Recipe Recognition for Indian Cuisine: Deep Learning vs Machine Learning

TIRIVEEDI HARI KIRAN[1], ANNAVARAPU GANESH[2], APPAJODU RAMA KRISHNA[3], DUVVURU MANOJ KUMAR[4], MUHAMMED SWALIH[5], MIRSAH VT[6]

[1, 2, 3, 4, 5, 6] *Scholar, School of Computer, Science and Engineering, Lovely Professional University, Punjab, India*

Abstract— In this research, we compared the Machine Learning and Deep Learning Algorithms on Indian Dataset. We Collected Indian food Images of more than 20 categories and applied Machine learning and Deep learning algorithms which includes MobileNet, KNN, CNN, and Random Forest, to the Indian food dataset and to train the algorithms and test them on that data. We measured the performance of each algorithm in the terms of accuracy, precision, recall and F1 score, while taking into account Identification of Indian food was the main priority. In contrast, to compare and analyze the computational efficiency and resource requirements for different algorithms, taking into account the model complexity, and the training model. Via thorough evaluation using the test data, each algorithm's accuracy in the food Recognition model is Comprehensively examined. impressively, MobileNet turned up as the top performer, achieving an impressive accuracy score of 92.1%, followed by CNN at 79.1%. On the other hand, Machine learning algorithms which are KNN and Randomforest display relatively lower accuracies of 25.5% and 33.3%, respectively. These results provide valuable insights into the efficacy of diverse machine learning and Deep learning techniques in food Recognition.

Index Terms— Image Classification, Food Recognition, CNN, MobileNet, KNN, Random Forest, Machine learning, Deep Learning

I. INTRODUCTION

Indian food is considered to be the most amazing in the whole world, diverse food culture, and the complicated nature of it. Due to the great variation of indigenous foods which could be found in every area automating the process of fusing culinary traditions of various regions is by no means an easy task for computer systems. Programs in food recognition merging computer vision and machine learning (ML) in the last while have caused major disruptions in multiple industries. Pictures often shown of the food items can be used for dietary monitoring, recipe recommendation, and preservation of culture with the accuracy to down to the pin point. This has enormous positive effects. The article states successfully the task realized under "Recipe Recognizer" an enterprise working on machine learning and computer vision and is basically do an Indian food item recognition.

1.1 Deep Learning:

Deep learning has its unique place in the domain of machine learning where it utilizes vast data with autonomy to develop the capability of machines for the purpose of making intelligent choices and deductions which can motivate machines make these decisions on their own even without explicit programming. In this case, it develops a separate organization (if we combine the working principles of the brain together) that can make itself study some patterns and insights from the large set of datasets and be able to perform this leads to the impartial problem-solving and autocracy decision-making process. The applications of Deep Learning might also vary from NLP to facial recognition and beyond. They can be used in decentralized driving, destination recommendation systems, and advanced medical solutions. Diverse types of neural networks show outstanding results for the specific liabilities. To point out, CNNs work excellent in picture treatment, RNNs deal the data with sequence and RL is for choice-making. In Deep Learning networks, network models amply mimic human brains learning processes, except that the parameters of such systems change with time and data constantly to achieve the precision and high level of complexity of these models (adaptive learning process). Studies of this kind do not wait for data to be understood in order to attribute it to the principal cause but rather in the process come great advancements also. Hence, the width and elegance of the Deep

Learning technologies reflect on how the problematic issues are nabbed from the jaws of failure

1.2 Machine Learning:

Machine learning is the domain of artificial intelligence (AI) which allows the computers to learn from the data and consequently make decisions or judgments without being explicitly programmed. It concentrates on design and implementation of algorithms and models that can extract essence of big data for patterns detection and empowered decision making. There is a wide range of applications for machine learning, such as speech and image classification, autonomous vehicles, online recommendation systems, and medical diagnosis. Machine learning includes separately such as supervised, unsupervised, and reinforcement learning each for its own type of approach for problem solving based on the constraints of their case.

