OPEN ACCESS
International Journal of Innovations in Science & Technology
RESEARCH & INNOVATION DIVISION
IJIST

Comparative Study of Food Image Classification Performance Using the Xception Architecture

Mian Jamal Shah[1], Aftab Alam[2]*, Ihsan Rabbi[3], Shah Khalid[4], Gohar Ali[5]
[1, 2, 4] Computer Science & IT, University of Malakand, Chakdara, Dir Lower, 18300, Khyber Pakhtunkhwa, Pakistan.
[3]Institute of Computer Science & Information Technology, University of Science and Technology Bannu, 28100, Khyber Pakhtunkhwa, Pakistan.
[5] College of Information Technology, Kingdom University Bahrain, 3903 Riffa Bahrain.
***Correspondence**: alam@uom.edu.pk;

Citation | Shah. M. J, Alam. A, Rabbi. I, Khalid. S, Ali. G, "Comparative Study of Food Image Classification Performance Using the Xception Architecture", IJIST, Vol. 7 Issue. 2 pp 741-754, May 2025
Received| April 17, 2025 **Revised|** May 10, 2025 **Accepted|** May 11, 2025 **Published|** May 12, 2025.

Food allergies remain a critical issue that needs more research. To identify and manage food allergies, the integration of complex computational approaches is becoming more and more important, opening the door to more individualized and efficient food safety solutions. Which aims to end hunger, achieve food security, improve nutrition, and promote sustainable agriculture. This research investigates the application of image classification techniques for allergen detection in food images. Specifically, we compare two models Model 1 serves as the baseline, trained on 11 classes. Two variations were explored: Model 2 focuses on Pakistani dishes, to investigate the impact of learning rate on the balance between adaptation speed and model precision. The objective is to determine the most effective model for classifying food images therefore *Model 2 achieves the highest accuracy of* ***94%***. These findings suggest that Model 2 is a promising candidate for real-world allergen detection applications. Future research will focus on creating a comprehensive new dataset of food images encompassing a wider variety of food items, as well as exploring the integration of a model similar to model 2 into mobile applications for consumer use.

Keywords Deep learning, Xception, CNN, Food Safety, Classifiers, Food-101.

Introduction:

Food safety has deep historical roots, with ancient civilizations and religious texts emphasizing the importance of consuming wholesome and uncontaminated food [1]. It can be simply defined as the practices and precautions put in place to ensure that the food people eat is safe and free from harmful substances or contaminants. It covers a range of topics, including how food is handled, produced, stored, and distributed to stop the spread of disease [2]. There is a dire need for policies and standards to ensure a common understanding of food safety [3]. Food safety is all about consumer health protection and well-being by making sure the food they eat is safe and good for them [1]. It has been noted that the application of optical sensors and biosensors makes it easier to detect contaminations and adulterant foodstuff, leading to an increase in speed, improvement in the accuracy of results, and a decrease in costs [4]. Breakthroughs in science during the 20th century led us to better understand foodborne pathogens and causes of contamination. A variety of new advanced food safety technologies including pasteurization, thermal processing, and refrigerated preservation resulted from this newly acquired knowledge and significantly enhanced the safety of the food supply [5].

Over the past three years, studies have focused on food safety, exploring everything from microscopic elements to safety products, laws, regulations, and related areas of study [6]. Ensuring food safety is crucial for preventing allergen contamination which has emerged as a significant food safety concern, affecting an estimated 5% of the global population. Allergies can trigger severe reactions, and individuals with such conditions are highly sensitive, even trace amounts of an allergen can lead to symptoms like itching, rashes, hives, or nausea. In more extreme cases, exposure can result in anaphylaxis, a life-threatening reaction. This heightened sensitivity adds significant complexity to food safety management [7]. A small group of eight foods including milk, peanuts, eggs, tree nuts, soy, fish, wheat, and shellfish are responsible for most food allergies. While some people may eventually outgrow certain allergies, others must manage them throughout their lives by strictly avoiding the trigger foods [8]. Historically, people with food allergies avoided the allergen and monitored themselves by observing reactions to named foods, using limited information about allergens that must be prioritized over food labels with little detailed ingredient information. As food allergies became known, labeling standards progressed more and more, making it possible to make better informed choices. Today, people with food allergies adopt a holistic strategy for managing their condition by reading ingredient labels thoroughly, especially with such enhanced labeling laws to specify potential allergens. Some individuals choose to prepare meals at home to have full control over the ingredients used, reducing the risk of accidental allergen exposure. They also consistently communicate their dietary needs in social and dining situations to ensure food is prepared safely. Epinephrine auto-injectors are also commonly carried for quick use in response to accidental exposure and emergency medications are also typically carried. With the development of allergen-free food alternatives, individuals now have more options to replace common allergens, making it easier to maintain their diets and achieve a balanced nutritional intake. Allergy management also provides connections to support groups and online resources to help effectively manage necessary lifestyle changes. This reflects the product of medical advancements as well as a better understanding of the disease, leading to a safer and more adaptable environment for those affected.

