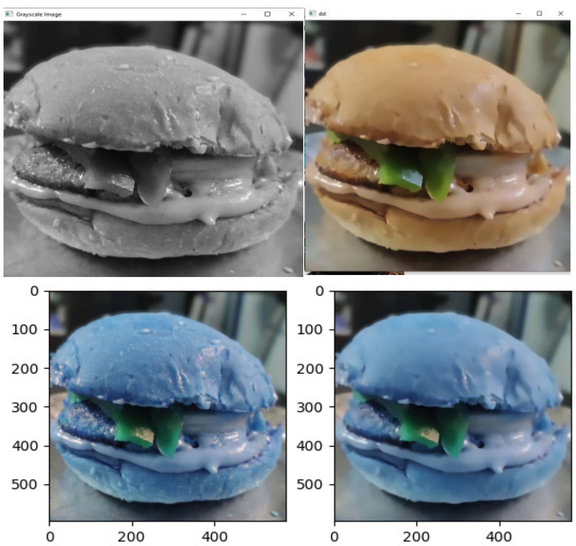

Figure 4:Preprocessing

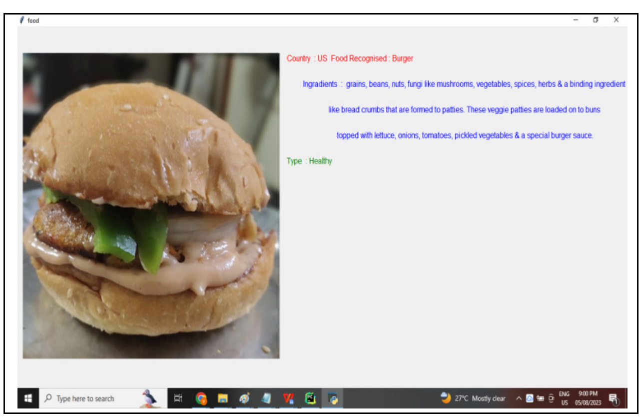

Figure5 :Classification And details

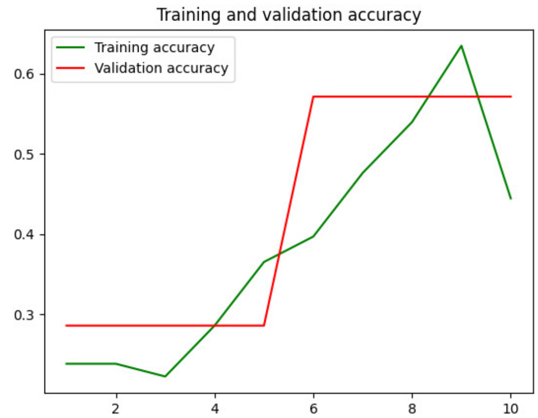

Figure 6: Accuracy graph

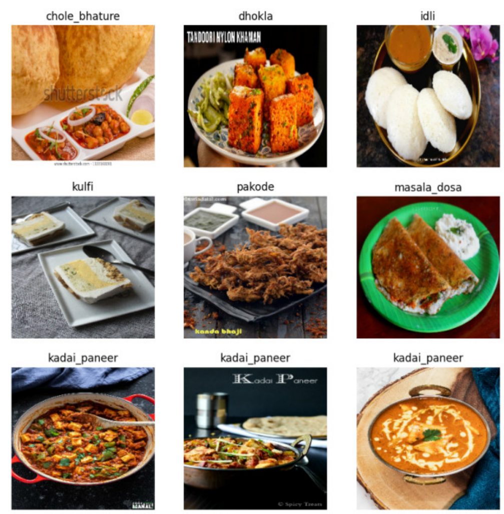

Figure 7: Food Dataset
Consists of 6269 files belonging to 20 classes.
Using 1880 files for validation.

Figure 8: Feature Extraction

The dimensionality reduction procedure includes feature extraction as one of its steps to minimize the size of the original dataset by grouping similar variables together. This will make processing it easy when the time comes. One key feature of these massive datasets is the abundance of variables they include. Processing these variables demands a significant amount of computational resources. Feature extraction is a useful tool for decreasing the amount of data needed to extract useful information from large datasets by choosing and combining variables into features. These characteristics accurately and uniquely characterize the real data collection while yet being straightforward to analyze.

VII. Conclusion

There is lot of room for growth and exploration in food identification dataset. Number of food categories and samples is constantly growing. Assuming the existence of all food classes and variants at the outset is the current premise of food recognition deep learning models. When taught in little chunks, they have awful memorization skills.

This work proposes a new paradigm for food identification open-ended continuous learning to address these challenges. For feature extraction, we use state-of-the-art deep learning networks; for feature selection and ranking, we employ Relief F; and for classification, we employ ARCIKELM. When extracting features from deep models, our study takes their great generalizability into account.

The results showed that Inception-Resnet-V2 outperformed the other two state-of-the-art deep networks. Features used in deep learning models, on the other hand, are more complex and time-consuming to categorize. The optimal length was determined by the framework using the Relief F technique, which resulted in a 52.14 percent decrease in the total learning time of projected classifier across all datasets. System used state-of-the-art Adaptive reduced class incremental kernel extreme learning machine to fix problems with data proportional learning as well as class incremental learning.