1.3 Image Classification:

An image classification is a way to recognize images through their visual elements in the computer vision which is sort by their content form. Machine learning algorithms, mainly ones distinction image patterns and CNNs, is used in this case to divide images into corresponding category. It is widely seen as a viable technique used to recognise and detect faces in many applications like facial recognition, object detection, image detection, and autonomous vehicles. It proves to be an indispensable element in the development and practical implementing artificial intelligence. Machine learning contains two modules in feature extraction and one module in classification. It can scrunch out some specific elements of the image yet it is unable to see the differentiating features from the training data. On the deep level, a deep learning may automatically identify significant attributes that are needed for certain applications. CNN is one of the used deep learning approaches often. A CNN is made of an input-layer, a hidden-layer, and a convolutional layer. The sequences of layers are made of convolutional, pooling and fully connected layers. The example network has the peak layer which is fully connected to classify the image. Another set of parameters which are known as as hyper-parameters are extensively studied and observed. The parameters are very imporgan in de imageklassificatieproces. An experiment is conducted in Deep Learning as a part of Tunning to understand the significance of various hyperparameter. At one level it can be put in to perception of a neural network which can imply the number of hidden layers, the quantity of batch, the number of epochs, filter size, the number of filters, learning rate, optimization method as well as other settings. Accordingly, we will use the supervisor machine learning technology too and it is a helpful construct for image annotating [17]. Modern advancements like convolutional neural networks (CNNs) and deep-learning can be infused with traditional machine learning methods towards better accuracy and efficiency of Recipe recognition systems [15]. The aim of this work is to attempt, form, and check the ML algorithms so that they will have the ability to produce an accurate result in predicting Indian cuisine. The comparison of various methods including transfer models, MobileNet, K-nearest neighbors (KNN), Convolutional Neural Networks (CNN), and Random Forest is set as our aim in determining which approach or set of approaches that will produce the most accurate prediction for the Indian food.

1.4 Mobile Net:

The MobileNet is built on linearinitiallziation. It then contains one Conv2D layer of (3x3) kernels and (32) filters interconnected to ReLU activation. This layer operates by combining the `DepthwiseConv2d` which is a `Conv2D` layer (kernel set to 1,1) to enhance. This imaging device is goes for pooling through the downsampling of the signal using MaxPooling2D (pool size (2,2)) and flattening the output to feed through the dense layers. `Dense` sets of 32 units each and `ReLU` activation are added by 2 with the respective layers. The sub-level for having 6 units having the softmax activation function has been created (the output layer has one unit). By virtue of having this layout, MobileNet gets the best of both worlds, by being fast, accurate and lightweight; in addition, this is what icing is for mobile and edge devices.

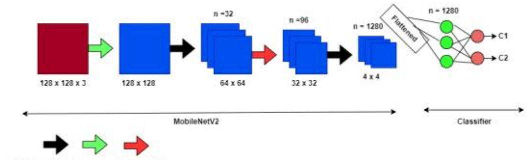

Figure 1: Mobile net Architecture

1.5 CNN
Convolutional Neural Network (CNN) is a deep learning (DL) algorithm which is wonderful at tracking images and sorting them where it is commonly used and considered as a sophisticated method in this field. Due to their specifics CNNs do not need sophisticated preparation because they can extract hierarchical feature representations from a visual input itself automatically. CNN model starts consecutively, with the Conv2D layer of the 32 filters used and the 3x3 kernel size as the single size followed by the ReLU activation function. This is immediately followed by a residual (Res) block and pooling layer (2x2) to avoid gradient vanishing and flatten spatial dimensions. More convolutions and pooling layers take place here, and after doing this, the data is flattened and goes through two dense layers made with 32 units each that use ReLU activation function. The output layer having 6 units pursuant to probabilities, assigned to the classes of classes, activates softmax function [20].

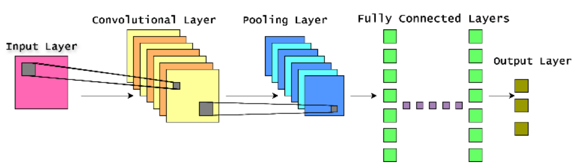

Figure 2: CNN Architecture

1.6 Random Forest:
Different from Random Forest, the Random Forest is actually the ensemble learning approach based on the decision trees. In training process, the Random Forest constructs many decision trees using the data samples with the mixed features of some of them randomly (the bagging method). Each decision tree responsible for prediction has separate predictions during testing and the final prediction is taken as a majority vote (for classification) or mean value (for regression) of individual tree results. The addition of the multiple decision trees to this growing bag of tricks reduces the overfitting which in turn enhances the prediction accuracy; hence leading Random Forest to being one of the popular methods used for classification and regression tasks.