Deep learning has been applied more often in various domains, proving its worth in enhancing outcomes in critical sectors such as food processing, pharmacogenomics, and food analysis [9][10]. Even though DL has great potential for solving a variety of issues related to food safety, there are still several important obstacles and opportunities to be addressed. To fully harness the potential of deep learning in enhancing food safety and protecting public health, several challenges must be addressed. These include the need for reliable model interpretability methods, access to standardized datasets, integration with emerging technologies such as the Internet of Things (IoT) and blockchain and careful consideration of ethical and legal issues [11]. Convolutional Neural Networks (CNNs) have recently received much attention for use in food science, such as the classification of foods, detection of allergens and evaluation of food safety. AlexNet, GoogleNet, VGGNet, ResNet, and DenseNet deep learning models are widely utilized for food classification tasks, due to their capacity to extract fine features from varying food items' common allergens. These architectures are becoming increasingly relevant for real-time applications and allergen detection on various food products. Deployment of deep learning models on small devices is often difficult since deep learning models often consume huge computational resources that are expensive in terms of memory and processing demand. For these reasons, implementing such models in food safety applications where low-power devices or real-time processing are desirable remains nontrivial. Although for deployment these systems require advanced computing resources such as GPUs and high-performance servers, their high cost precludes their practical use in many food safety contexts. However, this lack of research points out the need for innovative deep-learning approaches that scale down the parameter count power consumption and computational requirements to render food classification and allergen detection and verification much simpler and more efficient.

In this research, we adopted an enhanced Xception-based model architecture specifically tailored for food safety and food classification tasks using the Food-101 image dataset. We also contributed to the design of the model as well as to a specific application in the domain. We utilized the pre-trained Xception model as a feature extraction backbone, minifying a classification scheme by adding a few layers to allow a precise classification while still being computationally efficient. This model can achieve robust food image classification performance in food safety monitoring tasks, such as allergens identification and food category classification. by enhancing it with additional layers to further improve the accuracy of the model. Global Average Pooling layer, 128 units' Dense layer followed by ReLU activation and 0.5 Dropout after it is added to help prevent overfitting. The Xception-based model was trained over 100 epochs with data augmentation techniques (rotations, flips, and zooms) for generalization on various food image data. Finally, we compared this model against state-of-the-art architectures such as ResNet, EfficientNetB0, etc. on food classification tasks and showed that the model strikes a good balance between accuracy and efficiency. We also improved generalization and model stability through SGD optimization. Overall, this work is novel to the application of deep learning to food sciences, particularly in its application to a food safety problem focused on food image classification for allergen identification.

Related Work:

Zhidong Shen demonstrated a machine learning system that can identify food images and calculate their nutritional values, assisting in the fight against obesity and encouraging a healthy diet. Convolutional neural networks (CNNs) are used in the system; more precisely, the Inception V3 and V4 models are used, which have been optimized for the Food-101 dataset as well as a customized dataset of subcontinental dishes. With components for CNN-based classification, attribute estimates, and server management, this client-server-based system handles image classification requests on the server side. The proposed approach overcame the limitations of the best models at that time, which made a difficult to identify complicated or mixed foods including salads, prepared dishes, and liquids. Their system intended to raise awareness of nutrition and calorie consumption while achieving an accuracy of 85% using approaches like data augmentation [12].

The multimodal nature of today's digital environment is growing. Images and text often coexist in web browsing, therefore classification problems combining these two modalities are common. Using both textual data and visual representations of the same concept, Gallo et al [13]. Investigated multimodal categorization. They concentrate on two key strategies for multimodal fusion and modify them using stacking techniques to improve the handling of such problems. Gallo et al. used the highly complicated and noisy UPMC Food-101 dataset as an example of this class of multimodal problems. Their findings demonstrated that, on the dataset used, the proposed early fusion technique combined with a stacking-based approach achieved superior performance compared to existing state-of-the-art methods [13]. Food recommendations are quite important. To create a system, T. Tangpong[14] compiled a dataset of food images sourced from social media platforms, reflecting real user choices and behaviors shared online. The scalable design incorporated numerous Deep Neural Networks (DNNs) with a reliability score for each DNN and introduced it in the field of representation learning. Consequently, among the independently built networks, the combined DNN was able to choose the most suitable recognition result. To speed up the process of making food recommendations, the Apriori data mining algorithm was trained with the frequently occurring food sets that were taken from these photos [14].

Within computer vision, one of the most promising uses of visual object recognition is food image recognition. Many state-of-the-art approaches for a range of tasks are based on convolutional neural networks. A unique Deep Learning-based technique for identifying food photos was presented by M. Z. Amin and colleagues. Utilizing the TensorFlow platform's Inception V3 model, Google retrains the last layer of its well-known Inception V3 architecture for its categorization method through the use of transfer learning. Data augmentation methods based on geometric transformations were used to enhance the training dataset. This method showed encouraging results, with an overall accuracy of roughly 91% in accurately classifying food photos and a successful reduction of overfitting [15].

Objectives:

This research investigated the use of advanced deep learning methods to identify images associated with food safety, with a specific emphasis on the Xception architecture and compared its performance with other advanced models like ResNet-50 and EfficientNet using the Food-101 dataset, and analyzed potential constraints and future possibilities for enhancing food safety applications.

- Implemented an Xception-based classification model, achieving 94% accuracy on Food-101 through optimized training.
- Compared Xception's performance to ResNet-50 and EfficientNet, showing 10% higher accuracy.
- Identified limitations for food safety tasks, proposing data augmentation and domain adaptation as critical next steps.

Methodology:

This study follows a well-established machine learning pipeline, structured into four core phases: (1) data collection and preparation, (2) model selection and implementation, (3) training and evaluation and (4) result analysis. While the workflow itself is conventional, the novelty of this work lies in custom hyperparameter and augmentation tuning, Rescaling, rotation, width and height shifts, shear and zoom transformations, horizontal flipping, and nearest-neighbor fill mode are all part of its data augmentation method.