It dynamically enhances both hidden and output neurons. The preceding neurons' reduced flexibility reduces catastrophic forgetting. Using experimental data on four measures of classification performance and five metrics of catastrophic forgetting, the proposed classifier outperforms both the existing ACIELM and CIELM, as well as batch classifiers.

The proposed framework fulfils the needs for open-ended continuous learning and performs competitively when tested against other food identification architectures including PMTS, GTBB, and the supervised extreme learning committee.

Our hybrid open-ended learning technique aims to prevent catastrophic forgetting in future work. By clustering methods like self-organizing incremental neural networks, ARCIKELM may choose the mapping nodes that best represent its class and pick the closest nodes during live classification. This may prevent catastrophic forgetting and remain noise-invariant when categorization input is remote from neurons.Also, cloud computing allows for the automatic scalability of computing resources. Adding more new photos and inventive classes increases the amount of processing resources required.

In a similar spirit, user demands fluctuate during the various stages of categorization. Therefore, the ability to automatically increase computing resources is essential for an open-ended continual learning system in a cloud environment. Last but not least, projected framework for explainable AI is refined with further study. It can back up its categorization prediction with evidence.

References

[1] S. Rousseau, Food 'Porn' in Media. Dordrecht, The Netherlands: Springer, Nov. 2014, doi: 10.1007/978-94-007-0929-4_395.

[2] C. Chen, W. Min, X. Li, and S. Jiang, ''Hybrid incremental learning of new data and new classes for hand-held object recognition,'' J. Vis. Commun. Image Represent., vol. 58, pp. 138–148, Jan. 2019, doi: 10. 1016/j.jvcir.2018.11.009.

[3] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, ''Distance-based image classification: Generalizing to new classes at near-zero cost,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2624–2637, Nov. 2013, doi: 10.1109/TPAMI.2013.83.

[4] D. Maltoni and V. Lomonaco, ''Continuous learning in single-incrementaltask scenarios,'' CoRR, vol. abs/1806.08568, pp. 56–73, Apr. 2018.

[5] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, ''Learning in nonstationary environments: A survey,'' IEEE Comput. Intell. Mag., vol. 10, no. 4, pp. 12–25, Nov. 2015, doi: 10.1109/MCI.2015.2471196.

[6]Explainable deep learning ensemble for food image analysis on edge devices, panelGhalib Ahmed Tahir, Chu Kiong Loo,2021

[7] Food Recognition Using Extreme Learning Machines, Sanjeev T K, Merin Meleet, 2022

[8]Machine Learning Based Approach on Food Recognition and Nutrition Estimation Zhidong Shen , Adnan Shehzad , Si Chen , Hui Sun , Jin Liu · 2020

[9] P. Jackman, D.-W. Sun, P. Allen, N. A. Valous, F. Mendoza, and P. Ward, ''Identification of important image features for pork and turkey ham classification using colour and wavelet texture features and genetic selection,'' Meat Sci., vol. 84, no. 4, pp. 711–717, Apr. 2010, doi: 10. 1016/j.meatsci.2009.10.030.

[10] G. Ciocca, P. Napoletano, and R. Schettini, ''Food recognition and leftover estimation for daily diet monitoring,'' in New Trends in Image Analysis and Processing—ICIAP, V. Murino, E. Puppo, D. Sona, M. Cristani, and C. Sansone, Eds. Cham, Switzerland: Springer, 2015, pp. 334–341.

[11] P. Duan, W. Wang, W. Zhang, F. Gong, P. Zhang, and Y. Rao, ''Food image recognition using pervasive cloud computing,'' in Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Social Comput., Aug. 2013, pp. 1631–1637, doi: 10.1109/GreenCom-iThingsCPSCom.2013.296.

[12] W. Phetphoung, N. Kittimeteeworakul, and R. Waranusast, ''Automatic sushi classification from images using color histograms and shape properties,'' in Proc. 3rd ICT Int. Student Project Conf. (ICT-ISPC), Mar. 2014, pp. 83–86, doi: 10.1109/ICTISPC.2014.692322.

[13] F. Kong and J. Tan, ''DietCam: Automatic dietary assessment with mobile camera phones,'' Pervas. Mob. Comput., vol. 8, no. 1, pp. 147–163, Feb. 2012, doi: 10.1109/ICT-ISPC.2014.6923223.

[14] L. Bossard, M. Guillaumin, and L. Van Gool, ''Food-101—Mining discriminative components with random forests,'' in Proc. Eur. Conf. Comput. Vis., 2014, pp. 446–461.

[15] H. Wang, W. Min, X. Li, and S. Jiang, ''Where and what to eat: Simultaneous restaurant and dish recognition from food image,'' in Advances in Multimedia Information Processing—PCM, E. Chen, Y. Gong, and Y. Tie, Eds. Cham, Switzerland: Springer, 2016, pp. 520–528.