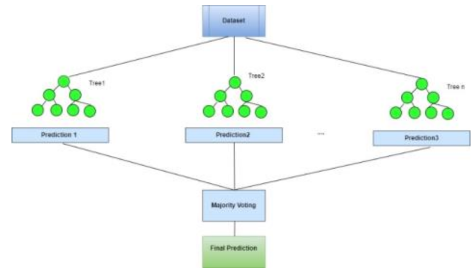

Figure 3: Random Forest Architecture

1.7 KNN:
The K-Nearest Neighbours (KNN) is a machine learning algorithm that although is seemingly simple, can achieve great results if used properly; therefore, despite the way it works differ significantly from neural networks or decision trees, it still uses the original training set to make predictions. Here is a combination of lots of features: the input is the vector of features, and the output is the matrix of features labels, using the heuristic of the Euclidean distance the matrix will find the appropriate group of this new data instance. The precondition for this prediction is "K" nearest neighbors of new instance (in case of multiple neighbors for instance). The gener9alized soft max function gives an average of the neighbors' approximation to the label to return the output label. KNN enjoys its popularity because of it easy-to-understand nature, and it is mostly employed with the classification task. Nevertheless, the right case and distance metrics choosing will, without doubt, bring better results, thus, cautious selection is also an important factor. [21]

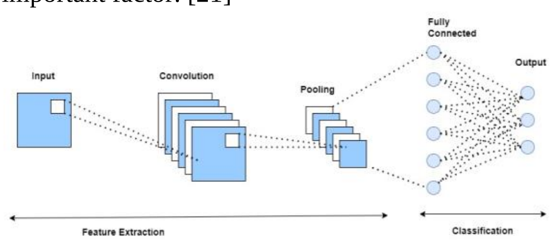

Figure 4: KNN architecture

II. LITERATURE REVIEW

The literature review includes the study of previous research that is associated with food classification and detection. the revised research highly contributed to the field of machine learning(ML), Deep Learning(DL) and computer vision which motivated us to select these articles to study the work of the other researchers to acquire a brief understanding of the procedure for such projects and standards to be satisfied for completion of the project. The reviewed papers are published over the years 2017 to 2023. This section also includes identifying major theories, or approaches that have been used, and summarizing the main findings of key studies to find the research gaps. Various techniques have been put forward in this literature on food recognition. In [1] [2] [4] [19] [21], these papers discuss the deep learning approach Convolutional Neural Network (CNN) algorithm on food/Cusine recognition. In [4] The authors trained the CNN model on Indian dataset with training accuracy of the model is 99.8% and precision, recall, and F1 score is 72%, 67% and 68%, respectively. In [23], the authors introduced a deep learning-based methods for a food recognition system that enables user to track the diet intake during the day. [18] talks about the Image Classification technique using Random Forest. [5][8] speaks about image recognition model in Machine Learning. [17] Speaks about Image Classification using Machine Leaning Models. The article [5] is machine learning and deep learning and these are where the article has reviewed almost 100 relevant works. These works portray itemised solution of machine learning & deep learning applied to the food computers with datasets as data and multimedia food related app development. After that, the journal thoroughly does some rearranging and putting together criteria of the data and put them up on websites which are Google-able and can be accessed anywhere, as it contains everything food-related, organized by locations. The study [19] introduces a new method for creating cooking recipes from food images using Convolutional Neural Networks (CNNs). The method uses deep learning to analyze food images and identify key ingredients and cooking steps. A unique dataset of food images and matching recipes is used, and a CNN model is developed specifically for this task. The model's understanding of the ingredients' visual context is improved by using attention mechanisms. The method's effectiveness is confirmed through tests, with the created recipes closely matching the input images. The study also includes comprehensive evaluations, such as recipe quality checks and user studies. This research could lead to new applications in the food industry, like automated recipe suggestions, real-time cooking help, and content creation for food-related platforms.