Proposed Method:

The FOOD-101 dataset was selected as the basis for our comparison, enabling us to evaluate our food classification results against those reported in various other studies. Its original aim is to support further research in the sphere of computer vision and machine learning, especially concerning the identification of food images. For image classification, we opted to use Xception which is short for "Extreme Inception"; a model introduced by François Chollet in 2017 [16] due to its strong performance and architectural efficiency. It expands on the concept of Inception modules, substituting depth wise separable convolutions for traditional convolutional layers at every network layer. Xception's unique application of depth-wise separable convolutions lowers computational complexity while retaining high accuracy in identifying fine-grained characteristics like textures and forms that are essential for differentiating between food items.

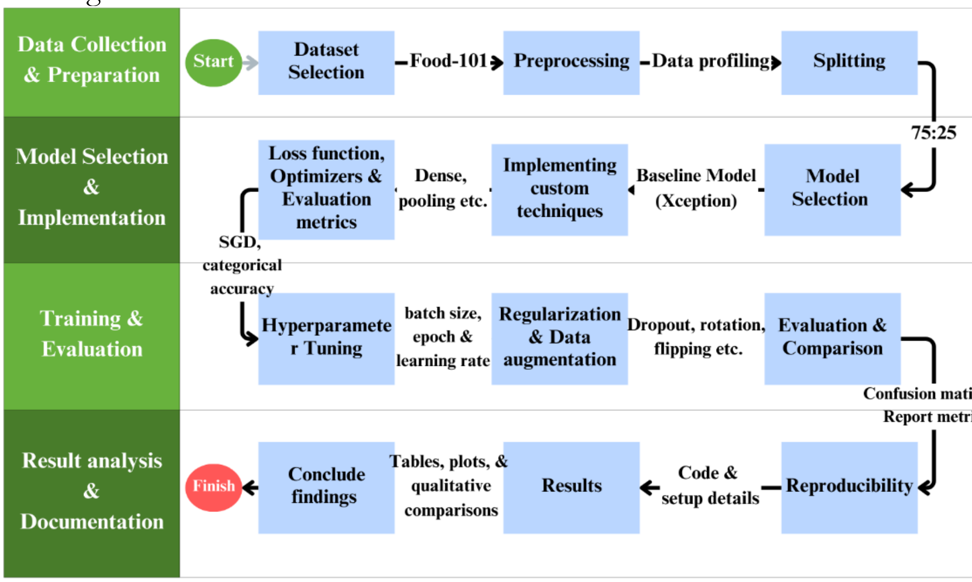

Figure 1. Dataflow diagram of methodology visualized

This architecture in Figure 1 is particularly good at capturing multi-scale elements, which is crucial for managing the variation in food appearances brought about by various display methods, lighting setups and cooking techniques. Transfer learning from pre-trained models such as Xception efficiently addresses challenges like dataset diversity, allowing for rapid adaptation to particular food image datasets and improving overall classification performance.

The Food-101 Dataset:

A large and readily available Food-101 dataset, created by Lukas Bossard et al., contains 101,000 photos spread throughout 101 food categories. The dataset was obtained from kaggle.com [17]. This dataset stands out as the first publicly accessible resource of its type, offering a significant advancement over previous research that primarily used small, closed datasets gathered under controlled conditions or from private sources [18]. Bossard et al. tackled the challenge of automated dish recognition using a new method called Random Forest Discriminant Components (RFDC). This approach employed Random Forests to identify discriminative regions within food images, enabling the simultaneous extraction of meaningful features across all categories and promoting knowledge transfer between classes. Achieving an average accuracy of 50.76%, the RFDC method surpassed other classification approaches except for Convolutional Neural Networks (CNNs) by margins of 11.88% and 8.13% when compared to SVM classification using Improved Fisher Vectors and existing part-mining techniques.

Data Collection:

The Food-101 dataset in Table 1, contains 101 food categories for fine-grained image recognition. Released by ETH Zurich in 2014, it is publicly available for non-commercial research use. Food photos were scraped from the food blogging platform foodspotting.com (now defunct) as shown in Figure 2.

Table 1. Dataset Statistics

Category	Details
Total Classes	101 (e.g., "apple pie," "chicken_curry," "ice_cream")
Images per Class	1,000 (750 train + 250 test)

Total Images	101,000 (75,750 train + 25,250 test)
Image Resolution	Variable (typically ~512×512 px)
Format	JPEG

Figure 2. Visualization of 100 classes out of 101 [18]

Data Preprocessing:

The raw dataset and images are not immediately suitable for direct sense in their original form. Therefore, the training dataset is first preprocessed by a series of transformations including random rotation, resized cropping, random horizontal flip and application of the ImageNet augmentation policy, followed by normalization. These preprocessing transformations serve multiple purposes: This reduces variations in image properties obtained from various backgrounds and also helps the model learn faster and improve its output accuracy. To form 'train mini' and 'test mini' directories, we took a subset of the original food image dataset by sampling representative images for each of the classes. We used image augmentation techniques, using the ImageDataGenerator in TensorFlow to help with training data, such as rescaling (1. /255), random rotations (up to 30 degrees), width and height shifts (20%), shearing (0.2 factor), zooming (20%), flipping horizontally and filling any missing pixels with nearest neighbor. We only applied rescaling for validation for resizing images to 299x299 pixels, with a batch size of 32, and set the data in categorical mode to handle multiple classes.