Table 1: Literature review Summary

Author	Task & Classifier	Results
Kiyoharu Aizawa, Makoto Ogawa, Hokuto Kagaya [1]	Food Detection and Recognition Using Convolutional Neural Network (CNN)	Accuracy – 93.8%
Haleema Begum, Akkanagamma [2]	Regenerating recipes from food images using CNN (Model)	Accuracy – 94%
Pawadea, D., Dalvia, A., Siddavatamb, I., Carvalhoc, M., Kotianc, P. and Georgec, H [4]	cuisine detection using the CNN network.	Accuracy – 92%
Zhu, B., Ngo, C.W. and Chan, W.K.[7]	Learning from web recipe image pairs for food recognition	Accuracy – 93%
Ann,E.T.L., Hao,N.S.,Wei, G.W. and Hee, K.C [8]	machine learning image recognition model	Accurarcy – 89%
Seon-Joo Park et. al. [16]	Korean Food Image detection (DCNN)	Accuracy – 91.3%

III. METHODOLOGY

The methodology of the Recipe Recognizer project involves several key steps (fig-5) Data Augmentation, Model Initialization, Training, and saving the model. Different machine learning models are selected for experimentation, which are MobileNet and CNN for deep learning, and Random Forest and KNN for traditional machine learning. These models are initialized with appropriate configurations and architectures. We have Collected, Data and Integrated into Directory Tree Format (fig-6) and Augmented for the new data points with available data. We have created train, test and valid data and Evaluated the models on validation and test sets to assess their generalization and performance on unseen data. We calculated and analyzed the accuracy, F1 score, precision, and recall to determine the effectiveness of each model in recipe recognition.

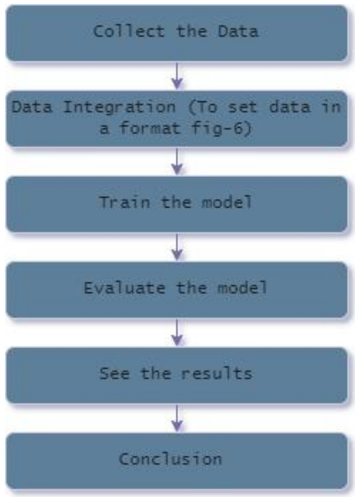

Figure 5: Methodology process

3.1 Data Collection:
After exploring a variety of sources, we gather a diverse collection of Indian food images from online sources, cookbooks, culinary databases, Google Images, Wikipedia, and Kaggle. Our dataset includes more than 20 Categories. We also included a decent number of entries of images per each dish so that we can train the model accurately. Our dataset includes more than 20 Categories.

3.2 Data Integration:
Further to data collection we cumulated all the data into a folder hierarchy structure. We have used this format to organize files into different categories (classes) for easier access and management. We labeled all the classes of food items by dish name. We created data folders for training, testing, and valid separately. The training folder consists of different classes of food items with dish names as a class, each class consists of specific food item images of an average of 100. The test set and valid consists of different classes of food items with dish name as class, Test set consists of an average of 50 images per class and valid set consists of 30 images per class. There are more than 20 categories of food items in the dataset.

3.3 Data Augmentation:
Data augmentation is a technique which artificially increasing the training data from the given data that is related to the same image. Data augmentation methods, including rotation, horizontal and vertical flips, as well as width and height shifts, applied and utilized in image recognition tasks, leading to enhanced accuracy performance. [22]

3.4 Model Training and Evaluation:
There are several steps in Model training (*fig-6*), Different machine learning models are selected for experimentation, which are MobileNet and CNN for deep learning, and Random Forest and KNN for traditional machine learning. These models are initialized with appropriate configurations and architectures.

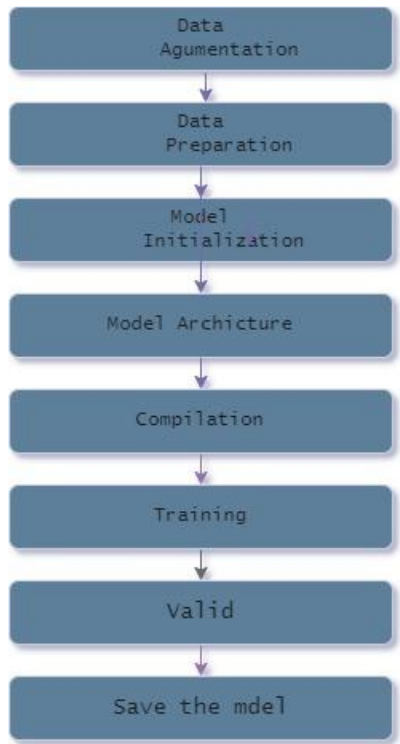

Figure 6: Model Training Process for deep Learning

3.5 Model Initialization and Training for Deep Learning:
First of all, we utilized TensorFlow and Keras libraries to initialize a CNN architecture and a mobile net. In the case of MobileNet, we used pretrained weights which were taken from the ImageNet dataset, so that the new classifier can take advantage of feature extraction that is provided by the pretrained weights. The training model cyclic process entails the generation of data augmentation (Fig-6) where data is rescaled, sheared, zoomed, horizontally flipped, and brightness adjusted. In pursuit of the aim to generate

many examples by distortion of the base images, such methods add more options to the training set in terms of its heterogeneity. Running it ahead the next stage is organizing the datasets using the ImageDataGenerator class from Keras by specifying target sizes, batch sizes, class-mode for training, validation and testing. For the model initialization, a sequential model is created using Tensorflow and Keras, normally starting with the convolutional feature extraction (conv2D) blocks and then the max-pooling (MaxPooling2D) layers which reduce the feature space. Through this architecture we achieve the possibility of reducing spatial dimensions and extracting image features efficiently. Another layers of compactness like flattening (Flatten), dense (Dense), and ultimately to dropout (Dropout) layers come into play to help achieve hierarchical representations and curb overfitting. (*fig-1*, *fig-2*)
The model commences initialization post that which is followed by model compilation done using an optimizer like adam and rmsprop. Also, a loss function like categorical cross-entropy is calculated along with performance metrics such as accuracy. This compilation of the model provides for training, exclusively, where the fit method is employed with parameters such as the batch size, epochs, and callbacks for early stopping and a check-pointing of the model. During the training stage, the model removes redundant parameters using the backpropagation procedure and optimizing algorithms that make it possible to minimize loss function and as a result enhance performance. The time trainer implement evaluation of the model through validation dataset via utilized evaluation method to measure the model accuracy and other assessment metrics. If the model is in compliance with the predicted parameters, can be used to save its trained weights and architecture in the way using the save method that can be employed later on for efficient deployment and predictions on new data.

3.6 Model Initialization and Training for Machine Learning:
In engineering machine learning pipelines that use models KNN and Random Forest, there is usually an initial step and the modeling process (fig-7). Firstly, the necessary libraries are imported, including 'scikit-learn' which is the one that affords access to the model classes and functionalities. The model instance is next created by calling the KNeighborsClassifier, RandomForestClassifier function with the arguments like .k for KNN and n estimators for Random Forest. The next step is launching the algorithm by which the data is then preprocessed by merging it into features and target variable, then after splitting into training and testing set using techniques like Cross validation and Train Test Split. Training process includes fitting the (training or adaptation) model to (by-means-of) the training data and finally calling the `fit` method on the model object. Additionality, the performance of the model can be checked by metrics like accuracy, precision, recall, F1 score, and confusion matrix which reveal the suitability of the model with regard to the new data it was never trained on. This generally approach of standardization is to psychologically create the method and system in establishing and training the KNN and random forest models in machine learning tasks.