Proposed CNN Architecture:

Convolutional Neural Networks (CNNs) are a specialized type of deep learning algorithm. Specifically, they represent a deep architecture and a convolution-based variant of feed-forward neural networks, designed to process and analyze visual data effectively. It can deliver low-level features from the original input data and assemble them into high-level features due to a multi-layered data structure. CNNs are a type of multilayer perceptual machines (MLPs) that were invented by Hubel and Wiesel in the 1960s [19] realized that there exists a special status of network convection for locally sensitive and direction-selective neurons that can minimize the number of neurons in the network [20]. The input, convolutional, ReLU, pooling and fully connected layers are the main layers of the convolutional neural networks. However, the ReLU layer is often a part of the convolutional layer. Staking these layers on one another provides another form of a complete convolutional neural network. The following Figure. 3 shows the structure of Convolutional Neural Network.

The paper presents a novel CNN architecture based on Xception, which consumes much computational effort Xception's depth wise separable convolutions assist in the efficient determination of such patterns.

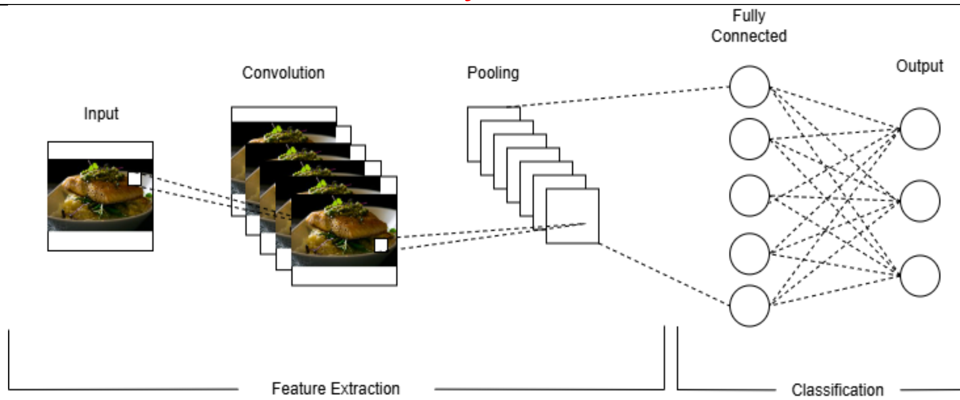

Figure 2. Schematic Diagram of CNN model structure

The architecture of the Proposed CNN Model:

To perform robust feature extraction, the proposed CNN model uses the Xception architecture (Figure 4) and it is pre-trained on the ImageNet dataset. Before passing inputs through a global average pooling layer to reduce dimensionality and increase computational efficiency, the network's output is first passed through a global average pooling layer. Next through dense layer with 128 units and ReLU activation to refine the features even further. After the dense layer, a dropout layer with 0.5 is used to avoid overfitting. Finally, we have the output layer, which with a softmax activation we want to classify images to one of n, meaning the layer data type is float32 to keep numerical stability. The model is trained with 0.0001 using SGD optimizer with momentum 0.9. We then utilized the ModelCheckpoint callback to monitor the training process and save the model with the best performance, while the CSVLogger was employed to record the training history for further analysis. We compiled the model with categorical cross-entropy loss and evaluated the metric as accuracy, and then the model was trained and saved the final model for later use.

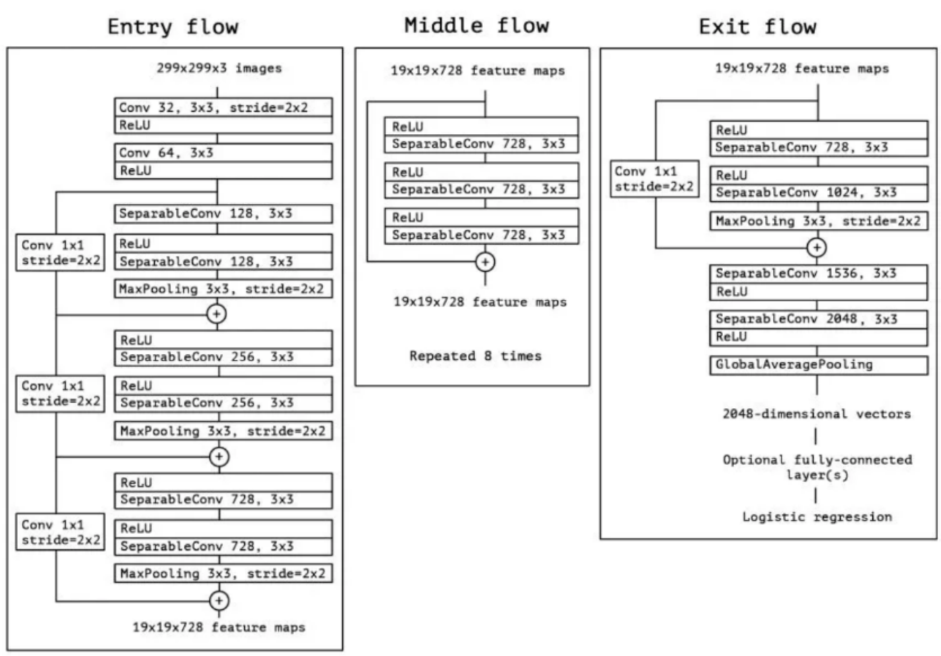

Figure 3. Xception: depth-wise separable convolutions [16].

The feature extraction backbone for the Xception-based model (Figure 5) is a pre-trained architecture, which does not include the top fully connected layers to permit custom classification. A Global Average Pooling layer follows this backbone and reduces spatial dimensions to a compact feature map. The next one is the 128 neurons Dense layer with ReLU activation for more feature refinement and Dropout layer (rate 0.5) which prevents overfitting by randomly inactivating neurons during training. The final output layer uses a softmax activation to classify an input to, say, some specified number of classes. We optimized the model using Stochastic Gradient Descent (SGD) with a learning rate of 0.0001 and momentum of 0.9 and used checkpoint and logging callbacks where we save the best model and write logs on the training progress.

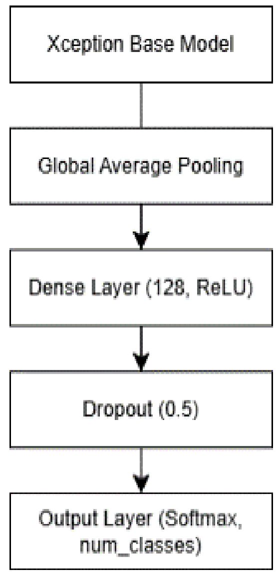

Figure 4. Structure of the Xception model with added layers

We used Kaggle kernels to experiment with various hyperparameter settings and model configurations. This platform enabled us to train and evaluate models efficiently, which was essential for comparing our methods to pre-established baselines. We duplicated and expanded on current study findings on image classification based on CNN Architectures using Kaggle kernels. The effectiveness of Kaggle kernels as a flexible platform for deep learning experiments is demonstrated in this work. This platform could be used in future research to investigate new designs and incorporate more datasets to improve the model's resilience.

Table 2. Kaggle kernels for Experimental configurations

Configuration	Information	Computational Constraints
Operating System	Environment using Docker, usually Linux.	-
CPU	Various configurations are available, often multi-core	12 hours for CPU
GPU	Options include NVIDIA Tesla P100, P4, T4, K80, or no GPU (CPU-only)	12 hours for GPU
Memory	Range from 4GB up to 16GB+ depending on kernel type.	30GB Max
Programming Language	Python	-
Python Version	3.9-3.12	-
IDE	Jupyter Notebook Interface	-
Computing Environment	Docker Containerized environment	-
Deep learning Framework	TensorFlow, Pytorch, Keras, and others.	-

Comparison of Accuracy and Loss Graphs of the Models:

Figure 6 shows 2 models namely (a,b),(c,d) with training and validation accuracy as well as training and validation loss for the proposed model, this study has been performed on 2 subsets of the food-101 dataset with different parameters for this study to evaluate how well they could categorize food images into the appropriate classes. The performance has been compared in Table 3-4 considering the size of the dataset, the authors tested models only on the first 11 classes and subset commonly available food images in my region and food items of 10 classes using various parameters utilizing the Xception architecture. The results of these analyses are shown below in the Figure. 6.

Experimental Results:

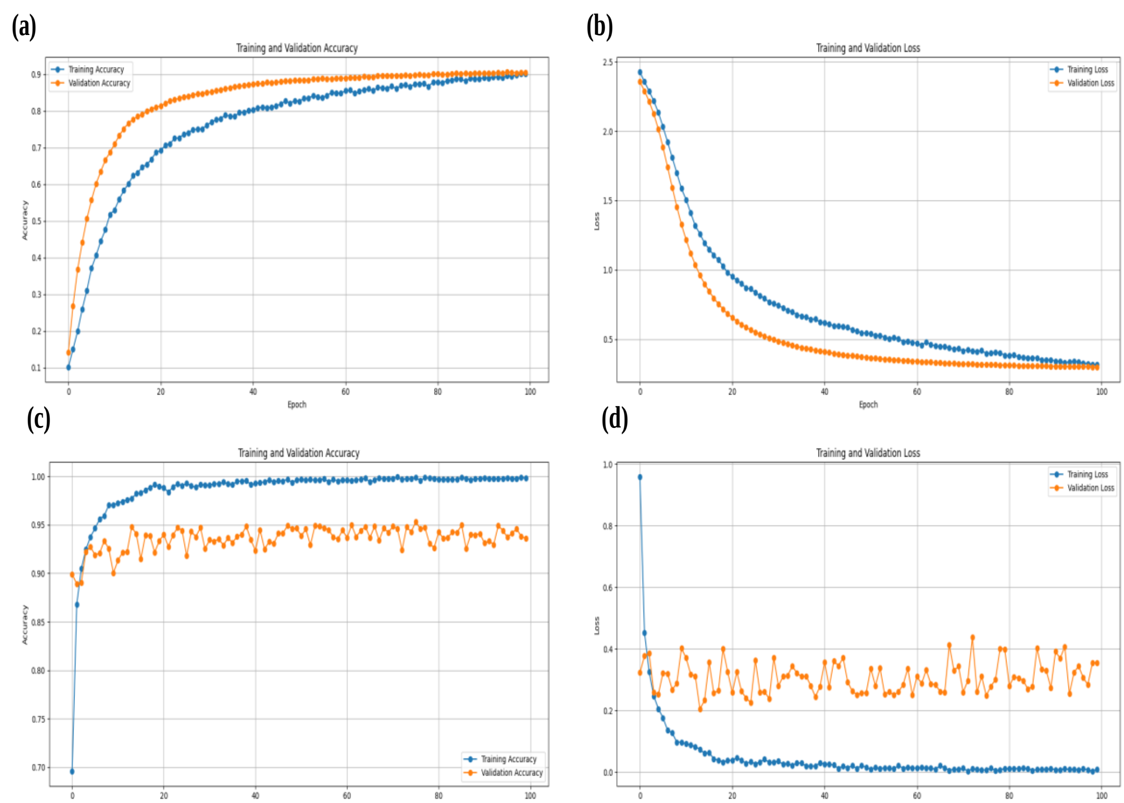

Figure 5. The accuracy and loss graphs of training the proposed model with different configurations on the food-101 dataset

Confusion Matrices of the Trained Models:

The following confusion matrices are provided (Table 3) to assess categorization models: To display the model's performance, it makes a comparison between the predicted and actual class labels. For binary classification, the matrix shows four important values.