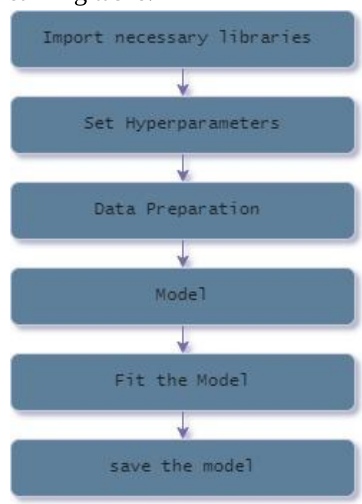

Figure 7: Model Training process for machine learning

3.7 Model Evaluation and Metrics:
We evaluated the models after training using valid data sets to determine their generalization and their accuracy in predicting future data. We went ahead and determined the accuracy, F1 score, precision, and recall to estimate how each model performed as regards recipe classifications.
Accuracy implies the general level of correctness of model predictions from the above statement and on the other hand the F1 score balances the precision and recall values during the multi-class classification tasks.

Precision measure the proportion of real positives in all the picked positives and recall measure the proportion of real positives in all the essential positives.

3.8 Evaluation Parameters:

3.8.1. Test Accuracy: Test accuracy assesses the number of cases in the test set that were classified correctly out of those taking the test set. It is determined by taking the number of right calls and total number of foreseeable events.

$$Test\ Accuracy = Number\ \boldsymbol{of\ Correct\ Predictions} / \boldsymbol{Total\ Number\ of\ Predictions}$$

3.8.2. F1 Score:

The F1 score comprises two parameters, precision and recall. Its utilization really makes sense when there are imbalanced classes and hence, number of samples in different classes is unequal. F1 score means harmonic mean of both precision and recall that guarantees its single score that is meant to have both false positives and false negatives into account.

$$F1\ Score = 2\ x\ (Precision\ \mathbf{x}\ \boldsymbol{Recall}) / (\boldsymbol{Precision} + \boldsymbol{Recall})$$

IV. OUTPUT

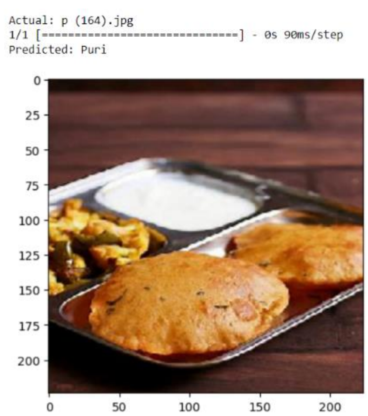

Figure 8: Output

Table 2: Results and Observations

Model	Train Accuracy	Test Accuracy	F1 Score	Epochs
CNN	0.924	0.791	0.755	25
Mobile Net	0.969	0.921	0.851	25
Random Forest	1.00	0.333	0.296	NA
KNN	0.738	0.255	0.698	NA

The results displayed in the Table-2 show quite clearly that Deep Learning type models, in particular the Mobile Net, have dramatically outperformed classical Machine Learning models which are Random Forest and K-Nearest Neighbours on the given Indian dataset. Having an test Accuracy of 92% and F1 Score of 85%, MobileNet performs the best among others which also shows clearly that the model is the most effective one among others compared. However, the Machine Learning models, in spite of the high Train Accuracy (Random Forest even got 100%), exhibit a marked fall in Test Accuracy and F1 Score, meaning overfitting problem. As a result, for this specific dataset, the Deep Learning models, as well as the MobileNet, present the highest accuracy rate in terms of predictions, In the specified Indian dataset.

REFERENCES

[1] Kagaya, H., Aizawa, K. and Ogawa, M., 2014, November. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1085-1088).

[2] [1] Haleema Begum,[2] Akkanagamma PIXEL TO PLATE: GENERATING RECIPES FROM FOOD IMAGES USING CNN.

[3] Mavrogiannis, A., Mavrogiannis, C. and Aloimonos, Y., 2023. Cook2LTL: Translating Cooking Recipes to LTL Formulae using Large Language Models. *arXiv preprint arXiv:2310.00163*.

[4] Pawadea, D., Dalvia, A., Siddavatamb, I., Carvalhoc, M., Kotianc, P. and Georgec, H., 2020. Cuisine detection using the convolutional neural network. *Int. J. Educ. Manage. Eng*, *10*(3), p.1.