Table 3. Configurations of trained models comparison

Layers	Model 1	Model 2
CNN	Xception	Xception
Input Size	299 x 299	299 x 299
Data Augmentation	rescale=1./255, rotation range=30, width shift range=0.2, height shift range=0.2, shear range=0.2, zoom range=0.2, horizontal flip=True,	rescale=1./255, rotation range=30, width shift range=0.2, height shift range=0.2, shear range=0.2, zoom range=0.2, horizontal flip=True,

	fill mode='nearest	fill mode='nearest
Pooling	GloabalAveragePooling	GloabalAveragePooling
Output	Dense(activation= Softmax)	Dense(activation= Softmax)
Learning Rate	1×10^{-4}	1×10^{-2}
Loss Function	Categorical_crossentropy	Categorical_crossentropy
Accuracy	91%	94%

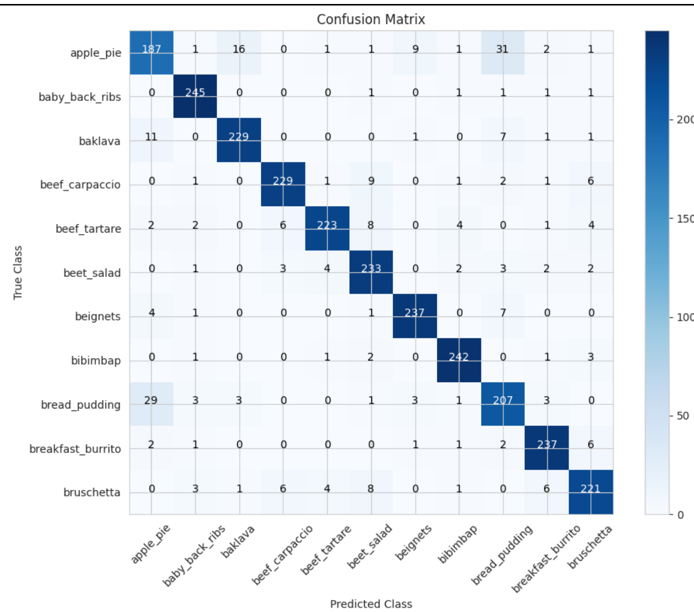

Figure 6. Confusion Matrix food-101 sample

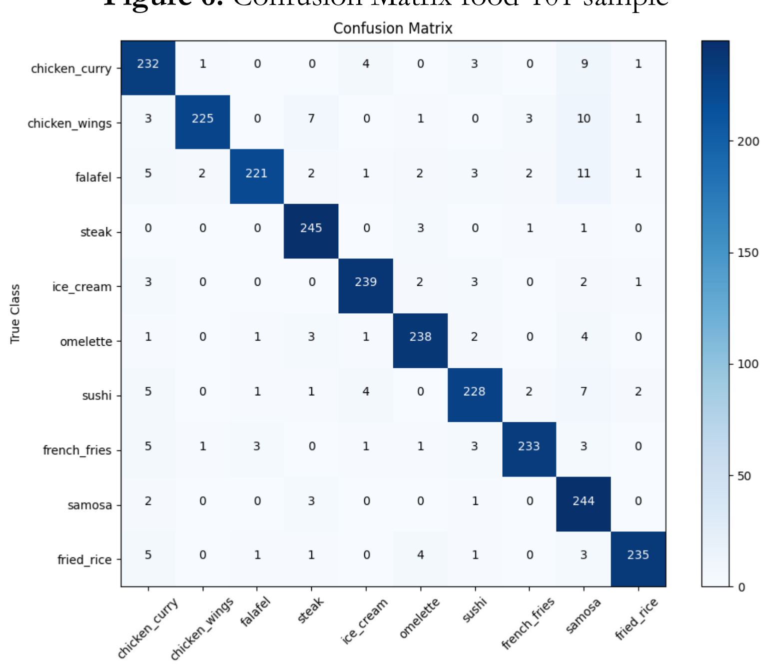

Figure 7. Confusion Matrix food-101 sample of Pakistani foods

Results Based on the Evaluation Metrix:

Model 1 used a 299x299 pixel input size and an Xception CNN architecture (Figure 7 and Figure 8 refers). Before the output layer, which employed a Dense layer with a Softmax activation function, the model integrates GlobalAveragePooling. It was trained with the Categorical Cross entropy loss function and a learning rate of 1×10^{-4}, achieving an accuracy of 91%. Model 2 follows a similar structure, utilizing GlobalAveragePooling, a Dense output layer with Softmax activation, and the same loss function as Model 1. However, Model 2 achieved a higher reported accuracy of 94% after being trained with a substantially greater learning rate of 1 × 10^-2. Table 4 refers below,

Table 4. Comparison of models using a standard evaluation matrix

Models	Classes	Precision	Recall	F1-Score
Model: 1	Initial 11 for the base architecture	90.55	90.55	90.51
Model: 2	Random 10	93.91	93.60	93.64

Discussion:

This research introduces a novel technique for sorting food images using the Xception architecture, which leverages depth-wise separable convolutions to achieve remarkable accuracy (91% top score: 94%), outperforming existing methods. The model's success stems from its ability to accurately capture intricate patterns in food images, making it particularly effective for classifying popular Pakistani dishes.

Table 5. Comparison with state-of-the-art methods

S.No	Authors	Year	Classification	Model	Accuracy
1	Lukas Bssard [18]	2014	Food-101	Random Forests	50.76
2	Igazio Gallo [13]	2020	Food-101	Inception V3	71.67
3	T Tangpong [14]	2021	Food-101	GoogLeNet	80.00
4	Chang Liu [21]	2022	Food-101	CNN	77.40
5	Prakhar Tripathi [22]	2022	Food-101	Random Forest	50.78
6	Vijaya Kumari G [23]	2022	Food-101	EfficientNetB0	80.00
7	Jo˜ao Louro [24]	2024	Food-101	ResNext-50	90.00
8	Proposed method	2024	Food-101	Xception	94.00

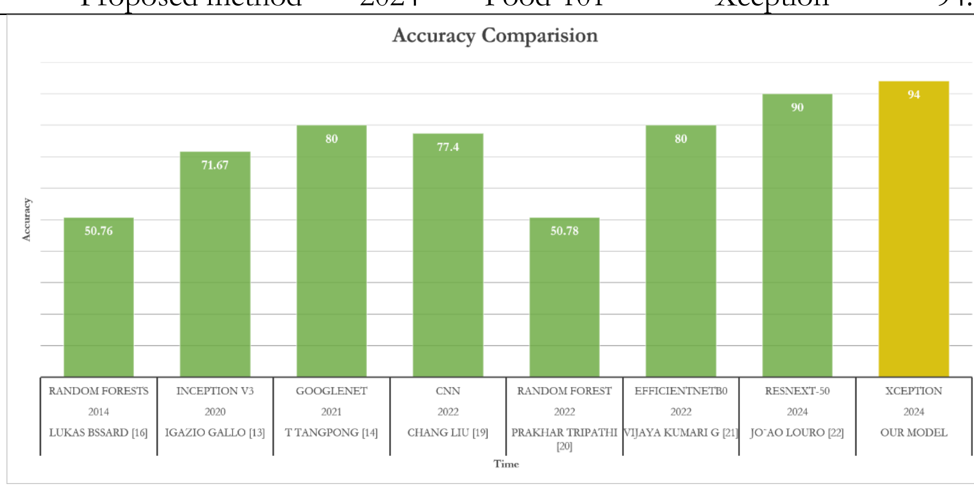

Figure 9. Accuracy comparison for Table 5

Accurate food identification is important for culinary research, restaurant management and food safety, improving efficiency, customer satisfaction and traceability. Some possible uses include

- Supply chain optimization
- Personalized diet planning

- Intelligent food assistance systems (e.g., meal recommendations)

To facilitate real-world testing, we deployed the model on an interactive web platform (https://foodcheck.streamlit.app/), where users can upload and classify food images. Our approach not only achieves high accuracy but also demonstrates the potential for expanding to regional and global cuisines, paving the way for improved food recognition systems in multiple domains.

Conclusion:

One of the most effective methods for simulating intricate processes and identifying patterns in applications with vast volumes of data is CNN, a deep feed-forward network that is simpler to train. Our method makes a great contribution to applications where accurate food product identification is crucial since it shows remarkable accuracy in food recognition. For example, in the realms of culinary research, restaurant management and food safety, the capacity to precisely classify food items can greatly improve both operational efficiency and user experience. Thus, the emergence of an efficient CNN-based model and the comparable results to the state-of-art methods, that have been offered in our work, opens the opportunity to investigate more complex architectures suitable for certain contexts of food recognition. Correct categorization of various meals is an example of how such techniques can be applied to other specialty or regional food categories hence improving the food safety processes.

Availability of Data and Material

The datasets used in this study, including Food-101, are publicly available at their respective repositories https://www.kaggle.com/datasets/dansbecker/food-101/code. Custom preprocessing scripts and model codes are available at https://github.com/mianjamalshah/Xception-food-101-data.git.

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

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions: The first author Mian Jamal Shah conceptualized the study and developed the methodology, implemented the models, and conducted experiments. Authors Dr. Aftab Alam, Dr. Ihsan rabbi, Dr. Shah Khalid, and Dr. Gohar Ali closely guided Author Mian Jamal Shah in conducting his study on the topic as part of his master's program in Computer Science at the University of Malakand. All authors contributed to writing, reviewing, and approving the final manuscript.

Acknowledgments: I would like to express my gratitude to Dr. Aftab Alam from the University of Malakand for their invaluable support and guidance throughout this study. I also appreciate the assistance of Dr. Ihsan rabbi, Gohar Ali, and Shah Khalid. Special thanks to my father for the financial support. Lastly, I acknowledge the use of publicly available datasets which was essential to this research.

References:

[1] Daniel Maestro, "Food safety–From pioneering steps to the modern scientific discipline," *J. Heal. Sci.*, vol. 12, no. 3, pp. 178–183, 2022, doi: https://doi.org/10.17532/jhsci.2022.1914.

[2] J. Karl Fink, "Food safety, plastics and sustainability: Materials, chemicals, recycling and the circular economy," *Food Safety, Plast. Sustain. Mater. Chem. Recycl. Circ. Econ.*, pp. 1–405, May 2023.