[5] Evaluating machine learning technologies for food computing from a dataset perspective Nauman Ullah Gilal1 · Khaled Al-Thelaya1 · Jumana Khalid Al-Saeed1 · Mohamed

Abdallah1· Jens Schneider1 · James She1 · Jawad Hussain Awan2· Marco Agus1

[6] Min, W., Wang, Z., Liu, Y., Luo, M., Kang, L., Wei, X., Wei, X. and Jiang, S., 2023. Large scale visual food recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[7] Zhu, B., Ngo, C.W. and Chan, W.K., 2021. Learning from web recipe-image pairs for food recognition: Problem, baselines and performance. IEEE Transactions on Multimedia, 24, pp.1175-1185.

[8] Ann, E.T.L., Hao, N.S., Wei, G.W. and Hee, K.C., 2021. Feast in: a machine learning image recognition model of recipe and lifestyle applications. In MATEC Web of Conferences (Vol. 335, p. 04006). EDP Sciences.

[9] Zhang, Y., Yamakata, Y. and Tajima, K., 2021, December. Mirecipe: A recipe dataset for stage-aware recognition of changes in appearance of ingredients. In *Proceedings of the 3rd ACM International Conference on Multimedia in Asia* (pp. 1-7).

[10] Yamaguchi, Y., Inuzuka, S., Hiramatsu, M. and Harashima, J., 2020, November. Non-ingredient Detection in User-generated Recipes using the Sequence Tagging Approach. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)* (pp. 76-80).

[11] Bień, M., Gilski, M., Maciejewska, M., Taisner, W., Wisniewski, D. and Lawrynowicz, A., 2020, December. RecipeNLG: A cooking recipes dataset for semi-structured text generation. In *Proceedings of the 13th International Conference on Natural Language Generation* (pp. 22-28).

[12] Kanazawa, N., Kawaharazuka, K., Obinata, Y., Okada, K. and Inaba, M., 2023. Recognition of Heat-Induced Food State Changes by Time-Series Use of Vision-Language Model for Cooking Robot. *arXiv preprint arXiv:2309.01528*.

[13] Zhang, Y., Yamakata, Y. and Tajima, K., 2021, September. Supplementing Omitted Named Entities in Cooking Procedural Text with Attached Images. In *2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 199-205). IEEE.

[14] Diallo, A., Bikakis, A., Dickens, L., Hunter, A. and Miller, R., 2024. Unsupervised Learning of Graph from Recipes. *arXiv preprint arXiv:2401.12088*.

[15] Arora, S., Chaware, G., Chinchankar, D., Dixit, E. and Jain, S., 2019. Survey of different approaches used for food recognition. In *Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017* (pp. 551-560). Springer Singapore.

[16] Park, S.J., Palvanov, A., Lee, C.H., Jeong, N., Cho, Y.I. and Lee, H.J., 2019. The development of food image detection and recognition model of Korean food for mobile dietary management. *Nutrition research and practice*, *13*(6), p.521.

[17] PERFORMANCE ANALYSIS AND EVALUATION OF IMAGE CLASSIFICATION MODELS USING MACHINE LEARNING 1MOHAMED NOUR, 2RASHA M. AL-MAKHLASAWY, 3MAYADA KHAIRY

[18] Chari, K.K., Babu, M.C. and Kodati, S., 2019. Classification of diabetes using random forest with feature selection algorithm. *Int. J. Innov. Technol. Explor. Eng*, *9*(1), pp.1295-1300.

[19] PIXEL TO PLATE: GENERATING RECIPES FROM FOOD IMAGES USING CNN 1Haleema Begum,2Asst.Prof.Akkanagamma, 1Student,2Assistant Professor 1Artificial Intelligence and Data Science,

[20] Hussain, M., Bird, J.J. and Faria, D.R., 2019. A study on CNN transfer learning for image classification. In *Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop on Computational Intelligence, September 5-7, 2018, Nottingham, UK* (pp. 191-202). Springer International Publishing.

[21] Fan, Z., Xie, J.K., Wang, Z.Y., Liu, P.C., Qu, S.J. and Huo, L., 2021, May. Image classification method based on improved KNN algorithm. In *Journal of physics: Conference series* (Vol. 1930, No. 1, p. 012009). IOP Publishing.

[22] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J. and Shen, F., 2022. Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*