[3] E. S. Bayarmaa Barkhuu, Munkhjargal Odonchimeg, Erdenechimeg Namjil, Damdindorj Mungunnaran, "Food Safety: The contamination of persistent organic pollutants in some food products," *Bull. Inst. Chem. Chem. Technol.*, vol. 10, no. 10, pp. 131–139, 2023, doi: https://doi.org/10.5564/bicct.v10i10.2606.

[4] M. T. I. Ansari, S. K. Raghuwanshi, and S. Kumar, "Recent Advancement in Fiber-Optic-Based SPR Biosensor for Food Adulteration Detection - A Review," *IEEE*

Trans. Nanobioscience, vol. 22, no. 4, pp. 978–988, Oct. 2023, doi: 10.1109/TNB.2023.3278468.
[5] Y. Z. Xiaoyan Jiang, Zuojin Hu, Shuihua Wang, "Deep learning for medical image-based cancer diagnosis," *Cancers (Basel).*, vol. 15, no. 14, p. 3608, 2023, doi: https://doi.org/10.3390/cancers15143608.
[6] Tingting Kuai, "Comprehensive Literature Review of Research on Online Food Safety," *Sci. J. Technol.*, 2023, doi: 10.54691/sjt.v5i4.4735.
[7] R. J. M. Mimi L. K. Tang, "Food allergy: is prevalence increasing?," *Intern. Med. J.*, 2017, doi: https://doi.org/10.1111/imj.13362.
[8] A. Toulin, "'Old Methods Not Up to New Ways': The Strategic Use of Advertising in the Fight for Pure Food After 1906," *J. Gilded Age Progress. Era*, vol. 18, no. 4, pp. 461–479, Oct. 2019, doi: 10.1017/S153778141900029X.
[9] J. K. Lisa-Carina Class, Gesine Kuhnen, Sascha Rohn, "Diving Deep into the Data: A Review of Deep Learning Approaches and Potential Applications in Foodomics," *Foods*, vol. 10, no. 8, p. 1803, 2021, doi: https://doi.org/10.3390/foods10081803.
[10] K. N. P. Lili Zhu, Petros Spachos, Erica Pensini, "Deep learning and machine vision for food processing: A survey," *Curr. Res. Food Sci.*, vol. 4, pp. 233–249, 2021, doi: https://doi.org/10.1016/j.crfs.2021.03.009.
[11] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of Deep Learning in Food: A Review," *Compr. Rev. Food Sci. Food Saf.*, vol. 18, no. 6, pp. 1793–1811, Nov. 2019, doi: 10.1111/1541-4337.12492.
[12] J. L. Zhidong Shen, Adnan Shehzad, Si Chen, Hui Sun, "Machine Learning Based Approach on Food Recognition and Nutrition Estimation," *Procedia Comput. Sci.*, vol. 174, pp. 448–453, 2020, doi: https://doi.org/10.1016/j.procs.2020.06.113.
[13] I. Gallo, G. Ria, N. Landro, and R. La Grassa, "Image and Text fusion for UPMC Food-101 using BERT and CNNs," *Int. Conf. Image Vis. Comput. New Zeal.*, vol. 2020-November, Nov. 2020, doi: 10.1109/IVCNZ51579.2020.9290622.
[14] A. H. and K. T. Thunchanok Tangpong, Somkiet Leanghirun, "A Food Recommender Based on Frequent Sets of Food Mining Using Image Recognition," *Artif. Intell. Adv. New Paradig. Nov. Appl.*, 2021, doi: 10.5772/intechopen.97186.
[15] M. Z. and A. A. Amin, "InceptFood: A CNN Based Classification Approach for Recognizing Food Images," *Natl. Conf. Emerg. Trends Comput. Satisics Math. Sci. SBBWU*, 2017, [Online]. Available: https://www.researchgate.net/publication/338095670_InceptFood_A_CNN_Based_Classification_Approach_for_Recognizing_Food_Images
[16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, Nov. 2017, doi: 10.1109/CVPR.2017.195.
[17] Dan Becker, "Food 101," *Kaggle*, 2024, [Online]. Available: https://www.kaggle.com/datasets/dansbecker/food-101
[18] M. G. & L. V. G. Lukas Bossard, "Food-101 – Mining Discriminative Components with Random Forests," *Comput. Vis. – ECCV 2014*, vol. 8694, pp. 446–461, 2014, doi: https://doi.org/10.1007/978-3-319-10599-4_29.
[19] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962, doi: 10.1113/JPHYSIOL.1962.SP006837.
[20] C. Liu *et al.*, "swTVM: Exploring the Automated Compilation for Deep Learning on Sunway Architecture," *arXiv.org*, 2019.
[21] Y. M. Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, "Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment,"

arXiv:1606.05675, 2016, doi: https://doi.org/10.48550/arXiv.1606.05675.
[22] Prakhar Tripathi, "TRANSFER LEARNING ON DEEP NEURAL NETWORK: A CASE STUDY ON FOOD-101 FOOD CLASSIFIER," *Int. J. Eng. Appl. Sci. Technol.*, vol. 5, no. 9, pp. 229–232, 2021, [Online]. Available: https://www.ijeast.com/papers/229-232,Tesma509,IJEAST.pdf
[23] V. P. VijayaKumari G., Priyanka Vutkur, "Food classification using transfer learning technique," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 225–229, 2022, doi: https://doi.org/10.1016/j.gltp.2022.03.027.
[24] Â. O. João Louro, Filipe Fidalgo, "Recognition of Food Ingredients—Dataset Analysis," *Appl. Sci.*, vol. 14, no. 13, p. 5448, 2024, doi: https://doi.org/10.3390/app14135448.